I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18307

Online Tutoring, School Performance, and School-to-Work Transitions: Evidence from a Randomized Controlled Trial

Silke Anger
Bernhard Christoph
Agata Galkiewicz
Shushanik Margaryan
Malte Sandner
Thomas Siedler

DECEMBER 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18307

Online Tutoring, School Performance, and School-to-Work Transitions: Evidence from a Randomized Controlled Trial

Silke Anger
IAB, University of Bamberg and IZA

Bernhard Christoph
IAB

Agata Galkiewicz
University of Potsdam and IAB

Shushanik Margaryan
University of Potsdam, IZA and BSoE

Malte Sandner
Nuremberg Institute of Technology, IAB and IZA

Thomas Siedler
University of Potsdam, IZA and BSoE

DECEMBER 2025

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

Online Tutoring, School Performance, and School-to-Work Transitions: Evidence from a Randomized Controlled Trial*

Tutoring programs for low-performing students, delivered in-person or online, effectively enhance school performance, yet their medium- and longer-term impacts on labor market outcomes remain less understood. To address this gap, we conduct a randomized controlled trial with 839 secondary school students in Germany to examine the effects of an online tutoring program for low-performing students on academic performance and school-to-work transitions. The online tutoring program had a non-significant intention-to-treat effect of 0.06 standard deviations on math grades six months after program start. However, among students who had not received other tutoring services prior to the intervention, the program significantly improved math grades by 0.14 standard deviations. Moreover, students in non-academic school tracks experienced smoother school-to-work transitions, with vocational training take-up 18 months later being 5 percentage points higher—an effect that was even larger (12 percentage points) among those without prior tutoring. Overall, the results indicate that tutoring can generate lasting benefits for low-performing students that extend beyond school performance.

JEL Classification: C93, I20, I24

Keywords: online tutoring, randomized controlled trial, disadvantaged youth, school grades, school-to-work transition

Corresponding author:
Malte Sandner
Nuernberg Institute for Technology
KA-Gebäude GSO
Wollentorstraße
90489 Nürnberg
Germany
E-mail: malte.sandner@th-nuernberg.de

* The experiment was preregistered in the AEA RCT registry, AEARCTR-0008937. This research was reviewed by the Institutional Review Board of the University of Bamberg. We would like to thank Jacopo Bassetto, Stefan Bauernschuster, Michael Grimm, Henning Hermes, Pascal Heß, Shihang Hou, Simon Janssen, Fabian Kosse, Adrian Lerche, Philipp Lergetporer, John List, Susanna Loeb, Frauke Peter, Simon Wiederhold, and participants at various seminars, workshops, and conferences (AEC, AFE, ESPE, Field Experiments in Economics and Business, VfS Jahrestagung, VfS Annual Conference) for their comments and suggestions. We are very grateful to Lisa Leschnig and Ulf-Michael Nützel for coordinating the fieldwork, to members of the Lern-Fair organization, and to all the student assistants and interns who supported our research. We gratefully acknowledge the financial support of the German Research Foundation (DFG, Project No. 518302089) within the DFG Research Unit "Labor Market Transformation: Labor Scarcity, Skill Mismatch, and Policy."

1 Introduction

An extensive literature has demonstrated the effectiveness of in-person tutoring in improving the test scores and grades of underachieving students (Kraft et al., 2024; Nickow et al., 2024; De Ree et al., 2023; Fryer Jr and Howard-Noveck, 2020; Dietrichson et al., 2017; Fryer Jr, 2014). More recently, particularly in response to the disruptions caused by the school closures during the COVID-19 pandemic, evidence has emerged showing that online tutoring can also be effective in improving students' school performance (Carlana and La Ferrara, 2024; Gortazar et al., 2024). Compared to traditional face-to-face tutoring, online tutoring offers more flexibility and potential cost savings, making it an increasingly popular policy tool for supporting low-performing students.[1]

Despite this evidence of the short-term academic gains of tutoring—be it in-person or online—there is still a lack of research on whether these academic improvements translate into smoother school-to-work transitions, better employment prospects, and higher future earnings. Understanding the medium- and longer-term economic returns to tutoring is crucial for assessing its broader societal and economic impact beyond immediate academic gains. Evidence of future returns could further justify investing in tutoring to reduce socioeconomic inequality, secure a skilled workforce, and drive economic growth. These insights help policymakers allocate resources efficiently and design programs with lasting benefits for individuals and society.

However, the question of whether tutoring—be that in-person or online—is effective in improving school-to-work transitions remains open. On the one hand, better grades can increase labor market chances either through improved academic skills or signals to em-

[1]Since the pandemic, governments have invested substantial sums in online tutoring programs. For instance, in July 2022, President Biden launched the American Rescue Plan (ARP) in the US. This is a three-year $122 billion federal program aimed at providing high-quality tutoring, summer learning, and after-school programs, which includes the National Partnership for Student Success (NPSS), a targeted initiative to recruit an additional 250,000 tutors and mentors for students. Similarly, in September 2022, the UK launched a new edition of the National Tutoring Programme, which originally had a budget of £1 billion, offering both face-to-face and online tutoring. In addition, as a response to the pandemic, the German federal government introduced the "Catching up after Corona for children and young people" program, which in 2021 and 2022, allocated €2 billion to fund various measures, including tutoring initiatives (for an overview, see BMFSFJ, 2024; Helbig et al., 2022).

ployers. On the other hand, disadvantaged youth may lack the behavioral and social skills—such as aspirations, grit, and search effort—that are crucial for successful transitions. If tutoring does not foster these skills, they are less likely to improve transitions. Recent empirical evidence supports this view: mentoring and coaching programs that aim at improving academic as well as behavioral and social skills have proven effective in facilitating school-to-work transitions (Resnjanskij et al., 2024; Oreopoulos et al., 2017). In contrast, many of the youth labor market programs that do not focus on behavioral and social skills but solely on retaking qualifications or obtaining certificates seem less successful (Card et al., 2018; Crépon and Van Den Berg, 2016). Since tutoring mainly targets academic skills, the expected effects on school-to-work transitions are ambiguous.

This paper investigates the effects of a one-on-one online tutoring program in Germany on both academic performance and school-to-work transitions. The program, operated by the nonprofit organization *Lern-Fair*, uses a web-based platform to connect university student volunteers with disadvantaged secondary school students in need of academic support. Founded in March 2020 in response to the COVID-19 pandemic, *Lern-Fair* has rapidly scaled up and by July 2022, had provided support to over 23,000 students across Germany. *Lern-Fair* targets low-performing secondary school students in grades 8 to 10 (ages 15–17), a juncture at which students in the German education system typically finish non-academic track schooling and may continue on the academic track or take up vocational training.

We conducted a randomized controlled trial to examine the causal effects of the *Lern-Fair* online tutoring on academic performance and school-to-work transitions. Our experimental sample includes 839 adolescents with low school performance, whose baseline characteristics were collected through a nationwide survey conducted in February 2022. Adolescents in the treatment group received an invitation to participate in the online tutoring program, while adolescents in the control group received no such invitation.

After randomization, we conducted two follow-up surveys. The first, six months after randomization, coincided with the start of the new school year; the second took place around 18 months later. At this point, students, particularly those on the non-academic school track (which finishes after grade 9 or 10), face crucial decisions regarding their educational and career paths. These decisions include whether to move to the academic track, pursue school-based or company-based vocational training, or remain in the post-school transition system, which mainly consists of preparatory programs for vocational training.[2]

In response to our invitation to participate in the online tutoring program, 30% of the students did in fact take part, usually attending weekly sessions for more than three months. Students with a migration background and those from households receiving welfare benefits were significantly more likely to take up the tutoring offer, underscoring the unmet needs for such services among disadvantaged groups. Conversely, uptake was significantly lower among students who were already receiving other tutoring services at baseline, i.e., when the randomization was conducted.

Beyond the lower uptake among students already receiving other tutoring services at baseline, there are additional reasons why the effectiveness of the *Lern-Fair* tutoring program may differ between students with and without prior tutoring. One explanation be the crowding out of more effective in-person tutoring by potentially less effective online tutoring (e.g., Behrman et al., 2024). Additionally, students may become overwhelmed when using too many support services simultaneously. For instance, Phipps and Amaya (2023) show that, particularly among low-performing students, additional training courses reduce student grade point average (GPA) and increase course failure rates. Finally, while prior research suggests that commitment mechanisms in education can increase both the effort invested and performance of low-performing students (Himmler et al., 2019), other

[2] In Germany, vocational training largely follows a two-pronged system, in which apprentices combine a practical work placement in a firm with theoretical education in a vocational college and receive a training salary from the firms they are placed with. Students typically apply directly to firms offering training positions. Vocational training is the most promising career path for students from non-academic track schools and offers substantial returns on the labor market (Fersterer et al., 2008; Piopiunik et al., 2017; Wolter and Ryan, 2011). The German vocational training system is explained in detail in Section 2 and Figure 1.

studies find that the presence of outside options can reduce the effectiveness of such commitments (Ek and Samahita, 2023). If regular tutoring functions as a commitment device that encourages investment in learning effort, the availability of a free alternative may crowd out commitment to paid tutoring—even if the free alternative is not used.[3]

We obtain three key sets of results. First, our analysis six months after the start of the program reveals that online tutoring improves math grades by 0.06 of a standard deviation (SD), although this effect is not statistically significant. However, among students with no other tutoring at baseline, the effect is more pronounced, with a significant 0.14 SD increase. Second, findings from the second follow-up show that, 18 months after the initial invitation, non-academic track students in the treatment group with no other tutoring at baseline were 12 percentage points (pp) more likely to start vocational training and 12 pp less likely to remain in the post-school transition system. The positive effect on vocational training also holds for the whole sample, with a statistically significant increase of 4.7 pp.[4]

Third, when analyzing mechanisms for the effects on school performance and transitions, we find no evidence that the online tutoring affected behavioral or socioemotional skills, such as grit or mental well-being. Similarly, we find no significant effects on student effort, study time, educational aspirations, or search effort, approximated by the number of applications submitted for vocational training. These results align with findings from other in-person and online tutoring programs, which tend to be more effective at improving academic performance than behavioral, social, or other non-academic skills (e.g., Guryan et al., 2023; Carlana and La Ferrara, 2024). However, we do find that, conditional on applying for a vocational training position, students in the treatment group were more successful at securing an apprenticeship contract.

[3]As evidence for the effectiveness of tutoring programs comes mainly from pilot studies with tight control over the use of additional services, the interaction between tutoring and other educational programs—including other tutoring programs—has been largely overlooked in the literature (see Nickow et al., 2024).

[4]All results remain robust to the use of augmented inverse probability weighting combined with lasso for selecting relevant control variables and across various outcome definitions.

Our findings contribute to at least three strands of literature. First, we contribute to the broader literature on the effects of tutoring programs—whether conducted online or in-person. This literature has primarily concentrated on the positive effects of tutoring on short-term school performance (for an overview, see Kraft et al., 2024; Nickow et al., 2024). Our study is the first to extend this research by exploring the impact of tutoring on labor market transitions, demonstrating that it can have positive effects on the career paths of low-performing students and possibly also on their life trajectories. This finding is important as many studies show that contemporaneous test-score gains from education policies diminish substantially over time (Gilraine and Penney, 2025).

Second, our study contributes to the emerging literature on the effectiveness of online tutoring programs, a field that has only recently gained attention. Before the COVID-19 pandemic, research focused almost exclusively on in-person tutoring, with little consideration of remote formats. During the pandemic, when schools were closed and teaching moved online, several studies examined the impact of online tutoring (e.g., Zoido et al., 2024; Hassan et al., 2024; Hardt et al., 2022; Carlana and La Ferrara, 2021). However, research on the effects of online tutoring in the post-pandemic context remains limited. Notable contributions comparable to our work include studies by Carlana and La Ferrara (2024), Gortazar et al. (2024), Kraft et al. (2022), and Fesler et al. (2023), which evaluated pilot programs in specific regional settings. While these studies generally report positive effects on test scores, their findings are not entirely consistent, underscoring the need for further research in this area.[5]

Besides expanding the scope beyond test scores, our study contributes to this emerging literature by providing evidence on the effectiveness of online tutoring through a nationwide experiment based on a fully scaled-up program. Unlike participants in pilot projects, the students in our study received a program that had already been delivered to over 23,000 students, reducing concerns about the diminished effectiveness that can arise when scal-

[5] Carlana and La Ferrara (2024) and Gortazar et al. (2024) find that online tutoring with compulsory participation improve math test scores by between 0.20 and 0.26 SD and grades by 0.49 SD. In contrast, Kraft et al. (2022) find positive but statistically insignificant effects of only 0.07 SD on math test scores.

ing up smaller pilot projects (Andersen and Hvidman, 2024; DellaVigna and Linos, 2022; Al-Ubaydli et al., 2017). Further, by drawing our sample from a nationwide survey, we reduce the risk of spillover effects from the treatment group to the control group, which is often a limitation in studies conducted within schools.

Lastly, we add to the literature on effective strategies for engaging disadvantaged, low-performing students in nonmandatory opt-in support programs. Previous studies have used similar strategies to ours to invite students to different types of in-person (offline) support programs. These studies often report lower take-up rates with only 15–20% of the invited students actually participating. Moreover, disadvantaged and struggling students are typically less likely to participate than their more engaged and higher-achieving peers (Munoz-Herrera, 2024; Robinson et al., 2022; Evans et al., 2020; Pugatch and Wilson, 2018). A likely explanation for this lower take-up is that the perceived costs of participating in certain support programs, such as in-person tutoring, are prohibitively high for disadvantaged students.

In contrast, the higher take-up rate of 30% in our study suggests that online services may be more appealing to this group. This increased attractiveness could be attributed to the high flexibility of online tutoring, which lowers financial and time barriers to participation, as well as to the reduced risk of stigmatization, as sessions are conducted privately on students' home computers, away from the eyes of their school peers. Our finding that the disadvantaged students in our study were more likely—and not less, as other studies show—to participate in the program further supports this argument.

The remainder of this paper proceeds as follows. Section 2 provides an overview of the intervention and the institutional setting. Section 3 lays out the study design and presents descriptive statistics of the sample. Section 4 explains the methods used for the empirical analyses. Section 5 presents the main results, along with analyses of the channels. Section 6 concludes.

2 Institutional Background

2.1 The Lern-Fair online tutoring program

The nonprofit organization *Lern-Fair* provides the online tutoring evaluated in this study. A group of university students founded the organization in March 2020, during the COVID-19 pandemic, to support low-performing students while schools were closed. During and after the pandemic, *Lern-Fair* has expanded rapidly and by July 2022 had successfully supported over 23,000 students. As *Lern-Fair* connects tutors and students online, it has no geographical restrictions, and as such serves students all over Germany.[6]

Lern-Fair recruits tutors from among university students who are pursuing teaching degrees and screens them to ensure they are qualified for tutoring. It also provides guidelines and structured frameworks to support the development of effective tutoring relationships, though the tutors are not obligated to follow these. In the event that tutors face challenges in their teaching, they can access support materials, participate in peer exchanges, and consult directly with the *Lern-Fair* team. Tutors work on a voluntary basis and receive a practical training certificate.[7]

Lern-Fair matches the tutors with low-performing students in need. To do this, *Lern-Fair* provides a web-based platform where tutors and students are assigned to each other based on the subject and grade of the student. The pair arranges their own meetings, deciding on the frequency and duration without *Lern-Fair* involvement. Tutors and students meet via video calls on their private devices. To reach out to students in need, *Lern-Fair* uses various channels: online (website, social media, mail campaigns), offline (flyers, posters, educational fairs, multiplier events), and by cooperating with institutions (university collaborations, NGO partnerships). The focus of the organization is explicitly on educational justice, meaning they offer their tutoring only to educationally disadvantaged students

[6] *Lern-Fair* is mainly funded through donations and public funds. It has won several prizes for its civic engagement.

[7] University students pursuing a teaching degree could use this certificate to gain credits as part of their required training practice.

who need help with learning and whose family is unable to support them. During the registration process, *Lern-Fair* verifies whether a student belongs to this target group.[8]

2.2 The German education system

The online tutoring program evaluated in this study focuses on students attending grades 8 to 10. Since academic tracking usually starts at grade 5 in Germany, tutored students are attending the academic track or one of two non-academic track school types.[9]

The non-academic secondary school tracks include lower secondary school (*Hauptschule*) and intermediate secondary school (*Realschule*), while the academic track schools are usually upper secondary schools (*Gymnasium*). In addition, there are comprehensive schools (*Gesamtschule*), which allow students to obtain any of the three types of school leaving qualification and at which some of the students will also obtain an upper secondary (academic track) qualification. The qualification received by students in the academic track school enables them to enroll at university after grade 12 or 13. Figure 1 summarizes the features of the German education system.[10]

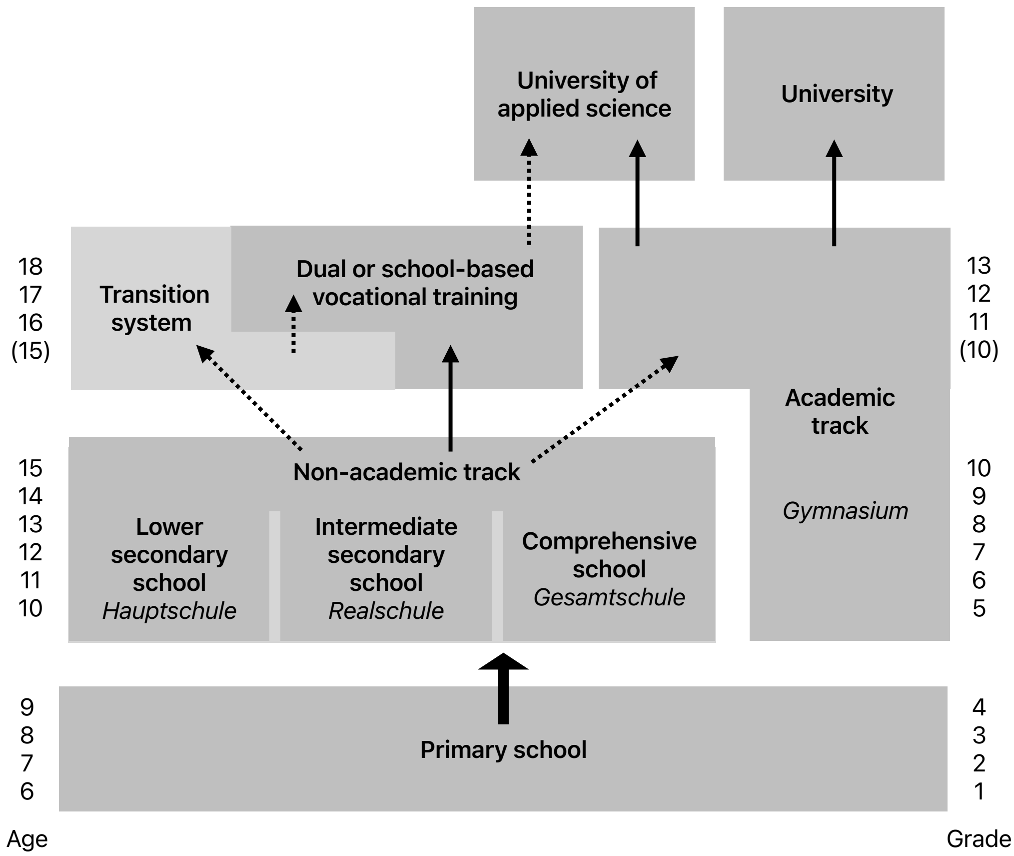
General education in the non-academic tracks lasts until grade 9 and 10, after which students may switch to the academic track if their grades are good enough or, more commonly, will transition to vocational training and enter the labor market. German vocational training typically follows a dual system, in which apprentices combine a practical work placement in a firm with theoretical education in a vocational school and receive a training salary from the firm they are placed in. Students typically apply directly to firms providing training positions.

