Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic *

Michela Carlana †, Eliana La Ferrara ‡

This version: February 2021

Abstract

In response to the COVID-19 outbreak, the governments of most countries ordered the closure of schools, potentially exacerbating existing learning gaps. This paper evaluates the effectiveness of an intervention implemented in Italian middle schools that provides free individual tutoring online to disadvantaged students during lock-down. Tutors are university students who volunteer for at least 3 hours per week. They were randomly assigned to middle school students, from a list of potential beneficiaries compiled by school principals. Using original survey data collected from students, parents, teachers and tutors, we find that the program substantially increased students' academic performance (by 0.26 SD on average) and that it significantly improved their socio-emotional skills, aspirations, and psychological well-being. Effects are stronger for children from lower socioeconomic status and, in the case of psychological well-being, for immigrant children.

Keywords: tutoring, COVID-19, education, achievement, aspirations, socio-emotional skills, well-being.

*We thank seminar participants at several universities and webinars for helpful comments. Micol Morellini, Vrinda Kapoor, Angelica Bozzi, Marco Cappelluti, Isabela Duarte, Agnese Gatti, Gaia Gaudenzi, Federica Mezza, Chiara Soriolo, Amy Tan and Monia Tommasella provided excellent research assistance. We are grateful to the schools that took part into the intervention for their collaboration, to the team of pedagogical experts guided by Giulia Pastori and Andrea Mangiatoridi and including Anna Maria Carletti, Paola Catalani, Silvia Negri, Doris Valente, Stefania Zacco and Monica Zanon. We also thank our team of tutor supervisors (Angela Caloia, Leila Pirbay, Ilaria Ricchi, and Giulia Zaratti). La Ferrara acknowledges financial support from the Invernizzi Foundation. Carlana acknowledges RAship support from the Malcolm Wiener Center for Social Policy at Harvard Kennedy School. AER RCT Registry ID: AEARCTR-0002148. The project has obtained IRB approval from Bocconi University and Harvard University.

†Harvard Kennedy School, CEPR, and IZA (e-mail: michela_carlana@hks.harvard.edu).

‡Department of Economics, IGIER and LEAP, Bocconi University, and CEPR (e-mail: eliana.laferrara@unibocconi.it).

1 Introduction

In response to the COVID-19 outbreak, schools have closed in over 190 countries (UNESCO, 2020). School closure has created massive learning losses for children (Grewenig et al., 2020; Psacharopoulos et al., 2020), estimated in up to 0.3 standard deviations in achievement test scores (Maldonado and De Witte, 2020) and 0.9 years of schooling for seven months of shuttered school buildings (Azevedo et al., 2020). The pandemic has also had adverse psychological and social effects for children and adolescents, leading to higher depression and lower development of socio-emotional skills (Orgilés et al., 2020; Golberstein et al., 2020). The combination of these effects risks having long term consequences on the human capital of the cohorts affected by school closures.

While many countries have tried to mitigate learning losses by switching to remote instruction and using asynchronous or synchronous platforms, the implementation of these tools has varied substantially. Even within the same country, schools in wealthier areas have showed higher prevalence of synchronous learning and online participation of students (Malkus, 2020; Chetty et al., 2020). High-income students also have access to better homeschooling inputs, including technology, help from parents (Agostinelli et al., 2020), and online learning resources (Engzell et al., 2020; Bacher-Hicks et al., 2020; Doyle, 2020), which exacerbates educational inequalities.

This paper reports the results of a novel policy experiment launched in Italy, the first country severely affected by the Covid-19 pandemic after China. In 2020, Italian schools were closed from the beginning of March until the summer – more than 1/3 of the entire school year. In response to this, our research team designed and implemented an innovative online tutoring program: TOP ("Tutoring Online Program"). The program targeted middle school students (grade 6 to 8) from disadvantaged background in terms of socioeconomic status, linguistic barriers, or learning difficulties, who were identified by school principals among those lagging behind during distance learning. The program was offered to middle schools from all over Italy on a voluntary basis, and it was completely free.

TOP has two defining features. First, tutoring is entirely *online*. While tutoring has shown promising results when done in person by teachers and paraprofessionals (Nickow et al., 2020), such mode of delivery was impossible during lockdown. All interaction in our program occurs through personal computers, tablets or smartphones. Second, the tutors in TOP are not trained professionals but *volunteer university students*, trained and supported by pedagogical experts. While teachers and professionals are certainly qualified, the skills required for tutoring differ from those for classroom interaction (Cook

et al., 2015). The choice of volunteer tutors has advantages in terms of budget (mobilizing resources to hire professionals may not allow for a rapid response and large scale implementation), and possibly also in terms of the quality of inter-personal interaction, as TOP leverages the intrinsic motivation of university students to be volunteers.

Four weeks after the announcement of the school shutdown by the Italian government, we emailed the principals of all Italian middle schools to introduce the program and ask for a list of potential beneficiary students. At the same time we emailed all students enrolled in three large universities in Milan, the second largest city in the country, offering them the possibility to volunteer for a minimum of three hours per week until the end of the school year. The response was extraordinary. Two weeks later, online tutoring activities started.

We received a total of 1,059 'valid' applications from 76 different middle schools from all over Italy.[1] For each student, the school had indicated which subject they needed help with, among math, Italian and English. 81 percent of the students needed help in more than one subject. We randomly assigned a tutor to 530 of the 1.059 applicants, conditioning on ten 'blocks' based on the timing of the valid application. Due to budgetary and administrative constraints, 530 was the maximum number of tutors to whom we could offer training and pedagogical support. In fact, in order to equip tutors with a basic set of pedagogical skills and to help with potential problems in the relationship with children, we worked with a team of education experts to design an online self-training module for tutors and to hold regular group meeting and on-demand one-to-one sessions with expert educators (see section 2.2.3).

We collected baseline data from students, parents and tutors before the start of the tutoring (first half of April), and follow-up data from students, parents, tutors and teachers at the end of the school year (June). Thanks to the over-subscription and random allocation of tutors to students, we can estimate the causal impact of the program on four sets of outcomes: academic performance, aspirations, socio-emotional skills, and psychological well-being.

We find sizeable and significant improvements for students who were assigned an online tutor compared to those who were not. Time devoted to homework and attendance to regular online classes increased, as reported by students as well as by teachers. Performance in a standardized test that we administered at endline and that covered math, Italian and English improved by 0.26 standard deviations (SD). The effects are particularly strong for math, which is the subject on which the majority of tutoring sessions focused. Teachers'

[1] An application was considered 'valid' when the parent had given informed consent and the child had given assent, and when both parent and child had completed their own (online) baseline questionnaire.

assessments of learning also improved for treated students compared to control ones, by 0.18 SD. These are remarkable effects given that the median length of the online tutoring was around 5 weeks.

Our second category of outcomes relates to educational aspirations, in particular, type of high school track the student plans to enroll in, and likelihood and perceived ability to attend university. We polled students, parents and teachers about these. The resulting index of educational aspirations shows a 0.15 SD increase for students in the program compared to the control group.

We also measured students' perseverance, grit and locus of control and we find that TOP increased the value of a composite index capturing these dimensions by 0.14 SD. The effect is driven by increases in treated students' perception that they can control what happens in their lives (locus of control).

Our fourth set of outcomes includes measures of psychological well-being. At the end of the tutoring treated students were happier and less depressed, as reported by themselves and by their parents. The effect corresponds to a 0.17 SD improvement in a composite psychological well-being index.

We examine treatment effect heterogeneity along several dimensions. The first is the intensity of treatment. While the vast majority of students received 3 hours of online tutoring per week, a random subset of those students who needed help in more than one subject (143 out of 427 treated students) were assigned a tutor who gave their availability for 6 hours per week. We find that performance gains double with the hours of tutoring; the other outcomes do not.

Another dimension of heterogeneity is access to technology. We know from the tutor endline survey that around 20 percent of the students connected using a smartphone, as opposed to a PC or tablet. While one may be concerned that this would diminish the effectiveness of tutoring –and that such decrease would disproportionately affect children from lower socio-economic status– we find that it did not. On the other hand, technical problems (e.g., problems with the internet connection during the tutoring), seem to qualitatively decrease the impact of the program, although the effect is imprecisely estimated. These aspects should be taken into account if one wanted to apply our online tutoring model to lower income or more remote settings.

In terms of demographics and socio-economic background, we do not detect significant differences in impact between boys and girls, nor immigrants and natives – except for the effect on psychological well-being which is entirely driven by immigrants. Improvements in learning outcomes are instead higher for students whose parents have less than college education, have a blue collar job and do not work from home. Interestingly, tu-

tor characteristics such as gender, GPA, degree program and pro-social attitudes do not systematically affect the effectiveness of the tutoring.

Finally, we can estimate how the experience of being a TOP tutor during the pandemic affected the tutors themselves. We can do so because we randomly selected the university students to whom we offered the job from the pool of those who applied to be volunteers.[2] Four months after the end of the program, we find that volunteers who were included in the TOP program have significantly higher empathy than those who were not. The effect corresponds to a 0.27 SD increase. We instead do not find significant effects on tutors' beliefs regarding the relative role of luck versus hard work in determining success in life.

Our paper contributes to several strands of literature. A robust body of empirical work shows that in-person tutoring is highly effective for improving academic outcomes. Recent meta analyses find that the impacts are sizeable (a pooled effect size of 0.37 SD in Nickow et al. (2020)), and robust across a wide array of contextual factors (Fryer Jr, 2017). The importance of small group or individual tutoring has been underlined for students who struggle (Ander et al., 2016) and in order to teach at the right level (Banerjee et al., 2015). Also, the tutor-student relationships is often close to a mentorship connection that may affect the development of cognitive as well as social skills, such as prosociality (Kosse et al., 2020). On the other hand, tutoring is much costlier than classroom instruction and it may not be easy to arrange individual, in person tutoring in the presence of geographical constraints. Also, tutoring may sometimes be attached with the stigma of being identified as a student in-need and pulled out from regular classes (Coie and Krehbiel, 1984; Richmond, 2015). We contribute to this literature by providing evidence on large-scale *online* tutoring, based on volunteer tutors supported and trained by pedagogical experts. Our model allows to substantially reduce the cost of tutoring –one of the biggest barriers to large-scale implementation– but also to efficiently reach students located in disadvantaged areas through virtual learning. Finally, online tutoring is less observable from peers than in-person tutoring, which may reduce the sense of stigma possibly attached with this intervention.

Our results are of course directly relevant to the debate on effective strategies to mitigate the effect of Covid-19 on education. The existing evidence suggests that students who lag behind the most during the pandemic are from low-income families with limited access to technology, and that they receive less support from parents and lower quality of remote learning from schools (Bacher-Hicks et al., 2020; Chetty et al., 2020). Different forms

[2]As mentioned above, we could not accept all tutor applicants because we were constrained in the number of hours of support for tutors that we could pay for. The randomization was conditional on the characteristics that we use to allocate tutors to students, notably subject of tutoring, number of hours per week, and previous tutoring experience and training.

of remote learning instruction have been adopted around the world and, although the evidence on interventions increasing access to computers and internet is mixed (Escueta et al., 2017; Malamud and Pop-Eleches, 2011), the impact of digital technology may differ during school closure compared to normal school years.

To the best of our knowledge, very few policy experiments have attempted to use remote tools to improve learning outcomes during the pandemic. Angrist et al. (2020) evaluate two low-tech interventions in Botswana that use SMS text messages and direct phone calls to support parents in the education of their children. The combined intervention resulted in a 0.12 SD improvement in student outcomes and led parents to update their beliefs about their children's learning level. Hardt et al. (2020) evaluate a remote peer mentoring intervention at a German university during the pandemic, where peers met online to discuss self-organization. They find positive effects on motivation and exam registration, though not on earned credits. Our work contributes to this literature by evaluating the effects of an innovative and low cost online tutoring program targeting teenage students who had been adversely affected by school closures, and by showing impacts on learning outcomes as well as soft skills and psychological well-being.

Finally, recent work on organizations highlights the power of intrinsic motivation and social recognition for improving public service delivery (e.g., Ashraf et al., 2014; Gauri et al., 2019). In particular, Levitt et al. (2016) underline that such behavioral aspects can be leveraged to improve educational performance. While we cannot directly speak to this question, as we did not vary the recruitment method or the incentives provided to tutors, the fact that our tutors self-selected into volunteering for the TOP program and their intrinsic motivation may have contributed to the effectiveness of our intervention. Furthermore, we do provide evidence that volunteering as a tutor increased empathy compared to those university students that applied but where not assigned to a student.

2 Intervention and Study Design

2.1 Institutional Background

Italy has been the first country after China hardly hit by the COVID-19 pandemic, with around 80,000 deaths as of January 2021, one third of which were concentrated in the region of Lombardy. All school buildings closed on March 5th, 2020. Since then, school reopening has been repeatedly postponed until September 2020 and, even during the Fall 2020, many schools had to offer remote learning instruction, depending on regional outbreaks.

The key components for effective remote school learning are the availability of infrastructure and of trained teachers with technological skills. Regarding infrastructure, on March 26th the Italian Ministry of Education allocated 70 million euro to buy tablets that students could temporarily borrow and 10 million to improve internet connection and online platforms of schools.[3] This intervention facilitated the access to devices and internet for disadvantaged students. However, not all students in need were offered a device due to bureaucratic delays.

As of March 2020, teachers' digital competences were still somewhat limited, with less than 50 percent of the teachers using any digital tool in their daily lectures (Agcom, 2019).[4] When schools closed due to COVID-19, the response from teachers was extremely heterogeneous: many students only received instructional packets with homework for the first few weeks. Training courses to improve teachers' technological knowledge were organized starting in Spring 2020 by the regional offices of the Ministry of Education, by private foundations and web platforms such as Google Classroom and WeSchool. Based on data we collected on 427 teachers in our 76 sample schools, by the month of June more than 96 percent of the teachers were providing synchronous online classes. Most of the teacher-student interaction was synchronous with the entire class of around 22 students. Around 85 percent of teachers provided some asynchronous videos, usually no more than one hour per week. Almost all teachers assigned some homework every week.

2.2 The Tutoring Online Program (TOP)

2.2.1 Timeline

Two weeks after the school closure in Italy, we started the process to design and implement an new program, the "Tutoring Online Program" (henceforth, TOP), as an attempt to provide immediate response to the emergency situation. We identified a team of pedagogical experts who could help us develop the curriculum for tutor training and support and we contacted the rectors of three large universities in Milan, asking for permission to advertise our program mong their students. We obtained IRB clearance and between March 30 and April 3 we sent out email invitations to university students and to the principals of all Italian middle schools. On April 14 tutoring activities started and they

[3]The same decree also allocated 5 million euro for the digital training of teachers (Ministerial Decree n. 187, 26 March 2020).

[4]The digital transition in Italian schools was promoted by the Italian law 107/2015 (the so-called "La Buona Scuola"). The first step toward digitalization was a tool called 'electronic class register', created to ease communication between teachers and parents. The register includes grades, absences, and other messages. By the end of the school year 2015-16, more than 90 percent of middle schools were using electronic class register (Agcom, 2019).

lasted until the beginning of June. Appendix Figure A.1 shows the timeline of the project, including the two rounds of data collection. The implementation was entirely supported by the research team with the help of student volunteers and research assistants. In what follows we describe the recruitment process and the key features of the program.

2.2.2 Recruiting schools and students

We sent a recruitment email to all Italian middle schools (grades 6 to 8), using publicly available email addresses. We informed school principals about the support we could provide with TOP, presenting it as "a free online individual tutoring service to students currently struggling" during the school closure. We explained that tutoring would be done by volunteer university students and that it would be for 3 to 6 hours per week. In order to participate in the program, each school principal had to complete a brief baseline survey expressing their interest in the project. In few days, more than 100 schools completed this first step.[5]

Second, school principals –possibly with the help of teachers– had to complete an application form with a list of students including up to three pupils for each class. We asked to select the students who "may need TOP the most in terms of their learning level and family environment". For each child, the form should indicate the preferred subjects of the tutoring (one or more among math, Italian, and English), and contact details of the relevant teachers. The school was in charge of contacting parents and ask for their authorization to share with us the name and surname of the child and the contact information (email and phone) of one of the parents. We asked schools to make sure the selected students had internet connection and a computer or tablet.[6] We clarified that we could not guarantee the tutoring to all applicants and that, if the number of requests exceeded the number of tutors that we could mobilze and support, we would randomly assign tutors to students, in order to give every applicant the same chances.

We received in total 1,594 names of students from 78 schools: 57 percent of these students were identified as needing help in all three subjects, 25 percent in two subjects, and 18 percent in one.[7] The research team contacted all parents to collect informed

[5]We also received some support from a few regional offices of the Italian Ministry of Education, which helped spreading the information on the project. However, the enrollment of schools was almost completed when we received this additional support.

[6]As clarified in Section 2.1, the Ministry provided to each Italian school resources to buy devices for students in need. Despite that, 36 parents among those selected from the school reveled that they had no internet connection or device and they were excluded from our experiment. On top of that, around 20 percent of students used only the phone for the tutoring, as we discuss below.

[7]According to the teachers, almost 90 percent of students needed support in math, 78 percent in Italian and 72 percent in English.

consent for the project and baseline surveys from parents and students. We sent the survey using email and text messages.[8]

Our final study sample comprises the 1,059 students from 76 schools: these are the students who completed the baseline survey and whose parents approved the informed consent and completed the baseline survey themselves by the end of enrollment period (i.e., April 25th). The geographical distribution with the number of students and schools for each region is reported in Appendix Figure A.2. To assess the representativeness of our self-selected sample of schools, in Appendix Table A.I we compare the provinces with and without schools that took part in the TOP program. We currently focus on the province level for data availability issues. While the breakdown by education level of the population is quite comparable (the differences are significant but extremely small), the provinces with schools in TOP tend to have slightly higher immigrant share, and lower unemployment rate. These differences are fully explained by the regional divide, with 65 percent of provinces in the North having at least one school included in the program, compared to 15 percent in the South and Islands, as shown in Appendix Figure A.2 (the Center is equally represented). The regional imbalance is not surprising, given that the North was by far the part of the country most hardly hit by COVID-19. It makes sense that schools in areas strongly hit by the pandemic early on were more likely to apply, most likely as they foresaw that schools would have not reopened until the end of the year.

2.2.3 Recruiting and training tutors

Thanks to the collaboration with the rectors of three large Italian universities in Milan, we sent a message to all students enrolled in undergraduate and graduate programs.[9] The message explained that a team of researchers was launching an online tutoring program and that we were recruiting "volunteers interested in helping middle school students who were struggling to keep up with their classes and with their homework". We required that volunteers should be currently enrolled in university and fluent in Italian. Applicant tutors had to complete a baseline survey, indicating among other things the subjects in which they would feel comfortable tutoring, and their availability for either 3 or 6 hours per week. The number of applications from volunteers reached 2,000 by the end of the enrollment period, far exceeding our expectations.

[8] If the parents did not respond within a few days, the research team followed-up with a phone call to check that they received the information and they eventually shared the consent form and baseline surveys to a new contact provided by the family.

[9] The three universities were Bicocca, Bocconi and Statale, which approximately enroll 33,000, 14,000 and 61,000 students respectively.

As our volunteers were not trained professionals, we hired a team of pedagogical experts to train and support the tutors.[10] Within a few weeks, they set up an online learning platform with a self-training program that included slides and videos. The topics included: how to approach students; tools and online platforms for effective online tutoring; learning disorders; and tips to help students in math, Italian and English. The platform also included a supervised forum where tutors could ask questions and share their experiences. Finally, the pedagogical team organized regular group meetings with around 20 tutors, as well as one-on-one meetings on demand to offer support in specific circumstances. Based on the information reported by our tutors at endline, around 80 percent of them used the training platform, 50 percent watched the videos and followed the online training, 8 percent used the forum, and 36 percent and 12 percent joined at least one small group or individual meeting, respectively.

The tutor training was an important component of TOP and it ensured that our volunteers, even without professional training, could offer a high quality service to their tutees and could receive professional advice and support in case of need. However, this was the most expensive part of TOP (see section 7 for cost estimates) and, given our budget constraints, it limited to 530 the number of tutors that could be trained and supported.