[8] Students can only access the registration page if they click to confirm that at least two of the following three conditions apply: 1) they need help with school, 2) their family cannot help them with their homework, 3) their family cannot afford to pay for private tutoring.

[9] In the federal states of Berlin, Brandenburg, and Mecklenburg-West Pomerania, tracking starts in grade 7.

[10] For further information about the school system in Germany, see, for example, Francesconi et al. (2010).

Figure 1: German education system



Notes: The figure shows the structure of the German education system starting from the first mandatory phase (primary school). The scale on the right-hand side indicates school grades and on the left-hand side age. Solid arrows indicate typical tracking and dashed arrows indicate a switch between tracks.

Obtaining a vocational training qualification offers high labor market returns over the life course[11] (Piopiunik et al., 2017; Wolter and Ryan, 2011; Fersterer et al., 2008) and strongly reduces the risk of unemployment, with unemployment rates comparable to those of university graduates, according to official figures from the German Federal Employment Agency.[12] Indeed, previous studies show that vocational education is beneficial for students close to dropping out of education (e.g., Matthewes and Ventura, 2022).

[11]For example, according to the German Federal Statistical Office, in April 2024, full-time employees with a completed vocational training qualification earned, on average, €3,870 euros per month, which was almost €900 more than employees without a vocational qualification.

[12]The unemployment rate in Germany in 2023 was 3.2% for those with a vocational qualification, compared to 2.5% for those with an academic degree, and 20.8% for those with neither a vocational nor academic qualification.

If students from the non-academic track cannot switch to the academic track and do not find a vocational training position, they can either voluntarily repeat the final grade in a non-academic track school or continue in the so-called transition system. The transition system in Germany is designed for school-leavers who are not successful in securing a vocational training placement and are mostly still subject to compulsory education. This phase usually includes preparatory classes aimed at helping participants gain the basic skills needed for vocational training or to enter the labor market. Although the transition system may benefit young people with particularly poor prospects, it does not improve the chances of success for those with more favorable conditions. For half of the participants, a transition system program does not lead to vocational training but to another transition program or—when they turn 18—low-skilled employment, unemployment, or inactivity (Ehlert et al., 2018). Additionally, all students who enter the transition system or voluntarily repeat a grade start employment one year later than those who start vocational training immediately, thus reducing overall lifetime income for those students. Hence, continuing in the transition system or voluntarily repeating a grade in a non-academic track school are less preferred outcomes in comparison to starting vocational training.

3 Study Design

3.1 Sample, randomization, and intervention

To analyze the effectiveness of the *Lern-Fair* online tutoring program, we conducted a randomized field experiment. We drew our experimental population from the online panel study "Corona & Du" (*CoDu*), which started in November 2020. This panel study surveys families with children aged 10 to 17 throughout Germany, with an oversampling of low-income and welfare-dependent families.[13] The study collected a range of socioeconomic and academic information from both children and parents, including household income, type of education of children, school grades, and children's educational aspirations. One

[13] For the purposes of this study, we define low-income households as those below the 60th percentile of the income equivalence scale.

major advantage of using participants from an existing panel study (compared to recruiting students via schools) is that they (and their schools) are distributed across various federal states in Germany, minimizing the risk of spillover effects between the treatment and control groups.[14]

Our baseline survey, which was part of the *CoDu* panel survey, took place between February and March 2022 with the participation of 2,225 students and their parents. To be eligible for the online tutoring program, students had to meet five conditions. First, they had to demonstrate willingness to participate in future follow-up surveys (panel readiness). Second, they had to attend a general school. Third, one grade in math, German, or English (as per the school report received after the first half of the 2021/2022 school year) had to be at most a 3 (equivalent to a C in the US grading system).[15] Fourth, students had to express at least some interest in tutoring.[16] Finally, students were required to complete the baseline survey. Appendix Table A1 presents an overview of the eligibility criteria and how many students were excluded based on each criterion. Of the 2,225 young people who participated in the baseline survey, 839 fulfilled the above conditions and entered the randomization process directly after they had completed the survey. These 839 students were stratified based on household welfare receipt and gender. Within each stratum, pairs were formed according to enrollment dates. The students were then pair-wise randomly assigned to either a treatment or a control group.

Students assigned to the treatment group received an invitation to participate in the *Lern-Fair* online tutoring program. The invitation was displayed on the exit screen of the online survey and later reiterated via invitation emails and letters sent to their home. Appendix Figure A2 presents the original invitation (in German) displayed on the survey

[14] Appendix Figure A1 shows the geographic spread of the treatment and control group and Appendix Table A2 presents the sample shares by German federal states.

[15] In the German school system, grades range from 1 to 6, with 1 being the best. Therefore, participants in the randomization process were those who received grade 3, 4, 5, or 6 in at least one of the subjects.

[16] Two questions were included to assess the respondent's interest in tutoring. The first asked whether participants agreed that tutoring could generally help improve school outcomes. The second inquired whether participants would be willing to participate in online tutoring provided by students, with response options being "definitely," "probably," or "don't know".

participant's screen. To join the tutoring program, students were required to register on the *Lern-Fair* website, which they could access via a personalized link in the invitation on the exit screen. In addition, students were sent the invitation by email, allowing them to register at a later time.

Using the link in the invitation, students could register after providing personal data such as their name, surname, email address, type of school attended, the grade they were in, the language in which they were able to communicate fluently, and subjects in which they would like to be tutored. Finally, the students had to agree to the study's data protection rules. The personalized link also enabled *Lern-Fair* to differentiate between *CoDu* study participants and those seeking tutoring in the regular way. *Lern-Fair* granted *CoDu* participants preferential access to the online tutoring, i.e., instant matching with tutors and no waiting list. After the registration had been completed, *Lern-Fair* matched registered students with available tutors using a predefined algorithm that paired them based on the subject offered and needed, as well as the student's grade as a proxy for difficulty. This process was run at least once a week to find the most suitable tutors for all the participating students.

To increase treatment take-up, we sent a series of reminders to students throughout the first month after the initial offer (Appendix Figure A3 shows the timeline of the reminders). The first and third reminders were emailed directly to students who had provided their email address in the *CoDu* survey. The second reminder was sent as a letter addressed to the student's parents. There were no disadvantages for families or students who did not register on the *Lern-Fair* website or who registered at a later point in time.

Students randomized into the control group received neither targeted information about the free online tutoring offered by *Lern-Fair*, nor easy and preferential access to the online tutoring program. However, we cannot exclude the possibility that some control group students found out about the *Lern-Fair* platform by other means and received tutoring during the treatment period through *Lern-Fair* or other tutoring providers.

3.2 Randomization balance and timeline

Table 1: Balance test

	Treatment mean	Control mean	Mean difference	p-value
	(1)	(2)	(3)	(4)
Panel A: Sociodemographic				
Female	0.618	0.620	0.001	0.965
Age	17.002	16.993	-0.010	0.903
Former West Germany	0.810	0.806	-0.004	0.890
Migration background	0.342	0.378	0.036	0.279
Welfare recipient	0.404	0.407	0.003	0.932
Panel B: School-related				
Academic track	0.580	0.620	0.040	0.237
Grade	10.271	10.325	0.055	0.464
Other tutoring	0.211	0.213	0.002	0.957
Math grade	-0.128	0.000	0.128*	0.068
German grade	0.050	0.000	-0.050	0.471
English grade	0.017	0.000	-0.017	0.811
Number of observations	*421*	*418*		

Notes: The table presents the balance test for the treatment and control groups. Variables are divided into sociodemographic characteristics, presented in Panel A, and school-related covariates, presented in Panel B. Columns 1–2 show the mean value of these variables for the treatment group (column 1) and the control group (column 2), respectively. Column 3 shows the difference in means, while column 4 displays the corresponding p-value. The variable "female" is equal to 1 if a student is a girl, and 0 otherwise. "Age" is a continuous variable measured on a scale from 15 to 21. The variable "former West Germany" is equal to 1 if a student lives in the former West Germany, excluding Berlin, and 0 otherwise. The variable "migration background" is equal to 1 if at least one parent was born outside Germany, and 0 otherwise. The variable "welfare recipient" is equal to 1 if the household receives welfare benefits, and 0 otherwise. Academic track includes students attending higher track schools (*Gymnasium* and *Gesamtschule* from grade 11 onward), compared to non-academic track which consists of students attending lower track (*Hauptschule*) and medium track (*Realschule* and *Gesamtschule* up to grade 11) schools. "Grade" is a continuous variable which takes a value from 8 to 13. The variable "other tutoring" is equal to 1 if a student is already receiving some form of tutoring, and 0 otherwise. Baseline grades in math, German, and English are the respective school grades received before the randomization. These are standardized by subtracting the control group mean and dividing by the control group standard deviation. Grades are measured such that higher values indicate better grades.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1 indicates that student characteristics were balanced between the treatment and the control group at baseline. Only the math grade averages show a statistically significant difference, prompting us to control for baseline math grades in the subsequent analyses (see Section 4.2). Further, Appendix Table A3 and Appendix Figure A4, Panel A, demonstrate that besides the student characteristics, parental and household characteristics, as well as the date of the baseline survey response, which may be a proxy for a student's procrastination, are also balanced between the treatment and control groups.

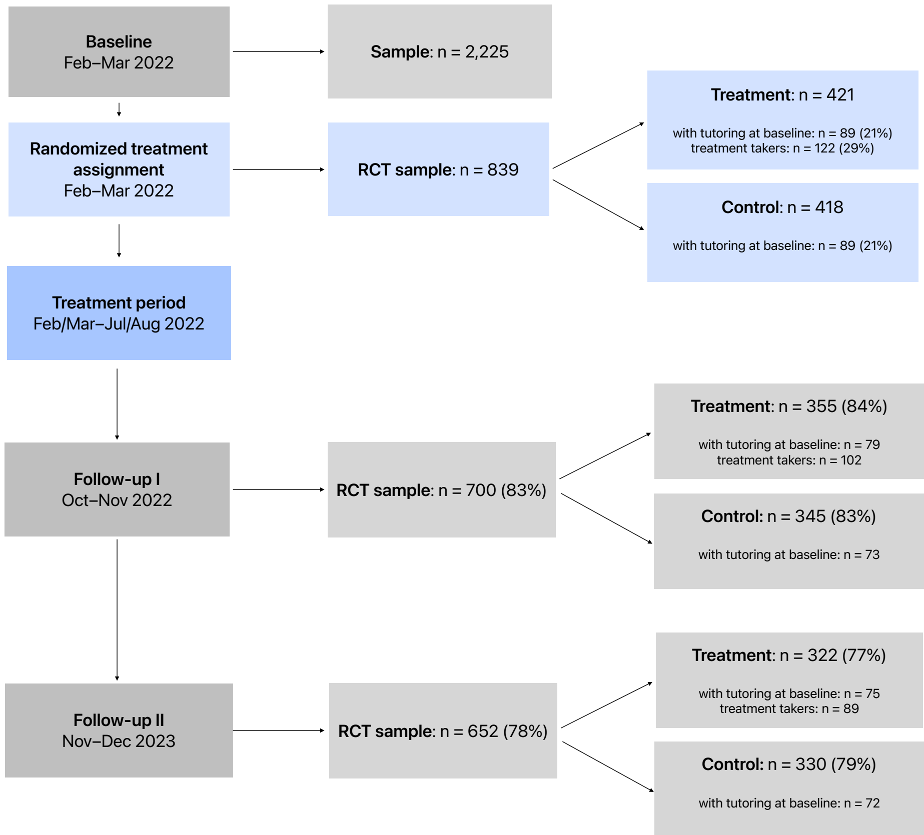
The sociodemographic characteristics show a high share of students with a migration background (35%) and living in households that receive welfare benefits (40%). Of the participating students, 58% attend an academic track school and the rest go to one of the three non-academic track schools. Appendix Table A4 provides baseline characteristics of the students by school track, highlighting significant differences in characteristics, particularly in welfare receipt, across tracks. Notably, 21% of all students had been using another tutoring service at baseline, which implies that some tutoring services are also available in the control group.[17]

To analyze the treatment effects of the *Lern-Fair* online tutoring program, we conducted two follow-up surveys. The first survey (follow-up I) took place in October and November 2022, during the new school year, after the treatment group had been invited to participate in *Lern-Fair*. The second survey (follow-up II) took place in November and December 2023. Both follow-up surveys included questions regarding school performance, aspirations, and socioemotional skills. In addition, in the follow-up II survey, we strongly focus on the transition from secondary to post-secondary schooling.

Figure 2 presents a detailed timeline of the study, along with participation rates and the number of students receiving tutoring at baseline. In follow-up I, the response rates were approximately 83% in both the treatment and control groups. In follow-up II, response rates were around 76% for both groups. Appendix Table A6 shows that the participation

[17]These services may include various forms of paid and unpaid tutoring provided by peers, schools, or companies.

Figure 2: Experimental timeline



Notes: The figure shows the experimental timeline of the study together with the number of participants at each stage and the response rates in the surveys. The first follow-up survey was conducted around eight months after the intervention and the second around 20 months after the intervention.

rates do not differ significantly between the two surveys. Importantly, attrition does not appear to be selective between the two groups, as neither baseline school grades nor sociodemographic characteristics differ significantly between the treatment and control groups in the follow-up surveys. Appendix Tables A5 and A7 present the distributions of the baseline characteristics between the two groups for follow-up I and follow-up II.

3.3 Take-up and implementation

Table 2 shows the take-up rates for the *Lern-Fair* online tutoring among those students allocated to the treatment group. Of the 421 students who received the invitation, 122 (29%) registered on the platform. Appendix Figure A5 illustrates that most students registered directly after the invitation or after receiving a reminder. According to the tracking data of the *Lern-Fair* registration platform, 80% of students registered for math tutoring, while the remaining 20% registered for English or German.

Lern-Fair collected information on tutoring duration and frequency from tutors working with students in the experimental sample.[18] Appendix Table A8 shows that student–tutor interaction was generally high. Panel A indicates that only 11% of pairs met for less than one month, while nearly 80% continued tutoring for more than three months. Panel B shows that two thirds of pairs met at least once a week, with more than half meeting multiple times per week.

An analysis of student characteristics related to registration reveals two key findings. First, the share of students with a migration background as well as the share of students from households receiving welfare benefits are each 19 pp higher in the group that registered for *Lern-Fair* than the group of students who did not register. This corresponds to take-up rates of 39% among the students from households receiving welfare (versus 22% among those who do not receive welfare) and of 40% among students with a migration

[18]The tutor reports cannot be linked to the students' survey data. Participation in the tutor survey was voluntary, which explains why not all tutors who provided tutoring to students in the experimental sample took part.

Table 2: Treatment take-up

	All treated	Tutoring takers	Tutoring non-takers	Difference	p-value
	(1)	(2)	(3)	(4)	(5)
Panel A: Sociodemographic					
Female	0.618	0.680	0.593	-0.088*	0.093
Age	17.002	17.172	16.933	-0.239*	0.051
Former West Germany	0.810	0.826	0.803	-0.024	0.576
Migration background	0.342	0.475	0.288	-0.188***	0.000
Welfare recipient	0.404	0.541	0.348	-0.193***	0.000
Panel B: School-related					
Academic track	0.580	0.598	0.572	-0.026	0.619
Grade	10.271	10.213	10.294	0.081	0.472
Other tutoring	0.211	0.123	0.247	0.125***	0.004
Math grade	-0.128	-0.237	-0.084	0.153	0.169
German grade	0.050	-0.040	0.086	0.126	0.234
English grade	0.017	-0.011	0.028	0.038	0.726
Number of observations	*421*	*122*	*299*		

Notes: The table presents the differences between the group of treatment takers and treatment non-takers. Variables are divided into sociodemographic characteristics, presented in Panel A, and school-related covariates presented, in Panel B. Column 1 shows the mean value of these variables for the whole treatment group, while columns 2–3 show this information for treatment takers and treatment non-takers, respectively. Column 4 shows the difference in means, while column 5 shows the corresponding p-value. The variable "female" is equal to 1 if a student is a girl, and 0 otherwise. "Age" is a continuous variable measured on a scale from 15 to 21. The variable "former West Germany" is equal to 1 if a student lives in the former West Germany, excluding Berlin, and 0 otherwise. The variable "migration background" is equal to 1 if at least one parent was born outside Germany, and 0 otherwise. The variable "welfare recipient" is equal to 1 if a household receives welfare benefits, and 0 otherwise. Academic track includes students in higher track (*Gymnasium* and *Gesamtschule* from grade 11 onward) schools, compared to non-academic track which consists of students in lower track (*Hauptschule*) and medium track (*Realschule* and *Gesamtschule* up to grade 11) schools. "Grade" is a continuous variable which takes a value from 8 to 13. The variable "other tutoring" is equal to 1 if a student is already receiving some form of tutoring, and 0 otherwise. Baseline grades in math, German, and English are the respective school grades received before the randomization. These are standardized by subtracting the control group mean and dividing by the control group standard deviation. Grades are measured such that higher values indicate better grades.
*** p<0.01, ** p<0.05, * p<0.1.

background (versus 23% with no migration background), respectively.[19] Second, the share of students who had already received tutoring at the time of the baseline survey differs significantly between those taking up online tutoring (12 pp) and the non-takers (25 pp).

[19] These take-up rates are calculated by dividing the number of registered students by the total number of students with this specific characteristic.

This corresponds to a take-up of 18% for those with previous tutoring, i.e., a much lower take-up rate than for those without previous tutoring (32%). Other characteristics are not or only slightly related to take-up; girls and older students are slightly more likely to use the tutoring program.

These findings on take-up rates present an initial set of important results. First, they reveal a significant demand for academic support among disadvantaged, low-performing students open to tutoring because the relatively high take-up rates resulted from a minimal intervention, i.e., a short and simple invitation letter. This high demand is surprising, given that students might access *Lern-Fair* through other channels and obtain free in-person tutoring via alternative routes,[20] and signals informational frictions. These may stem from a lack of awareness about tutoring alternatives, unclear communication regarding the benefits of or eligibility for different tutoring programs, or the absence of nudges to help students navigate the available resources.

Second, students from disadvantaged backgrounds (migration background and households receiving welfare), in particular, are attracted by the invitation for online tutoring despite there often being low take-up of support programs among these students (e.g., Robinson et al., 2022). Online tutoring may be attractive to this group in particular because of the flexibility, the virtually nonexisting financial or time costs, and the low risk of stigmatization. Our results on take-up indicate that invitations to online tutoring programs for individuals in need can be a successful way of providing services to disadvantaged, low-performing students who may otherwise lack access to or resist support services.

Third, students who were already receiving another type of tutoring at the time of randomization were less likely to take up the *Lern-Fair* offer. Nevertheless, some students in the treatment group enrolled in *Lern-Fair* despite using another tutoring service. This may have led to a substitution effect, with one service replacing the other; a "double service" effect, with students using both services simultaneously; or a crowding-out effect,

[20] School-based support programs may provide in-person tutoring for low-performing students. In addition, students from households receiving welfare benefits may be eligible for tutoring vouchers under the German federal government's education package.

with students reducing their use of alternative tutoring simply because free online tutoring was offered—even if they ultimately did not use the latter. Each of these mechanisms could reduce the overall effectiveness of the *Lern-Fair* program. Given these expectations of lower effectiveness among students already receiving tutoring at the time of randomization, we present our main findings separately for students with and without other tutoring at baseline, even though this split was not preregistered.[21]

3.4 External validity

One strength of our study is that we examine a project implemented at a large scale, which reduces the risk of effects fading out when the program is scaled up (List, 2022). However, two factors may limit the generalizability of the results from our experiment to the regular *Lern-Fair* sample: (1) the experimental sample might have different characteristics to the regular *Lern-Fair* population, or (2) the implementation of the tutoring (e.g., tutors' motivation or tutoring content) during the experiment may have differed from the regular *Lern-Fair* program.

To address these concerns, Table 3 presents two comparisons. The upper part of the table compares students' place of residence, migration background, academic track, and grade level in the experimental sample with those of students participating in the regular *Lern-Fair* program. Data for both groups are drawn from the *Lern-Fair* registration platform, where students provide this information when they enroll. The lower panel of Table 3 compares tutors' reports on the implementation of tutoring sessions in the experimental setting with those from tutors offering regular *Lern-Fair* tutoring.[22] The first column in the upper panel of Table 3 displays information on compliers in the treatment group. The second column refers to the population of regular *Lern-Fair* users during the experimental period (February–July 2022), excluding individuals from the experimental sample. Comparing the treatment compliers with the regular *Lern-Fair* participants, we

[21] Appendix Table A9 shows that the baseline characteristics are also balanced between the treatment and control group for the subgroup without other tutoring at baseline. In the subsequent analyses, we refer to these groups as "with tutoring at baseline" and "with no tutoring at baseline."

[22] Questions on duration and frequency are not available for the regular sample.

observe broadly similar characteristics. In both groups, around 80% of students come from the western Germany, indicating that the nationwide sampling in the experimental setting is comparable to the regular *Lern-Fair* program. A substantial share of students in both samples has a migration background (between 40 and 50%), and more than half attend an academic track. Although the experimental sample includes students in slightly higher grades, on average, *Lern-Fair's* general focus is also on supporting students who are reaching the end of secondary school.

The lower part of Table 3 shows, first, that the tutoring content is largely comparable across both settings. Only the share of sessions focused on covering and revising lesson content is higher in the regular *Lern-Fair* program. Second, tutors' assessments of students' cooperativeness and motivation are also similar between the two groups, suggesting that the students themselves are comparable, as tutors perceive them to be equally cooperative and motivated.

Overall, there is no indication that selection into the sample differs systematically between the two settings, and the parameter estimated in the experiment appears close to the effect of the regular *Lern-Fair* program. However, this does not imply that all disadvantaged students in need would benefit equally. As participation in the *Lern-Fair* program is voluntary, students who choose to participate are probably more motivated than the average student. However, this holds for both participants in the regular *Lern-Fair* program and those from our experimental setting, and as such, we would not expect any differences in motivation between these two groups.

Table 3: Sample comparison: Experimental sample vs. regular *Lern-Fair*

	Experimental sample	**Regular *Lern-Fair***
I. Student characteristics	(1)	(2)
Panel A: Sociodemographic		
Former West Germany	0.83	0.85
Migration background	0.48	0.39
Panel B: School-related		
Academic track	0.60	0.52
Grade level	10.21	8.77
Number of students	*122*	*611*
II. Tutor ratings		
Panel A: Tutoring content (0–1)		
Covering and revising lesson content	0.74	0.91
Exam and test preparation	0.63	0.65
Completing homework	0.59	0.55
Mentoring	0.13	0.17
Panel B: Student's attitudes[a]		
Student's cooperativeness (1–5)	4.26	4.30
Student's motivation (1–100)	72.29	79.33
Number of tutors[b]	*32*	*1,100*

Notes: Part I of the table shows the comparison between students in our experimental sample who took up *Lern-Fair* tutoring (column 1) and regular *Lern-Fair* participants in the treatment period [February–July 2022] (column 2). Panel A presents the sociodemographic characteristics and Panel B the school-related characteristics of the students. Part II of the table shows the comparison between *Lern-Fair* tutors involved in the experiment (column 1) and the overall pool of *Lern-Fair* tutors (column 2). Panel A presents the answers to the questions on tutoring content. Panel B presents the tutors' perspective on the students' attitudes during the tutoring process.

[a]In the survey for the regular *Lern-Fair* tutors, the student's cooperativeness (item 1 in Panel B) was measured on a 1–7 scale (mean outcome: 5.95), while the student's motivation (item 2 in Panel B) was measured on a 1–4 scale (mean outcome: 3.38). The results presented in the table stem from the linear transformation of these scales.

[b]Deviating from these numbers, in the experiment (column 1), 23 tutors answered a question on their student's cooperativeness, while 35 answered a question on their student's motivation.

4 Analysis plan

4.1 Outcomes

To examine the effectiveness of the online tutoring, we analyze preregistered school performance and labor market transition outcomes, as well as measures on students' behavioral and socioemotional skills (AEA RCTR-0008937).

To measure school performance, we preregistered two main outcomes: grades in math and grade retention. We focus on grades in math for two reasons. First, math grades are directly associated with higher wages and better employment opportunities in Germany (Resnjanskij et al., 2024). Second, the registration for math tutoring was significantly higher (80% of registered students) than for all other subjects. All grades are self-reported by the young participants and relate to the final grades from school years 2021/2022 and 2022/2023, respectively. All grades are standardized by subtracting the control group mean and dividing them by the control group standard deviation separately in each of the survey waves. Grades were recoded so that higher values indicate better grades.

We preregistered involuntary grade retention or grade repetition as an outcome to measure school performance as it is a popular remedial practice in the German school system.[23] However, studies show that grade retention has limited positive effects on school performance and, in the long term, leads to earnings losses due to reduced labor market experience (ter Meulen, 2023; Cockx et al., 2019). Therefore, it would be beneficial if tutoring were to reduce grade retention. To measure grade retention, we asked the students in both survey waves whether they had repeated a grade.

To measure the preregistered outcome school-to-work transition, we asked the students in follow-up surveys I and II about their current status. Specifically, we inquired whether: (i) they had moved to the academic school track, (ii) they had started vocational training, or (iii) they did not belong to either group (i) or (ii). The last group (iii) is further divided

[23] ter Meulen (2023) shows that 10.6% of 15-year-old students across OECD countries have repeated a grade at some point, with the rate being around 15% in Germany.

into two categories. The first category includes students who are either in the transition system or still attending a non-academic school, mostly due to voluntarily repeating the final grade. These students typically lacked the grades required to move to the academic track. The second category includes students who are over the age of 18 and have left the education system. These individuals are mainly employed as unskilled workers or unemployed.

Close personal mentoring relationships appear to be effective in changing the behavioral and social skills of disadvantaged youth (Kosse et al., 2020) and these changes offer important channels for improved school-to-work transitions (Resnjanskij et al., 2024). To investigate whether similar improvements can be achieved through less personal—i.e., online—tutoring relationships, and whether these also act as key channels, we assess the impact of *Lern-Fair* on these skills.

In follow-up survey I, we measured school-related inputs and effort, including the average time students spent learning after school and their active participation in lessons on the previous school day. We also used survey batteries to measure students' mental well-being and grit. In addition, we used students' educational and occupational aspirations, measured by questions on whether a student plans to pursue vocational training or a university degree.

Finally, in addition to grades in math, we also asked for grades in German (national language) and English (most common foreign language) to analyze spillover effects from tutoring in math on other subjects.

4.2 Estimation

Our estimation strategy follows standard procedures. In the first step, we estimate the intention-to-treat (ITT) effects. The regression specification is as follows:

$$Y_i = \alpha + \beta\, treat_i + \lambda\, X_i + \epsilon_i \tag{1}$$

where Y_i is the outcome of interest, $treat_i$ indicates the initial assignment to treatment or control group, X_i includes strata (household welfare receipt and gender), several predetermined characteristics, and the baseline grades in math, German, and English, and ϵ_i is the error term.[24]

To further explore whether selective attrition may influence our results, we apply augmented inverse probability weighting (AIPW) in combination with lasso to select relevant control variables from those listed in Appendix Table A18 (e.g., Chernozhukov et al., 2018). To explain the approach, let $Y_i(d)$ denote the outcome under the treatment status $d \in \{0, 1\}$, with the expectation modeled as $m_d(X_i) = E(Y_i(d)|X_i)$. Let D_i be a binary variable denoting the treatment assignment, whose propensity score is $p(W_i) = \Pr(D_i = 1|W_i)$. W_i is a set of control variables potentially equal to X_i. Let $\widehat{m}_d(X_i)$ and $\widehat{p}_d(W_i)$ be the estimated models of $m_d(X_i)$ and $p(W_i)$. The AIPW estimator is then given by:

$$\widehat{ATE}_{AIPW} = \frac{1}{n}\sum_{i=1}^{N}\left\{\left[\frac{D_i Y_i}{\widehat{p}(W_i)} - \frac{(1-D_i)Y_i}{1-\widehat{p}(W_i)}\right] - \frac{(D_i - \widehat{p}(W_i))}{\widehat{p}(W_i)(1-\widehat{p}(W_i))} \times [(1-\widehat{p}(W_i))\widehat{m}_1(X_i) + \widehat{p}(W_i)\widehat{m}_0(X_i)]\right\} \quad (2)$$

$\widehat{m}_d(X_i)$ and $\widehat{p}_d(W_i)$ are specified as linear and logit functions, respectively. X_i and W_i are selected through a lasso procedure. The strata variables are always included in $\widehat{p}_d(W_i)$. We will present the unweighted and weighted (AIPW) specifications in the results section, although they yield very similar coefficients.

[24] We also estimate regressions with pair fixed effects. The results are similar, although the sample size is smaller due to attrition, which means that pair information is not available for all participants in the follow-up surveys.

5 Results

5.1 Follow-up I: School performance

This section investigates the effects of the tutoring invitation on math grades and grade retention in fall 2022, approximately six months after randomization.

Figure 3 shows the intention-to-treat (ITT) effects on grades in math and on grade retention for the whole sample (first bar), for the sample with tutoring at baseline (second bar), and for the sample with no tutoring at baseline (third bar).[25]

The effects for the whole sample are positive but not statistically significantly different from zero. If we only concentrate on the students who were not receiving tutoring at baseline, grades in math significantly improve through tutoring. The effects of tutoring for those students who had received tutoring at baseline are diminished. A similar pattern is visible for grade retention, which, overally, becomes less likely due to tutoring but with a stronger effect for those students with no tutoring at baseline.

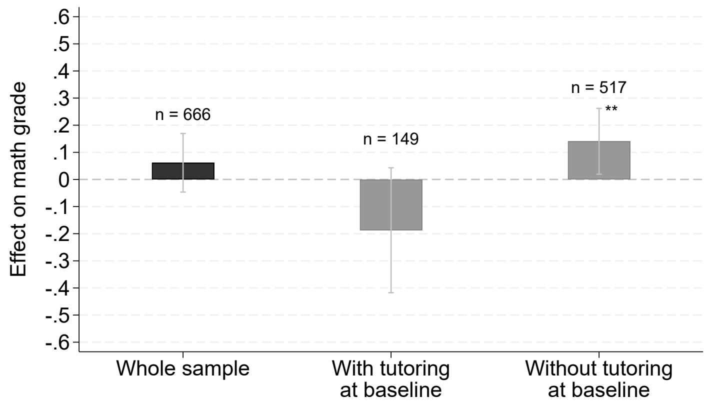
Table 4 shows the point estimates for the two outcomes for the whole sample (Panel A), the students with tutoring at baseline (Panel B), and only for those students with no tutoring at baseline (Panel C). In the first rows of each panel, we control for gender, household welfare recipient status, age, living in former West Germany, migration background, school type attended, and baseline grades in math, German, and English. In the second row of each panel, we show the results from the AIPW specification with the control variables selected by the lasso method (Appendix Table A19 presents the control variables selected by the lasso method).

The results indicate that the math grades did not improve statistically significantly, with a 0.06 SD improvement in the unweighted specification for the whole sample. For those students without prior tutoring at baseline, there is a statistically significant improvement

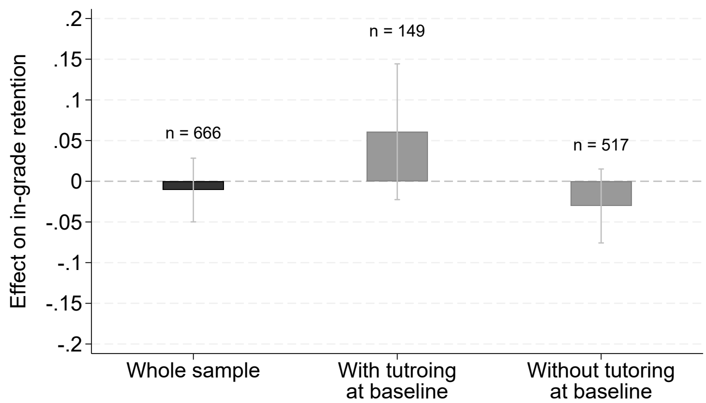
[25]In Figure 3, the sample size for the whole sample is 666 instead of the 700 indicated in Table A6. This difference arises because 34 individuals have missing data in one of the outcomes and are excluded from the analysis sample.

Figure 3: Effects of online tutoring on school outcomes depending on tutoring at baseline

Panel A: Math grade average



Panel B: Grade retention



Notes: The figure shows the estimated effects of the treatment (online tutoring offer) on the school outcomes, together with 95% confidence intervals. In Panel A, the outcome is grade average in math. In Panel B, the outcome is in-grade retention (repetition). The effects are estimated within the whole randomized sample (first black bars with n = 666), within the subsample of students who had tutoring at baseline (second gray bars with n = 149), and within the subsample of students who had no tutoring at baseline (third gray bars with n = 517). The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. The math grade average is standardized by subtracting the control group mean and dividing by the control group standard deviation. It is measured such that higher values indicate better grades. In-grade retention is a binary outcome equal to 1 if a student has to repeat the grade, and 0 otherwise. All regressions use robust standard errors and control for gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: Effects of online tutoring on school outcomes

	Math grade average	Grade retention
	(1)	(2)
Panel A: Whole sample		
Tutoring effects (unweighted)	0.061	-0.011
	(0.055)	(0.020)
Tutoring effects (AIPW)	0.046	-0.008
	(0.053)	(0.020)
Control group mean	0	0.076
Number of observations	*666*	
Panel B: With tutoring at baseline		
Tutoring effects (unweighted)	-0.187	0.061
	(0.116)	(0.042)
Tutoring effects (AIPW)	-0.195	0.046
	(0.119)	(0.040)
Control group mean	0	0.043
Number of observations	*149*	
Panel C: Without tutoring at baseline		
Tutoring effects (unweighted)	0.142**	-0.030
	(0.062)	(0.023)
Tutoring effects (AIPW)	0.102*	-0.024
	(0.058)	(0.023)
Control group mean	0	0.086
Number of observations	*517*	

Notes: The table shows the intention-to-treat (ITT) effects of the treatment (online tutoring offer) on the school outcomes. In Panel A, analysis is performed on the whole sample (n = 666), in Panel B, on the sample of students with tutoring at baseline (n = 149), and in Panel C, on the sample of students without tutoring at baseline (n = 517). In column 1, the outcome is grade average in math. In column 2, the outcome is in-grade retention (repetition). The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. The math grade average is standardized by subtracting the control group mean and dividing by the control group standard deviation. It is measured such that higher values indicate better grades. In-grade retention is a binary outcome equal to 1 if a student has to repeat the grade, and 0 otherwise. In unweighted estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. In the AIPW estimations, the control variables are chosen by the lasso method (see Appendix Table A19). Robust standard errors are presented in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

in math grade of 0.14 SD. The coefficients in the AIPW specifications are slightly smaller. Effects on grade retention are not significant in any specification, although the point estimates for grade retention are large for the whole sample (3 pp), given that the control group mean is 8.4%. Overall, the results do not vary much between the unweighted and the AIPW specifications, suggesting that variable choice or selective attrition is unlikely to drive the results.

We present the treatment-on-the-treated (TOT) results using registration on the *Lern-Fair* platform as a take-up proxy (see Appendix Table A11). For the whole sample, the results show a TOT effect size of 0.22 SD on math grades which is exactly the same estimate as for all types of tutoring in the meta-analysis by Kraft et al. (2024). For those students without tutoring, the TOT effect size is 0.43 SD, which represents the effect sizes for highly effective programs.

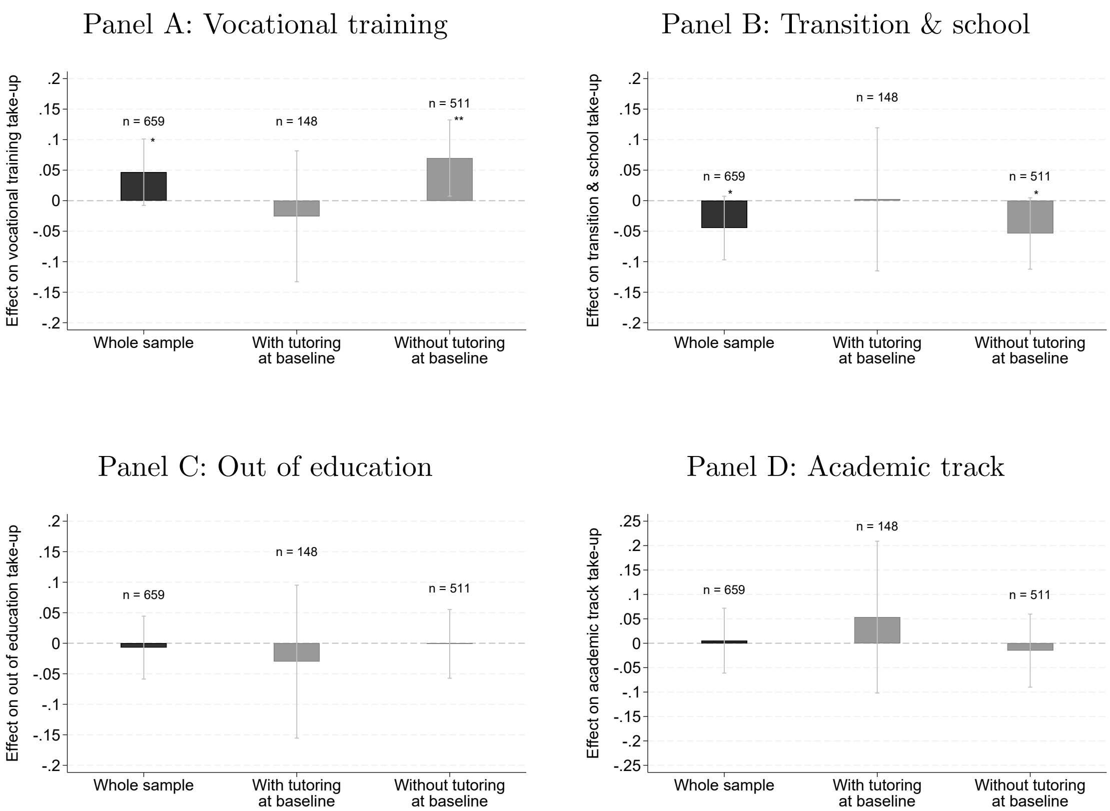
5.2 Follow-up II: School-to-work transitions

After showing that the tutoring improved math grades at follow-up I (six months after randomization), we now investigate whether this improvement has lasting effects on school-to-work transitions a year after follow-up I and 18 months after randomization.