2.3 Experimental Design

2.3.1 Randomization

We randomized the allocation of the 1,059 students in our sample into two groups: a treatment group that received tutoring (530 students) and a control group that did not (529 students). In order to guarantee that students could start as soon as possible, we processed applications on a rolling basis by creating 'blocks' of around 100 student applicants. We stratified the randomization at the block level, where blocks were created depending on the timing of baseline completion.[11] Appendix Table A.II shows that the treatment and control groups do not differ according to baseline characteristics collected from students and parents, including gender, immigration status, learning disorders, grade, interest in the subjects taught, and parental education and occupation.

Of the 1,059 students included in the experimental design, 712 completed the endline

[10] The team was led by prof. Giulia Pastori and prof. Andrea Mangiatordi , both at Bicocca University, and included six other members with teaching and pedagogical expertise.

[11] After reaching around 100 completed applications including parental consent, baseline survey of parents and students, we created a 'block' and randomly assigned 50 percent of students to the treatment and 50 percent to the control group. Within each block, we ordered the observations by school ID and grade.

test. Attrition rates were different for the treatment and control group, which is not surprising given that students who received a tutor remained engaged with the program until the month of June, while control students had to be contacted after not receiving a tutor. As shown in the Appendix Table A.III, on average 67 percent of students completed the endline test score: 46 percent of the control group and 88 percent of the treatment group. We find that children with college educated fathers or with higher familiarity with computers are more likely to complete the endline test, with the effect being driven mainly by students in the control group (column 4, Appendix Table A.III). Compared to students in grade 8, students in grade 6 were more likely to complete the endline in the control group, which may depend on the fact that, during the period of the endline, grade 8 was involved in the final middle school exam, and control students may have been relatively less motivated to devote time to the survey.

[Insert Table I]

Table I reports the balance Table restricting the sample only to students who completed the final survey. Overall, most characteristics are balanced between treatment and control group. If anything, compared to the full sample shown in Appendix Table A.II, the control group is marginally positively selected in terms of parental education (as highlighted above). Given the direction of imbalance in response rates, one may expect an underestimate of the treatment effect. Nonetheless, we will present different robustness checks, including inverse probability-weighted estimates of treatment effects and the inclusion of different sets of controls (Appendix Table A.XII).

Among the 530 treated students, teachers identified 427 as needing help in more than one subject. We randomly assigned one third of these 427 students to an 'intense' version of the program with 6 hours of tutoring per week instead of 3. This will allow us to estimate the impact of treatment intensity in section 5.1. We present the balance Table for random assignment to the intense tutoring in Appendix Table A.IV.[12]

2.3.2 Tutor allocation

We assigned tutors to students following a step-by-step procedure. First, we restricted the sample of tutors to those currently enrolled in university and fluent in Italian. Given the high number of volunteers, we decided to further restrict the sample to tutors with previ-

[12]The table shows some imbalances in the education level of the mother, with a higher share of mothers with at least high-school diploma among students in the 6h treatment vs. 3h treatment group. We control for these baseline characteristics in all regressions.

ous tutoring experience and/or specific training (e.g., to support students with learning disorders or immigrants).

Second, we divided tutors into different groups depending on their expertise in the various subjects (math, Italian, English or combinations of these), their time availability (3 vs. 6 hours per week), and their training (general, specific for immigrants, specific for students with learning disorders). Within each group, we randomly ordered the tutors.

Third, we randomly assigned treated students to tutors taken from the relevant group, considering the subjects they needed help with, whether they needed intense tutoring, and their characteristics (learning disorders and immigration status). Note that only 4 percent of tutors had specific training on learning disorders, while 32 percent of the students in our sample have learning disorders. Hence, the great majority of students with learning disorders were supported by a tutor who had no training other than the support provided by our pedagogical team. Similarly, only 1 percent of tutors had studied specifically to work with immigrant children, who constitute 22 percent of our sample.

As expected, given the allocation procedure, tutors assigned to students differ from the overall sample of tutors that applied. Appendix Table A.V reports the differences in characteristics of assigned tutors with those who applied and where not assigned to a student.

Column 1 of Appendix Table A.VI provides summary statistics for the tutors who were assigned to students, from the tutor baseline survey. Notice that 530 students were assigned to the treatment, but 7 dropped out before starting the tutoring, therefore we only assigned 523 tutors. The great majority of tutors are female (70 percent), born in Italy (98 percent), they were moved by a desire to help others when applying to TOP (83 percent) and they have previous experience as volunteers (83 percent). In terms of degree program, about 34 percent of the tutors attend a STEM major or medical school, 28 percent an economics/business major, 14 percent a humanities major, and only 7 percent a major in education.[13]

Columns 2 and 3 of Appendix Table A.VI show summary statistics separately for tutors that offered their availability for 3 vs. 6 hours per week, and the last two columns report the p-value on the null that the difference is zero and the standardized difference. Tutors who made themselves available for 6 hours are less likely to come from economics or business and more likely to come from humanities; they are also more likely to be born outside Italy and to have training to work with immigrants. Although students are randomly assigned to a high vs low-intensity treatment, we should keep in mind that

[13]The relatively high share of students from economics and business is due to the fact that one of the three universities from which we recruited, Bocconi, specializes in those subjects.

they get a 'package' of different tutor characteristics when assigned to 6 vs. 3 hours of tutoring.

2.3.3 Implementation

We matched all tutors with students by April 25, as shown in the timeline (Appendix Figure A.1). The tutoring lasted from mid-April to the beginning of June 2020.[14] In June 2020, we collected the endline surveys.

After each meeting, tutors were required to record some information about the session using a management tool prepared by the research team. The information included the day and time of the meeting, whether the student had done the homework assigned, and whether he/she had exerted effort during the tutoring session. On average, treated students had 14 tutoring meetings over the course of the program, for a total of 17 hours over 34 days. The distribution of the number of meetings and tutoring days is presented in Appendix Figure A.3. Less than 5 percent of students chose not to start the tutoring (hence have zero meetings). During the tutoring, the subject covered by the great majority of students was math, which was covered by 78 percent of the students. The entire distribution on subjects covered in the meetings, as reported in tutors' registries, is displayed in Appendix Table A.4.

3 Data and Empirical Strategy

We build a unique dataset merging the baseline surveys of parents, students, and tutors with endline data coming from (i) the results of a standardized test administered by us and taken by the students, and (ii) surveys of parents, students, tutors and teachers of the classes in which our treated and control students were enrolled. We report the summary statistics of our main outcomes measured at endline in Appendix Table A.VII.[15] The relevant survey questions are reported in Online Appendix B.

3.1 Student achievement

One of our main outcomes of interest is student learning. In normal years, standardized test scores are collected in May/June from all Italian students in grade 8 by the Institute for the Evaluation of the Italian Schooling System (INVALSI). However, due to the

[14]Some tutors voluntarily decided to support the students during the summer and in the following academic year.

[15]The variables labeled as 'outcomes reported by parent' or 'outcomes reported by teachers' do not refer to parents or teachers themselves, but to the answers that parents/teachers gave about a given child.

pandemic, these tests were not administered in 2020. In collaboration with two expert middle school teachers, we designed a (shorter) standardized test very close in format to the national standardized one. Our test included seven multiple choice questions in math, seven in Italian, and five in English.

The test was administered to treatment and control students by enumerators. The research team sent to each student the link to complete the test score, but they needed a password to access it. The enumerator called each parent to set a time for the test. During the test, the student was on a video call with the enumerator, he/she opened the link with the questionnaire in his/her own device and entered the password given in real time by the enumerator: at that point the test could start. Enumerators were clearly instructed *not* to help children during the test. Once the student completed and submitted the test online, the enumerators were available to discuss any doubts and answer potential questions.

By design, during the course of our program TOP tutors did not follow a specific curriculum but they helped students with the homework assigned by school teachers. For this reason, the test we administered covered the basic achievement expected from students of each grade. On average, treated and control students answered correctly 56 percent of the questions (as shown in Appendix Table A.VII, line 1): 67 percent in math, 48 percent in Italian, and 50 percent in English. The assessment covered a wide range of competencies and very few students reached a ceiling in terms of correct answers.

3.2 Student, parent, and teacher surveys

We asked students, parents, and teachers to complete a questionnaire that we sent by email and/or SMS. The questions covered four main sets of outcomes: academic achievement and beliefs, educational aspirations, socio-emotional skills, and psychological well-being. Teachers were asked to complete the same question for each child in their class that was either treated or control in TOP. The main outcomes in our empirical analysis will be indexes built extracting the first principal component from the variables in each category, standardized to have mean zero and standard deviation one in the control group.

Academic outcomes and beliefs. We asked children and parents their beliefs on the number of correct questions for each subject of the test described in Section 3.1.[16] only 56

[16]Children were asked about this at the end of the test. Parents were asked this question in their endline survey, which typically took place after the kid had taken the test (neither the test not the child's answers were shared with the parent). Indeed, we expect an impact of TOP on academic outcomes, but also on the beliefs and expectations of parents and teachers (Rosenthal, 1973). Overall, the data show that students and their parents are overconfident on their performance, with an average expected share of correct answers equal to 67 percent (for students) and 71 percent (for parents), against an actual share

percent in the test. Notably, 64 percent of the students and 71 percent of the parents are 'overconfident', in the sense that they expect a higher number of correct answers than one actually obtained in the test. Teachers' beliefs tend to be closer to the actual performance: teachers expect their students to correctly answer 49 percent of the questions on average, and only 36 percent of them are overconfident about children's performance.[17]

To obtain a measure of achievement different from the standardized test score, we asked each teacher to assign a grade from 1 to 10 to every child in our study (treated or control) that was in one of their classes. The average grade was 5.65, that is just below the pass grade of 6 in the Italian school context. This is consistent with the target of our intervention being children who were struggling to keep up with school work. We also asked children how they would rate their own school performance on a 1 to 10 scale, and the average was 6.29, somewhat more optimistic than the teachers but definitely not high.

Aspirations. Low goals and ambitions may lead students into an "aspiration trap" (Genicot and Ray, 2017; La Ferrara, 2019). Children from disadvantaged background may underinvest in their education, dropping out from school or choosing easier and less profitable high school tracks (Carlana et al., 2021). We hypothesized that TOP may have a direct effect on students' aspirations, by providing an alternative role model (the tutor) that may induce them to revise their goals. In our survey, we collected information from students on their long term educational goals (e.g., attend university), and on their short term plans (e.g., the type of high school they wanted to enroll in).[18] Among the students in our sample, only 15 percent are interested in a top-tier academic high school, while around 1/3 are planning to attend a vocational high school.[19] As for long-term goals, 39 percent of the students at endline tell us that they are considering university education, and the Figure is similar for parents (35 percent). The share is instead much lower when we ask teachers until what level the student should continue to study: only 14 percent say 'university'. Finally, we also collected a measure of self-efficacy (Bandura et al., 1999), asking students (and parents) whether, aside from what they would like to do in the future, they think they (their children) would be capable of successfully attending

of correct answers of

[17]We tried to interview the math, Italian and English teachers for each child. For cases where one of the teachers did not reply, we calculate the beliefs (and grade, to be described below) as the average for the subjects for which data is available.

[18]In Italy, after grade 8 students need to choose their high school track. The schooling system is organized in top-tier academic tracks (scientific and classical *lyceum*), other academic tracks (linguistic, pedagogical, and other types of *lyceum*), technical tracks (with technological or economic focus, e.g., accounting), and vocational tracks.

[19]On average, in Italy 32 percent of students are enrolled in a top tier track and 14 percent in a vocational track. As expected from the targeting, at baseline, the sample of students who applied to TOP tends to include more low-achieving and low-aspiring students.

university if they wanted to.

Socio-emotional skills. Social distancing and school closure can result in a lack of opportunities to develop not only cognitive, but also socio-emotional skills in the classroom (Alan et al., 2019). In our endline survey we collected several outcomes to capture socio-emotional skills. First, in order to measure perseverance, we asked students to answer a logic question. At the end of the question, we asked them whether they wanted to answer a new question with the same level of difficulty, with a higher level of difficulty or whether they wanted to give up. We use their choice as an outcome measure of perseverance in a real effort task. Second, we measure 'grit' following the Short Grit Scale developed by Duckworth and Quinn (2009). Starting from 8 questions on a 5-point scale, we add up all the points and divide by 40. The maximum score on this scale is 1 (extremely gritty), and the lowest is 0 (not at all gritty). We asked the same questions to children and parents, finding a high correlation among their answers (0.64). Third, we collected a measure of 'locus of control' to capture the extent to which students believe they can control the outcome of events in their lives or whether fate and luck determine the course of action (Rotter, 1966). To calculate the final score, we start from 4 questions on a 5-point scale, add up all the points and divide by 20. Also for this outcome, the maximum score is 1 (high locus of control), and the lowest is 0 (low locus of control).

Well-being. Last but not least, we want to understand if the interaction with the tutor may have helped students to feel less isolated, possibly overcoming depression, and happier. For this purpose, we collected two measures of psychological well-being from students and their parents. The first is the Children's Depression Screener (ChilD-S) developed by Frühe et al. (2012), which is calculated aggregating a battery of 9 questions.[20] The answers are given on a 4-point likert scale; we add up all the points and divide by 36. Also on this outcome, the maximum score is 1 (high level of depression), and the minimum is 0 (no depression). The second measure is a proxy for happiness: we asked whether students were feeling happy or unhappy during the lockdown, on a scale from 1 to 10 (10 being the maximum happiness). The correlation between the depression measure reported by parents and the one reported by students is 0.56, while for happiness it is 0.53.

3.3 Tutor Survey

On top of the baseline information, we asked all the volunteers that had applied to be tutors to complete a very short endline survey in September 2020, six months after the

[20]For a detailed list, see Online Appendix B.

start of the program. Almost all the tutors who were recruited into TOP completed the endline survey, while only around one third of those who were not assigned a student did so. Appendix Table A.VIII shows the difference in observable characteristics among the tutors who participated in TOP ('treated' tutors) and the others ('control' tutors). Once we account for the criteria used to assign students to tutors (e.g., tutors from STEM are over-represented in treatment because math was the subject most in demand by the students), very few significant differences appear. This will allow us to investigate how participation in TOP affected some outcomes measured at the tutor level.

The first outcome is empathy. We collected two standard questions on a 4-point likert scale, asking respondents if they (i) "find it easy to put themselves in somebody else's shoes"; and (ii) "are able to make decisions without being influenced by people's feelings". We sum all points and divide by 8 to obtain a variable ranging from 0 to 1.

The second set of outcomes concerns views on the role of hard work and effort to achieve success in life. The index we build aggregates answers to three separate questions on (i) income differences and effort; (ii) the importance of hard work versus luck and connections; and (iii) the prospects of getting a well-paid job after studying hard, independent of family background. We aggregate the variables in a similar way as described above.

Finally, in addition to the short endline, tutors recruited into TOP also completed some further information on their experience during tutoring, e.g., how satisfied they were, etc.[21]

3.4 Empirical strategy

To assess the impact of TOP on the various outcomes we collected, we estimate the following OLS regression:

$$Y_{ir} = \alpha_r + \beta Treated_i + \gamma X_i + \varepsilon_{ir} \quad (1)$$

where Y_{ir} is the relevant outcome for student i who was assigned to treatment or control in randomization round r; α_r denotes randomization round fixed effects; $Treated_i$ is an indicator for whether the student was assigned a tutor in the TOP program; X_i is a vector of student level controls measured at baseline, including: gender, immigrant status, grade in which the student is enrolled, mother and father's education, mother and father's employment type, learning disability, interest for the different subjects, perseverance, belief on the importance of luck, and familiarity with computers; ε_{ir} is an error term. We

[21]TOP tutors received a longer questionnaire in June that included the questions on their experience during TOP, and then again in September the same short questionnaire that control students received.

estimate robust standard errors. We also correct for multiple hypothesis testing using the Westfall-Young stepdown adjusted p-values, which also control the family-wise error rate (FWER) and allow for dependence amongst p-values.

4 Results

4.1 Online classes and homework

We start by assessing how participation in the program affected key 'inputs' in the learning process on the part of the students. In particular, we consider the time devoted to homework and the quality of their participation in regular (online) classes offered by their schools.

[Insert Figure 1]

Figure 1 shows the distribution of time devoted to homework (in minutes) during the last month of school, as reported by students (panel a) and parents (panel b), as well as the teachers' assessment of how regularly the student handed in their homework (panel c). For each graph, blue bars refer to students in the control group and red ones to students in the TOP program.

Panel (a) shows that the majority of the students report doing between 30 minutes and two hours of homework each day, with a small fraction reporting less than 30 minutes and about 20 percent reporting more than 2 hours. Importantly, the distribution for treated students is clearly skewed to the right compared to that for control ones, with a marked reduction in those that report less than 1 hour and a clear increase in those that report more than 1.5 hours.

When we consider parents' reports (panel b), we see some discrepancy in the levels reported: parents are more likely to report very low values (30 minutes or less) and less likely to report more than 150 minutes. However, it is true also in this case that parents of students enrolled in TOP report comparatively more time devoted to homework by their children.

The bottom panel in Figure 1 shows how school teachers perceive students' commitment to homework. For control students, 12 percent of the teachers report that they never hand in any homework, 28 percent say sometimes, 31 percent most of the time and 29 percent always. The corresponding figures for students in TOP are 4 percent, 23 percent, 35 percent and 38 percent. This confirms that our program did induce students to exert more effort in homework than they would otherwise have exerted.

[Insert Table II]

In Table II we consider a broader set of outcomes which includes not only homework, but also attendance to online classes, behavior during classes and students' liking of the subjects. Each outcome is regressed on the treatment dummy and on the controls detailed in equation (1). We have different sources reporting on the various outcomes, namely students (columns 1-4), parents (columns 5-6) and teachers (columns 7-9).

Consistent with the data in Figure 1, we find that treatment increased the time devoted to homework: the average effect is about 10 minutes per day (column 1) or 9 minutes per day (column 5), depending on whether it is reported by students or by parents.[22] This represents approximately an 11 percent increase over the mean for the control group. Also the regularity of homework completion as reported by teachers is significantly higher for treated students. This is shown in column 7, where we estimate an ordered logit model using as an outcome the categorical variable described in Figure 1(c).

During lockdown, classes were offered online and students were supposed to connect every day and attend them. Compliance with this requirement was not always full, though: sometimes less motivated students pretended to have internet problems and did not connect, or connected for part of the class and then left. In columns 2 and 6, we find that the probability of regularly attending online classes, as reported by the children and the parents, respectively, is uncorrelated with treatment.[23] This is not true, however, when we consider teachers' reports (column 8). In this case, students in the TOP program are 9.4 percentage points more likely to attend classes regularly – a 16 percent increase over the control group mean. The discrepancy is not surprising if one observes the difference in average values of the dependent variable reported by the three categories of respondents: children and parents report regular attendance in 83 and 88 percent of the cases, respectively, while –for the same student– teachers only report it in 57 percent of the cases. It is possible that reporting bias by children and parents may introduce too much noise for us to detect a treatment effect, while the positive impact of TOP is clear if one takes teachers' reports as more reliable –which makes sense given that teachers have no incentive to over-report good behavior.

Column 3 of Table II shows that treated students are 8 percentage points less likely to report that they found it difficult to follow classes online and use their school's online

[22]The continuous dependent variable expressed in minutes per day and used in columns 1 and 5 is constructed by assigning midpoint values to the intervals displayed in Figure 1, panels (a) and (b).

[23]We asked students whether in the last month of school they had been following online classes regularly, and we posed the same question to parents regarding their children. The dependent variable in columns 2 and 6 is a dummy taking value 1 if the answer is "Yes, every time there was an online class".

platform during the last month of school, representing a 10 percent increase over the mean.

Column 9 shows that treated students also behaved better during school hours. While for 83 of the students in the control group teachers report behavioral problems during the last month of school, this fraction is 6.4 percentage points lower among students in the TOP program.

Overall, these results indicate that both the 'quantity' dimension of class attendance and the 'quality' of learning from classwork were positively affected by our program, suggesting a potential complementarity between the work done by the tutor after school and that done by the teachers during school hours.