Figure 4 shows the intention-to-treat (ITT) effects on vocational training take-up (Panel A), being in the transition system or still in non-academic schooling (Panel B), being out of education and mainly in low-qualified employment (Panel C), and switching to the academic track (Panel D), for the whole sample (first bar), for the sample with tutoring at baseline (second bar), and for the sample with no tutoring at baseline (third bar).[26]

[26]In Figure 4, the sample size for the whole sample is 659 instead of 652 as indicated for follow-up II in Table A6. This difference arises because, in a small number of cases, we utilized information from follow-up I about academic track or vocational training uptake when sample attrition occurred in follow-up II. The results change only marginally when we restrict the analysis to individuals with data solely from follow-up II.

Figure 4: Effects of online tutoring on career decisions depending on tutoring at baseline



Notes: The figure shows the estimated effects of the treatment (online tutoring offer) on career decisions, together with 95% confidence intervals. In Panel A, the binary outcome is being enrolled in a vocational training program. In Panel B, the binary outcome is being in the transition system or being enrolled in any other form of general school (non-academic track schooling). In Panel C, the binary outcome is dropping out of the education system. In Panel D, the binary outcome is being enrolled at university or attending a general school in preparation for higher education. The effects are estimated within the whole randomized sample (first black bars with n = 659), within the subsample of students who had tutoring at baseline (second gray bars with n = 148), and within the subsample of students who had no tutoring at baseline (third gray bars with n = 511). The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. All regressions use robust standard errors and control for gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English.
*** p<0.01, ** p<0.05, * p<0.1.

Table 5: Effects of online tutoring on career decisions

	Vocational training	Transition & school	Out of education	Academic track
	(1)	(2)	(3)	(4)
Panel A: Whole sample				
Tutoring effects (unweighted)	0.047*	-0.045*	-0.007	0.005
	(0.028)	(0.027)	(0.026)	(0.034)
Tutoring effects (AIPW)	0.056**	-0.036	-0.003	-0.001
	(0.027)	(0.027)	(0.027)	(0.034)
Control group mean	0.141	0.169	0.147	0.543
Number of observations	*659*			
Panel B: With tutoring at baseline				
Tutoring effects (unweighted)	-0.026	0.002	-0.030	0.053
	(0.054)	(0.059)	(0.063)	(0.079)
Tutoring effects (AIPW)	-0.014	-0.017	-0.066	0.060
	(0.061)	(0.059)	(0.063)	(0.080)
Control group mean	0.188	0.159	0.203	0.449
Number of observations	*148*			
Panel C: Without tutoring at baseline				
Tutoring effects (unweighted)	0.070**	-0.054*	-0.001	-0.015
	(0.032)	(0.030)	(0.029)	(0.038)
Tutoring effects (AIPW)	0.073**	-0.042	-0.003	-0.014
	(0.031)	(0.031)	(0.030)	(0.038)
Control group mean	0.128	0.171	0.132	0.568
Number of observations	*511*			

Notes: The table shows the intention-to-treat (ITT) effects of the treatment (online tutoring offer) on career decisions. In Panel A, analysis is performed on the whole sample (n = 659), in Panel B, on the sample of students with tutoring at baseline (n = 148), and in Panel C, on the sample of students without tutoring at baseline (n = 511). In column 1, the binary outcome is being enrolled in a vocational training program. In column 2, the binary outcome is being in the transition system or being enrolled in any other form of general school (non-academic track schooling). In column 3, the binary outcome is dropping out of the education system. In column 4, the binary outcome is being enrolled at university or attending a general school in preparation for higher education. The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. In unweighted estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. In the AIPW estimations, the control variables are chosen by the lasso method (see Appendix Table A19). Robust standard errors are shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 6: Effects of online tutoring on career decisions, split by school track

	Vocational training	Transition & school	Out of education	Academic track
	(1)	(2)	(3)	(4)
Without tutoring at baseline				
C1: Non-academic track				
Tutoring effects (unweighted)	0.117*	-0.124**	-0.052	0.059
	(0.066)	(0.062)	(0.048)	(0.059)
Tutoring effects (AIPW)	0.121*	-0.117*	-0.052	0.050
	(0.062)	(0.065)	(0.048)	(0.061)
Control group mean	0.237	0.376	0.161	0.226
Number of observations	*198*			
C2: Academic track				
Tutoring effects (unweighted)	0.034	-0.009	0.036	-0.061
	(0.031)	(0.026)	(0.036)	(0.046)
Tutoring effects (AIPW)	0.032	-0.008	0.034	-0.070
	(0.031)	(0.025)	(0.038)	(0.045)
Control group mean	0.067	0.055	0.116	0.762
Number of observations	*313*			

Notes: The table shows the intention-to-treat (ITT) effects of the treatment (online tutoring offer) on career decisions in the sample of students without tutoring at baseline (n = 511), which is further split into a subsample of students in non-academic track (n = 198) and in academic track (n = 313). In column 1, the binary outcome is being enrolled in a vocational training program. In column 2, the binary outcome is being in the transition system or being enrolled in any other form of general school (non-academic track schooling). In column 3, the binary outcome is dropping out of the education system. In column 4, the binary outcome is being enrolled at university or attending a general school in preparation for higher education. The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. In unweighted estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. In the AIPW estimations, the control variables are chosen by the lasso method (see Appendix Table A19). Robust standard errors are shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

The treatment significantly increases vocational training take-up for the whole sample. The other outcomes are not significantly affected by the treatment for the entire sample. However, the increase in vocational training take-up and decrease in continuing in the transition system or attending school are larger for the sample with no tutoring at baseline.

Table 5 shows the point estimates for all students (Panel A), for those with tutoring at baseline (Panel B), and for those students with no tutoring at baseline (Panel C), for the four outcomes. In each panel, the first row shows the unweighted results and the second row shows the AIPW results with the controls (see Appendix Table A19 for the variables selected by the lasso method).

The results in Table 5 indicate that participation in vocational training increased significantly by 4.7 pp for all students, and by 7 pp for those without tutoring at baseline. In contrast, being in the transition system or attending a non-academic track school was reduced by 4.5 pp for the first and 5.4 pp for the second group. Table 6 shows that the effects for students without tutoring at baseline are primarily concentrated among those enrolled in non-academic track schools.[27] For these students, participation in vocational training increased by 11.7 pp. In line with this strong increase in the take-up of vocational training, we find a strong decline in the share of non-academic track students in the transition system or still attending a non-academic school (12.4 pp). The tutoring offer has no significant effects on switching from the non-academic track to the academic track, being out of the education system, or on any career decision among the academic track students, who mostly continue schooling. Tables A12 and A13 present the TOT effects for the ITT estimates shown in Tables 5 and 6.

Results for the unweighted and weighted specifications are similar in Tables 5 and 6, demonstrating that neither selective attrition between the treatment and control groups in comparison with the baseline, nor the choice of control variables is driving the effects. When we investigate the effects on grades in follow-up II, we find that tutoring has lasting

[27] We divide the sample by academic track because for students in the academic track, starting vocational training is rarely a relevant alternative. This split was not preregistered. Appendix Table A10 also presents the results for the full sample and for students who received tutoring at baseline.

positive effects on grades, with a 0.11 SD improvement in math grades observed 18 months after randomization. However, as several non-academic track students had left school by follow-up II, the sample size is smaller for this analysis than at follow-up I. The estimates of the coefficients are therefore also less precise than in follow-up I (see Appendix Table A14).

Overall, the results show that the labor market transition of those students who attended a non-academic track school at baseline is strongly affected by the tutoring offer. Specifically, more students from the non-academic track manage to find their way directly into the labor market through vocational training. Notably, the increase in vocational training take-up does not reduce the rate of non-academic track students progressing to the academic track—a critical finding, as vocational training has been shown to have negative effects for these students (Matthewes and Ventura, 2022). Instead, the online tutoring reduces the rate of students taking detours through preparatory classes for vocational training or additional years of non-academic schooling. For these students, Matthewes and Ventura (2022) show that vocational training leads to higher lifetime earnings and reduces the likelihood of low-skilled employment, consistent with previous evidence that vocational education is particularly beneficial for students who would otherwise enter the labor market directly (e.g., Bertrand et al., 2021).

5.3 Channels

In this section, we use our rich data to examine the channels through which the online tutoring might influence math grades and labor market transitions. We start our analysis by investigating whether the tutoring also affected students' behavioral and socioemotional skills beyond grades and labor market transition. Specifically, we looked at effort, health and behavior, and aspirations.

Table 7 shows the effects on study time, class activity, mental well-being, grit, and aspirations for students with no tutoring at baseline. Given the large difference between students on the academic and non-academic track when it comes to their preferred qualifications, we show aspiration results separately for the two groups. The tutoring had no significant effect on any of these outcomes. In line with these results, the tutoring also did not affect the non-preregistered outcome "self-efficacy" (not shown in a table) or the survey response timing patterns at follow-up I and II (Appendix Figure A4, Panels B and C), which suggests no difference in procrastination between students in the treatment and control groups. Observing no effects on students' behavioral and socioemotional skills is not uncommon for tutoring programs, with, e.g., Guryan et al. (2023) and Carlana and La Ferrara (2024) finding no effect of tutoring on these outcomes either, although they do find large effects on grades.

Table 7: Potential mechanisms of the effects of online tutoring, students without tutoring at baseline

	School input		Health & behaviour		Aspirations			
					Non-academic track		Academic track	
	Studying time	Class activity	Mental well-being	Grit	Vocational training	Uni degree	Vocational training	Uni degree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Tutoring effects (unweighted)	-0.057	0.040	-0.012	0.012	-0.013	-0.021	-0.036	-0.013
	(0.093)	(0.184)	(0.066)	(0.069)	(0.071)	(0.069)	(0.054)	(0.036)
Tutoring effects (AIPW)	-0.083	-0.051	0.003	0.004	-0.059	-0.009	-0.034	-0.025
	(0.094)	(0.178)	(0.066)	(0.068)	(0.078)	(0.070)	(0.055)	(0.037)
Control group mean	1.816	2.921	0	0	0.710	0.754	0.589	0.901
Number of observations	*374*	*374*	*463*	*463*	*147*	*147*	*286*	*286*

Notes: The table shows the intention-to-treat (ITT) effects of the treatment (online tutoring offer) on other outcomes, seen as the potential mechanisms, in the sample of 517 students with no tutoring at baseline (out of the whole sample of 666 students with available data). "Studying time" (column 1) is a continuous variable indicating how much time a student spends studying, on average, other than school classes. It is measured on a scale from 0 to 6 hours. "Class activity" (column 2) is a continuous variable indicating how many times, on average, a student raised their hand in their last math class. It is measured on a scale from 0 to 6. "Mental well-being" (column 3) is a continuous variable measured by the HCL-10 question on a scale from 1 to 4, with higher values indicating less hypomanic symptoms. The outcome is standardized by subtracting the control group mean and dividing by the control group standard deviation. "Grit" (column 4) is a continuous variable measured on a scale from 1 to 5, with higher values indicating more perseverance. The outcome is standardized by subtracting the control group mean and dividing by the control group standard deviation. "Vocational training" (columns 5 and 7) is a binary variable equal to 1 if a student plans to obtain vocational training qualifications in the future, and 0 otherwise. "Uni degree" (columns 6 and 8) is a binary variable equal to 1 if a student plans to obtain a university degree in the future, and 0 otherwise. Outcomes in columns 1–2 are measured within the subsample of students who are still at school. Outcomes in columns 5–8 are measured within the subsample of students who are still in school or who have left school but have not started their vocational training or university education yet. Outcomes in columns 3–4 are measured within the whole sample. "Treatment" (online tutoring offer) is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. In unweighted estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. In the AIPW estimations, the control variables are selected by the lasso method (see Appendix Table A19). Robust standard errors are shown in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Since changes in behavioral or socioemotional skills are unlikely to be the channels through which online tutoring enhances grades, more plausible explanations may include improved learning efficiency—more measurable academic progress given a certain amount of study time—or a deeper understanding of subject content. The finding that tutoring also improved English grades (see Appendix Table A15), despite significantly fewer students having registered for tutoring in English, suggests that improved learning efficiency, which is transferable across multiple subjects, may have contributed to the observed improvements in school performance.[28]

Next, we investigate potential channels for the improved labor market transition. To this end, Table 8 examines the application behavior of the students, also reporting the results from Tables 5 and 6 for reference. In line with the finding that besides grades no other outcomes are significantly affected, students do not appear to be more motivated in their application process either.[29] The rate of application submissions remains unchanged (column 2). However, if we consider only those students who applied for vocational training, this reveals that those who applied are more likely to end up in vocational training (column 3). This finding indicates that these students are more successful in the application process.

Taken together, the results suggest that, most likely, math grades improved because the tutoring increases learning efficiency. Alternative explanations, such as changes in socioemotional skills or increased effort, are less plausible, as no significant changes were observed in these areas. Similarly, it is unlikely that tutors focus solely on preparing students for specific exams, given the observed spillover effects on other subjects, such

[28]Appendix Table A15 also shows no effect on German grades, which may cast doubt on the hypothesis that improved learning efficiency led to higher grades, as we would have then expected an effect on German as well. However, several studies on educational interventions suggest that native-language performance is less malleable. For example, Carlana and La Ferrara (2024) find positive effects of online tutoring on math achievement and smaller effects on native-language performance. Moreover, they report that the effects on foreign-language skills are larger than those on the native language. This evidence suggests that the absence of effects on German grades is not necessarily inconsistent with the learning efficiency hypothesis.

[29]The results in Appendix Table A16 show that the effects on grades are positive in both tracks, albeit not significantly because of the smaller sample size.

Table 8: Vocational training application behavior,
students without tutoring at baseline

	Voc. train. enrollment	Application submission	Successful application
	(1)	(2)	(3)
Panel A: All students			
Tutoring effects (unweighted)	0.070**	0.009	0.182**
	(0.032)	(0.039)	(0.078)
Tutoring effects (AIPW)	0.073*	0.021	0.214***
	(0.031)	(0.076)	(0.076)
Control group mean	0.128	0.266	0.500
Number of observations	*511*	*501*	*139*
Panel B: Non-academic track			
Tutoring effects (unweighted)	0.117*	0.013	0.217**
	(0.066)	(0.071)	(0.102)
Tutoring effects (AIPW)	0.121*	0.022	0.239**
	(0.062)	(0.067)	(0.096)
Control group mean	0.237	0.391	0.647
Number of observations	*198*	*192*	*78*
Panel C: Academic track			
Tutoring effects (unweighted)	0.034	0.000	0.063
	(0.031)	(0.044)	(0.166)
Tutoring effects (AIPW)	0.032	-0.003	0.153
	(0.031)	(0.045)	(0.126)
Control group mean	0.067	0.199	0.344
Number of observations	*313*	*309*	*61*

Notes: The table shows the intention-to-treat (ITT) effects of the treatment (online tutoring offer) on the vocational training application behavior in the sample of 511 students with no tutoring at baseline (out of the whole sample of 659 students with available data). In column 1, the binary outcome is being enrolled in a vocational training program. In column 2, the binary outcome is having submitted an application for a vocational training program. In column 3, the binary outcome is equal to 1 if an application was successful, and 0 otherwise, conditional on having applied for a vocational training program. The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. In unweighted estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. In the AIPW estimations, the control variables are selected by the lasso method (see Appendix Table A19). Robust standard errors are presented in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

as English. While better understanding of math may not directly improve grades in English, the enhanced learning efficiency fostered by tutoring could have wider benefits, including in English. These improvements in grades are the most likely explanation for tutored students in the non-academic track experiencing a smoother labor market transition through better signals to the employer. As we do not observe that the tutoring changed aspirations, socioemotional skills, or the number of job applications submitted, it is unlikely that students applied for different types of vacancy or performed better in job interviews. This assumption is supported by tutor reports indicating low levels of mentoring, which often focus on topics such as applications and behavior in job interviews (see Table 3).

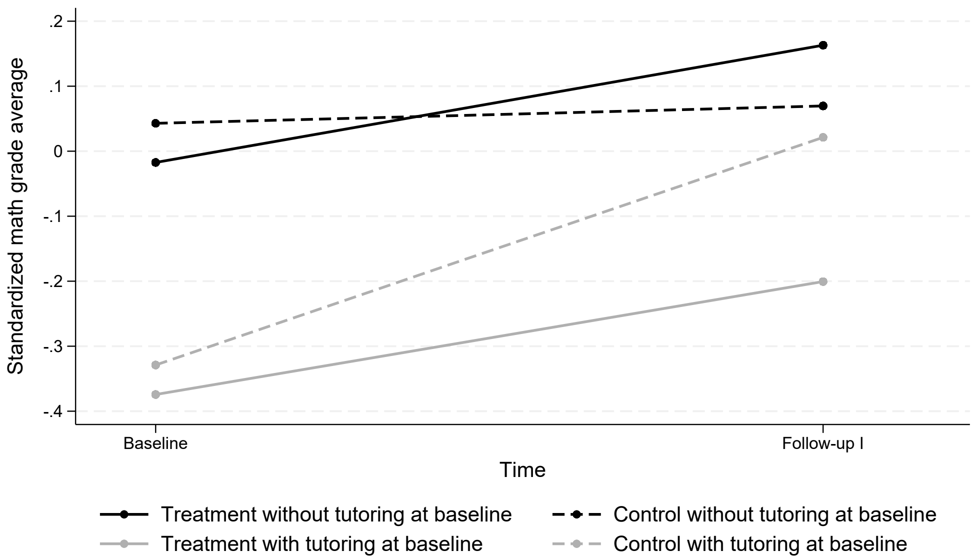
5.4 Heterogeneity

So far, our analyses have focused on students with no additional tutoring at baseline, demonstrating that tutoring is more effective for this group. We concentrated on these students because those with other tutoring at baseline had a lower take-up rate, and we expected lower tutoring effectiveness for this subgroup. However, other student characteristics, such as welfare receipt or migration background, are also associated with tutoring take-up. Moreover, having additional tutoring at baseline may itself be correlated with other student characteristics. If this is the case, we may observe the effect of these other student characteristics if we separate the sample by having tutoring at baseline.

To explore this possibility, Appendix Figures A6 and A7 examine whether heterogeneous effects are observed for other characteristics associated with tutoring take-up, such as gender, age, migration background, and welfare receipt. When we split the sample based on these characteristics, we do not find different effects across any subgroup. The only factor consistently linked to the effectiveness of online tutoring is whether students had other tutoring services at baseline, as shown in sections 5.1 and 5.2.

Next, we investigate possible reasons why the tutoring effectiveness is lower for those students with tutoring—probably in-person tutoring—at baseline. To address this, Figure

Figure 5: Change in math grade average from baseline to follow-up I dependending on tutoring at baseline



Notes: The figure shows the change in the standardized math grade average between baseline and the first follow-up survey. The averages are shown separately for the treatment and control groups as well as for the sample with and without tutoring at baseline. Follow-up I = six months after baseline.

5 analyzes the development of math grades between baseline and follow-up I for the four groups:[30] (1) treatment group with no tutoring at baseline, (2) control group with no tutoring at baseline, (3) treatment group with tutoring at baseline, and (4) control group with tutoring at baseline. At baseline, students with tutoring at baseline perform worse (0.3–0.4 SD) than those with no tutoring at baseline in the treatment and control groups. Analyzing the change in grades over time shows that only students in the control group with no tutoring at baseline fail to improve their grades between baseline and follow-up I. All other groups exhibit grade improvements, with students in the control group with tutoring at baseline showing the most substantial improvement.

These findings suggest that the control group with tutoring at baseline shows stronger grade improvements than all other groups, indicating that other tutoring services were effective. However, as students in the treatment group with prior tutoring developed less favorably, these other tutoring services appear to have been less effective within the

[30] The patterns are similar for German and English.

treatment group. Interestingly, students in the treatment group who already had tutoring still improved their grades to a similar extent as those in the treatment group without prior tutoring, suggesting that the online tutoring itself was not completely ineffective, even among students already receiving other tutoring.

We discuss three possible mechanisms that could explain why tutoring becomes less effective in the treatment group when other tutoring services are already in use: (1) substitution of effective in-person tutoring, (2) overload when two services are used simultaneously, and (3) crowding out of existing tutoring due to the availability of an additional offer. Appendix Figure A8 helps to identify the mechanisms. Like Figure 5, it illustrates the development for math grades between baseline and follow-up I. However, unlike Figure 5, Appendix Figure A8 only shows the grades for students with prior tutoring in the treatment and control groups. To rule out either the substitution/overload mechanism or the crowding-out mechanism, the trajectories would need to develop in parallel with the control group. However, we do not observe such a pattern, suggesting that all three mechanisms may be relevant to some extent.

5.5 Robustness

This section investigates the robustness of our outcome variables. In Appendix Table A17, we test whether our results are robust for different specifications of our grade measures and we analyze which grades improve most across the grade distribution. In column 1, we use grades on a 0–15-point scale as outcomes and still observe positive tutoring effects.[31] In columns 2 to 5, we transform our grade outcomes into binary variables. In column 2, students with "very good" grades are assigned a value of 1, while all others are assigned a value of 0. Similarly, column 3 includes students with "good" or better grades as 1, column 4 includes those with "satisfactory" or better grades as 1, and column 5

[31] In the German education system, students in the academic track from grade 11 onward receive their grades either in the form of full points on a 0–15-point scale (with 15 points being the best grade) or in the form of grades on a scale of 1–6 (with 1 being the best grade), which also include half grades in the form of "+" or "-" for grades between 1 and 5. It is therefore always possible to transform grades into points and vice versa. Students in the lower school years, where grades are not officially reported on the 15-point scale in school reports, are usually familiar with their equivalent 15-point grades.

includes those with "adequate" or better grades as 1. Our analysis reveals that the effect is concentrated among students with "good" or better grades.

One limitation of our study is the reliance on self-reported outcomes. As German data restrictions did not allow us to obtain grades from administrative sources, we cannot test the reliability of the students' self-reports using an external source. However, Resnjanskij et al. (2024) elicited math grade information from German secondary school students in an experimental context and checked the validity of the reports by comparing them to the administrative grades they had access to. The correlation between administrative and self-reported math grades in their survey was r = 0.86, indicating a high level of reliability for self-reported grades. For the transition outcomes, we also anticipate high reliability in the self-reports, as self-reported employment spells are generally considered highly reliable (e.g., Wahrendorf et al., 2019).

Additional support for the reliability of self-reported data comes from Anger et al. (2024), a related project that also draws on households from the *CoDu* study (though involving younger children, and with no household participating in both studies). The study examines the effects of providing e-readers on children's reading skills, where interviewer demand effects, social desirability, and stigma associated with self-reported reading behavior are also potential concerns. In that study, reports of reading behavior independently collected from children and parents show a high degree of concordance, suggesting low bias in self-reported measures within the same survey framework, albeit for a younger population.

Another limitation of our study is the sample size and the relatively low number of treatment takers. The sample size of 839 individuals (421 in the treatment group and 418 in the control group) enables us to detect effects with a size of 0.19 SD with an alpha of 0.1 and power of 0.8, given full compliance and no attrition. If we consider our take-up rate of 30% and follow-up attrition of 20%, the corresponding minimum detectable TOT (LATE) effect size for a continuous variable such as math grade is d = 0.63. Given that TOT effect sizes in the tutoring literature are around 0.4, our study is slightly underpowered.

We also adjusted our p-values for multiple hypothesis testing (MHT) to assess whether the significant results might have occurred by chance. The effects on math performance remain significant, as do the effects on vocational training for the full sample. To further demonstrate that differences in observable characteristics between the experimental sample and the regular *Lern-Fair* population do not substantially affect the program's effectiveness, we reweighted the experimental sample using the distribution of characteristics of the regular *Lern-Fair* participants. The results remain similar. Finally, following Ghanem et al. (2023), we formally tested whether attrition between the treatment and control groups was selective. The results provide no evidence for selective attrition. The corresponding tables for these robustness checks are available upon request.

6 Conclusion

This paper studied the effect of the *Lern-Fair* online tutoring program on school performance and labor market transitions. Our results show that online tutoring enhances both outcomes, with higher effects for those students who were not receiving any other tutoring service at baseline.

The effects on labor market transitions provide a novel perspective on tutoring programs in general—both in-person and online—since we are the first to present evidence that tutoring, as a support service during a young person's school years, affects this important longer-term outcome. While life cycle outcomes—such as long-term wages and employment in adulthood—remain beyond the scope of this study, existing evidence corroborates the long-term effects of vocational training for students close to dropping out of education (e.g., Matthewes and Ventura, 2022; Bertrand et al., 2021). Our results therefore indicate that tutoring can provide lasting benefits for individuals, while also delivering potentially substantial returns for society as a whole. This is also true in societies where vocational training is less prevalent, as the findings demonstrate, for the first time, that improvements in education extend beyond the school setting.

The positive effects of online tutoring on grades are also remarkable, further highlighting its effectiveness relative to in-person tutoring. Given that the take-up rate is about 30% in our study, the treatment-on-the-treated (TOT) effects of online tutoring are about three times larger than the ITT effects, resulting in TOT effect sizes for the whole sample of 0.22 SD and for those without tutoring at baseline of 0.41 SD. This effect sizes falls exactly in the range of successful in-person tutoring programs (e.g., Nickow et al., 2024; Fryer Jr, 2014), highlighting that online tutoring can be as effective as in-person tutoring. These effects on grades are relevant as many governments are leveraging online tutoring to address the COVID-19 achievement gap caused by school closures. This gap has been estimated at between 0.08 and 0.14 SDs, with more negative effects for low-performing students (Engzell et al., 2021; Betthäuser et al., 2023). Thus, the substantial effects of online tutoring have the potential to fully close the reported achievement gaps stemming from school closures.

Moreover, the program's design—using volunteer tutors and decentralized organization—results in very low costs, making this a highly cost-effective intervention for supporting disadvantaged youth. A cost-benefit analysis yields a very favorable result, even when accounting for the opportunity costs of volunteer university students who provide tutoring free of charge. We calculate these costs together with the administrative expenses that arose when the nonprofit organization *Lern-Fair* began hiring salaried support staff funded by donations and public programs after the regular program had been scaled up. This results in total costs of about €300 per tutoring relationship. For the benefit analysis, we base our calculations on students in the treatment group entering the labor market as qualified employees one year earlier than those in the control group, thereby gaining an additional year of earnings valued between €26,665 (2025 minimum wage in Germany) and €46,440 (average salary with vocational training in 2024, BA Statistics). Based on our point estimate (TOT) of a 16.1 pp increase in vocational training take-up, we estimate an expected average earnings gain of approximately €4,266 to €7,430, which yields a cost-benefit ratio of about 14 to 25.[32]

These results are comparable with those from randomized mentoring interventions for disadvantaged youth in Germany which focus on influencing socioemotional skills. Resnjanskij et al. (2024) report a cost-benefit ratio of 18, whereas a mentoring intervention targeting slightly younger children in elementary school yielded a ratio of 31 (Kosse et al., 2020). The cost-benefit ratio of *Lern-Fair* tutoring suggests that programs focused only on improving educational outcomes can also be highly cost-efficient.

Beyond the high cost efficiency, the *Lern-Fair* tutoring program possesses two additional features that make it attractive for policymakers. First, our findings show that the online tutoring is more effective at reaching students in need compared to similar in-person support services. Second, the program has already been scaled up, which is a critical advantage, as many successful pilot programs face challenges in scaling and maintaining quality at larger scales (List, 2022).

[32] This is very likely a lower bound, as we do not account for potential additional long-term gains through improved academic performance. Additionally, we do not consider further educational qualifications (e.g., a master craftsman certificate), which younger graduates may be more likely to obtain, or from the counterfactual scenario in which non-academic track students enter the labor market without vocational training, resulting in lower earnings and higher unemployment, and thus additional costs for society as a whole.

References

Al-Ubaydli, O., J. A. List, and D. L. Suskind (2017). What can we learn from experiments? Understanding the threats to the scalability of experimental results. *American Economic Review 107*(5), 282–286.

Andersen, S. C. and U. Hvidman (2024). Implementing home-based educational interventions at scale. *Journal of Human Resources 59*(6).

Anger, S., B. Christoph, A. Galkiewicz, S. Margaryan, F. Peter, M. Sandner, and T. Siedler (2024). A library in the palm of your hand? a randomized field experiment with low-income children.

Behrman, J. R., C. S. Fan, N. Guo, X. Wei, H. Zhang, and J. Zhang (2024). Tutoring efficacy, household substitution, and student achievement: Experimental evidence from an after-school tutoring program in rural china. *International Economic Review 65*(1), 149–189.

Bertrand, M., M. Mogstad, and J. Mountjoy (2021). Improving educational pathways to social mobility: Evidence from norway's reform 94. *Journal of Labor Economics 39*(4), 965–1010.

Betthäuser, B. A., A. M. Bach-Mortensen, and P. Engzell (2023). A systematic review and meta-analysis of the evidence on learning during the covid-19 pandemic. *Nature human behaviour 7*(3), 375–385.

BMFSFJ (2024). Billions to support children and young people in the pandemic. *Ministry for Family, Seniors, Women and Youth; Press Release*(https://www.bundesregierung.de/breg-en/service/archive/post-covid-catch-up-programme-1911200).

Card, D., J. Kluve, and A. Weber (2018). What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association 16*(3), 894–931.

Carlana, M. and E. La Ferrara (2021). Apart but connected: Online tutoring and student outcomes during the COVID-19 pandemic. *CEPR Discussion Paper No. DP15761*.

Carlana, M. and E. La Ferrara (2024). Apart but connected: Online tutoring, cognitive outcomes, and soft skills. Technical report, National Bureau of Economic Research.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Cockx, B., M. Picchio, and S. Baert (2019). Modeling the effects of grade retention in high school. *Journal of Applied Econometrics 34*(3), 403–424.

Crépon, B. and G. J. Van Den Berg (2016). Active labor market policies. *Annual Review of Economics 8*(1), 521–546.

De Ree, J., M. A. Maggioni, B. Paulle, D. Rossignoli, N. Ruijs, and D. Walentek (2023).

Closing the income-achievement gap? Experimental evidence from high-dosage tutoring in Dutch primary education. *Economics of Education Review 94*, 102383.

DellaVigna, S. and E. Linos (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica 90*(1), 81–116.

Dietrichson, J., M. Bøg, T. Filges, and A.-M. Klint Jørgensen (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research 87*(2), 243–282.

Ehlert, M., A. C. Holtmann, L. Menze, and H. Solga (2018). Besser als ihr Ruf. Übergangsmaßnahmen erhöhen Ausbildungschancen bei leistungsschwachen Jugendlichen. *WZB Mitteilungen 162*, 41–43.

Ek, C. and M. Samahita (2023). Too much commitment? an online experiment with tempting youtube content. *Journal of Economic Behavior & Organization 208*, 21–38.

Engzell, P., A. Frey, and M. D. Verhagen (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences 118*(17), e2022376118.

Evans, W. N., M. S. Kearney, B. Perry, and J. X. Sullivan (2020). Increasing community college completion rates among low-income students: Evidence from a randomized controlled trial evaluation of a case-management intervention. *Journal of Policy Analysis and Management 39*(4), 930–965.

Fersterer, J., J.-S. Pischke, and R. Winter-Ebmer (2008). Returns to apprenticeship training in austria: Evidence from failed firms. *The Scandinavian Journal of Economics 110*(4), 733–753.

Fesler, L., A. Gu, and G. Chojnacki (2023). Air tutors' online tutoring: Math knowledge impacts and participant math perceptions. middle years math grantee report series. *Mathematica.*

Francesconi, M., S. Jenkins, and T. Siedler (2010, June). Childhood family structure and schooling outcomes: evidence for Germany. *Journal of Population Economics 23*(3), 1073–1103.

Fryer Jr, R. G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics 129*(3), 1355–1407.

Fryer Jr, R. G. and M. Howard-Noveck (2020). High-dosage tutoring and reading achievement: Evidence from New York City. *Journal of Labor Economics 38*(2), 421–452.

Ghanem, D., S. Hirshleifer, and K. Ortiz-Becerra (2023). Testing attrition bias in field experiments. *Journal of Human resources.*

Gilraine, M. and J. Penney (2025). Focused interventions and test score fade-out. *Review of Economics and Statistics 107*(2), 570–579.

Gortazar, L., C. Hupkau, and A. Roldán (2024). Online tutoring works: Experimental evidence from a program with vulnerable children. *Journal of Public Economics 232*, 105082.

Guryan, J., J. Ludwig, M. P. Bhatt, P. J. Cook, J. M. Davis, K. Dodge, G. Farkas, R. G. Fryer Jr, S. Mayer, H. Pollack, et al. (2023). Not too late: Improving academic outcomes among adolescents. *American Economic Review 113*(3), 738–765.

Hardt, D., M. Nagler, and J. Rincke (2022). Can peer mentoring improve online teaching effectiveness? an rct during the covid-19 pandemic. *Labour Economics 78*, 102220.

Hassan, H., A. Islam, A. Siddique, and L. C. Wang (2024). Telementoring and homeschooling during school closures: a randomised experiment in rural Bangladesh. *The Economic Journal 134*(662), 2418–2438.

Helbig, M., B. Edelstein, D. Fickermann, and C. Zink (2022). *Aufholen nach Corona?: Maßnahmen der Länder im Kontext des Aktionsprogramms von Bund und Ländern*, Volume 19. Waxmann Verlag.

Himmler, O., R. Jäckle, and P. Weinschenk (2019). Soft commitments, reminders, and academic performance. *American Economic Journal: Applied Economics 11*(2), 114–142.

Kosse, F., T. Deckers, P. Pinger, H. Schildberg-Hörisch, and A. Falk (2020). The formation of prosociality: causal evidence on the role of social environment. *Journal of Political Economy 128*(2), 434–467.

Kraft, M. A., J. A. List, J. A. Livingston, and S. Sadoff (2022). Online tutoring by college volunteers: Experimental evidence from a pilot program. In *AEA Papers and Proceedings*, Volume 112, pp. 614–618. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Kraft, M. A., B. E. Schueler, and G. Falken (2024). What impacts should we expect from tutoring at scale? exploring meta-analytic generalizability. edworkingpaper no. 24-1031. *Annenberg Institute for School Reform at Brown University*.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale*. Crown Currency.

Matthewes, S. H. and G. Ventura (2022). On track to success?: Returns to vocational education against different alternatives. *Centre for Vocational Education Research (CVER) Discussion Paper, London School of Economics and Political Science*.

Munoz-Herrera, M. (2024). Publicly targeting disadvantaged groups triggers stigma and limits take-up of educational opportunities. *mimeo, https://manumunoz.github.io/files/Targeting*$_M$*anuMunoz.pdf*.

Nickow, A., P. Oreopoulos, and V. Quan (2024). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence. *American Educational Research Journal 61*, 74–107.

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

Phipps, A. and A. Amaya (2023). Are students time constrained? course load, gpa, and failing. *Journal of Public Economics 225*, 104981.

Piopiunik, M., F. Kugler, and L. Wößmann (2017). Einkommenserträge von bildungsabschlüssen im lebensverlauf: Aktuelle berechnungen für deutschland. *ifo Schnelldienst 70*(07), 19–30.

Pugatch, T. and N. Wilson (2018). Nudging study habits: A field experiment on peer tutoring in higher education. *Economics of Education Review 62*, 151–161.

Resnjanskij, S., J. Ruhose, S. Wiederhold, L. Woessmann, and K. Wedel (2024). Can mentoring alleviate family disadvantage in adolescence? A field experiment to improve labor market prospects. *Journal of Political Economy 132*(3), 1013–1062.

Robinson, C. D., B. Bisht, and S. Loeb (2022). The inequity of opt-in educational resources and an intervention to increase equitable access. *Annenberg Institute at Brown University EdWorkingPaper* (22-654).

ter Meulen, S. (2023). Long-term effects of grade retention. Technical report, CESifo Working Paper.

Wahrendorf, M., A. Marr, M. Antoni, B. Pesch, K.-H. Jöckel, T. Lunau, S. Moebus, M. Arendt, T. Brüning, T. Behrens, et al. (2019). Agreement of self-reported and administrative data on employment histories in a german cohort study: a sequence analysis. *European Journal of Population 35*, 329–346.

Wolter, S. C. and P. Ryan (2011). Apprenticeship. In *Handbook of the Economics of Education*, Volume 3, pp. 521–576. Elsevier.

Zoido, P., I. Flores-Ceceña, M. Székely, F. J. Hevia, and E. Castro (2024). Remote tutoring with low-tech means to accelerate learning: Evidence for El Salvador. *Economics of Education Review 98*, 102506.