Tutors also seem to have contributed to making the subjects more interesting for their tutees. Column 4 shows that treated students have a 5 percentage points higher probability of liking the subjects of math, literature or English relative to control students.[24] Given how little our target population likes these subjects (only 28 percent answer in the affirmative in the control group) this is a sizeable increase.

4.2 Academic outcomes and beliefs

In Table III we study the impact of TOP on academic performance and beliefs.[25]

[Insert Table III]

The dependent variable in column 1 is our key measure of performance, that is, the fraction of correct answers given by the student in the standardized test we administered at the end of the program, which covered the subjects of math, Italian and English (see Section 3.1 for a detailed description). We find that the share of correct answers in the test is 4.5 percentage points higher for treated students, a 9 percent increase over the average of 53 percent correct answers in the control group. The effect is highly significant (p-value 0.013) and corresponds to a 0.26 SD increase in the index of performance. This is an impressive result if we take into account two factors. First, the median duration of tutoring was five weeks. Second, tutors did not specifically prepare the students for this type of test (multiple choice tests are not typically assigned as homework in Italian

[24] We asked students how much they liked the three subjects in which tutoring was offered, on a 5-point scale from "Not at all" to "Very much". The dependent variable in column 4 is the mean of three dummies taking value 1 if the answer is 4 or 5 in math, Italian, and English, respectively.

[25] The outcomes in this Table are average values in all three subjects: math, Italian, and English. For *Beliefs* and *Overconfidence* there are few cases for which we have missing information for one subject. For those cases, we take the average over the subjects for which we have information.

schools), but rather focused on helping students find a method for studying and doing regular homework.

In columns 2 to 4 we estimate the effect of the program on the beliefs held by students, parents and teachers about the number of correct responses given by students in the test. In all three cases we find a positive impact, which remains significant for teachers (at the 5 percent level) and for students (at the 10 percent level) also after accounting for multiple hypothesis testing.

Given that TOP led to actual performance improvements, the positive effect on expectations is not surprising. However, as discussed in section 3.2, students and parents on average tend to over-estimate the number of correct answers to the test, while teachers tend to under-estimate it. This can also be seen in the means of the dependent variables for the control group reported at the bottom of Table III. In columns 5 to 7, we test whether the program helped re-align individual beliefs with actual performance, using as dependent variable the dummy 'Overconfidence', which takes value 1 if the *expected* number of correct answers exceeds the *actual* one.[26] The estimated coefficients point in the right direction for students and parents (although they are not significant), while the effect on teachers is a precise zero.

Finally, in the last two columns of Table III we use an alternative measure of academic performance. In the endline survey for students (implemented a few days before the students took the standardized test) we asked them how they would rank themselves compared to their classmates, on a scale from 1 to 10. The 10-point scale is akin to the grading scale used in Italian schools, where 6 indicates a pass. In the teacher endline survey, we posed the same question to the teacher for each of the students in our sample that was in a given teacher's class. We see that students and teachers' evaluations do not diverge much on average: the control group mean in the students' answers is 6.2, while in teachers' answers is 5.5. Treatment significantly increases the two outcomes by 0.25 and 0.33, respectively, corresponding to 0.18 SD for both outcomes.

While the results so far represent average impacts pooling math, Italian and English, Appendix Table A.IX reports the effects separately by subject. Impacts are positive across the board, but in terms of significance the most robust effects are detected on math performance (panel A, columns 1, 4 and 5). This is not surprising, as most students focused on math during the tutoring sessions (see Appendix Figure A.4).

[26]The number of observations is lower in columns 6 and 7 compared to columns 3 and 4 because we have to restrict the sample to cases in which both the students and their parents (or teachers) completed the endline survey.

4.3 Aspirations

In Table IV we estimate the impact of the program on students' aspirations and perceived ability to achieve educational goals, as reported by the students (columns 1-4), their parents (columns 5-6) and their teachers (column 7).

[Insert Table IV]

The direction of the effects suggests that TOP had a small positive impact. Starting from long term goals, TOP students and their parents appear more likely to report that in the future they plan to enroll in university (columns 1 and 5) and teachers are more likely to say that they should do so (column 7). None of these effects is statistically significant, though. A similar consideration applies to self-efficacy: we asked students and their parents, aside from their intentions, how much they thought the student would be *able* to attend university.[27] Perceptions here are low on average (only 21 percent of students and 29 percent of parents in the control group respond in the affirmative), and the positive coefficient on the treatment dummy is not significant at conventional levels (columns 2 and 6).

Finally, a more immediate choice for our students (in terms of time horizon) concerns the high school track in which they plan to enroll after middle school: vocational, technical or academic. Treated students are 6 percentage points less likely to say that they plan to attend the least prestigious track, that is, vocational (column 3). The effect corresponds to almost a 20 percent decrease compared to students in the control group, although it is not statistically significant at conventional levels once we adjust for multiple hypothesis testing.

Overall, the results in Table IV do not show robust evidence of a significant effect of TOP on individual aspirations, although they qualitatively point in a positive direction.

4.4 Socio-emotional skills

We next test whether the program affected students' socio-emotional skills, in particular their reactions in the face of obstacles and their perceived ability to control what happens in their lives.

[Insert Table V]

[27] The original scale for the response was from 1 to 5, where 1 indicated "not at all" and 5 "very much". The dependent variable in columns 2 and 6 is a dummy taking value 1 if the original response was 4 or 5. Results are very similar when estimating an ordered logit model with the original question.

As explained in section 3.2, to measure perseverance we gave students a logic task and after they completed it we asked if they wanted to get another one of the same level of difficulty, a more difficult one, or if they wanted to stop. The direction of the effects in columns 1 and 2 of Table V (panel A) point to an increase in the probability of asking for the more difficult task and a decrease in the probability of giving up among treated students, but neither is statistically significant. The effect is insignificant, and quantitatively negligible, also when the outcome is the index of 'grit' proposed by Duckworth and Quinn (2009) (columns 3-4, panel A).

While perseverance and grit do not appear significantly affected by the interaction with a tutor, students' 'locus of control' does. Column 5 shows that students in TOP believe to a greater extent that they (rather than fate or luck) can control the outcome of events in their lives. The magnitude of the treatment effect corresponds to a 0.19 SD increase over the control group mean. A possible interpretation of this finding is that students who worked with a tutor saw positive results on the academic front (as shown in Table III), thus understanding that success in school was not a matter of luck. They may have then extrapolated this belief to life in general.

4.5 Psychological Well-being

An important goal of our program, in addition to the academic component, was to help students navigate the psychological difficulties that the lockdown and isolation from their friends may have created. The tutor represented, among other things, someone to talk to outside one's own immediate family –a different voice and a connection with the outside world.

In panel B of Table V we estimate the impact of the program on two measures of psychological well being: Frühe et al. (2012) Children's Depression Screener (columns 1 and 3) and a self-reported index of happiness (columns 2 and 4).[28] We construct both measures using the student's own answers (columns 1-2) and using the parent's answer about their child (columns 3 and 4).

Column 1 shows that students in TOP report less symptoms of depression. The magnitude of the effect corresponds to a 0.16 SD decrease. The effect is qualitatively similar but insignificant when reported by the parents. Correspondingly, happiness increases: this time the effect from parents responses is more precisely estimated (but not very different in magnitude from that based on students' answers). The coefficient in column 4 corresponds to a 0.16 SD increase.

[28]Both variables have been normalized so that they range from 0 to 1, as explained in Section 3.2.

These results suggest that TOP played an important role not only in improving learning outcomes of students who would have otherwise lagged behind, but also in mitigating potential mental health problems associated with the pandemic and with the strict regime of lockdown.

4.6 Summary of Main Results and Robustness

Our main outcomes are related to four dimensions, and in all previous tables we reported p-values adjusted for multiple hypothesis testing within each family of outcomes. To summarize the key results, we now report the impact of TOP on standardized test performance and on three summary indexes, constructed using principal component analysis and standardizing the outcome to have mean zero and standard deviation one for students in the control group.

[Insert Figure 2]

The main results on the impact of TOP are reported in Figure 2 (and in the corresponding Appendix Table A.XI). The strongest improvement for the treatment group is in test performance, with an increase of 0.26 SD compared to the control group. This impact is comparable in magnitude to the average impact of large-scale in-person tutoring interventions, as reported in the meta-analysis by Nickow et al. (2020).[29] The overall impact on aspirations, socio-emotional skills, and well-being is also positive with an improvement between 0.14 and 0.17 SD and a p-value of about 0.10 when adjusted for multiple hypothesis testing across the four summary indexes.

Appendix Table A.X provides a robustness analysis of the main results presented so far. First, in columns 1 and 2, we present the OLS estimates and standard errors without including the baseline controls: we find that the results are very similar to the main results reported in Tables III, IV, V, and in Appendix Table A.XI.

Second, in columns 3 and 5, we re-estimate the effect of treatment on our main outcomes choosing the set of control variables in a systematic was with double post LASSO procedure, following Belloni et al. (2012). We include all baseline characteristics that are sufficiently correlated with treatment (after imposing the LASSO penalty given that the regression includes many variables) and the variables that are sufficiently correlated with

[29] We consider large-scale tutoring interventions involving more than 400 observations and implemented mainly by non-professional tutors. Compared to our intervention, most of the previous tutoring experiments that were causally evaluated and included in the meta-analysis by Nickow et al. (2020) focus on elementary school children.

control (after imposing the LASSO penalty) (Ludwig et al., 2017).[30] We list the controls selected using LASSO for each outcome in Appendix Table A.XII. Including the variables picked by LASSO in the second step makes no substantial difference in most results, with the exception of the Aspiration index, where the estimated effect is still positive but smaller in magnitude and not significant at conventional levels.

Finally, in the last two columns we present inverse probability-weighted estimates of treatment effects. The estimated effects are almost unchanged and, if anything, slightly higher due to the minor unbalances presented in Table I, with the control group being more positively selected compared to the treatment group in terms of parental background.

These different robustness checks provide a consistent and overall positive picture of the impact of TOP on student outcomes.

5 Mechanisms and Heterogeneous Treatment Effects

In this section we explore treatment heterogeneity along several dimensions, to better understand the ways in which the program had an impact. We start by considering features such as the number of hours of tutoring and the technology used to connect virtually, and then we move to study the role of students' and tutors' characteristics. We present evidence on heterogeneity in two ways: (i) augmenting our benchmark specification with an interaction term between treatment and the relevant characteristic; and (ii) applying an honest causal forest algorithm (Wager and Athey, 2018).

5.1 Treatment Intensity

As described in section 2.3, students received a different number of hours of tutoring depending on how many subjects they needed help in and on the availability of tutors. Among all the students who needed help in more than one subject, we randomly chose which students would be matched with a tutor who had offered to help for 3 versus 6 hours per week. We thus have exogenous variation that we can exploit to understand how the impact of the program varies with the intensity of the tutoring.

[Insert Table VI]

[30] The double post LASSO procedure is based on three steps. First, we fit a LASSO regression predicting the dependent variable and we select all variables with non-zero estimated coefficient after the introduction of a penalty term that shrinks the estimated regression coefficients towards zero to reduce over-fitting. Second, we fit a LASSO regression predicting the treatment variable and following the same procedure of step one. Finally, we fit a linear regression of the outcome variable on the treatment variable including the covariates selected in the first or second step.

In Table VI we regress our main outcome indexes on the treatment dummy and on an indicator for whether the student got a 6-hour tutor ("Higher treatment intensity").[31] Column 1 shows that the impact of the standard, 3-hours/week version of the program is a 0.2 SD increase in academic performance, as measured by our standardized test score. Having a tutor for 6 hours a week adds another 0.22 SD, leading to an overall impact of 0.42 SD. This is a remarkable effect, and shows that –in this range of hours– the impact of additional hours of tutoring is linear. The magnitude of the effect of the intense treatment compared to the less intense treatment is not surprising in light of the meta-analysis in (Nickow et al., 2020): they show that the average effect of tutoring on learning almost doubles going from 1-2 days per week to 4-5 days (from 0.24 to 0.41 standard deviations).

While doubling the hours of tutoring had a sizeable impact on learning, when we look at other outcomes it did not generate significant gains on top of the gains from the basic version of the program. In columns 2 to 4 of Table VI the coefficient on 'Intense treatment' is positive but never significant. Based on this evidence, it seems that three hours of interaction with the tutor were already enough to generate the bulk of the improvements on these soft skills.

A different dimension of intensity of the tutoring is how productive the hours were, in terms of attention and effort exerted by the student. This is, of course, an endogenous variable, hence we cannot provide any causal evidence in this direction. However, we can descriptively examine the patterns that emerge by exploiting information from the tutor registries. For each session, tutors recorded whether the effort exerted by their student during that session was 'poor', 'fair' or 'very good'. Tutors also recorded the same information for the homework that the student was supposed to complete before each session. In Appendix Table A.XIII we find that treated students who exerted higher than average effort during the session and in completing assignments have higher performance and educational aspirations at endline, while they are not significantly different from the rest of the treated students in socio-emotional skills and psychological well-being. It should be stressed once more that these are *not* causal impacts.

5.2 Devices and Internet Connection

The key feature of TOP is the virtual nature of the interaction between tutor and student. By definition, the program requires a minimum technological input, namely an internet connection and a device that the two can use to have a video call. When we recruited

[31] Overall, 27 percent of the *treated* students received a tutor for 6 hours per week (the mean of this variable in the full sample of treated and control students is 0.135). 427 students out of the 530 treated were in need of help in more than one subject, and 1/3 of them received the 6-hours tutoring.

middle school students, we told school principals that the beneficiaries should have access to a tablet or PC and to an internet connection for at least 3 hours per week. We have information on these aspects in the tutor endline survey and the registries, where tutors recorded for each meeting the type of device used by the student and the occurrence of technical issues. Based on these sources, 22 percent of the students mainly used a smartphone to connect, and 77 percent of the students had technical issues during at least one of the meetings.

[Insert Table VII]

In Panel A of Table VII we test whether the impact of the program was different for students who connected using a smartphone, compared to those who used a PC or a tablet. We find that it was not, except for aspirations, where the effect on students who used a smartphone is zero. Importantly, column 1 shows that, compared to an increase in test score of 0.27 SD for the students who connected with better devices, the impact for students who connected through a smarpthone was 0.22 SD, significant at the 5 percent level. Considering that this may be a more disadvantaged group of students, the sizeable effect is particularly encouraging. It also leaves room for optimism in considering online tutoring as a tool that could be applied in contexts where the population may have lower income and/or only have access to smartphones.

In Panel B of Table VII we consider the occurrence of technical issues during the tutoring sessions. While the coefficient of the variable 'Technical issues' is never statistically significant, its sign points to a lower effectiveness of the treatment in the presence of technical problems. This may be due to disruptions in the learning process as well as to shortening of the duration of the sessions.

5.3 Students' and Parents' Characteristics

An important dimension of heterogeneity pertains to student demographics and socioeconomic background.

[Insert Figure 3]

In Figure 3 we show the impact of treatment separately for different sub-groups of students, split according to predetermined characteristics: gender, immigrant status, and learning disorders. The figure shows the estimated impact (relative to the control group) and associated 95 percent confidence interval, for our four main outcome indexes. All

indexes are standardized so that the coefficients represent the effect of treatment in units of standard deviations.

Starting from the top-left panel, we find that boys and girls benefited from the program to the same extent, as did native and immigrant students. One category that benefited significantly more was that of students with learning disorders: for them, performance in the standardized test increased by over 0.5 SD, significantly more than for treated students without learning disorders. This is a group of students that may have faced particular difficulties with the learning methods and materials that schools provided during distance learning, in most cases not tailored to the needs of students with dyslexia, dyscalculia, or other disorders. Our program helped alleviate such difficulties. This is noteworthy because only 20 out of 523 tutors had specific training on learning disorders, so the vast majority of the 149 students in the treatment group who had a learning disorder got a tutor without specific training.[32]

When we look at aspirations and socio-emotional skills, the treatment effect appears to be similar across all these subgroups – the only pattern worth mentioning is that aspirations increase for natives but not for immigrants, possibly because the latter face different types of barriers when planning their future education (Carlana et al., 2021).

The outcome for which heterogeneity in treatment effects is most striking is psychological well being (bottom-right panel of Figure 3). TOP worked equally well for boys and girls and for students with and without learning disorders (with a marginally higher benefit for students with learning disorders). But when we compare native and immigrant students, it is clear that the increased happiness and reduced depression we detected in Table V is entirely driven by immigrant students. The magnitude of the effect for this group is a striking 0.77 SD increase in well being. One possible interpretation is that immigrant students have a less dense network of friendships, hence felt more isolated during the lockdown: they were prevented from meeting classmates in school and they may not have been included in conversations that were happening online through WhatsApp or other groups. In fact, among students in the control group, immigrants have on average a 0.52 SD lower well-being compared to natives. Meeting regularly with a tutor proved particularly beneficial for the psychological well being of these students.

Next, we examine impact heterogeneity by socioeconomic status of the family, as mea-

[32]We did prepare a module on how to teach students with learning disorders as part of our online support for tutors, so this was probably a useful resource for them. Indeed, we know from the endline survey that 62 percent of tutors assigned to students with learning disorders watched the videos compared to 47 percent of other tutors. They were also 12 percent more likely to participate in the group meetings for tutors organized by the pedagogical experts and 20 percent more likely to ask for a one-on-one meeting with an expert to get recommendations on how to effectively help their student.

sured in the parents' questionnaire. We focus on three characteristics: education (less than high school, high school diploma, or higher); type of employment (none, blue collar job, or white collar job); and whether at least one parent worked from home during the lockdown. The results are reported in Figure 4.[33]

[Insert Figure 4]

While the effects are not very precisely estimated, it appears that the gains in academic performance are concentrated among students whose parents have a high school degree or less (which is the case for about 90 percent of our sample, anyway). Treated students whose parents did not complete high school (45 percent of the sample) also appear to have benefited more from the program in terms of aspirations and socio-emotional skills. Impacts on psychological well-being are constant across the parents' education gradient.

Turning to parental employment, we see that having a tutor improved students' performance significantly more for the children of blue collar mothers compared to white collar ones. These children also benefit significantly more in terms of happiness and reduced depression.[34]

What seems to emerge as a consistent pattern, at least qualitatively, is that TOP had a bigger impact on students whose parents both worked outside the home. These are the students who may have received the least support from parents in terms of schoolwork, and also may have been monitored less during distance learning. Regular meetings with a tutor had particular relevance in these cases.

5.4 Tutor characteristics

After discussing impact heterogeneity in terms of students' and parents' characteristics, we investigate whether tutors' characteristics played a significant role in explaining the effects of the program.

[Insert Table VIII]

In Table VIII we explore three sets of tutor baseline characteristics: gender, academic performance, and pro-social attitudes. For each of these characteristics, we report the

[33]The Figure shows the effect using mothers' education and occupation. The results are quantitatively and qualitatively very similar using fathers' education and occupation.

[34]The category 'no job' typically shows an intermediate coefficient, which may result from the fact that it pools families where the parent is (involuntarily) unemployed and families that have chosen to keep one parent (typically the mother) at home. Also, parents who do not work can stay home and help their children with homework, as we discuss shortly.

coefficient of the treatment dummy interacted with the relevant subgroup, and a p-value for the null hypothesis that the two coefficients are the same.

Panel A shows that the effect of TOP on our four main outcomes of interest did not differ on the basis of the gender of the tutor. Also when we distinguish possible combinations of gender of the tutor and gender of the student (panel B), we fail to detect significant pairwise differences: same gender pairs do not perform significantly better or worse than mixed gender ones.

We also find that tutors' GPA (a proxy for their academic ability) did not significantly affect the impact of the program: treated students benefited equally from interacting with a tutor above and below the median GPA in their faculty.[35]

One possibility is that what matters is not how well a tutor does in his/her university exams, but rather the type of program they are enrolled in. For example, tutors who are enrolled in a STEM degree may be more effective if the subject which the student needs help is math, rather than Italian, etc. In Appendix Table A.XIV we estimate the impact of the program on students' performance, disaggregating students' average test score into separate scores for math (top panel), Italian (intermediate panel) and English (bottom panel). The sample only includes treated students, since by design tutor characteristics are only available for students who are in the TOP program. For each subject, we consider two proxies for tutors' proficiency in that subject. The first is whether the tutor expressed a preference for that subject when they signed up for our program (variable 'Volunteer in [subject]' in the table). The second proxy is an objective measure, specifically: being enrolled in a STEM degree, for math performance; being enrolled in a Humanities degree, for Italian; and having an international certification in English (e.g., TOEFL, IELTS, etc.) for English. Interestingly, we do not find significant differences for math and Italian, nor a consistent pattern, when we look at what they volunteered to teach and the faculty they are enrolled in. For English, both proxies point to a higher effectiveness of more 'competent' tutors, although neither difference is statistically significant.

Finally, the last two panels of Table VIII capture tutors' pro-social attitudes and motivation. We compare the impact of tutors with and without previous volunteering experience (panel D) and of tutors who, when asked at baseline what motivated them to take part in the project, replied "To make myself useful" (variable 'Help others' in panel E). Note that our tutors are generally highly pro-social: 82 percent had previous experience as a volunteer and 83 percent joined TOP to be useful to others. For this reason, it is not too surprising that we do not detect significant differences in the outcomes of students

[35] We standardize the GPA within faculty to account for potential differences in grading criteria, number of credits, etc. across programs.

who were assigned different types of tutors. The one outcome in which tutors' motivation seems to make a difference is aspirations (column 2), where the positive impact is entirely driven by the more pro-social tutors.

5.5 Heterogeneous Treatment Effects using Causal Forest

To complement the above analysis with a more systematic approach, we estimate heterogeneous treatment effects using a causal forest algorithm. We follow Athey and Imbens (2016) and Wager and Athey (2018) and apply their method to understand who benefits most from the tutoring. Online Appendix C describes our methodology in detail.

In a nutshell, we estimate the Conditional Average Treatment Effect (CATE), including in the causal forest demographics (e.g., gender, immigrant dummy, parental education and occupation, etc.) and other controls (e.g., school grades and interest for different subjects, familiarity with computers, etc.). We use the predictions on the expected treatment effect for each individual, given the covariates, to investigate treatment heterogeneity. We divide the sample in two groups: top and bottom half of the predictions.

Appendix Table A.XV reports the mean of each baseline characteristics for the students above and below the median of predicted impact. Overall the results are consistent with our analysis in the previous sub-sections. Students who have learning disorders, lower initial grades and parents with less skilled occupations (e.g., blue collar mothers) are over-represented among the students with the highest predicted impact on performance. Immigrants and students with blue collar mothers are over-represented among the students with the highest enhancement in their well-being.

Overall, the most disadvantaged children seem to have benefited the most from the tutoring. However, heterogeneity depending on parents' or students' characteristics is not stark. It is worth emphasizing that the sample of students included in TOP had been already selected by school principals and teachers on the basis of their being most in need of the tutoring intervention. This may have potentially led to lower heterogeneity in treatment effects.

Finally, Appendix Table A.XVI reports the characteristics of the most and least effective tutors for all four main outcomes.[36] Overall, also for tutor characteristics, we do not find evidence of strong heterogeneity based on observable characteristics.

[36]In this table we restrict the sample to treated students, as students in the control group are not assigned a tutor.

6 The Impact of TOP on Tutors

The primary purpose of the project was to improve outcomes for students who where the direct beneficiaries of the intervention. However, the volunteering experience of being a TOP tutor may have affected tutors' capacity to empathize, as well as their perception of the relative importance of hard work versus luck to achieve success in life. We collected a short questionnaire (as described in Section 3.3) from volunteers who applied to the TOP tutoring program, independently on whether they were assigned a student or not. As mentioned in Section 2.3.2, the assignment of tutors to students was random, conditional on a set of baseline characteristics used for the allocation of tutors (e.g., subject and time availability). Around half of the respondents who completed the endline questionnaire had been randomly assigned to a student. Appendix Table A.VIII shows that the characteristics of tutors who experienced TOP and who did not experience TOP are overall balanced once we take into account the allocation criteria.

[Insert Table IX]

Table IX shows the impact of tutoring on the two key outcomes (the Empathy index and the Hard Work index) described in Section 3.3, when controlling for the factors used in the assignment of tutors to students (time and subject availability, previous training and tutoring experience, and regular enrollment in university).

We find that participating in TOP increased tutors' empathy by 3.4 percentage points, a 0.27 SD increase compared to volunteers who were not assigned a student to supervise. We do not find any economically or statistically significant effect on tutors' perceptions of the role of hard work to achieve success in life.[37]

Finally, we also asked tutors whether they where satisfied with their tutoring experience and whether they would be interested in volunteering again during the following academic year.[38] Appendix Table A.XVIII shows the baseline characteristics of students and tutors correlated with a higher level of satisfaction (column 1-2) and willingness to tutor again (column 3-4). Tutors matched to students whose parents have a bluecollar job report higher satisfaction and willingness to repeat the experience. Being enrolled in STEM (and

[37]For completeness, we present the ordered logit results for the individual questions used to build the empathy and hard work indexes in Appendix Table A.XVII. Panel A shows the results using the endline collected in September, six months after the beginning of the intervention, while Panel B shows the results when imputing the value for the first endline collected in June for those tutors who did not complete the second endline. The results are qualitatively and quantitatively very similar with both samples.

[38]The original scale for the response on satisfaction was from 1 to 5, where 1 indicated "not at all satisfied" and 5 "very satisfied". For the question on whether tutors would like to volunteer during the following academic year the answer was from 1 to 3, where 1 indicates "no", 2 "I need to think about it", and 3 "yes".

marginally in Economics) negatively correlates with the intention to tutor for an extra year, possibly due to exam/program requirements rather than dissatisfaction (in fact, column 1 shows that students from Economics report higher satisfactions than others). Similarly, higher GPA is negatively associated with the intention to tutor for an extra year, but not with satisfaction.

7 Conclusions

School closure due to the COVID-19 outbreak has created massive learning losses and adverse psychological effects for children, especially the most vulnerable and those from low socioeconomic background (Agostinelli et al., 2020; Azevedo et al., 2020; Orgilés et al., 2020; Golberstein et al., 2020). In this paper, we show that online tutoring can be an effective tool to help students during the pandemic, improving their academic outcomes but also their psychological well-being and development of socio-emotional skills.

We exploit over-subscription by schools and students to an innovative online tutoring program in Italy, "TOP", to evaluate its impact using a randomized control trial. We find that the one-on-one support provided virtually by volunteer university students for around 5 weeks increased performance in a standardized test by 0.26 SD, psychological well-being by 0.17 SD, and aspirations and socio-emotional skills by 0.15 and 0.14 SD, respectively.

In-person tutoring, especially when implemented by professionals and teachers and/or for several days per week, has proved highly effective in several contexts (Nickow et al., 2020; Fryer Jr, 2017). However, these programs are widely viewed as "too costly to be undertaken on a large scale" (Ander et al., 2016). Our TOP intervention allows to achieve sizeable results on learning and other life outcomes, keeping the costs extremely contained. The program leverages volunteer university students as tutors, mainly moved by intrinsic motivation and supported by a team of pedagogical experts. Volunteer tutors represent a viable and effective solution to reach a large number of students in need of support. The overall cost of the program per pupil was around 50 euro, covering the organizational and pedagogical support.[39]

Even when schools re-open after the COVID-19 outbreak, virtual tutoring implemented by volunteer university students may provide an effective tool to help vulnerable children and prevent inequalities to emerge, in a cost-effective way. Indeed, the design of the intervention easily adapts to 'normal' school times, when the role of tutors may be that of

[39]This excludes the research costs, namely the incentives to complete the endline survey for families, and the salaries of the enumerators who supervised the endline test score data collection.

helping to target learning at the right level for students. Future evidence from other countries and time periods may help better understand the scope for exploiting this versatile educational tool.

References

Agcom (2019). Educare digitale lo stato di sviluppo della scuola digitale un sistema complesso ed integrato di risorse digitali abilitanti. *Studio del Servizio Economico-Statistico Agcom*.

Agostinelli, F., Doepke, M., Sorrenti, G., Zilibotti, F., et al. (2020). When the great equalizer shuts down: Schools, peers, and parents in pandemic times. *IZA Discussion Paper No. 13965*.

Alan, S., Boneva, T., and Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics*, 134(3):1121–1162.

Ander, R., Guryan, J., and Ludwig, J. (2016). Improving academic outcomes for disadvantaged students: Scaling up individualized tutorials. *The Hamilton Project – Brookings*.

Angrist, N., Bergman, P., and Matsheng, M. (2020). School's out: Experimental evidence on limiting learning loss using "low-tech" in a pandemic. *NBER Working Paper*, (w28205).

Ashraf, N., Bandiera, O., and Jack, B. K. (2014). No margin, no mission? a field experiment on incentives for public service delivery. *Journal of Public Economics*, 120:1–17.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Azevedo, J. P., Hasan, A., Goldemberg, D., Iqbal, S. A., and Geven, K. (2020). Simulating the potential impacts of covid-19 school closures on schooling and learning outcomes: A set of global estimates. *Policy Research Working Paper Series 9284, The World Bank*.

Bacher-Hicks, A., Goodman, J., and Mulhern, C. (2020). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, 193:104345.

Bandura, A., Freeman, W., and Lightsey, R. (1999). *Self-efficacy: The exercise of control*. Springer.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., and Walton, M. (2015). Teaching at the right level: Evidence from randomized evaluations in India. *NBER Working Paper*, 22746.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Carlana, M., La Ferrara, E., and Pinotti, P. (2021). Goals and gaps: Educational careers of immigrant children. *Econometrica (Forthcoming)*.

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M., et al. (2020). How did covid-19 and stabilization policies affect spending and employment? A new real-time economic tracker based on private sector data. *NBER Working Paper*.

Coie, J. D. and Krehbiel, G. (1984). Effects of academic tutoring on the social status of low-achieving, socially rejected children. *Child Development*, pages 1465–1478.

Cook, P. J., Dodge, K., Farkas, G., Fryer, R. G., Guryan, J., Ludwig, J., Mayer, S., Pollack, H., and Steinberg, L. (2015). Not too late: Improving academic outcomes for disadvantaged youth. *Institute for Policy Research Northwestern University Working Paper WP-15*, 1.

Davis, J. and Heller, S. B. (2017). Using causal forests to predict treatment heterogeneity: An application to summer jobs. *American Economic Review*, 107(5):546–50.

Doyle, O. (2020). Covid-19: Exacerbating educational inequalities? *Working Paper*.

Duckworth, A. L. and Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment*, 91(2):166–174.

Engzell, P., Frey, A., and Verhagen, M. D. (2020). Learning inequality during the covid-19 pandemic. *SocArXiv*.

Escueta, M., Quan, V., Nickow, A. J., and Oreopoulos, P. (2017). Education technology: An evidence-based review. *NBER Working Paper*.

Frühe, B., Allgaier, A.-K., Pietsch, K., Baethmann, M., Peters, J., Kellnar, S., Heep, A., Burdach, S., von Schweinitz, D., and Schulte-Körne, G. (2012). Children's depression screener (ChilD-S): development and validation of a depression screening instrument for children in pediatric care. *Child Psychiatry & Human Development*, 43(1):137–151.

Fryer Jr, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of Economic Field Experiments*, volume 2, pages 95–322. Elsevier.

Gauri, V., Jamison, J. C., Mazar, N., and Ozier, O. (2019). Motivating bureaucrats through social recognition: External validity—a tale of two states. *Organizational Behavior and Human Decision Processes*.

Genicot, G. and Ray, D. (2017). Aspirations and inequality. *Econometrica*, 85(2):489–519.

Golberstein, E., Wen, H., and Miller, B. F. (2020). Coronavirus disease 2019 (covid-19) and mental health for children and adolescents. *JAMA pediatrics*.

Grewenig, E., Lergetporer, P., Werner, K., Woessmann, L., and Zierow, L. (2020). Covid-19 and educational inequality: How school closures affect low-and high-achieving students. *CESifo Working Paper 8648*.

Hardt, D., Nagler, M., and Rincke, J. (2020). Can peer mentoring improve online teaching effectiveness? an rct during the covid-19 pandemic. *CESifo Working Paper 8671*.

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

La Ferrara, E. (2019). Aspirations, social norms, and development. *Journal of the European Economic Association*, 17(6):1687–1722.

Levitt, S. D., List, J. A., Neckermann, S., and Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, 8(4):183–219.

Ludwig, J., Mullainathan, S., and Spiess, J. (2017). Machine-learning tests for effects on multiple outcomes. *arXiv preprint arXiv:1707.01473*.

Malamud, O. and Pop-Eleches, C. (2011). Home computer use and the development of human capital. *The Quarterly Journal of Economics*, 126(2):987–1027.

Maldonado, J. and De Witte, K. (2020). The effect of school closures on standardised student test. *FEB Research Report Department of Economics*.

Malkus, N. (2020). School districts' remote-learning plans may widen student achievement gap. *Education Next*, 20(3).

Nickow, A., Oreopoulos, P., and Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence. *NBER Working Paper*, (w27476).

Orgilés, M., Morales, A., Delvecchio, E., Mazzeschi, C., and Espada, J. P. (2020). Immediate psychological effects of the covid-19 quarantine in youth from italy and spain. *PsyArXiv*.

Psacharopoulos, G., Collis, V., Patrinos, H. A., and Vegas, E. (2020). Lost wages: The covid-19 cost of school closures. *Available at SSRN 3682160*.

Richmond, A. D. (2015). Academic task avoidance and achievement as predictors of peer status during the early primary school years. *Educational Psychology*.

Rosenthal, R. (1973). The pygmalion effect lives. *Psychology Today*.

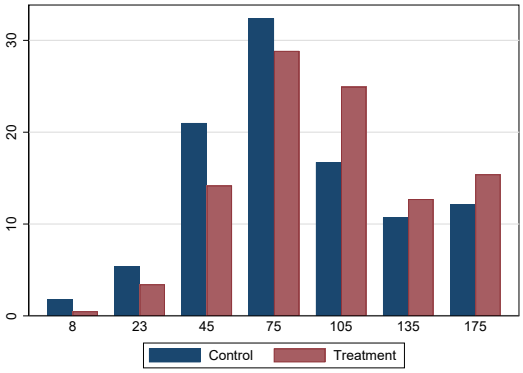
Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, 80(1):1.

UNESCO (2020). Covid-19 impact on education.

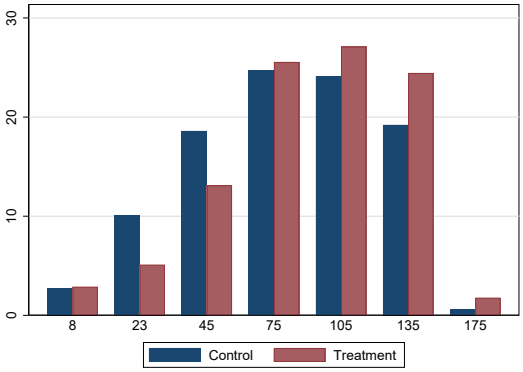
Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Figure 1. Time devoted to HW by students at endline

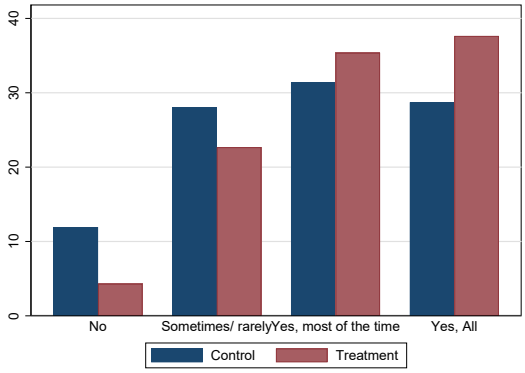
(a). Students



(b). Parents

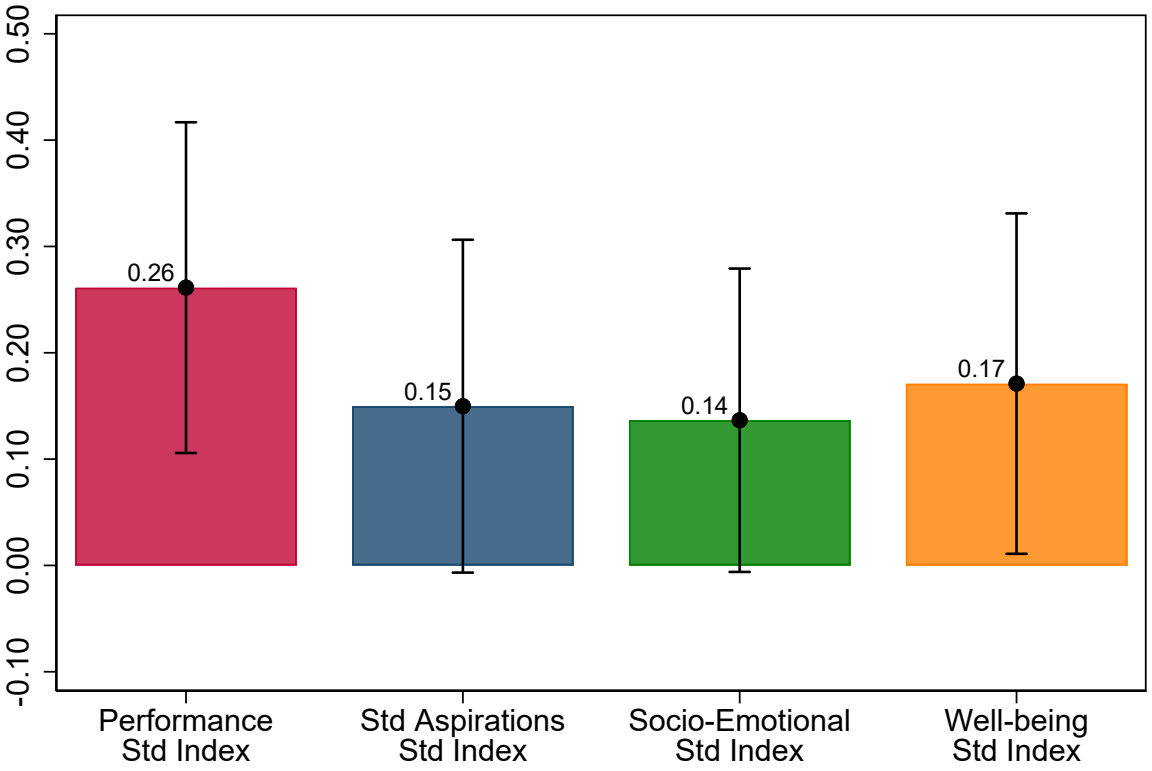


(c). Teachers



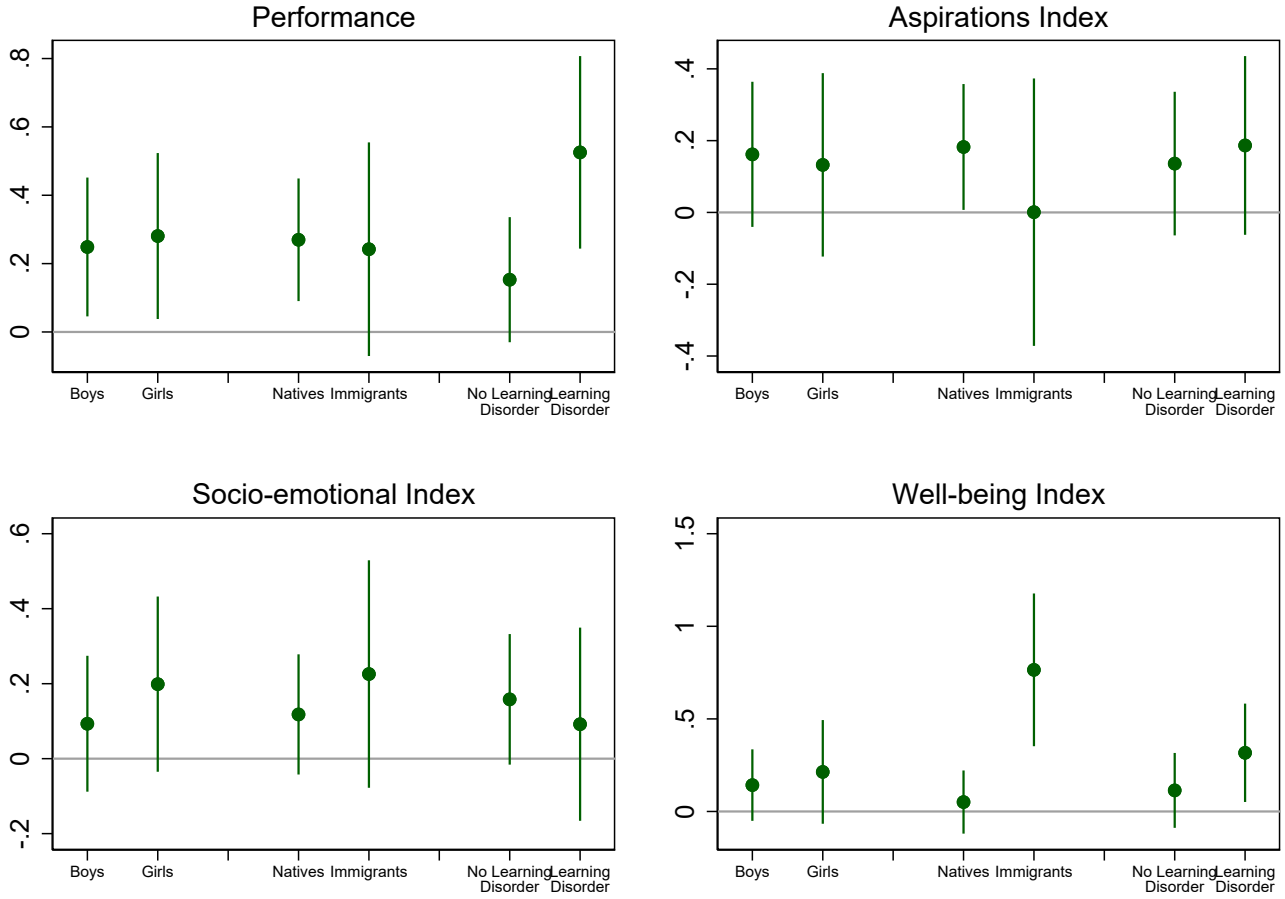
Notes: This Figure shows the time devoted to homework for students in the treatment and control group, as reported by students (panel a), parents (panel b), and teachers (panel c).