Appendix A

Figure A1: Spatial distribution of experimental households

Notes: The map shows the spatial distribution of experimental households. Households with students in the treatment group are marked with circles, households with students in the control group are marked with triangles. The size of the circle/triangle indicates how many experimental households are located in a given area.

Figure A2: Treatment offer

Fortschritt 95%

W2_Jugendliche2

CoDu hat zum Schluss noch ein tolles Angebot für Dich!

Wir freuen uns, dass wir Dir einen kostenlosen Online-Nachhilfe-Platz bei „Lern-Fair" anbieten können!

CoDu kann einer begrenzten Anzahl von Schülerinnen und Schülern ***kostenlose*** *Online-Nachhilfe anbieten. Diese Online-Nachhilfe wird über die* ***digitale Plattform "Lern-Fair"*** *von jungen Menschen ehrenamtlich angeboten, die bis vor kurzem noch selbst in der Schule waren und nun studieren oder eine Ausbildung machen.*

Du hast damit bis zum Schuljahresende die Möglichkeit, Dich für etwa ***1,5 Stunden pro Woche*** *mit einer/m erfahrenen und motivierten Lernhelferin oder Lernhelfer zum* ***Online-Lernen*** *zu treffen.*
Deine Lieblingstage und -zeiten für die Online-Nachhilfe lassen sich ***individuell planen****.*

Du kannst Dich hier direkt unverbindlich für die kostenlose Nachhilfe anmelden: Anmeldung für die kostenlose Nachhilfe

Über einen Link an Deine E-Mail-Adresse kannst Du Deine Anmeldung zur Online-Nachhilfe bestätigen. Danach geht es los und Lern-Fair sucht eine **Lernhelferin oder einen Lernhelfer** für Dich. Ihr vereinbart dann gemeinsam ein erstes Treffen zum Kennenlernen.

Noch ein wichtiger Hinweis: Dein persönlicher Link ist nach dem Schließen des Browserfensters nicht mehr sichtbar. Wir empfehlen Dir daher, Dich jetzt direkt anzumelden. Zur Sicherheit kannst Du hier (nochmals) Deine ***E-Mail-Adresse*** *angeben, damit wir Dir die Zugangsdaten erneut zusenden können:*

Für nähere Informationen zur digitalen Plattform "Lern-Fair" kannst Du Dir den Lernfair_Flyer herunterladen.

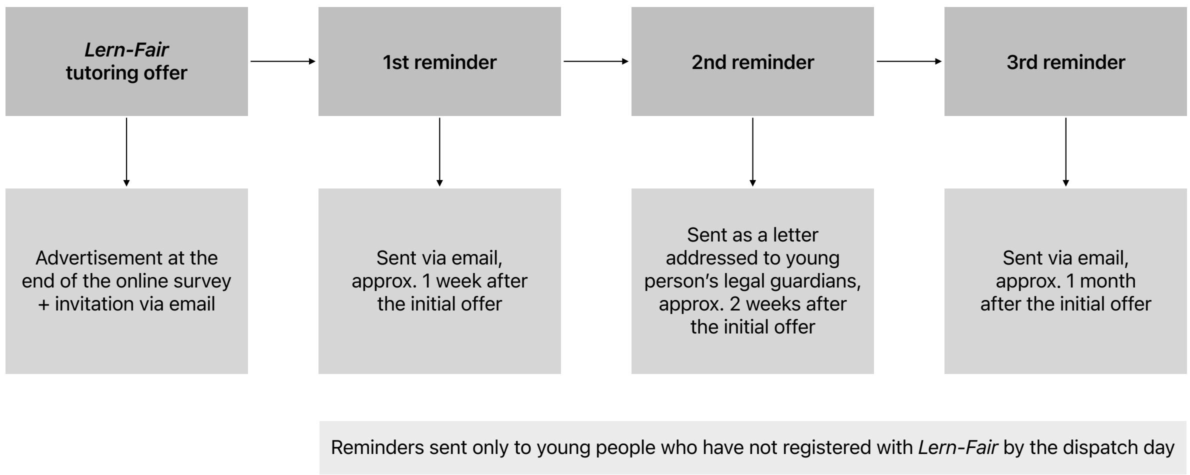
Wir wünschen Dir **viel Erfolg und Freude beim Lernen mit der Online-Nachhilfe!** Wir sind uns sicher, dass Du damit Deine Noten bis zum Schuljahresende verbessern kannst!

Kehre bitte nach Deiner Registrierung bei Lern-Fair hierher zurück und schließe Deine Befragung und ggf. Übermittlung der E-Mail-Adresse mit einem Klick auf "Weiter" ab.

... zurück | weiter ...

Notes: The figure shows the online tutoring offer (in German) that was displayed on the screen after the students who were randomized into the treatment group completed the baseline survey (February–March 2022). Students could access the *Lern-Fair* registration page by directly clicking on the hyperlink, or they could request a copy of the invitation to be sent to the email address they entered in the empty field.

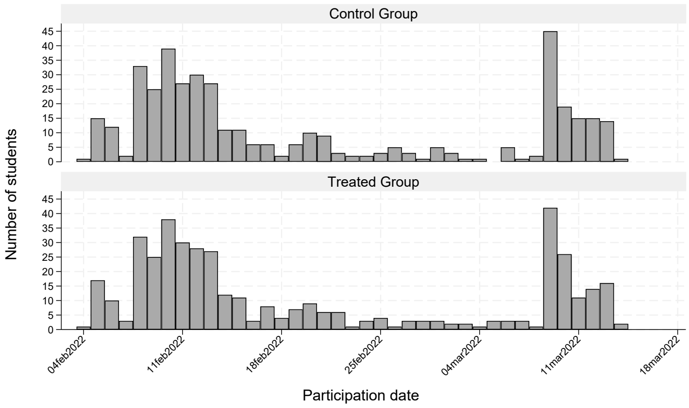
Figure A3: Timeline of reminders



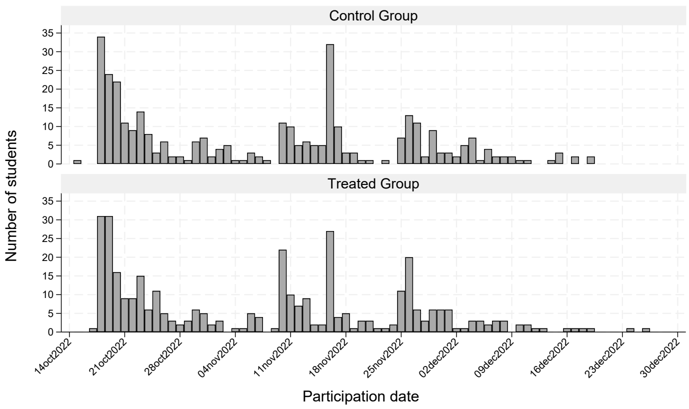
Notes: The figure shows the timeline of reminders that were sent to students from the treatment group in order to encourage them to take up the treatment (online tutoring offer).

Figure A4: Survey response behavior

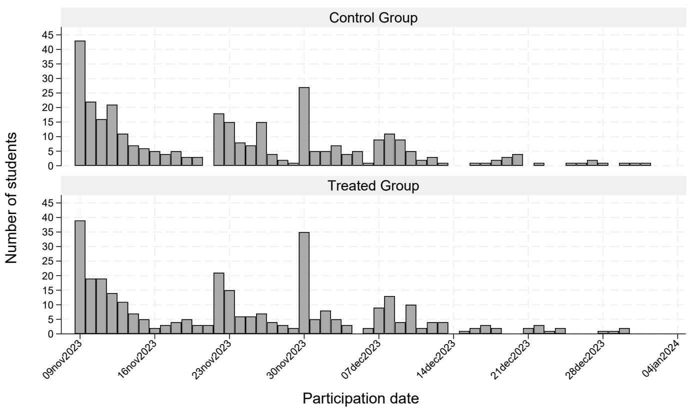
Panel A: Baseline



Panel B: Follow-up I

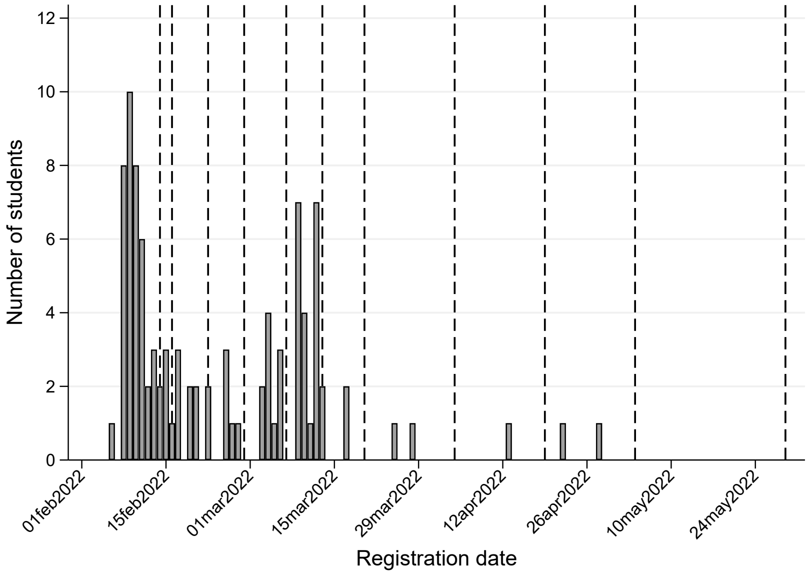


Panel C: Follow-up II



Notes: The figures show the number of students (y-axis) who took part in the survey on a given day of the field period (x-axis). In Panel A, the response behavior in the baseline survey is presented (February–March 2022). Panel B shows the response behavior in the first follow-up survey (October–November 2022) and Panel C in the second follow-up survey (November–December 2023). On each graph, the top panel represents the control group and the bottom panel the treatment group.

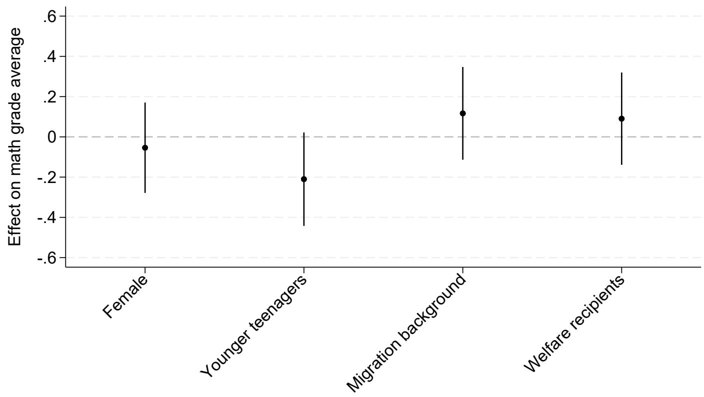
Figure A5: Registration with *Lern-Fair*



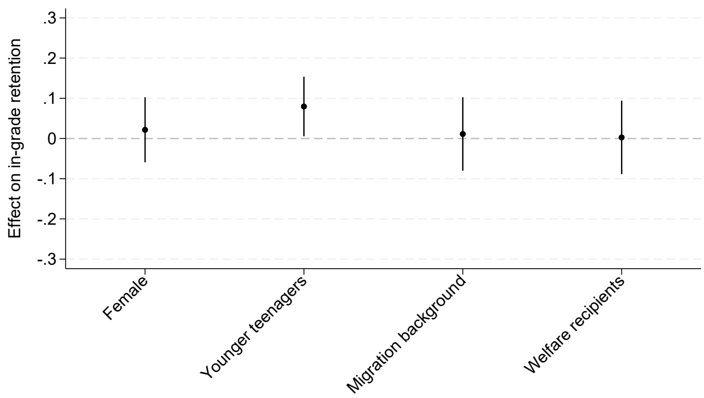
Notes: The figure shows the timing of the registrations on the *Lern-Fair* website and the subsequent student-tutor matches within the sample of treated students. The x-axis displays the time from approximately the point of online survey distribution to approximately the point that the registration was closed. The y-axis indicates the number of students who registered with *Lern-Fair*. The gray bars show the exact number of students who registered with *Lern-Fair* on a particular day. The dashed lines indicate days on which matches between students and tutors were made. Matches are run manually by *Lern-Fair* and can show whether it is the first time a student is matched to a tutor or whether it is a subsequent match, in the event that the first was not successful.

Figure A6: Heterogeneity: School outcomes,
all students

Panel A: Math grade average

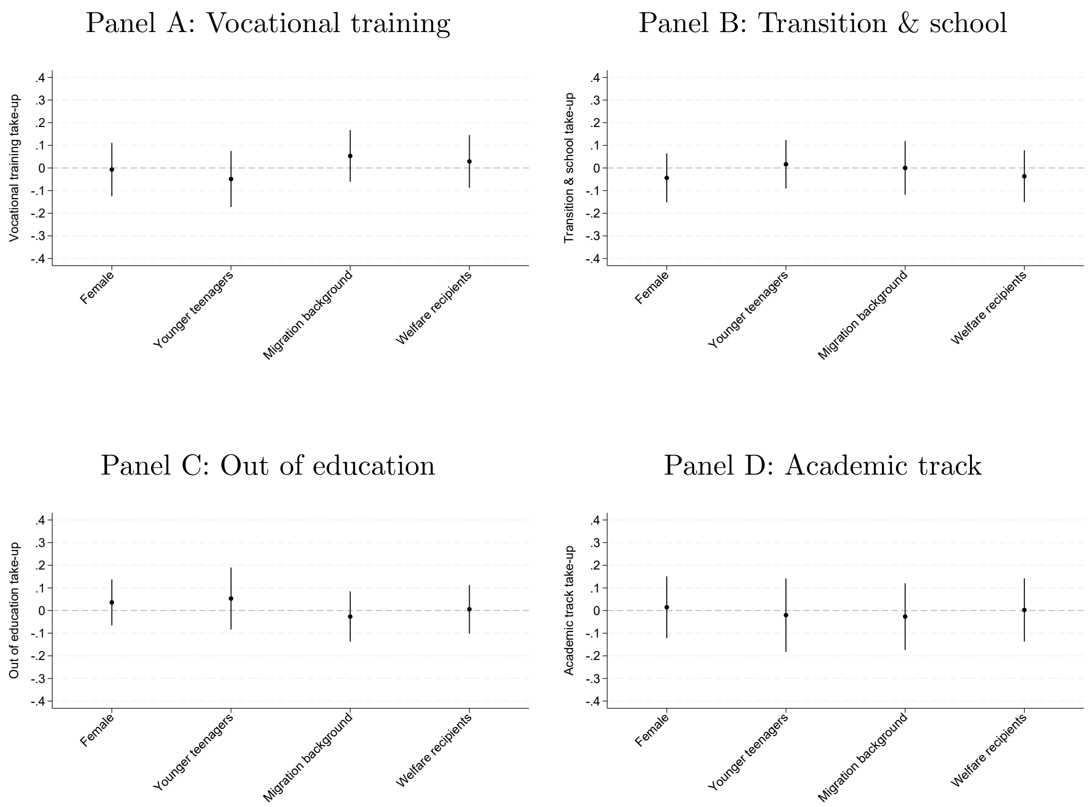


Panel B: Grade retention



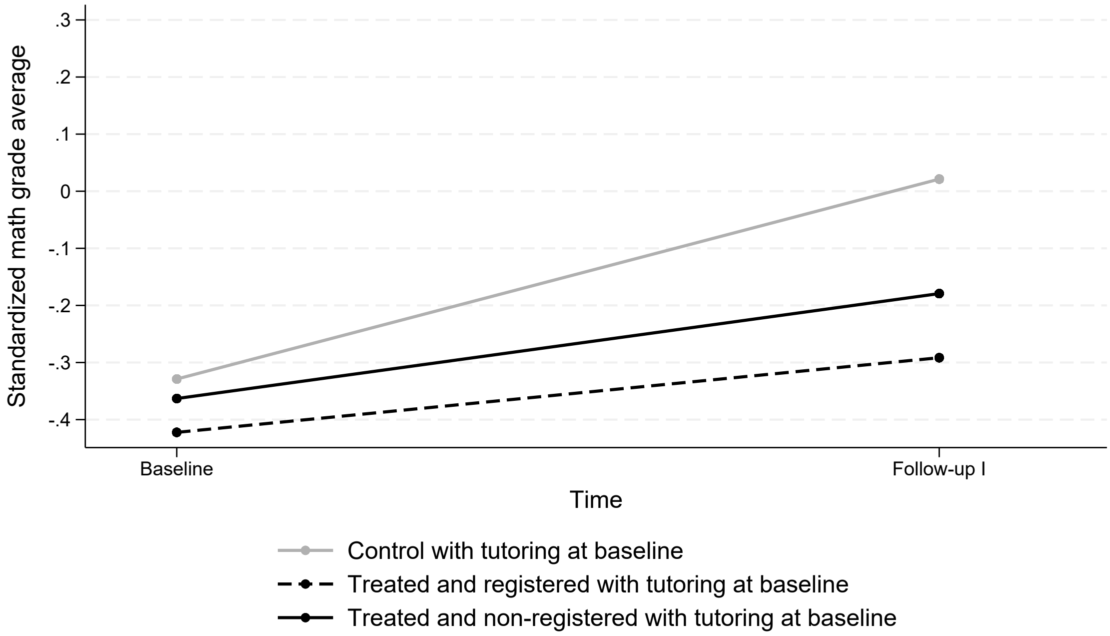
Notes: The figures show the heterogeneous effects of the treatment (online tutoring offer) on the school outcomes together with 95% confidence intervals. In Panel A, the outcome is the grade average in math. In Panel B, the outcome is in-grade retention (repetition). The point estimates are the interaction terms between the stratifying variables and the treatment offer dummy variable. The stratifying variables are as follows: female (=1 if a girl), younger students (=1 if 15–17 years old, =0 if 18–21 years old), migration background (=1 if at least one parent was born outside Germany), welfare recipients (=1 if the household receives welfare benefits). The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. School grades are standardized by subtracting the control group mean and dividing by the control group standard deviation. They are measured such that higher values indicate better grades. In-grade retention is a binary outcome equal to 1 if a student has to repeat the grade, and 0 otherwise. All regressions use robust standard errors and control for gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. The number of observations is 666.

Figure A7: Heterogeneity: Career decisions,
all students



Notes: The figures show the heterogeneous effects of the treatment (online tutoring offer) on the career decisions together with 95% confidence intervals. In Panel A, the binary outcome is being enrolled in a vocational training program. In Panel B, the binary outcome is being in the transition system or being enrolled in any other form of a general school (non-academic track schooling). In Panel C, the binary outcome is dropping out of the education system. In Panel D, the binary outcome is being enrolled at university or attending a general school in preparation for higher education. The point estimates are the interaction terms between the stratifying variables and the treatment offer dummy variable. The stratifying variables are as follows: female (=1 if a girl), younger students (=1 if 15–17 years old, =0 if 18–21 years old), migration background (=1 if at least one parent was born outside Germany), welfare recipients (=1 if the household receives welfare benefits). The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. All regressions use robust standard errors and control for gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. The number of observations is 659.