Figure 2. Summary of main results



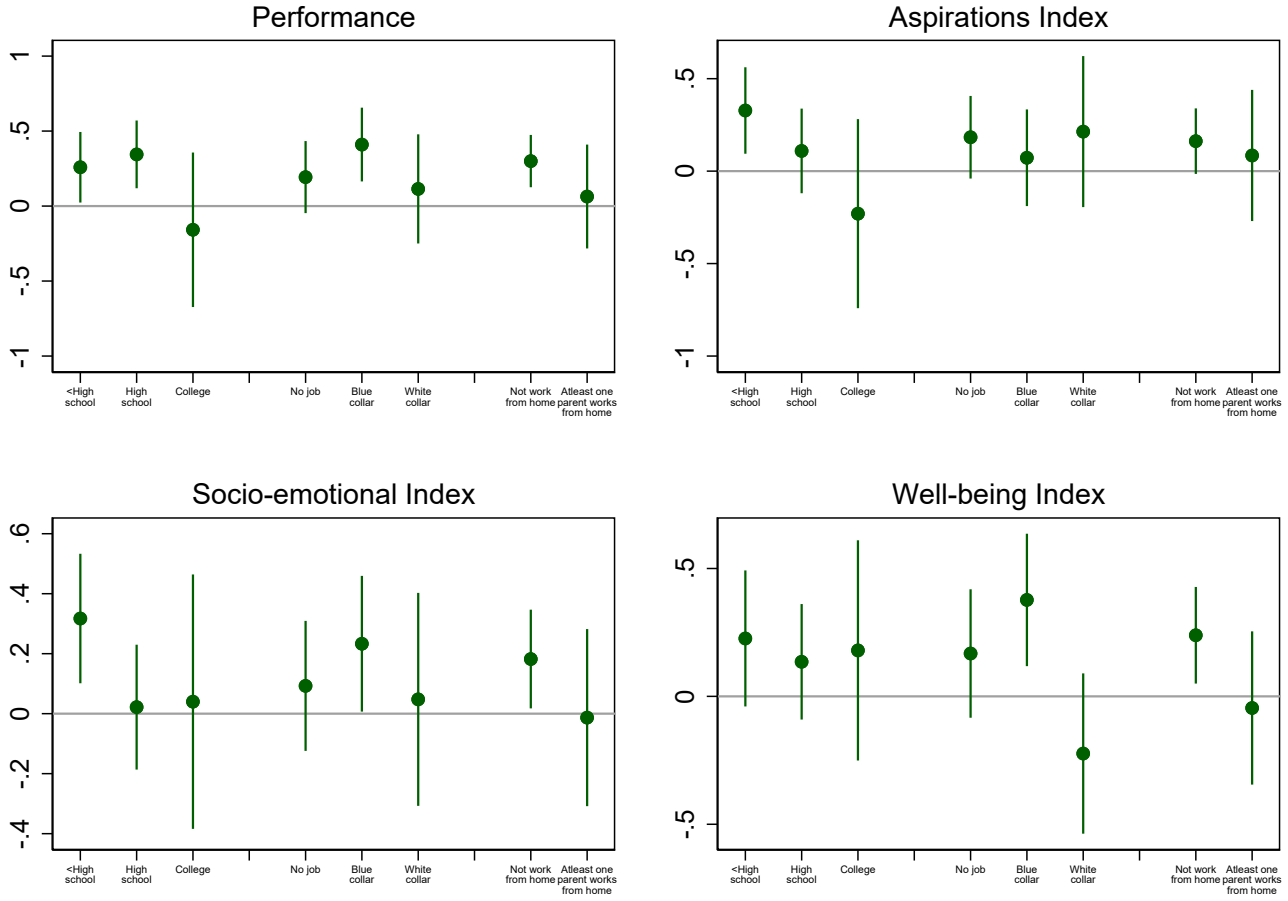
Notes: This Figure reports four OLS estimates of the assignment to the TOP tutoring treatment on the main outcomes in the paper, reported in the headings.The bar shows 95% confidence intervals. The mean of the control group for each index is 0 and the standard deviation is 1. Controls included are the same as in Table II.

Figure 3. Heterogeneity on main indices by student characteristics



Notes: This Figure reports OLS estimates of the assignment to the TOP tutoring treatment by student characteristics. Randomization round fixed effects included in all regressions. Ex-ante student baseline controls include gender, immigrant, grade, parental education for each parent, employment type for each parent, learning disability, interest for the different subjects, perseverance, importance of luck, and familiarity with computers. The mean of the control group for each index is 0 and the standard deviation is 1. The bar shows 95% confidence intervals.

Figure 4. Heterogeneity on main indices by mothers' characteristics



Notes: This Figure reports OLS estimates of the assignment to the TOP tutoring treatment by parent characteristics. Randomization round fixed effects included in all regressions. Ex-ante student baseline controls include gender, immigrant, grade, parental education for each parent, employment type for each parent, learning disability, interest for the different subjects, perseverance, importance of luck, and familiarity with computers. The mean of the control group for each index is 0 and the standard deviation is 1. The bar shows 95% confidence intervals. The results are similar using fathers' characteristics.

Table I. Balance Table

	Control	Treatment	P-value	Std diff.
Students				
Male	0.620	0.576	0.270	-0.089
Immigrant	0.237	0.223	0.662	-0.033
Learning disorders	0.322	0.319	0.459	-0.006
Grade in Math	6.282	6.274	0.847	-0.007
Grade in Italian	5.939	5.906	0.456	-0.025
Grade in English	6.318	6.225	0.549	-0.067
Grade 6	0.367	0.317	0.677	-0.106
Grade 7	0.335	0.336	0.914	0.002
Grade 8	0.298	0.347	0.763	0.104
How much do you like Math?	2.788	2.704	0.553	-0.081
How much do you like Italian?	2.943	3.077	0.056	0.153
How much do you like English?	3.114	3.120	0.389	0.005
Perseverance	0.816	0.814	0.805	-0.005
Importance Luck (vs. Effort)	0.058	0.056	0.971	-0.011
Familiarity with computers	3.163	3.101	0.606	-0.062
Parents				
Child lives with single parent	0.249	0.212	0.262	-0.089
Edu Mother: High-School	0.490	0.428	0.117	-0.124
Edu Mother: Degree	0.110	0.099	0.625	-0.036
Edu Father: High-School	0.359	0.364	0.899	0.010
Edu Father: Degree	0.102	0.062	0.056	-0.151
Mother has blue collar job	0.363	0.373	0.807	0.021
Mother has white collar job	0.163	0.161	0.927	-0.005
Father has blue collar job	0.584	0.520	0.107	-0.128
Father has white collar job	0.224	0.212	0.701	-0.029
Observations	245	467		

Notes: This Table shows the characteristics of treated and control students in the final sample used for the analysis of performance. P-values for difference in means are reported in the third column. The last column also reports the standardized difference between group averages. In the Appendix Table A.II, we report the same characteristics for the entire sample of 529 control students and 530 treated students.

Table II. Estimation of the impact of TOP tutoring on online classes and homework

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Students				Parents		Teachers		
	Time devoted to HW	Attendance Online Classes	Difficult Online Classes	Like Subjects	Time devoted to HW	Attendance Online Classes	HW	Attendance Online Classes	Issues Behavior
Treatment	9.884***	0.018	-0.081**	0.049**	8.714***	-0.004	0.611***	0.094***	-0.064**
	(3.436)	(0.029)	(0.035)	(0.019)	(2.795)	(0.024)	(0.134)	(0.032)	(0.026)
	[0.022]	[0.746]	[0.059]	[0.052]	[0.014]	[0.866]	[0.000]	[0.022]	[0.052]
Mean Dep:	88.26	0.83	0.79	0.28	81.71	0.88	2.77	0.57	0.83
Obs	690	687	688	688	778	777	852	859	841
R^2	0.107	0.072	0.043	0.318	0.089	0.083	0.049	0.108	0.126

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading. Coefficients estimated with OLS, except for col. 7 where it is ordered logit. Randomization round fixed effects included in all regressions. Baseline controls included but not shown: gender, immigrant, grade, parental education for each parent, employment type for each parent, learning disability, interest for the different subjects, perseverance, importance of luck, and familiarity with computers. Time devoted to homework (HW) by students (col. 1) and parents (col. 5) is reported in minutes, while by teachers (col. 7) on a scale from 1 "Never" to 4 "Always". "Like subjects" (col. 4) is the average of three indicators for whether the student likes math, Italian, and English. "Mean Dep" at the bottom of the table is the mean of the dependent variables for students in the control group.

Table III. Estimation of the impact of TOP tutoring on academic outcomes and self-perception

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Performance	Beliefs			Overconfidence			Grade	
		Student	Parent	Teacher	Student	Parent	Teacher	Student	Teacher
All subject									
Treatment	0.045***	0.028**	0.019*	0.039***	-0.052	-0.044	0.000	0.250**	0.335***
	(0.014)	(0.012)	(0.010)	(0.014)	(0.039)	(0.038)	(0.048)	(0.105)	(0.126)
	[0.013]	[0.084]	[0.174]	[0.040]	[0.418]	[0.447]	[0.998]	[0.084]	[0.060]
Mean Dep:	0.53	0.65	0.69	0.48	0.68	0.72	0.37	6.15	5.49
Obs	712	705	756	792	705	618	520	680	792
R^2	0.112	0.153	0.158	0.150	0.053	0.073	0.065	0.188	0.162

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading and it refers to the average for the three subjects of math, Italian and English. Controls included are the same as in Table II. The number of observations is lower in cols. 6-7 compared to cols. 3-4 because we restrict the sample to students and parents/teachers who both completed the endline surveys. "Mean Dep" at the bottom of the table is the mean of the dependent variables for students in the control group.

Table IV. Estimation of the impact of TOP tutoring on aspirations and self-efficacy

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Students				Parents		Teachers
	Aspirations	Self-efficacy	High-School Goal		Aspirations	Self-efficacy	Aspirations
	University	University	Vocation	Top Academic	University	University	University
Treatment	0.041	0.037	-0.060*	-0.002	0.017	0.055*	0.019
	(0.036)	(0.032)	(0.034)	(0.027)	(0.033)	(0.032)	(0.023)
	[0.720]	[0.720]	[0.407]	[0.930]	[0.841]	[0.407]	[0.794]
Mean Dep:	0.36	0.21	0.31	0.16	0.34	0.29	0.14
Obs	674	682	681	681	765	772	839
R^2	0.173	0.147	0.161	0.161	0.175	0.163	0.163

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading. Controls included are the same as in Table II. "Mean Dep" at the bottom of the table is the mean of the dependent variables for students in the control group.

Table V. Estimation of the impact of TOP tutoring on socio-emotional skills and well-being

	(1)	(2)	(3)	(4)	(5)
Panel A: Socio-Emotional Skills					
	Logic Task		Grit		Locus of Control
	Difficult	Give-up	Child	Parent	Child
Treatment	0.038	-0.040	0.012	-0.010	0.021**
	(0.037)	(0.025)	(0.010)	(0.010)	(0.009)
	[0.521]	[0.320]	[0.521]	[0.521]	[0.077]
Mean Dep:	0.56	0.14	0.68	0.67	0.71
Obs	685	685	673	736	685
R^2	0.143	0.180	0.142	0.145	0.120
		(1)	(2)	(3)	(4)
Panel B: Well-being					
		Reported from			
		Students		Parents	
		Depression	Happiness	Depression	Happiness
Treatment		-0.019**	0.027	-0.010	0.035**
		(0.009)	(0.018)	(0.009)	(0.018)
		[0.146]	[0.261]	[0.254]	[0.146]
Mean Dep:		0.55	0.61	0.59	0.60
Obs		642	645	642	645
R^2		0.122	0.061	0.057	0.055

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading. Controls included are the same as in Table II. "Mean Dep" at the bottom of the table is the mean of the dependent variables for students in the control group.

Table VI. Heterogeneity by treatment intensity

	(1)	(2)	(3)	(4)
	Performance	Aspirations Index	Socio-emotional Index	Wellbeing Index
Treatment	0.199**	0.130	0.137*	0.151*
	(0.083)	(0.085)	(0.079)	(0.087)
Intense treatment (6h)	0.219*	0.062	0.008	0.071
	(0.119)	(0.118)	(0.108)	(0.117)
Treat+Intense Treat=0	0.001	0.107	0.175	0.062
Obs	712	523	636	614
R^2	0.117	0.321	0.212	0.080

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading. Controls included are the same as in Table II.

Table VII. Heterogeneity by whether student used phone for tutoring

	(1)	(2)	(3)	(4)
	Performance	Aspirations Index	Socio-emotional Index	Wellbeing Index
Panel A: Heterogeneous effects based on using phone reported by parents				
Treatment	0.272***	0.183**	0.122	0.177**
	(0.084)	(0.084)	(0.076)	(0.084)
Tutored on phone only	-0.048	-0.181	0.073	-0.030
	(0.111)	(0.113)	(0.112)	(0.131)
Treat+Tutored on phone==0	0.049	0.983	0.088	0.284
Obs	712	523	636	614
R^2	0.112	0.316	0.211	0.079
Panel B: Heterogeneous effects by technical isssues reported by tutors				
Treatment	0.313***	0.210**	0.143*	0.216**
	(0.093)	(0.092)	(0.087)	(0.093)
Technical issues during tutoring	-0.160	-0.141	-0.042	-0.142
	(0.108)	(0.106)	(0.101)	(0.107)
Treat+Technical Issues==0	0.139	0.508	0.286	0.478
Obs	712	523	636	614
R^2	0.115	0.316	0.211	0.082

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading. Controls included are the same as in Table II. The mean of technical issues for students in the treatment group is 0.336, while the mean of phone usage is 0.22.

Table VIII. Heterogeneity on main indices by tutor characteristics

	(1)	(2)	(3)	(4)
	Performance	Aspirations Index	Socio-emotional Index	Wellbeing Index
Panel A				
Male	0.200*	0.184*	0.164	0.192*
	(0.108)	(0.108)	(0.100)	(0.110)
Female	0.284***	0.138	0.126	0.165*
	(0.087)	(0.087)	(0.078)	(0.088)
p-value (diff.=0):	[0.444]	[0.661]	[0.695]	[0.798]
Panel B				
Female tutor, Female student	0.294**	0.095	0.150	0.144
	(0.132)	(0.138)	(0.127)	(0.149)
Male tutor, Male student	0.136	0.149	0.046	0.071
	(0.137)	(0.136)	(0.124)	(0.128)
Male Tutor, Female Student	0.264	0.239	0.335**	0.416**
	(0.177)	(0.180)	(0.170)	(0.203)
Female Tutor, Male Student	0.290**	0.167	0.114	0.173
	(0.114)	(0.114)	(0.100)	(0.108)
p-value (diff.=0):				
Female Tut, Male Stu - Female Tut, Female Stu	[0.982]	[0.691]	[0.826]	[0.876]
Female Tut, Male Stu - Male Tut, Male Stu	[0.265]	[0.895]	[0.566]	[0.423]
Female Tut, Male Stu - Male Tut, Female Stu	[0.900]	[0.736]	[0.265]	[0.296]
Panel C				
Low standardized GPA	0.319*	0.084	0.013	0.108
	(0.168)	(0.179)	(0.144)	(0.163)
High standardized GPA	0.390**	0.044	-0.038	0.042
	(0.180)	(0.195)	(0.162)	(0.187)
p-value (diff.=0):	[0.384]	[0.649]	[0.487]	[0.457]
Panel D				
No volunteer exp.	0.368***	0.021	0.238**	0.331***
	(0.135)	(0.139)	(0.115)	(0.122)
Volunteer exp.	0.235***	0.182**	0.112	0.135
	(0.082)	(0.082)	(0.076)	(0.085)
p-value (diff.=0):	[0.309]	[0.233]	[0.257]	[0.093]
Panel E				
Not help others	0.310**	-0.084	0.135	0.174
	(0.141)	(0.132)	(0.128)	(0.139)
Help others	0.246***	0.188**	0.138*	0.175**
	(0.082)	(0.082)	(0.075)	(0.084)
p-value (diff.=0):	[0.641]	[0.033]	[0.982]	[0.995]

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading. Controls included are the same as in Table II. GPA scores have been standardized for each faculty: education, humanities, stem, economics, other faculties. Additionally, we control for missing GPA values and faculty in the regression. The p-value reported is the p-value of this post-estimation test of the differences between the two groups (e.g., male vs. female tutors).

Table IX. Treatment Effect on Tutors' Outcomes

	(1)	(2)	(3)	(4)
	Emphaty Index		Hardwork Index	
TOP Tutors	0.034***	0.029**	-0.010	-0.015
	(0.012)	(0.012)	(0.012)	(0.012)
	[0.012]	[0.039]	[0.425]	[0.223]
Randomization controls:	Yes	Yes	Yes	Yes
Tutor controls:	No	Yes	No	Yes
Mean Dep:	0.67	0.67	0.65	0.65
Obs	740	740	735	735

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading . The randomization controls include whether the volunteer has tutoring experience and specific training (to support students with learning disorders or immigrants), their expertise in the subjects (math, Italian, English), their time availability (3 hours per week or 6 hours per week), whether they are on time in their university enrollment and if they confirmed their availability. The additional tutor controls include gender, university faculty, whether they are enrolled in a undergraduate or master, GPA, previous volunteering activities, whether they applied to TOP to help others (motivation), parental education, and familiarity with the computer. "Mean Dep" at the bottom of the table is the mean of the dependent variables for students in the control group.

Online Appendix - Not for publication

A Additional Tables and Figures

Figure A.1. Timeline of TOP

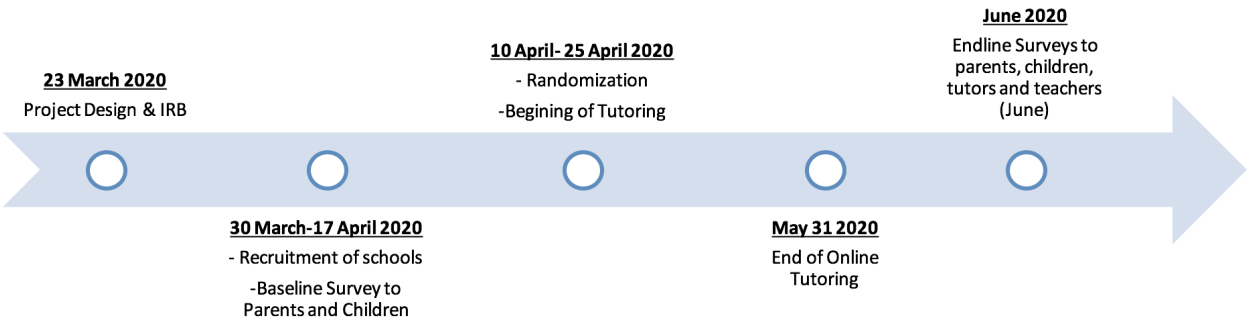
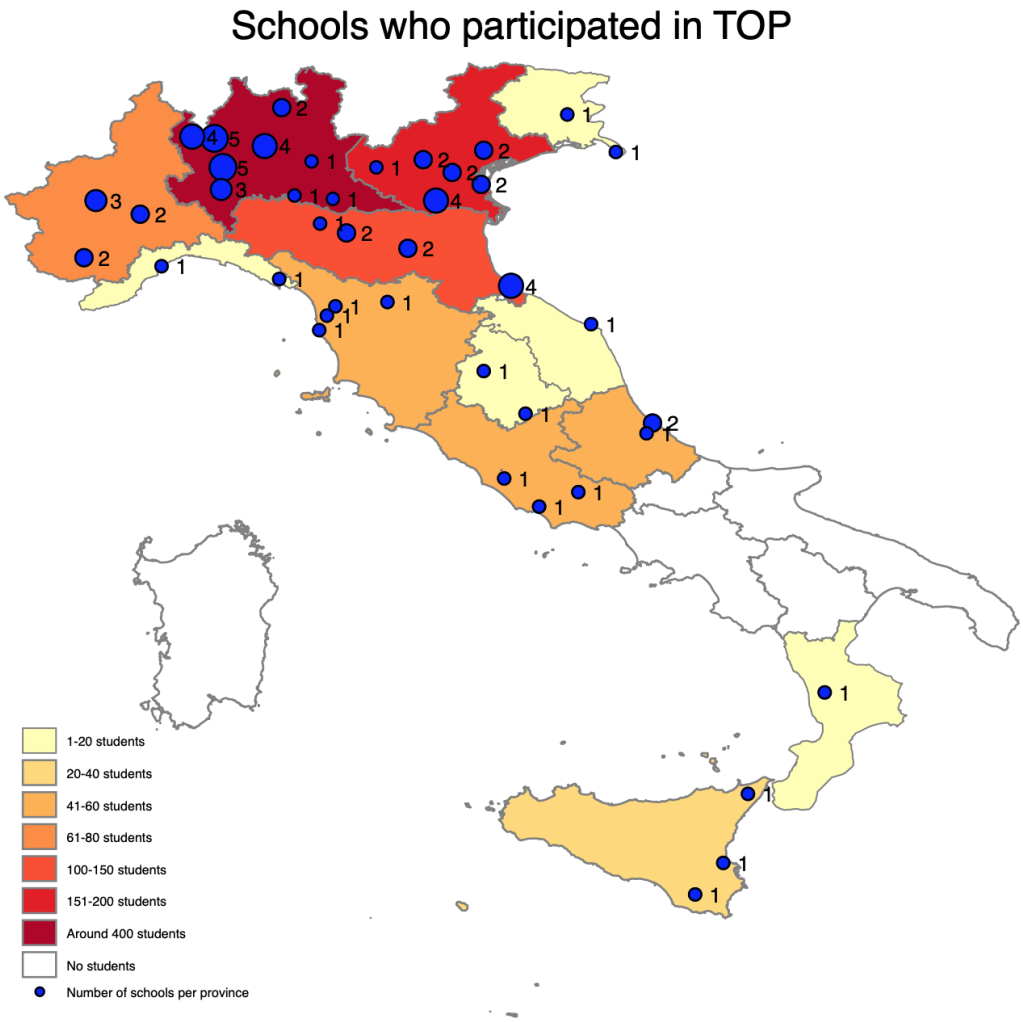
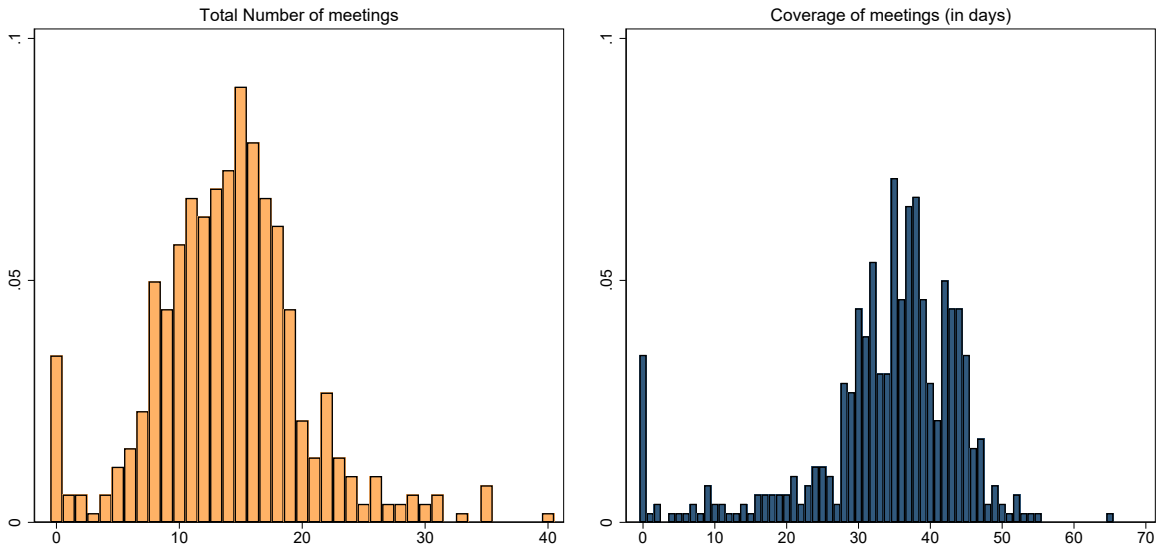


Figure A.2. Maps of schools and students participating in TOP



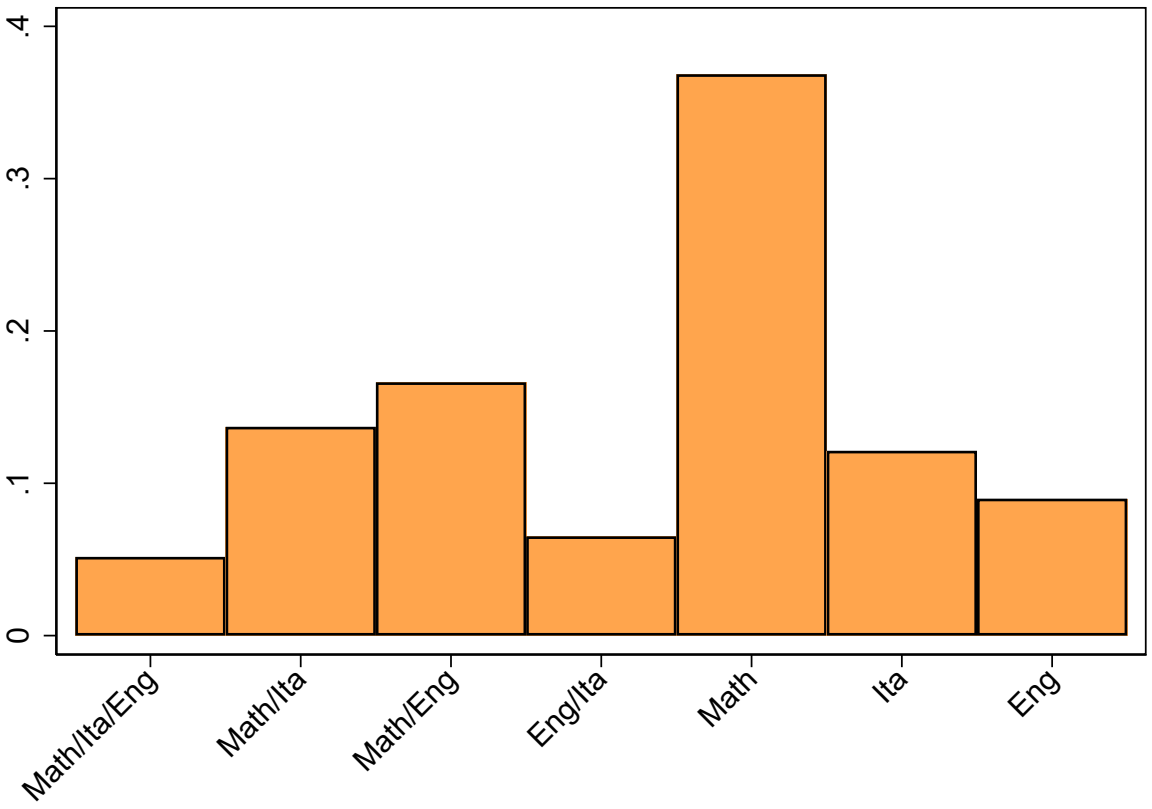
Notes: This Figure shows the total number of students involved in the research project of TOP for each Italian region. Darker colors indicate a higher number of students in the region. The size of the dots (and the number close to them) indicates the number of schools involved in the project for each province.

Figure A.3. Number of tutoring meetings and coverage in days of TOP



Notes: This Figure shows the total number of tutoring meetings (left panel) and the number of days from the beginning to the end of the tutoring (right panel). The data used in these graphs are reported by tutors in the registry. The Figure shows data from 522 treated students: 8 students did the tutoring but we do not have precise information from the tutors on the number of meetings.

Figure A.4. Main Subjects During Tutoring



Notes: This Figure shows the main subjects done during the tutoring meetings, as reported by tutors at endline.

Table A.I. Balance Table at province level

Variable	(1) No TOP	(2) TOP	(3) Diff.	(4) Std. Diff.
Macro-area: North	0.313	0.650	0.337***	0.501
	(0.467)	(0.483)	(0.095)	
Macro-area: Center	0.209	0.200	-0.009	-0.016
	(0.410)	(0.405)	(0.082)	
Macro-area: South and Islands	0.478	0.150	-0.328***	-0.529
	(0.503)	(0.362)	(0.091)	
Level of education: Elementary	0.305	0.287	-0.017***	-0.453
	(0.026)	(0.027)	(0.005)	
Level of education: Middle school	0.300	0.297	-0.003	-0.079
	(0.024)	(0.023)	(0.005)	
Level of education: Diploma	0.297	0.309	0.012**	0.354
	(0.028)	(0.020)	(0.005)	
Level of education: University	0.101	0.109	0.008**	0.296
	(0.015)	(0.023)	(0.004)	
Covid-19 cases March'20 (1000 inhabitants)	1.477	2.234	0.757*	0.266
	(1.799)	(2.199)	(0.391)	
Covid-19 cases April'20 (1000 inhabitants)	2.889	4.399	1.511**	0.333
	(3.024)	(3.376)	(0.631)	
Covid-19 cases May'20 (1000 inhabitants)	3.209	4.961	1.752**	0.343
	(3.418)	(3.800)	(0.712)	
Immigrants 2020	0.070	0.096	0.026***	0.592
	(0.033)	(0.030)	(0.006)	
Unemployment rate (2019)	11.678	8.316	-3.362***	-0.438
	(5.908)	(4.912)	(1.111)	
Observations	67	40	107	

Notes: This Table shows the characteristics of provinces that had at least one treated school ("TOP province") compared to provinces with no treated schools ("No TOP Province"). The first two columns show the mean for the two groups, the third column shows the difference in means, while the last column provides the standardized difference between group averages. In parenthesis, the first two columns show the standard deviations of the mean, while the third column shows the standard errors of the difference between treatment and control groups.

Table A.II. Balance Table (initial sample)

	Control	Treatment	P-value	Std diff.
Students				
Male	0.616	0.583	0.270	-0.067
Immigrant	0.204	0.215	0.662	0.027
Learning disorders	0.340	0.319	0.459	-0.045
Grade in Math	6.272	6.285	0.847	0.012
Grade in Italian	5.953	5.894	0.456	-0.046
Grade in English	6.278	6.228	0.549	-0.037
Grade 6	0.325	0.313	0.677	-0.026
Grade 7	0.338	0.342	0.914	0.008
Grade 8	0.336	0.345	0.763	0.019
How much do you like Math?	2.741	2.704	0.553	-0.036
How much do you like Italian?	2.983	3.087	0.056	0.118
How much do you like English?	3.060	3.119	0.389	0.053
Perseverance	0.800	0.806	0.805	0.015
Importance Luck (vs. Effort)	0.052	0.052	0.971	0.000
Familiarity with computers	3.074	3.106	0.606	0.032
Parents				
Child lives with single parent	0.253	0.217	0.164	-0.086
Edu Mother: High-School	0.456	0.425	0.309	-0.062
Edu Mother: Degree	0.110	0.104	0.757	-0.020
Edu Father: High-School	0.346	0.364	0.536	0.037
Edu Father: Degree	0.081	0.060	0.185	-0.079
Mother has blue collar job	0.391	0.379	0.687	-0.025
Mother has white collar job	0.157	0.160	0.877	0.008
Father has blue collar job	0.560	0.511	0.116	-0.098
Father has white collar job	0.219	0.215	0.869	-0.010
Observations	529	530		

Notes: This Table shows the characteristics of treated and control students in the initial sample at baseline. P-values for difference in means are reported in the third column. The last column also reports the standardized difference between group averages. This Table includes the entire baseline sample of students who was randomized to treatment and control groups. In the Appendix Table I, we report the same characteristics for the sample of students who completed the endline achievement test score.

Table A.III. Attrition between baseline and endline

Dependent variable: Dummy for endline completion				
	Full	Full	Treatment	Control
	Surveyed at endline	Surveyed at endline	Surveyed at endline	Surveyed at endline
	(1)	(2)	(3)	(4)
Treatment	0.418***	0.427***		
	(0.026)	(0.026)		
Learning disorders		-0.013	-0.012	-0.036
		(0.029)	(0.031)	(0.049)
Immigrant		0.053	0.017	0.066
		(0.033)	(0.034)	(0.058)
Male		-0.017	-0.039	0.003
		(0.027)	(0.030)	(0.047)
Edu Mother: High-School		0.018	0.004	0.032
		(0.030)	(0.034)	(0.053)
Edu Mother: Degree		-0.068	-0.083	-0.082
		(0.051)	(0.062)	(0.084)
Edu Father: High-School		0.014	0.010	0.015
		(0.029)	(0.031)	(0.051)
Edu Father: Degree		0.138**	0.106	0.176*
		(0.057)	(0.065)	(0.096)
Mother has blue collar job		-0.046	-0.045	-0.048
		(0.029)	(0.033)	(0.049)
Mother has white collar job		0.012	0.008	0.010
		(0.040)	(0.043)	(0.067)
Father has blue collar job		0.062**	0.041	0.076
		(0.031)	(0.034)	(0.056)
Father has white collar job		0.015	-0.009	0.030
		(0.041)	(0.048)	(0.069)
Grade 6		0.057*	0.021	0.094*
		(0.032)	(0.034)	(0.056)
Grade 7		0.014	-0.008	0.043
		(0.032)	(0.035)	(0.055)
How much do you like Math?		0.010	0.000	0.019
		(0.013)	(0.015)	(0.025)
How much do you like Italian?		-0.029*	-0.018	-0.055*
		(0.016)	(0.015)	(0.028)
How much do you like English?		0.007	0.002	0.014
		(0.013)	(0.015)	(0.022)
Perseverance		0.052	0.036	0.050
		(0.034)	(0.044)	(0.057)
Importance Luck (vs. Effort)		0.125*	0.113*	0.171
		(0.073)	(0.064)	(0.137)
Familiarity with computers		0.023	0.002	0.052**
		(0.015)	(0.018)	(0.025)
Dep var mean	.672	.672	.881	.463
Obs.	1059	1059	530	529
R^2	0.202	0.239	0.075	0.091

Notes: This Table reports the coefficients from a OLS regression. The outcome is a dummy which assumes value 1 if the student completed the endline survey. The sample is this Table includes all students who completed the baseline survey and therefore are included in the TOP randomization sample. Column 3 restricts the sample to students assigned to the treatment group, while column 4 to students assigned to the control group. Robust standard errors in parentheses. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively.

Table A.IV. Balance Table: 3 hours vs. 6 hours tutoring

	3h Tutoring	6h Tutoring	P-value	Std diff.
Students				
Male	0.570	0.559	0.829	-0.022
Immigrant	0.250	0.224	0.551	-0.061
Learning disorders	0.331	0.364	0.503	0.069
Grade in Math	6.232	6.238	0.960	0.006
Grade in Italian	5.820	5.867	0.730	0.036
Grade in English	6.222	6.014	0.158	-0.145
Grade 6	0.342	0.315	0.579	-0.057
Grade 7	0.338	0.315	0.629	-0.049
Grade 8	0.320	0.371	0.301	0.108
How much do you like Math?	2.680	2.748	0.527	0.064
How much do you like Italian?	3.067	3.098	0.734	0.035
How much do you like English?	3.151	2.951	0.075	-0.182
Perseverance	0.785	0.783	0.962	-0.005
Importance Luck (vs. Effort)	0.049	0.065	0.387	0.092
Familiarity with computers	3.081	3.098	0.871	0.017
Parents				
Child lives with single parent	0.246	0.210	0.399	-0.086
Edu Mother: High-School	0.384	0.497	0.026	0.228
Edu Mother: Degree	0.102	0.070	0.277	-0.115
Edu Father: High-School	0.352	0.357	0.927	0.010
Edu Father: Degree	0.060	0.056	0.871	-0.017
Mother has blue collar job	0.349	0.399	0.312	0.104
Mother has white collar job	0.176	0.112	0.084	-0.179
Father has blue collar job	0.518	0.524	0.894	0.012
Father has white collar job	0.211	0.175	0.375	-0.091
Observations	284	143		

Notes: This Table shows the characteristics of treated students randomly assigned to the 3 hours or 6 hours tutoring in the initial sample at baseline. The sample in this Table is restricted only to students identified by the school as in need for help in more than one subject (427 observations). P-values for difference in means are reported in the third column. The last column also reports the standardized difference between group averages.