Figure A8: Change in **math** grade average from baseline to follow-up I, students with tutoring at baseline



Notes: The figure shows the change in the standardized math grade average between baseline and the first follow-up survey among the sample with tutoring at baseline (n = 149). The averages are shown separately for the treatment (n = 79) and control (n = 70) groups. Additionally, the treatment group is divided into thoese who have registered with *Lern-Fair* (n = 15) [treatment takers] and those who have not (n = 64) [treatment non-takers]. Follow-up I = six months after baseline.

Table A1: RCT Sample Selection

	Sample reduction	Sample rize	Share of initial sample
Baseline survey participants		2,225	100.0%
1. Panel readiness (agreement to receive invitations to participate in further surveys)	-6	2,219	99.7%
2. Enrolled in general school (being enrolled in general school in school year 2021/2022)	-343	1,876	84.3%
3. Low school grades (at least one grade in math, German, or English on the school report received mid-school year 2021/2022 as low as 3 (satisfactory))	-526	1,350	60.7%
4. Positive attitude towards online tutoring (hypothetical willingness to take part in free of charge, online tutoring provided by the students or agreement with the statement that tutoring can generally contribute to improvements at school)	-491	859	38.6%
5. Survey completion	-20	839	37.7%
RCT sample		**839**	
Treatment		**421**	
with preexisting tutoring		89 (21%)	
treatment takers		122 (29%)	
Control		**418**	
with preexisting tutoring		89 (21%)	

Notes: The table lists the criteria which individuals participating in the baseline survey had to fulfill in order to be included in the RCT sample. Column 2 shows how many individuals are excluded for not meeting this condition. Column 3 indicates the remaining sample. Column 4 displays the share of the initial sample size. The summary at the bottom of the table shows how many observations from the RCT sample belong to the treatment group (n = 421) and to the control group (n = 418). The table also shows the number and the share of young people in these groups who have had tutoring at baseline (n = 89 in treatment group, n = 89 in control group). Finally, the number and the share of treatment takers in the treatment group are also indicated (n = 122).

Table A2: Share of participants by federal states

	CoDu sample	Germany
	(1)	(2)
Baden-Württemberg	14.18	13.37
Bavaria	17.04	15.85
Berlin	4.77	4.45
Brandenburg	3.10	3.05
Bremen	0.95	0.81
Hamburg	1.79	2.24
Hesse	6.91	7.58
Mecklenburg-Western Pomerania	1.55	1.93
Lower Saxony	9.65	9.65
North Rhine-Westphalia	20.26	21.50
Rhineland-Palatinate	6.20	4.93
Saarland	0.36	1.18
Saxony	4.77	4.84
Saxony-Anhalt	2.62	2.59
Schleswig-Holstein	3.34	3.50
Thuringia	2.50	2.52

Notes: The table compares the spatial distribution of the experimental *CoDu* sample to the spatial distribution of the entire German population. Column 1 shows the share of students from the experimental sample (n = 839) that report living in particular federal states in the baseline survey (February–March 2022). Column 2 shows the share of the entire German population living in particular federal states as of 31.12.2022.
Source: Federal Statistical Office of Germany (2023).

Table A3: Balance test: Parental and household characteristics

	Treatment mean	Control mean	Mean difference	p-value
	(1)	(2)	(3)	(4)
Panel A: Parental characteristics				
Female	0.986	0.988	0.002	0.771
Age	46.996	46.754	-0.242	0.548
Migration background	0.238	0.268	0.030	0.311
Disability	0.005	0.005	0.000	0.994
Education				
No school leaving qualification	0.052	0.050	-0.002	0.895
High school leaving qualification	0.297	0.280	-0.017	0.587
Employment				
Currently employed	0.570	0.560	-0.010	0.765
Ever employed	0.831	0.794	-0.037	0.169
Current daily wage	66.037	67.838	1.801	0.632
Most common occupational groups				
Office clerks and secretaries	0.093	0.100	0.008	0.701
Education and social work	0.081	0.069	-0.011	0.532
Cleaning services	0.059	0.048	-0.012	0.459
Retail trade (without specialization)	0.055	0.041	-0.014	0.343
Nursing and emergency medical services	0.048	0.041	-0.007	0.630
Social assistance				
Currently receiving social assistance	0.114	0.127	0.013	0.570
Ever received social assistance	0.468	0.495	0.027	0.430
Panel B: Household characteristics				
Net monthly income	3491.034	3463.386	-27.648	0.809
BuT package	0.173	0.191	0.018	0.500
Single household	0.166	0.151	-0.016	0.538
Number of children	1.977	2.039	0.062	0.317
Number of individuals	4.057	4.101	0.049	0.626
Number of observations	421	418		

Notes: The table presents the balance test for the treatment and control groups with respect to parental (Panel A) and household (Panel B) characteristics. Columns 1–2 show the mean value of these variables for the treatment group (column 1) and the control group (column 2), respectively. Column 3 shows the difference in means, while column 4 shows the corresponding p-value. Information on parental characteristics is taken from the administrative data of the Institute for Employment Research (IAB). In most cases, these characteristics relate to the mother of a student. Information on household characteristics comes from the baseline parental surveys. Here, information is provided by a parent filling out the survey. BuT package is a dummy variable equal to 1 if a household receives any benefits from the government *Bildung und Teilhabe* (education and participation) package, and 0 otherwise. Single household is a binary variable equal to 1 if a parent does not have a partner or has a partner who does not live in the same household, and 0 otherwise. In case of missing values on age, current daily wage, net monthly income, number of children, and number of individuals, sample means are imputed. In case missing values on the BuT package dummy and single household dummy, the missing information is coded as zero.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A4: Baseline students characteristics by school track

	Non-academic mean	Academic mean	Mean difference	p-value
	(1)	(2)	(3)	(4)
Panel A: Sociodemographic				
Female	0.539	0.672	-0.134***	0.000
Age	16.679	17.211	-0.532***	0.000
Former West Germany	0.836	0.789	0.047*	0.091
Migration background	0.440	0.306	0.134***	0.000
Welfare recipient	0.613	0.266	0.347***	0.000
Panel B: School-related				
Grade	9.693	10.702	-1.008***	0.000
Other tutoring	0.176	0.237	-0.061**	0.034
Math grade	0.007	-0.005	0.013	0.859
German grade	-0.195	0.130	-0.325***	0.000
English grade	-0.247	0.165	-0.412**	0.034
Number of observations	336	503		

Notes: The table presents baseline characteristics of the students by school track. School track is coded as non-academic and academic. Non-academic school track includes students in lower track (*Hauptschule*) and medium track (*Realschule* and *Gesamtschule* up to grade 11) schools. Academic school track includes students in higher track (*Gymnasium* and *Gesamtschule* from grade 11 onward) schools. Variables are divided into sociodemographic characteristics (presented in Panel A) and school-related covariates (presented in Panel B). Columns 1–2 show the mean value of these variables for students enrolled in a non-academic track (column 1) and academic track (column 2) school, respectively. Column 3 shows the difference in means, while column 4 shows the corresponding p-value. The variable "female" is equal to 1 if a student is a girl, and 0 otherwise. "Age" is a continuous variable measured on a scale from 15 to 21. The variable "former West Germany" is equal to 1 if a student lives in the former West Germany, excluding Berlin, and 0 otherwise. The variable "migration background" is equal to 1 if at least one parent was born outside Germany, and 0 otherwise. The variable "welfare recipient" is equal to 1 if the household receives welfare benefits, and 0 otherwise. "Grade" is a continuous variable which takes a value from 8 to 13. The variable "other tutoring" is equal to 1 if a student is already receiving some form of tutoring, and 0 otherwise. Baseline grades in math, German, and English are the respective school grades received before the randomization. These are standardized with a mean of zero and a standard deviation of one in the whole sample. Grades are measured such that higher values indicate better grades.
*** p<0.01, ** p<0.05, * p<0.1.

Table A5: Balance test at follow-up I

	Treatment mean	Control mean	Mean difference	p-value
	(1)	(2)	(3)	(4)
Panel A: Sociodemographic				
Female	0.629	0.623	-0.006	0.876
Age	16.966	16.965	-0.001	0.991
Former West Germany	0.800	0.806	0.006	0.847
Migration background	0.313	0.339	0.026	0.456
Welfare recipient	0.372	0.386	0.014	0.710
Panel B: School-related				
Academic track	0.606	0.643	0.038	0.302
Grade	10.321	10.339	0.018	0.821
Other tutoring	0.223	0.212	-0.011	0.726
Math grade	-0.080	0.000	0.080	0.297
German grade	0.096	0.000	-0.096	0.195
English grade	0.041	0.000	-0.041	0.588
Number of observations	355	345		

Notes: The table presents the balance test for the treatment and control groups among students who participated in the follow-up I survey (October–November 2022). Variables are divided into sociodemographic characteristics, presented in Panel A, and school-related covariates, presented in Panel B. Columns 1–2 show the mean value of these variables for the treatment group (column 1) and the control group (column 2), respectively. Column 3 shows the difference in means, while column 4 shows the corresponding p-value. The variable "female" is equal to 1 if a student is a girl, and 0 otherwise. "Age" is a continuous variable measured on a scale from 15 to 21. The variable "former West Germany" is equal to 1 if a student lives in the former West Germany, excluding Berlin, and 0 otherwise. The variable "migration background" is equal to 1 if at least one parent was born outside Germany, and 0 otherwise. The variable "welfare recipient" is equal to 1 if the household receives welfare benefits, and 0 otherwise. Academic school track includes students in a higher track (*Gymnasium* and *Gesamtschule* from grade 11 onward) school, compared to non-academic track which consists of students in lower track (*Hauptschule*) and medium track (*Realschule* and *Gesamtschule* up to grade 11) schools. "Grade" is a continuous variable which takes a value from 8 to 13. The variable "other tutoring" is equal to 1 if a student is already receiving some form of tutoring, and 0 otherwise. Baseline grades in math, German, and English are the respective school grades received before the randomization. These are standardized by subtracting the control group mean and dividing by the control group standard deviation. Grades are measured such that higher values indicate better grades.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A6: Response rates in follow-ups I and II

	Total	Treatment	Control	p-value
	(1)	(2)	(3)	(4)
Panel A: Baseline				
Number of participants	839	421	418	
Panel B: Follow-up I				
Response rate	0.834	0.843	0.825	0.487
Number of participants	700	355	345	
Panel C: Follow-up II				
Response rate	0.777	0.765	0.790	0.392
Number of participants	652	322	330	

Notes: The table shows the response rates and the number of participants in the baseline, follow-up I, and follow-up II surveys. In Panel A, the baseline survey—the survey wave during which the randomization process was conducted—is considered. In Panel B, information about the follow-up I survey is summarized, and in Panel C, the same is done for the follow-up II survey. Column 1 indicates the total number of participants as well as the overall response rate. Columns 2–3 present the number of participants and response rates for the treatment and control groups, respectively. Column 4 shows the p-value corresponding to the difference in the response rates between the treatment and control groups.
*** p<0.01, ** p<0.05, * p<0.1

Table A7: Balance test at follow-up II

	Treatment mean	Control mean	Mean difference	p-value
	(1)	(2)	(3)	(4)
Panel A: Sociodemographic				
Female	0.613	0.630	0.018	0.641
Age	16.978	16.994	0.016	0.858
Former West Germany	0.804	0.812	0.007	0.816
Migration background	0.304	0.342	0.038	0.300
Welfare recipient	0.360	0.364	0.003	0.928
Panel B: School-related				
Academic track	0.612	0.658	0.046	0.225
Grade	10.314	10.370	0.056	0.503
Other tutoring	0.233	0.218	-0.015	0.653
Math grade	-0.095	0.000	0.095	0.229
German grade	0.060	0.000	-0.060	0.437
English grade	0.047	0.000	-0.047	0.554
Number of observations	322	330		

Notes: The table presents the balance test for the treatment and control groups among students who participated in the follow-up II survey (November–December 2023). Variables are divided into sociodemographic characteristics, presented in Panel A, and school-related covariates, presented in Panel B. Columns 1–2 show the mean value of these variables for the treatment group (column 1) and the control group (column 2), respectively. Column 3 shows the difference in means, while column 4 shows the corresponding p-value. The variable "female" is equal to 1 if a student is a girl, and 0 otherwise. "Age" is a continuous variable measured on a scale from 15 to 21. The variable "former West Germany" is equal to 1 if a student lives in the former West Germany, excluding Berlin, and 0 otherwise. The variable "migration background" is equal to 1 if at least one parent was born outside Germany, and 0 otherwise. The variable "welfare recipient" is equal to 1 if the household receives welfare benefits, and 0 otherwise. Academic school track includes students in higher track (*Gymnasium* and *Gesamtschule* from grade 11 onward) schools, compared to non-academic track which consists of students in lower track (*Hauptschule*) and medium track (*Realschule* and *Gesamtschule* up to grade 11) schools. "Grade" is a continuous variable which takes a value from 8 to 13. The variable "other tutoring" is equal to 1 if a student is already receiving some form of tutoring, and 0 otherwise. Baseline grades in math, German, and English are the respective school grades received before the randomization. These are standardized by subtracting the control group mean and dividing by the control group standard deviation. Grades are measured such that higher values indicate better grades.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A8: *Lern-Fair* implementation

	Mean	n
	(1)	(2)
Panel A: Tutoring duration		
Less than a month	0.11	27
1–3 months	0.11	27
3–5 months	0.59	27
More than 5 months	0.19	27
Panel B: Tutoring frequency		
Less than once a week	0.36	31
Once a week	0.13	31
1–2 times a week	0.45	31
Twice or more a week	0.06	31
Panel C: Reasons for volunteering		
Contribution to educational equality	1.00	16
Societal contribution during coronavirus crisis	0.94	16
Enjoyment from teaching	0.81	16
Enjoyment from working with children and young people	0.63	16
Engagement in long-term volunteering	0.56	16
Monetary reward	0.06	16

Notes: The table shows the statistics on *Lern-Fair* tutoring implementation. All information is taken from from the surveys conducted among *Lern-Fair* tutors of the *CoDu* students. In Panel A, information on average tutoring duration is presented, with the sum of the shares being one. In Panel B, information on average tutoring frequency is presented, with the sum of the shares being one. In Panel C, reasons for volunteering for the *Lern-Fair* organization are presented. Tutors were asked to indicate all the reasons that apply.

Table A9: Balance test at baseline,
students without tutoring at baseline

	Treatment Mean	Control Mean	Mean Difference	p-value
	(1)	(2)	(3)	(4)
Panel A: Sociodemographic				
Female	0.631	0.642	0.011	0.791
Age	16.912	16.977	0.065	0.510
Former West Germany	0.769	0.813	0.044	0.219
Migration background	0.323	0.354	0.031	0.457
Welfare recipient	0.415	0.416	0.001	0.982
Panel B: School-related				
Academic track	0.554	0.576	0.022	0.614
Grade	10.277	10.323	0.046	0.620
Math grade	0.027	0.081	0.054	0.544
German grade	0.186	0.057	-0.129	0.133
English grade	0.083	0.039	-0.044	0.616
Number of observations	260	257		

Notes: The table presents the balance test for the treatment and control groups for the students without other tutoring at baseline. Variables are divided into sociodemographic characteristics, presented in Panel A, and school-related covariates, presented in Panel B. Columns 1–2 show the mean value of these variables for the treatment group (column 1) and the control group (column 2), respectively. Column 3 shows the difference in means, while column 4 shows the corresponding p-value. The variable "female" is equal to 1 if a student is a girl, and 0 otherwise. "Age" is a continuous variable measured on a scale from 15 to 21. The variable "former West Germany" is equal to 1 if a student lives in the former West Germany, excluding Berlin, and 0 otherwise. The variable "migration background" is equal to 1 if at least one parent was born outside Germany, and 0 otherwise. The variable "welfare recipient" is equal to 1 if the household receives welfare benefits, and 0 otherwise. Academic track includes students attending higher track schools (*Gymnasium* and *Gesamtschule* from grade 11 onward), compared to non-academic track which consists of students attending lower track (*Hauptschule*) and medium track (*Realschule* and *Gesamtschule* up to grade 11) schools. "Grade" is a continuous variable which takes a value from 8 to 13. Baseline grades in math, German, and English are the respective school grades received before the randomization. These are standardized by subtracting the control group mean and dividing by the control group standard deviation. Grades are measured such that higher values indicate better grades. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A10: Effects of online tutoring on career decisions, split by school track

	Vocational training	Transition & school	Out of education	Academic track
	(1)	(2)	(3)	(4)
Panel A: Whole sample				
A1: Non-academic track				
Tutoring effects (unweighted)	0.072	-0.078	-0.052	0.058
	(0.059)	(0.055)	(0.049)	(0.053)
Tutoring effects (AIPW)	0.076	-0.071	-0.046	0.037
	(0.057)	(0.057)	(0.045)	(0.051)
Control group mean	0.328	0.291	0.143	0.238
Number of observations	*244*			
A2: Academic track				
Tutoring effects (unweighted)	0.024	-0.028	0.015	-0.012
	(0.026)	(0.025)	(0.032)	(0.041)
Tutoring effects (AIPW)	0.023	-0.025	0.022	-0.028
	(0.026)	(0.024)	(0.032)	(0.040)
Control group mean	0.077	0.067	0.145	0.711
Number of observations	*415*			
Panel B: With tutoring at baseline				
B1: Non-academic track				
Tutoring effects (unweighted)	-0.321**	0.346***	-0.062	0.037
	(0.136)	(0.146)	(0.048)	(0.148)
Tutoring effects (AIPW)	-0.099	0.156	-0.055	-0.002
	(0.145)	(0.120)	(0.113)	(0.106)
Control group mean	0.435	0.217	0.196	0.152
Number of observations	*46*			
B2: Academic track				
Tutoring effects (unweighted)	-0.001	-0.078	-0.057	0.136
	(0.051)	(0.077)	(0.036)	(0.091)
Tutoring effects (AIPW)	-0.003	-0.089	-0.033	0.124
	(0.047)	(0.064)	(0.076)	(0.094)
Control group mean	0.059	0.117	0.186	0.637
Number of observations	*102*			

Notes: The table shows the intention-to-treat (ITT) effects of the treatment (online tutoring offer) on career decisions. In Panel A, analysis is performed on the whole sample, while in in Panel B, it is performed on the sample of students with tutoring at baseline. Further, in both panels, the sample is split into sub-samples of students attending non-academic track (Panel A1 and B1) and in academic track (Panel A2 and B2). In column 1, the binary outcome is being enrolled in a vocational training program. In column 2, the binary outcome is being in the transition system or being enrolled in any other form of general school (non-academic track schooling). In column 3, the binary outcome is dropping out of the education system. In column 4, the binary outcome is being enrolled at university or attending a general school in preparation for higher education. The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. In unweighted estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. In the AIPW estimations, the control variables are chosen by the lasso method (see Appendix Table A19). Robust standard errors are shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table A11: Effects of online tutoring on school outcomes: instrumental variable approach