Table A.V. Balance Table Tutors (sample of all tutors at baseline)

Variable	(1) No TOP	(2) TOP Tutor	(3) Diff.	(4) Std. Diff.
Number of modules	1.120	1.325	0.205	0.300
	(0.367)	(0.575)	(0.022)***	
Training Learning Disorders	0.003	0.038	0.036	0.179
	(0.051)	(0.192)	(0.005)***	
Training Immigrants	0.007	0.013	0.007	0.049
	(0.081)	(0.115)	(0.005)	
Subjects: Math	0.151	0.201	0.050	0.093
	(0.358)	(0.401)	(0.019)***	
Subjects: Eng	0.191	0.034	-0.157	-0.362
	(0.393)	(0.182)	(0.018)***	
Subjects: Ita	0.230	0.052	-0.178	-0.375
	(0.421)	(0.221)	(0.019)***	
Subjects: Math, Eng	0.071	0.149	0.078	0.177
	(0.257)	(0.357)	(0.014)***	
Subjects: Math, Ita	0.046	0.140	0.093	0.230
	(0.210)	(0.347)	(0.013)***	
Subjects: Ita, Eng	0.249	0.134	-0.116	-0.210
	(0.433)	(0.341)	(0.021)***	
Subjects: Math, Ita, Eng	0.063	0.291	0.228	0.443
	(0.242)	(0.454)	(0.016)***	
Tutoring before	0.717	0.958	0.241	0.488
	(0.450)	(0.201)	(0.020)***	
Female	0.723	0.700	-0.024	-0.037
	(0.447)	(0.459)	(0.023)	
Tutor's Major - STEM	0.158	0.337	0.179	0.299
	(0.365)	(0.473)	(0.020)***	
Tutor's Major - Humanities	0.158	0.138	-0.020	-0.040
	(0.365)	(0.345)	(0.018)	
Tutor's Major - Education	0.058	0.065	0.007	0.020
	(0.234)	(0.247)	(0.012)	
Tutor's Major - Economics	0.307	0.287	-0.021	-0.032
	(0.462)	(0.453)	(0.023)	
Degree Type: Bachelor's	0.510	0.472	-0.038	-0.053
	(0.500)	(0.500)	(0.025)	
GPA	26.895	26.727	-0.168	-0.054
	(2.196)	(2.228)	(0.113)	
Volunteering Before	0.777	0.822	0.045	0.080
	(0.417)	(0.383)	(0.021)**	
Hours studying per day	5.353	5.106	-0.247	-0.084
	(2.158)	(2.012)	(0.108)**	
Familiarity with computer	3.247	3.323	0.076	0.082
	(0.677)	(0.632)	(0.034)**	
Do you have younger siblings?	0.502	0.550	0.048	0.067
	(0.500)	(0.498)	(0.025)*	
Tutor's Mother: Completed high school	0.431	0.483	0.052	0.073
	(0.495)	(0.500)	(0.025)**	
Tutor's Mother: Completed College	0.423	0.387	-0.036	-0.052
	(0.494)	(0.487)	(0.025)	
Tutor's Father: Completed high school	0.407	0.437	0.030	0.043
	(0.492)	(0.497)	(0.025)	
Tutor's Father: Completed College	0.402	0.387	-0.015	-0.021
	(0.490)	(0.488)	(0.025)	
Motivation TOP: help others	0.790	0.831	0.041	0.075
	(0.407)	(0.375)	(0.020)**	
Observations	1,532	523	2,055	

Notes: Column 1 and 2 shows the mean for the control group and treatment group, respectively. Column 3 shows the difference between the treatment and the control group. The last column reports the standardized difference between group averages. The sample includes all tutors who applied to be part of TOP as volunteers. If a tutor was assigned but then decided not to participate, he/she is considered among the "No TOP" tutors.

Table A.VI. Balance Table Tutors (by availability 3h vs. 6h)

	All Tutors	3h Tutor	6h Tutor	P-value	Std diff.
Female	0.700	0.688	0.730	0.353	0.092
Born in Italy	0.983	0.992	0.957	0.007	-0.269
Faculty: Education	0.065	0.055	0.092	0.126	0.150
Faculty: Humanities	0.138	0.113	0.206	0.006	0.270
Faculty: Economics	0.287	0.327	0.177	0.001	-0.331
Faculty: STEM+ Medical	0.337	0.356	0.284	0.121	-0.152
English language certificate	0.666	0.646	0.721	0.104	0.159
University GPA	26.727	26.823	26.455	0.101	-0.165
Voluntering	0.822	0.817	0.837	0.594	0.052
Motivation TOP: help others	0.831	0.845	0.794	0.169	-0.136
Tutoring before	0.958	0.971	0.922	0.013	-0.244
Training Immigrants	0.013	0.003	0.043	0.000	0.348
Training Learning Disorders	0.038	0.037	0.043	0.755	0.031
Observations	523	382	141		

Notes: This Table shows the characteristics of all tutors (column 1), those who offered their availability for 3 hours of tutoring per week (column 2) vs. 6 hours of tutoring per week (column 3). P-values for difference in means between column 2 and 3 are reported in the fourth column. The last column reports the standardized difference between group averages.

Table A.VII. Summary statistics of main outcome variables

	count	mean	sd	min	max
1. Academic and Beliefs (average of all subjects)					
Outcomes reported by Students					
Performance	712	0.56	0.18	0.05	1.00
Beliefs	705	0.67	0.15	0.00	1.00
Overconfidence	705	0.64	0.48	0.00	1.00
Grade - Self Rate	680	6.29	1.38	1.00	10.00
Outcomes reported by Parents					
Beliefs	756	0.71	0.14	0.16	1.00
Overconfidence	618	0.70	0.46	0.00	1.00
Outcomes reported by Teachers					
Beliefs	792	0.49	0.20	0.00	1.00
Overconfidence	520	0.36	0.48	0.00	1.00
Grade	792	5.65	1.87	1.00	10.00
2. Aspirations					
Std Aspirations Index	523	0.07	1.00	-1.08	2.35
Outcomes reported by Students					
Aspirations: University	674	0.39	0.49	0.00	1.00
Self efficacy: University	682	0.23	0.42	0.00	1.00
High school goal: Vocational	681	0.28	0.45	0.00	1.00
High school goal: Top track	681	0.15	0.36	0.00	1.00
Outcomes reported by Parents					
Aspirations: University	765	0.35	0.48	0.00	1.00
Self efficacy: University	772	0.33	0.47	0.00	1.00
Outcomes reported by Teachers					
Aspirations: University	839	0.14	0.34	0.00	1.00
3. Socio-Emotional Skills					
Std Socio-Emotional Index	636	0.07	0.95	-2.90	2.66
Outcomes reported by Students					
Perseverance - Difficult	685	0.59	0.49	0.00	1.00
Perseverance - Give up	685	0.12	0.32	0.00	1.00
Grit	673	0.69	0.13	0.32	1.00
Locus of control	685	0.72	0.11	0.25	1.00
Outcomes reported by Parents					
Grit	736	0.67	0.13	0.25	1.00
4. Well-being					
Std Depression Index	614	0.10	0.95	-3.57	2.79
Outcomes reported by Students					
Depression	669	0.54	0.12	0.25	0.97
Happiness	665	0.62	0.22	0.00	1.00
Outcomes reported by Parents					
Depression	731	0.58	0.10	0.28	0.92
Happiness	741	0.62	0.21	0.00	1.00

Notes: This Table shows the summary statistics of all outcome variables, as reported in the endline questionnaire from students, parents, and teachers. All outcomes refer to children even when reported by parents or teachers. The table also includes the mean of the indices in the entire sample, standardized to have mean 0 and standard deviation 1 for the control group.

Table A.VIII. Balance Table Tutors (endline survey)

Variable	(1) Control	(2) Treat	(3) Diff.	(4) Std. Diff.
Female	0.723	0.710	0.020	-0.020
	(0.448)	(0.454)	(0.046)	
Tutor's Major - STEM	0.126	0.333	0.057	0.360
	(0.332)	(0.472)	(0.040)	
Tutor's Major - Humanities	0.172	0.140	0.017	-0.063
	(0.378)	(0.347)	(0.036)	
Tutor's Major - Education	0.053	0.068	0.049	0.044
	(0.224)	(0.252)	(0.025)**	
Tutor's Major - Economics	0.342	0.286	-0.073	-0.086
	(0.475)	(0.452)	(0.047)	
Degree Type: Bachelor's	0.541	0.471	-0.064	-0.099
	(0.499)	(0.500)	(0.052)	
GPA	27.195	26.744	-0.648	-0.149
	(2.065)	(2.213)	(0.226)***	
Volunteering Before	0.753	0.825	0.020	0.126
	(0.432)	(0.380)	(0.042)	
Hours studying per day	5.529	5.122	-0.174	-0.140
	(2.085)	(2.021)	(0.215)	
Familiarity with computer	3.296	3.319	-0.145	0.025
	(0.692)	(0.635)	(0.069)**	
Do you have younger siblings?	0.534	0.549	-0.023	0.022
	(0.499)	(0.498)	(0.052)	
Tutor's Mother: Completed high school	0.427	0.489	0.030	0.088
	(0.495)	(0.500)	(0.052)	
Tutor's Mother: Completed College	0.400	0.381	-0.011	-0.028
	(0.491)	(0.486)	(0.051)	
Tutor's Father: Completed high school	0.396	0.442	-0.031	0.066
	(0.490)	(0.497)	(0.052)	
Tutor's Father: Completed College	0.405	0.382	0.008	-0.033
	(0.491)	(0.486)	(0.051)	
Motivation TOP: help others	0.806	0.835	0.022	0.054
	(0.396)	(0.372)	(0.040)	
Observations	453	486	939	

Notes: Columns 1 and 2 show the means (and standard deviations) for tutor applicants who were not assigned a student (Control) and those who were assigned a student (treatment), respectively. Column 3 shows the coefficient and standard error of a regression that includes the controls we used to assign tutors to students, that is: weekly availability for 3 or 6 hours, specific training for learning disabilities, subject availability, tutoring experience, whether the tutor was born before 1994, and whether the tutor confirmed their availability. The sample includes all tutors who completed the endline survey.

Table A.IX. Estimation of the impact of TOP tutoring on academic outcomes and self-perception

	(1)	(2)	(3)	(4)	(5)	(6)
	Performance		Beliefs		Grade	
	Student	Student	Parent	Teacher	Student	Teacher
Panel A: Math						
Treatment	0.041**	0.027*	0.022*	0.040	0.365***	0.494**
	(0.017)	(0.014)	(0.013)	(0.024)	(0.126)	(0.214)
	[0.084]	[0.172]	[0.172]	[0.172]	[0.022]	[0.084]
Mean Dep:	0.65	0.66	0.67	0.46	5.93	5.29
Obs	712	704	746	355	679	355
R^2	0.167	0.151	0.201	0.184	0.274	0.171
Panel B: Italian						
Treatment	0.033*	0.022	0.002	0.020	0.176	0.168
	(0.018)	(0.015)	(0.011)	(0.019)	(0.124)	(0.177)
	[0.361]	[0.520]	[0.885]	[0.617]	[0.520]	[0.617]
Mean Dep:	0.46	0.64	0.72	0.50	6.32	5.59
Obs	712	701	743	439	679	439
R^2	0.077	0.116	0.167	0.197	0.171	0.213
Panel C: English						
Treatment	0.046**	0.037**	0.035***	0.058	0.218*	0.315
	(0.023)	(0.019)	(0.013)	(0.035)	(0.130)	(0.312)
	[0.220]	[0.220]	[0.053]	[0.264]	[0.264]	[0.327]
Mean Dep:	0.46	0.67	0.69	0.44	6.19	5.35
Obs	514	701	748	193	680	193
R^2	0.189	0.173	0.224	0.300	0.242	0.241

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading. Controls included are the same as in Table II. Unfortunately, in column 1 of Panel C, we have do drop the first 218 questionnaire completed (196 treated and 22 control) because the English test was automatically translated in Italian from own student devices in the surveys administered in the first few days. "Mean Dep" at the bottom of the table is the mean of the dependent variables for students in the control group.

Table A.X. Robustness checks with different sets of controls

	(1)	(2)	(3)	(4)	(5)	(6)
	No controls		LASSO controls		Inverse Probability Weighting	
	Coeff	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
1. Academic and Beliefs (average of all subjects)						
Std Performance	0.248***	0.081	0.246***	0.078	0.277***	0.080
Outcomes reported by Students						
Beliefs	0.026**	0.012	0.026**	0.011	0.031***	0.012
Overconfidence	-0.053	0.038	-0.053	0.038	-0.055	0.039
Grade - Self Rate	0.225**	0.109	0.199**	0.101	0.260**	0.107
Outcomes reported by Parents						
Beliefs	0.024**	0.010	0.024**	0.010	0.019*	0.010
Overconfidence	-0.030	0.039	-0.030	0.039	-0.052	0.038
Outcomes reported by Teachers						
Beliefs	0.035**	0.014	0.026*	0.014	0.037***	0.014
Overconfidence	-0.023	0.045	-0.023	0.045	-0.007	0.048
Grade	0.302**	0.131	0.284**	0.124	0.345***	0.128
2. Aspirations						
Std Aspirations Index	0.111	0.090	0.045	0.051	0.152*	0.079
Outcomes reported by Students						
Aspirations: University	0.039	0.038	0.022	0.027	0.040	0.036
Self efficacy: University	0.041	0.033	0.028	0.030	0.032	0.031
High school goal: Vocational	-0.047	0.036	-0.006	0.028	-0.057	0.035
High school goal: Top track	-0.012	0.028	0.005	0.020	0.002	0.026
Outcomes reported by Parents						
Aspirations: University	0.014	0.035	0.012	0.023	0.021	0.033
Self efficacy: University	0.058*	0.034	0.051*	0.030	0.059*	0.032
Outcomes reported by Teachers						
Aspirations: University	0.014	0.024	0.019	0.022	0.018	0.022
3. Socio-Emotional Skills						
Std Socio-Emotional Index	0.125	0.079	0.124*	0.069	0.159**	0.072
Outcomes reported by Students						
Perseverance - Difficult	0.038	0.039	0.023	0.036	0.037	0.038
Perseverance - Give up	-0.034	0.026	-0.035	0.025	-0.040	0.026
Grit	0.013	0.010	0.011	0.009	0.013	0.010
Locus of control	0.024***	0.009	0.021**	0.009	0.023***	0.009
Outcomes reported by Parents						
Grit	-0.005	0.010	-0.007	0.009	-0.007	0.010
4. Well-being						
Std Depression Index	0.156*	0.080	0.141*	0.076	0.214**	0.084
Outcomes reported by Students						
Depression	-0.019**	0.009	-0.020**	0.009	-0.022**	0.009
Happiness	0.023	0.018	0.023	0.018	0.033*	0.018
Outcomes reported by Parents						
Depression	-0.011	0.008	-0.009	0.007	-0.013*	0.008
Happiness	0.035**	0.016	0.035**	0.016	0.040**	0.016

Notes: Randomization round fixed effects included in all regressions. The controls included for each regression and selected with LASSO are listed in Appendix Table A.XII. The results with inverse probability weighting include all standard controls as in our main Tables.

Table A.XI. Estimation of the impact of TOP tutoring on indices

	(1)	(2)	(3)	(4)
	Performance	Aspirations Index	Socio-emotional Index	Wellbeing Index
Treatment	0.261***	0.150*	0.137*	0.171**
	(0.079)	(0.080)	(0.073)	(0.082)
	[0.007]	[0.108]	[0.108]	[0.090]
Mean Dep:	-0.00	0.00	-0.00	-0.00
Obs	712	523	636	614
R^2	0.112	0.313	0.211	0.079

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading. Controls included are the same as in Table II. "Mean Dep" at the bottom of the table is the mean of the dependent variables for students in the control group.

Table A.XII. LASSO selected variables

(a). Academic Outcomes and Beliefs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Performance	Students			Parents		Teachers		
		Beliefs	Overconfidence	Grade	Beliefs	Overconfidence	Beliefs	Overconfidence	Grade
Learning disorders	✓	✓			✓		✓		
How much do you like English?		✓		✓	✓				
Self-efficacy: university		✓		✓	✓		✓		✓
Goal: university (parent)		✓		✓	✓		✓		✓
High school preference 1: vocational high-school		✓					✓		
How much do you like Math?				✓					✓
Goal: vocational high-school				✓					
Goal: university (child)				✓			✓		✓
Are you following classes online and doing your homework? - Yes, everyday									✓

(b). Aspirations Index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Aspirations Index	Students				Parents		Teachers
		Aspirations	Self-efficacy	High school vocational	High school top track	Aspirations	Self-efficacy	Aspirations
Self-efficacy: university	✓		✓			✓	✓	✓
Barriers: own ability	✓							
Learning disorders	✓						✓	
Goal: Professional institute	✓			✓			✓	
Goal: university (parent)	✓	✓	✓	✓	✓	✓	✓	✓
Child does HW on his/her own	✓							
High school: scientific high-school	✓				✓			✓
High school preference 1: vocational high-school	✓			✓				
Goal: vocational high-school	✓			✓		✓		
Goal: university (child)	✓	✓	✓			✓		✓
Achieve goals - Ability (Square-Root)		✓						
High school preference 2: vocational high-school				✓				
High school preference 1: humanities high-school					✓			
High school preference 2: scientific high-school					✓			
How much do you like Math?								✓

(c). Socio-emotional Skills

	(1)	(2)	(3)	(4)	(5)	(6)
	Socio-Emotional Index	Students				Parents
		Perservance: difficulty	Perseverance: give up	Grit	Locus of control	Grit
How much do you like Math?	✓	✓				
Self-efficacy: university	✓	✓		✓	✓	✓
Barriers: own ability	✓			✓	✓	✓
Learning disorders	✓	✓				
Logic question: correct	✓	✓	✓			
Goal: university (parent)	✓					✓
Child does HW on his/her own	✓			✓		
High school preference 1: vocational high-school		✓				
Logic question: correct+Perseverance		✓	✓			
Goal: vocational high-school		✓				
Locus of control 1					✓	
How much do you like Italian?						✓

(d). Well-being

	(1)	(2)	(3)	(4)	(5)
	Well-being Index	Students		Parents	
		Depression	Happiness	Depression	Happiness
Barriers: own ability	✓	✓		✓	
Locus of control 1	✓			✓	
Self-efficacy: university				✓	

Notes: This Table shows the controls selected using LASSO for each outcome variable.

Table A.XIII. Heterogeneous Treatment Effect by Effort (reported in tutors' registry)

	(1)	(2)	(3)	(4)
	Performance	Aspirations Index	Socio-emotional Index	Wellbeing Index
Panel A: Heterogeneous effects based on effort during tutoring reported by tutors				
Treatment	0.037	0.009	0.044	0.196*
	(0.108)	(0.101)	(0.101)	(0.105)
High Effort Tutoring	0.349***	0.276**	0.157	-0.060
	(0.112)	(0.107)	(0.105)	(0.109)
Treat+High Effort Tutoring==0	0.000	0.003	0.019	0.155
Obs	712	523	636	614
R^2	0.126	0.323	0.214	0.080
Panel B: Heterogeneous effects based on effort in homeworks reported by tutors				
Treatment	0.003	-0.017	0.112	0.170
	(0.110)	(0.101)	(0.103)	(0.114)
High Effort Homework	0.404***	0.373***	0.027	-0.013
	(0.120)	(0.114)	(0.113)	(0.123)
Treat+High Effort Homework==0	0.000	0.001	0.138	0.128
Obs	712	523	636	614
R^2	0.129	0.330	0.211	0.079

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading. Controls included are the same as in Table II. "Mean Dep" at the bottom of the table is the mean of the dependent variables for students in the control group. "High Effort Tutoring" is a dummy variable that assumes value 1 if the mean effort of the treated student during tutoring reported in the registry by the tutor is higher than the mean. "High Effort Homeworks" is a dummy variable that assumes value 1 if the mean effort of the treated student in doing homeworks reported in the registry by the tutor is higher than the mean. We also include a dummy variable which assumes value 1 if "High Effort Tutoring" or "High Effort Homeworks" is missing in Panel A and B respectively.

Table A.XIV. Heterogeneity on subject performance by tutor characteristics

	(1)	(2)
	Dep. var.: student's performance in a given subject	
Panel A: Math		
Not volunteer in math	0.048*	
	(0.026)	
Volunteer in math	0.039**	
	(0.018)	
Degree not STEM		0.034*
		(0.018)
Degree is STEM		0.054**
		(0.023)
P-value of difference:	0.726	0.364
Mean Dep:	0.65	0.65
Obs	712	711
R^2	0.167	0.168
Panel B: Italian		
Not volunteer in italian	0.034	
	(0.023)	
Volunteer in italian	0.032	
	(0.020)	
Degree not Humanities		0.035*
		(0.018)
Degree is Humanities		0.015
		(0.034)
P-value of difference:	0.962	0.558
Mean Dep:	0.46	0.46
Obs	712	711
R^2	0.077	0.078
Panel C: English		
Not volunteer in english	0.047	
	(0.030)	
Volunteer in english	0.044*	
	(0.027)	
No english certificate		0.027
		(0.030)
English certificate		0.055**
		(0.026)
P-value of difference:	0.927	0.373
Mean Dep:	0.46	0.46
Obs	516	514
R^2	0.192	0.193

Notes: Robust standard errors in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. FWER adjusted p-values in square brackets, based on 10,000 replications. The dependent variable for each regression is listed in the column heading. Controls included are the same as in Table II. "Mean Dep" at the bottom of the table is the mean of the dependent variables for students in the control group.