	Math grade average	Grade retention
	(1)	(2)
Panel A: Whole sample		
ITT effects	0.063	-0.011
	(0.055)	(0.020)
TOT effects	0.217	-0.037
	(0.189)	(0.069)
Control group mean	0	0.076
Number of observations	*666*	
Share of treatment takers	*28.91%*	
Panel B: With tutoring at baseline		
ITT effects	-0.187	0.061
	(0.116)	(0.042)
TOT effects	-0.995	0.323
	(0.636)	(0.208)
Control group mean	0	0.043
Number of observations	*149*	
Share of treatment takers	*18.99%*	
Panel C: Without tutoring at baseline		
ITT effects	0.141**	-0.030
	(0.062)	(0.023)
TOT effects	0.433**	-0.093
	(0.194)	(0.071)
Control group mean	0	0.086
Number of observations	*517*	
Share of treatment takers	*31.92%*	

Notes: The table shows the intention-to-treat (ITT) and treatment-on-the-treated (TOT) effects of the treatment (online tutoring offer) on the school outcomes. In Panel A, analysis is performed on the whole sample (n = 666), in Panel B, it is peformed on the sample of students with tutoring at baseline (n = 149), and in Panel C, it is performed on the sample of students without tutoring at baseline (n = 517). In column 1, the outcome is grade average in math. In column 2, the outcome is in-grade retention (repetition). The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. The math grade average is standardized by subtracting the control group mean and dividing by the control group standard deviation. It is measured such that higher values indicate better grades. In-grade retention is a binary outcome equal to 1 if a student has to repeat the grade, and 0 otherwise. In the estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. Robust standard errors are presented in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table A12: Effects of online tutoring on career decisions: instrumental variable approach

	Vocational training	Transition & school	Out of education	Academic track
	(1)	(2)	(3)	(4)
Panel A: Whole Sample				
ITT effects	0.047*	-0.045*	-0.007	0.005
	(0.028)	(0.027)	(0.026)	(0.034)
TOT effects	0.161*	-0.154*	-0.025	0.018
	(0.096)	(0.091)	(0.089)	(0.115)
Control group mean	0.141	0.169	0.147	0.543
Number of observations	*659*			
Share of treatment takers	*28.83%*			
Panel B: With tutoring at baseline				
ITT effects	-0.026	0.002	-0.030	0.053
	(0.054)	(0.059)	(0.063)	(0.079)
TOT effects	-0.136	0.012	-0.161	0.285
	(0.275)	(0.300)	(0.323)	(0.408)
Control group mean	0.188	0.159	0.203	0.449
Number of observations	*148*			
Share of treatment takers	*18.99%*			
Panel C: Without tutoring at baseline				
ITT effects	0.070**	-0.054*	-0.001	-0.015
	(0.032)	(0.030)	(0.029)	(0.038)
TOT effects	0.216**	-0.166*	-0.003	-0.047
	(0.099)	(0.092)	(0.087)	(0.116)
Control group mean	0.128	0.171	0.132	0.568
Number of observations	*511*			
Share of treatment takers	*31.89%*			

Notes: The table shows the intention-to-treat (ITT) and treatment-on-the-treated (TOT) effects of the treatment (online tutoring offer) on career decisions. In Panel A analysis is performed on the whole sample (n = 659), in Panel B, it is performed on the sample of students with tutoring at baseline (n = 148), and in Panel C, it is performed on the sample of students without tutoring at baseline (n = 511). In column 1, the binary outcome is being enrolled in a vocational training program. In column 2, the binary outcome is being in the transition system or being enrolled in any other form of general school (non-academic track schooling). In column 3, the binary outcome is dropping out of the education system. In column 4, the binary outcome is being enrolled at university or attending a general school in preparation for higher education. The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. In the estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. Robust standard errors are shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table A13: Effects of online tutoring on career decisions, students without tutoring at baseline: instrumental variable approach

	Vocational training	Transition & school	Out of education	Academic track
	(1)	(2)	(3)	(4)
Panel A: Non-academic track				
ITT effects	0.117*	-0.124**	-0.052	0.059
	(0.066)	(0.062)	(0.048)	(0.059)
TOT effects	0.361*	-0.380**	-0.161	0.181
	(0.203)	(0.189)	(0.142)	(0.174)
Control group mean	0.303	0.308	0.131	0.258
Number of observations	*198*			
Panel B: Academic track				
ITT effects	0.034	-0.009	0.036	-0.061
	(0.031)	(0.026)	(0.036)	(0.046)
TOT effects	0.102	-0.026	0.107	-0.184
	(0.094)	(0.076)	(0.108)	(0.139)
Control group mean	0.083	0.051	0.131	0.735
Number of observations	*313*			

Notes: The table shows the intention-to-treat (ITT) and treatment-on-the-treated (TOT) effects of the treatment (online tutoring offer) on career decisions among students without tutoring at baseline (N = 511). In Panel A, analysis is performed on the sample of non-academic track students (n = 198), and in Panel B, it is performed on the sample of academic track students (n = 313). In column 1, the binary outcome is being enrolled in a vocational training program. In column 2, the binary outcome is being in the transition system or being enrolled in any other form of general school (non-academic track schooling). In column 3, the binary outcome is dropping out of the education system. In column 4, the binary outcome is being enrolled at university or attending a general school in preparation for higher education. The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. In the estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. Robust standard errors are shown in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A14: Long-term effects of online tutoring on school outcomes, students without tutoring at baseline

	Math grade average	In-grade retention
	(1)	(2)
Tutoring effects (unweighted)	0.111	-0.031
	(0.105)	(0.027)
Control group mean	0	0.087
Number of observations	350	350

Notes: The table shows the long-term intention-to-treat (ITT) effects of the treatment (online tutoring offer) on the school outcomes for 517 students with no tutoring at baseline (out of the whole sample of 666 students with available data). The long-term effects are measured in the follow-up II survey (November–December 2023). In column 1, the outcome is grade average in math. In column 2, the outcome is in-grade retention (repetition). The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. Math grade average is standardized by subtracting the control group mean and dividing by the control group standard deviation. It is measured such that higher values indicate better grades. In the unweighted estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. In the AIPW estimations, the control variables are selected by the lasso method (see Appendix Table A19). Robust standard errors are shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table A15: Effects of online tutoring on English and German grade average, students without tutoring at baseline

	Grade average	
	English	German
	(1)	(2)
Tutoring effects (unweighted)	0.142**	0.010
	(0.061)	(0.071)
Tutoring effects (AIPW)	0.119**	-0.023
	(0.060)	(0.066)
Control group mean	0	0
Number of observations	517	517

Notes: The table shows the intention-to-treat (ITT) effects of the treatment (online tutoring offer) on the English and German grade average for 517 students with no tutoring at baseline (out of the whole sample of 666 students with available data). Column 1 shows the grade average in English and column 2 shows the grade average in German. The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. School grades are standardized by subtracting the control group mean and dividing by the control group standard deviation. They are measured such that higher values indicate better grades. In the unweighted estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. In the AIPW estimations, the control variables are selected using the lasso mechanism (see Table A19). Robust standard errors are shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table A16: Effects of online tutoring on school outcomes by school track, students without tutoring at baseline

	Math grade average	In-grade retention
	(1)	(2)
Panel A: All students		
Tutoring effects (unweighted)	0.142**	-0.030
	(0.062)	(0.023)
Tutoring effects (AIPW)	0.102*	-0.024
	(0.058)	(0.023)
Control group mean	0	0.086
Number of observations	517	517
Panel B: Non-academic track		
Tutoring effects (unweighted)	0.095	-0.030
	(0.108)	(0.039)
Tutoring effects (AIPW)	0.072	-0.014
	(0.102)	(0.039)
Control group mean	0.019	0.097
Number of observations	201	201
Panel C: Academic track		
Tutoring effects (unweighted)	0.148**	-0.030
	(0.074)	(0.029)
Tutoring effects (AIPW)	0.129*	-0.028
	(0.068)	(0.028)
Control group mean	-0.011	0.079
Number of observations	316	316

Notes: The table shows the intention-to-treat (ITT) effects of the treatment (online tutoring offer) on the school outcomes in the sample of 517 students with no tutoring at baseline (out of the whole sample of 666 students with available data) by school track. In Panel A, all students are considered. In Panel B, only students in non-academic track schools are considered and in Panel C, only students in academic track schools are considered. In column 1, the outcome is grade average in math. In column 2, the outcome is in-grade retention (repetition). The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. The math grade average is standardized by subtracting the control group mean and dividing by the control group standard deviation. It is measured such that higher values indicate better grades. In the unweighted estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. In the AIPW estimations, the control variables are selected by the lasso method (see Appendix Table A19). Robust standard errors are presented in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table A17: Robustness checks: Different definitions of math grade average, students without tutoring at baseline

	Points 0–15	Dummies for specific grades			
		"Very good"	"Good" or better	"Satisfactory" or better	"Adequate" or better
	(1)	(2)	(3)	(4)	(5)
Tutoring effects (unweighted)	0.461**	0.030	0.078**	0.032	0.013
	(0.184)	(0.022)	(0.034)	(0.032)	(0.021)
Tutoring effects (AIPW)	0.345**	0.021	0.070**	0.028	0.004
	(0.171)	(0.022)	(0.033)	(0.029)	(0.022)
Control group mean	8.152	0.070	0.311	0.708	0.922
Number of observations	517	517	517	517	517

Notes: The table shows the intention-to-treat (ITT) effects of the treatment (online tutoring offer) on the math grade average in the sample of 517 students with no tutoring at baseline (out of the whole sample of 666 students with available data). In column 1, the math grade average is measured in points on a scale of 0–15, where 15 points is the maximum a student can obtain. In columns 2–5, dummy variables indicating specific grades in the German education system are considered. In column 2, having a "very good" grade is defined as 1 if a student receives 13 or more points, and 0 otherwise. In column 3, having a "good" or better grade is defined as 1 if a student receives 10 or more points, and 0 otherwise. In column 4, having a "satisfactory" or better grade is defined as 1 if a student receives 7 or more points, and 0 otherwise. In column 5, having an "adequate" or better grade is defined as 1 if a student receives 5 or more points, and 0 otherwise. The treatment offer is a dummy variable equal to 1 if a student was randomized into the treatment group, and 0 otherwise. In the unweighted estimations, the following control variables are used: gender, age, school type, school grade, living in former West Germany, migration background, welfare recipient, and baseline grades in math, German, and English. In the AIPW estimations, the control variables are selected by the lasso method (see: Appendix Table A19). Robust standard errors are presented in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table A18: List of potential control variables
for augmented inverse-probability weighting lasso estimations

Panel A: Basic characteristics
Gender
Age
Grade
West vs. East Germany
Migration background
Socioeconomic status (SES)
Type of school attended
Number household members
Life satisfaction
Grit
Mental health (HCL-10)
Panel B: Aspirations
Vocational training
University degree
Panel C: School performance
Baseline German/math/English grade
Having bad baseline German/math/English grade
Activity during German/math/English lessons
Learning German/math/English because of pleasure content
Learning German/math/English because of personal interest
Learning German/math/English because of its importance
Learning German/math/English because of interesting content
Amount of time spent of learning after school (daily average/yesterday)
Being satisfied with one's one grades
Worrying about one's own school performance
Struggling to improve or maintain grades on the same level
Panel D: Proxies for financial situation
Having own desk at home
Having own room at home
Having own vs. shared computer at home
Having e-book reader at home
Having helpful books at home
Having internet connection at home
Panel E: Proxies for pandemic severity experienced
Having been sick with COVID-19
Parents employed on short-term (*Kurzarbeit*) contract
Parents being affected financially
Having to stay at home because of own COVID-19 infection in respective school years
Changing future plans because of pandemic
Having to wear medical mask during lessons
Frequency of contact with teachers during school closure
Frequency of receiving homework during school closure
Frequency of online lessons during school closure
Frequency of receiving interactive content during school closure
Panel F: Characteristics of main caregiver:
Gender
Age
Migration background
Disability status
Education attainment
Employment history
History of social benefits
Daily wage

Notes: The table shows all potential control variables the lasso mechanism can choose from in the estimation of augmented inverse-probability weighting effects. The controls are divided into baseline characteristics (Panel A), aspirations (Panel B), school performance (Panel C), proxies for financial situation (Panel D), proxies for pandemic severity experienced (Panel E), and characteristics of main caregiver (Panel F). All control variables are measured in the baseline survey (February–March 2022). Missing values are imputed with the sample mean. Additionally, dummy variables indicating imputations are created.

Table A19: List of control variables chosen by lasso

Panel A: School outcomes

	Whole sample		With tutoring at baseline		Without tutoring at baseline					
					All		Non-academic track		Academic track	
	Math grade	In-grade retention	Math grade	In-grade retention	Math grade	In-grade retention	Math grade	In-grade retention	Math grade	In-grade retention
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Gender	X	X	X	X	X	X	X	X	X	X
Socioeconomic status (SES)	X	X	X	X	X	X	X	X	X	X
Baseline math grade	X		X		X		X		X	
Baseline German grade			X							
Activity during math lessons					X					
Learning math because of interesting content									X	
Learning English because of interesting content							X			
Daily average time spent of learning after school			X							
Being satisfied with one's one grades			X							
Main caregiver: disability status	X				X					

Panel B: Career decisions

	Whole sample				With tutoring as baseline				Without tutoring as baseline			
	Voc. train.	Trans. & school	Out of edu.	Acad. track	Voc. train.	Tran. & school	Out of edu.	Acad. track	Voc. train.	Tran. & school	Out of edu.	Acad. track
	(1)	(2)	(3)	(4)								
Gender	X	X	X	X	X	X	X	X	X	X	X	X
Socioeconomic status (SES)	X	X	X	X	X	X	X	X	X	X	X	X
Grade			X									
Type of school attended	X			X					X			X
Baseline math grade			X									
Having bad baseline English grade											X	
Activity during German lessons							X					
Learning math for pleasure												X
Aspiration: vocational training	X			X					X			X
Aspiration: university degree				X								
Worrying about one's own school performance										X		
Having own room at home												X
Having to stay at home due own COVID-19 infection: school year 2020/2021												X
Main caregiver: educational attainment									X			

Panel B1: Career decisions, whole sample

	Non-academic track				Academic track			
	Vocational training	Transition & school	Out of education	Academic track	Vocational training	Transition & school	Out of education	Academic track
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender	X	X	X	X	X	X	X	X
Socio-economic status (SES)	X	X	X	X	X	X	X	X
Grade							X	X
Baseline math grade				X				
Aspiration: vocational training	X			X				
Learning English because of interesting content	X							
Having to wear medical mask during lessons								X

Panel B2: Career decisions, with tutoring at baseline

	Non-academic track				Academic track			
	Vocational training	Transition & school	Out of education	Academic track	Vocational training	Transition & school	Out of education	Academic track
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender	X	X	X	X	X	X	X	X
Socio-economic status (SES)	X	X	X	X	X	X	X	X

Panel B3: Career decisions, without tutoring at baseline

	Non-academic track				Academic track			
	Vocational training	Transition & school	Out of education	Academic track	Vocational training	Transition & school	Out of education	Academic track
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender	X	X	X	X	X	X	X	X
Socioeconomic status (SES)	X	X	X	X	X	X	X	X
Age								X
Grade								X
Aspiration: vocational training	X							
Having to stay at home due to own COVID-19 infection: school year 2020/2021								X

Panel C: Potential mechanisms

	Studying time	Class activity	Mental well-being	Grit	Non-Academic track: Vocational training	Non-Academic track: Uni degree	Academic track: Vocational training	Academic track: Uni degree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender	X	X	X	X	X	X	X	X
Socio-economic status (SES)	X	X	X	X	X	X	X	X
Grit				X				
Mental health (HCL-10)			X					
Aspiration: vocational training							X	
Baseline math grade		X						
Activity during math lessons		X						
Activity during English lessons	X							
Daily average amount of time spent of learning after school	X							

Panel D: Application behavior

	All: Application submission	All: Successful application	Non-academic track: Application submission	Non-academic track: Successful application	Academic track: Application submission	Academic track: Successful application
	(1)	(2)	(3)	(4)	(5)	(6)
Gender	X	X	X	X	X	X
Socioeconomic status (SES)	X	X	X	X	X	X
Type of school attended		X				
Aspiration: vocational training	X		X			
Aspiration: university degree	X					
Baseline math grade		X				
Baseline English grade		X				
Learning English because of personal interest			X			
Struggling to improve or maintain grades on the same level				X		

Panel E: Robustness checks

	Points 0–15	Dummies for specific grades: "Very good"	"Good" or better	"Satisfactory" or better	"Sufficient" or better
	(1)	(2)	(3)	(4)	(5)
Gender	X	X	X	X	X
Socioeconomic status (SES)	X	X	X	X	X
Baseline math grade	X	X	X	X	
Having bad baseline math grade				X	
Activity during math lessons	X				
Learning math for pleasure	X				X
Worrying about one's own school performance		X			
Struggling to improve or maintain grades on the same level			X		X
Frequency of contact with teachers during school closure				X	
Main caregiver: disability status	X			X	

Panel F: English and German grade average

	English grade average	German grade average
	(1)	(2)
Gender	X	X
Socio-economic status (SES)	X	X
Baseline German grade	X	X
Baseline English grade	X	X
Activity during English lessons	X	
Learning math for pleasure		X

Notes: The table shows which control variables the lasso mechanism chooses for each of the considered outcomes in the estimation of augmented inverse-probability weighting effects. In Panel A, school outcomes for the whole sample, students with tutoring at baseline and students without tutoring at baseline (with a division into non-academic and academic track) are considered. In Panel B, career decisions for the whole sample, students with tutoring at baseline and students without tutoring at baseline, are considered. In Panel B1, career decisions for the whole sample, with a division into non-academic and academic track, are considered. In Panel B2, career decisions among students with tutoring at baseline, with a division into non-academic and academic track, are considered. In Panel B3, career decisions among students without tutoring at baseline, with a division into non-academic and academic track, are considered. In Panel C, potential mechanisms among students without tutoring at baseline are considered. In Panel D, application behavior among students without tutoring at baseline are considered. In Panel E, robustness checks among students without tutoring at baseline are considered. In Panel F, English and German grade averages among students without tutoring at baseline are considered. The variables "gender" and "socioeconomic status (SES)" were enforced in each of the estimated regressions.

Appendix B

This part of the Appendix outlines two deviations from the preregistered analysis plan (AEARCTR-0008937) and provides explanations for each case. In general, our deviations relate to the creation of subgroups to interpret the results better and gain new insights. All outcomes are analyzed as registered.

1. Inclusion of additional subgroup analyses
In addition to the preregistered analyses, we present results separately for students with and without other tutoring services at baseline. This split was not preregistered. We decided to conduct these analyses after observing that the take-up rate of the online tutoring program was substantially lower among students already receiving other tutoring services. This raised concerns about potential interaction or substitution effects between the online tutoring offer and existing tutoring arrangements. To provide a more nuanced interpretation of the treatment effects, we therefore included these additional subgroup analyses.

2. Presentation of results by school track
Our pre-analysis plan did not explicitly mention separate analyses by school track. During the analysis, we decided to present results separately for academic and non-academic tracks because the transition to vocational training is largely irrelevant for students in the academic track who typically remain in school. This distinction allows for a more meaningful interpretation of the treatment effects on school-to-work transitions.