Table A.XV. Causal Forest Heterogeneous Treatment Effect: Median of Predicted Impact by Student and Parents Characteristics

Variables	Performance Index		Aspirations Index		Socio-Emotional Index		Well-being Index	
	p(50)	p(100)	p(50)	p(100)	p(50)	p(100)	p(50)	p(100)
Students								
Male	0.578	0.604	0.530	0.625	0.614	0.553	0.629	0.552
Immigrant	0.241	0.215	0.188	0.189	0.176	0.182	0.093	0.267
Learning disorders	0.243	0.397	0.320	0.301	0.331	0.324	0.304	0.344
Grade in math	6.456	6.095	6.375	6.396	6.467	6.228	6.370	6.343
Grade in Italian	6.048	5.784	5.957	6.052	6.250	5.737	6.178	5.832
Grade in English	6.494	6.018	6.360	6.279	6.562	6.079	6.373	6.218
Grade 6	0.301	0.367	0.340	0.332	0.370	0.246	0.322	0.285
Grade 7	0.336	0.338	0.304	0.350	0.275	0.430	0.361	0.363
Grade 8	0.363	0.296	0.355	0.317	0.355	0.324	0.316	0.353
How much do you like math?	2.785	2.680	2.657	2.883	2.921	2.572	2.879	2.636
How much do you like Italian?	3.138	2.923	3.038	3.106	3.036	3.025	2.943	3.123
How much do you like English?	3.259	2.977	3.123	3.158	3.276	2.999	3.123	3.117
Perseverance	0.827	0.801	0.817	0.826	0.848	0.804	0.846	0.818
Importance Luck (vs. Effort)	0.056	0.058	0.048	0.048	0.042	0.049	0.038	0.052
Familiarity with computers	3.156	3.086	3.150	3.217	3.249	3.147	3.275	3.114
Parents								
Child lives with single parent	0.223	0.225	0.223	0.194	0.191	0.248	0.198	0.211
Edu Mother: High-School	0.430	0.469	0.451	0.471	0.522	0.409	0.502	0.452
Edu Mother: Degree	0.125	0.080	0.112	0.120	0.134	0.128	0.164	0.097
Edu Father: High-School	0.367	0.359	0.379	0.393	0.428	0.330	0.395	0.387
Edu Father: Degree	0.085	0.067	0.068	0.081	0.079	0.085	0.093	0.066
Mother has blue collar job	0.313	0.426	0.367	0.329	0.313	0.418	0.237	0.476
Mother has white collar job	0.179	0.144	0.161	0.183	0.191	0.168	0.271	0.102
Father has blue collar job	0.559	0.525	0.554	0.539	0.537	0.550	0.498	0.587
Father has white collar job	0.206	0.226	0.193	0.235	0.242	0.217	0.287	0.182
At least one parent works from home	0.218	0.188	0.208	0.224	0.228	0.205	0.278	0.158

Notes: This Table presents the average value of students and parents' characteristics for students with below median conditional treatment effect and above median conditional treatment effect. Student level conditional treatment effects are estimated using the Causal Forest methodology.

Table A.XVI. Causal Forest Heterogeneous Treatment Effect: Median of Predicted Impact by Tutor Characteristics

Variables	Performance Index		Aspirations Index		Socio-Emotional Index		Well-being Index	
	p(50)	p(100)	p(50)	p(100)	p(50)	p(100)	p(50)	p(100)
Baseline Characteristics of Tutors								
Female	0.706	0.702	0.699	0.685	0.703	0.696	0.712	0.690
Volunteered Before	0.814	0.838	0.821	0.807	0.812	0.806	0.819	0.802
Tutoring before	0.949	0.956	0.947	0.954	0.948	0.949	0.951	0.945
Training immigrants	0.009	0.013	0.012	0.013	0.009	0.013	0.010	0.018
Training Learning Disorders	0.027	0.050	0.023	0.033	0.038	0.032	0.034	0.032
Motivation TOP: help others	0.837	0.826	0.850	0.853	0.849	0.841	0.834	0.857
Hard work more important than Luck	0.505	0.551	0.493	0.504	0.493	0.518	0.509	0.495
Faculty: Humanities	0.134	0.130	0.133	0.135	0.134	0.131	0.136	0.119
Faculty: STEM + Medical	0.342	0.325	0.340	0.339	0.324	0.322	0.321	0.326
Faculty: Economics	0.297	0.291	0.301	0.301	0.301	0.301	0.294	0.313
Faculty: Education	0.056	0.077	0.054	0.050	0.068	0.064	0.057	0.072
Tutor volunteered to teach math	0.777	0.769	0.780	0.758	0.767	0.789	0.742	0.794
Tutor voluteered to teach Italian	0.607	0.620	0.612	0.594	0.601	0.626	0.624	0.589
Tutor voluteered to teach English	0.603	0.619	0.612	0.615	0.627	0.622	0.632	0.638
Tutor has English language certificate	0.663	0.665	0.663	0.663	0.688	0.684	0.684	0.682
GPA (normalized by faculty)	0.062	0.033	0.070	0.064	0.070	0.060	0.070	0.039
Tutor is familiar with computers	0.916	0.925	0.909	0.913	0.925	0.920	0.917	0.917
Gender match Tutor-Student: Male	0.178	0.181	0.170	0.209	0.190	0.175	0.199	0.171
Gender match Tutor-Student: Female	0.312	0.291	0.354	0.274	0.295	0.318	0.287	0.324
Male Tutor, Female Student	0.113	0.114	0.126	0.102	0.105	0.125	0.086	0.134

Notes: This Table presents the average value of tutors' characteristics for students with below median conditional treatment effect and above median conditional treatment effect. Student level conditional treatment effects are estimated using the Causal Forest methodology.

Table A.XVII. Treatment Effect on Tutors

	(1)	(2)	(3)	(4)	(5)	(6)
	Income as Incentive vs. Income equality	Hard work vs. Luck	Work to Natives over Immigrants	If Effort Well-paid job	Easy to put in others' shoes	Make decisions irrespective others' feelings
Panel A: Sample –second endline for all tutors						
Treatment	0.151	0.118	0.327	-0.279	0.499**	-0.200
	(0.195)	(0.200)	(0.232)	(0.223)	(0.238)	(0.228)
Mean Dep:	4.46	3.47	1.81	2.98	3.15	2.79
Obs	739	742	738	738	740	740
Panel B: Sample – second endline with imputed values from first endline						
Treatment	-0.026	-0.020	0.377*	-0.485**	0.373*	-0.361*
	(0.181)	(0.183)	(0.213)	(0.209)	(0.219)	(0.206)
Mean Dep:	4.46	3.47	1.81	2.98	3.15	2.79
Obs	933	936	933	936	937	937

Notes: This Table reports the coefficients from an ordered logit regressions. Panel A restricts the sample only to tutors that replied to the second short endline. Panel B imputes the values from the first endline for those who did not reply to the second endline. All columns include the controls that were used to assign the tutors to students: whether the volunteer has tutoring experience and specific training (to support students with learning disorders or immigrants), their expertise in the subjects (math, Italian, English), their time availability (3 hours per week or 6 hours per week), whether they are on time in their university enrollment and if they confirmed their availability. We also include the additional tutor controls (gender, university faculty, whether they are enrolled in a undergraduate or master, GPA, previous volunteering activities, whether they applied to TOP to help others (motivation), parental education, and familiarity with the computer). The mean of the dependent variable reported in the Table is for the control group of tutors. Robust standard errors in parentheses. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively.

Table A.XVIII. Tutors' Satisfaction, Students' and Tutors' Characteristics

	(1)	(2)	(3)	(4)
Dep. Var.	**Satisfaction**		**Tutoring again?**	
	Coeff.	SE	Coeff.	SE
Student Characteristics				
Learning Disorders	-0.054	(0.202)	0.045	(0.207)
Immigrant	-0.370	(0.234)	-0.058	(0.251)
Male	0.100	(0.184)	0.446**	(0.199)
Grade 6	0.359	(0.220)	0.046	(0.246)
Grade 7	0.072	(0.231)	0.266	(0.227)
Edu Mother: High-School	0.084	(0.212)	-0.231	(0.226)
Edu Mother: Degree	-0.107	(0.388)	-0.147	(0.404)
Edu Father: High-School	-0.243	(0.215)	-0.069	(0.230)
Edu Father: Degree	-0.124	(0.422)	-0.566	(0.365)
Mother has blue collar job	0.092	(0.195)	0.553**	(0.219)
Mother has white collar job	-0.240	(0.322)	0.119	(0.301)
Father has blue collar job	0.441**	(0.207)	0.506**	(0.227)
Father has white collar job	0.261	(0.273)	0.666	(0.287)
Tutor Characteristics				
Female	0.218	(0.211)	0.070	(0.215)
Faculty: Education	0.210	(0.391)	0.172	(0.441)
Faculty: Economics	0.522**	(0.242)	-0.434 *	(0.260)
Faculty: STEM+ Medical	-0.197	(0.234)	-0.830***	(0.256)
University GPA	-0.032	(0.041)	-0.131***	(0.046)
Voluntering	0.221	(0.245)	0.136	(0.265)
Motivation TOP: help others	-0.251	(0.229)	-0.157	(0.293)
Tutoring before	0.365	(0.465)	-0.594	(0.452)
Training Immigrants	-0.386	(0.772)	0.357	(1.240)
Training Learning Disorders	-0.096	(0.412)	-0.197	(0.628)
Importance Hardwork	0.393 **	(0.190)	0.148	(0.190)
Mean Dep:	3.69		2.15	
Obs	451		451	

Notes: This Table reports the coefficients from an ordered logit regression. The sample is limited to tutors who completed the questions in the endline. Robust standard errors in parentheses. ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively.

B Questionnaires

B.1 Student Test Score

- **Achievement Test:**
  - **Example of math question for grade 8:** a is a odd number grater than 3. Which of the following expression represents the first odd number following a?
    * a+1
    * 2a+1
    * 2a-1
    * a+2
  - **Example of Italian question for grade 8:** which of the following words corresponds to the grammar analysis: name, male, singular, derivative
    * Libreria
    * Libresco
    * Libraio
    * Libricini
  - **Example of English question for grade 8:** Correct the following sentence: "You go to the swimming pool in Sunday".
    * You go the swimming pool in Sunday
    * You goes to the swimming pool in Sunday
    * You go to the swimming pool on Sunday
    * You go on swimming pool on Sunday
- **Beliefs on academic outcomes:** How many questions do you expect to have answer correctly in MATH/ITALIAN/ENGLISH?

B.2 Student Questionnaire

- **Beliefs on academic outcomes:**
  - **Self-grade** Overall, considering your school performance in all assignments (homework, oral test, written test) in the month of May, how would you rate yourself compared to your classmates for each of the following subjects (Math/Italian/English)? Consider a scale from 1 to 10, where 10 are the high-performing students (top 2-3 students) in the class and 1 are the low-performing students in the class (bottom 2-3 students).
- **Aspirations:**
  - **Education Goals.** Thinking about your future, how long do you think you will continue to study? Multiple choice options: (1) I think I will start working

as soon as I complete this school (2) I think I will continue studying and enroll in high school, and start working after obtaining a diploma (3) I think I will continue studying and enroll in a technical institute, and start working after obtaining a diploma (4) I think I will continue studying and enroll in a professional/vocational institute (such as cosmetology, auto mechanic, etc.) and then start working (5) I think I will continue studying and reach university.

- **High-school goal.** Which high-school would you like to do? Up to two choices are possible. Multiple choice options with all sub-tracks of high school including the two top tier tracks (humanistic and scientific) and vocational high-school
- **Self-efficacy.** Apart from what you would like to do in the future, do you think you will be able to go to university when you are older if you wish to do so? Multiple choice options: (1) Very much (2) Much (3) Somewhat (4) Slightly (5) Not at all

- **Socio-emotional skills:**
  - **Perseverance.** First, we ask students to answer a first logic question. Second, if they want to persevere, we ask them a second logic question.
    - Would you like to try and answer another logic question? Multiple choice options: (1) Yes, I'd like to try with a question as difficult as this one (2) Yes, but I'd like to try an easier question (3) No
  - **Grit (following Duckworth and Quinn (2009)).** Here are a number of statements that may or may not apply to you. There are no right or wrong answers, so please answer truthfully, considering how you compare to most people. (5-points likert scale)
    1. I like schoolwork best which makes me think hard, even if I make a lot of mistakes.
    2. Setbacks discourage me.
    3. If I think I will lose in a game, I do not want to continue playing.
    4. If I set a goal and see that it's harder than I thought I easily lose interest.
    5. When I receive a bad result on a test I spend less time on this subject and focus on other subjects that I'm actually good at.
    6. I work hard in tasks.
    7. I prefer easy homework where I can easily answer all questions correctly.
    8. If I'm having difficulty in a task, it is a waste of time to keep trying. I move on to things which I am better at doing.
  - **Locus of control.** For each of the following statements, give a score from 1 to 5 indicating whether you agree or disagree with the statement.
    1. Many of the unhappy things in people's lives are partly due to bad luck
    2. Trusting in fate has turned out better for me than making a decision to take a definite course of action.

3. In the case of the well-prepared student, there is rarely, if ever, such a thing as an unfair test.
4. When I make plans, I am almost certain that I can make them work

- **Well-being:**
  - **Depression (following Frühe et al. (2012)).** For each item please mark whether you agree or disagree with the statement. (4 points likert scale)
    1. I am happy
    2. I worry a lot
    3. I feel sad
    4. I get upset quickly
    5. I am not in the mood for anything
    6. I often think I did something wrong
    7. It's often hard for me to concentrate
    8. I feel lonely
    9. I enjoy a lot of things
  - **Happiness.** Think about the period of lockdown during Covid-19. During this period, how happy or unhappy have you been overall? 1-10 scale going from very unhappy to very happy
- **Additional outcomes:**
  - **Homework.** Think about the month of May this year. On average, how much time did you devote to doing homework every day? Multiple choice options: (1) Less than 15 minutes (2) 15 30 minutes (3) 30 - 60 minutes (4) 1 hour - 1 hour and a half (5) 1 hour and a half - 2 hours (6) 2 hours - 2 hours and a half (7) More than 2 hours and a half
  - **Following online classes.** In the month of May, have you been following classes online? Multiple choice options: (1) Yes, everytime there was an online class (2) Yes, but not always (3) Sometimes (4) No.
  - **Like subjects** How much do you like the following subjects (Math/Italian/English)? Check one box for each subject. Multiple choice options: Very much/ Much/ Somewhat/ Slightly/ Not at all
  - **Difficult online classes.** How difficult do you find it to follow classes online and use your school's online platform during the month of May? Multiple choice options: Extremely difficult /Very difficult / Moderately difficult /Slightly difficult / Not at all difficult
- **Tutoring experience and satisfaction:** we included few questions only for treated students.

B.3 Parent Questionnaire

- **Beliefs on academic outcomes.** As part of the final questionnaire for the project, we will ask your child 7 (7/5) questions in math (Italian/English). These are multiple choice questions prepared by middle school teachers that collaborate with us. How many correct answers do you expect your child to get? We will not share your answers with your child.

- **Aspirations:**

  - **Education Goals.** Thinking about your child's future, how long do you think he/she will continue to study? Multiple choice options: (1) I think he/she should start working as soon as he/she completes compulsory schooling (2)I think he/she should continue studying and enroll in high school, and start working after obtaining a diploma (3) I think he/she should continue studying and enroll in a technical institute, and start working after obtaining a diploma (4) I think he/she should continue studying and enroll in a vocational high-school (such as cosmetology, auto mechanic, etc.) and then start working (5) I think he/she should continue studying and reach university.
  - **Self-efficacy.** Do you think your child has the capability to attend and successfully graduate from university if he/she wanted to? Multiple choice options: (1) Very much (2) Much (3) Somewhat (4) Slightly (5) Not at all

- **Socio-emotional skills:**

  - **Grit (following Duckworth and Quinn (2009)).** Here are a number of statements that may or may not apply to your child. There are no right or wrong answers, so please just answer truthfully. Think mainly about your perception from the last month. (5 points likert scale)
    1. He/she likes schoolwork best which makes him/her think hard, even if he/she makes a lot of mistakes.
    2. Setbacks discourage him/her.
    3. If he/she thinks he/she will lose in a game, he/she does not want to continue playing.
    4. If he/she sets a goal and sees that it's harder than he/she thought he/she easily loses interest.
    5. When he/she receives a bad result on a test he/she spends less time on this subject and focuses on other subjects that he/she is actually good at.
    6. He/she works hard in tasks.
    7. He/she prefers easy homework where he/she can easily answer all questions correctly.
    8. If he/she is having difficulty in a task, he/she thinks it is a waste of time to keep trying. He/she moves on to things which he/she is better at doing.

- **Well-being:**

- **Depression (following Frühe et al. (2012)).** For each item please mark whether you believe the statement is true for your child. (4 points likert scale)
  1. is happy
  2. worries a lot
  3. feels sad
  4. gets upset quickly
  5. is not in the mood for anything
  6. often thinks he/she did something wrong
  7. is often hard for him/her to concentrate
  8. feels lonely
  9. enjoys a lot of things
- **Happiness.** Think about the period of lockdown during Covid-19. During this period, how happy or unhappy would you say your child has been overall?

- **Additional outcomes:**
  - **Homework.** Think about the month of May. On average, how much time did your child devote to studying and doing homework every day? Multiple choice options: (1) Less than 15 minutes (2) 15 30 minutes (3) 30 - 60 minutes (4) 1 hour - 1 hour and a half (5) 1 hour and a half - 2 hours (6) 2 hours - 2 hours and a half (7) More than 2 hours and a half
  - **Following online classes.** In the month of May, did your child follow classes online? Multiple choice options: (1) Yes, everytime there was an online class (2) Yes, but not always (3) Sometimes (4) No .

- **Tutoring experience and satisfaction:** we included few questions only for treated students.

B.4 Teacher Questionnaire

- **Beliefs on academic outcomes:**
  - **Beliefs on academic outcomes.** As part of the final questionnaire for the project, we will ask 7 (7/5) questions in math (Italian/English). These are multiple choice questions prepared by middle school teachers that collaborate with us. How many correct answers do you expect student X to get? We will not share your answers with your students.
  - **Grade.** Overall, considering the performance of your students in all assignments (homework, oral tests, written tests) in the month of May, how would you rate student X? Consider a scale from 1 to 10, where 10 are the best-performing students (top 2-3 students) in the class and 1 are the least-performing students in the class (bottom 2-3 students).

- **Aspirations:**

- **Education Goals.** Thinking about the future of the student, how long do you think he/she should continue to study? Multiple choice options: (1) I think he/she should start working as soon as he/she completes compulsory schooling (2)I think he/she should continue studying and enroll in high school, and start working after obtaining a diploma (3) I think he/she should continue studying and enroll in a technical institute, and start working after obtaining a diploma (4) I think he/she should continue studying and enroll in a vocational high-school (such as cosmetology, auto mechanic, etc.) and then start working (5) I think he/she should continue studying and reach university.

- **Additional outcomes:**
  - **Homework.** Did the student X do his/her homework during the month of May 2020? Multiple choice options: (1)Yes, regularly did all assigned homework (2) Yes, did the assigned homework most of the times, but not always (3) Sometimes/rarely (4) No
  - **Issue Behavior.** Especially thinking about the month of May, do you think that the student X had behaviour problems? Multiple choice options: (1) No (2) Yes - minor difficulties (3) Yes - some difficulties (4) Yes - serious difficulties

- **Tutoring experience and satisfaction:** we included few questions only for treated students.

B.5 Tutor Questionnaire

- **Empathy.** Below is a list of statements. Please read each statement carefully and rate how strongly you agree or disagree with it. There are no right or wrong answers. (4-points likert scale)
  1. I find it easy to put myself in somebody else's shoes.
  2. I am able to make decisions without being influenced by people's feelings.

- **Hard work.**
  1. We would like to start by asking your views on a few issues. How would you place your views on this 1-10 scale? 1 means you agree completely with the statement on the left; 10 means you agree completely with the statement on the right; and if your views fall somewhere in between, you can choose any number in between.
     - Incomes should be made more equal vs. We need larger income differences as incentives for individual effort
     - In the long run, hard work usually brings a better life vs. Hard work doesn't generally bring success – it's more a matter of luck and connections
  2. How much do you agree with the following statement? If students put effort in studying, they can get a well-paid job, independent of their family background. (4-points likert scale)

- **Tutoring experience and satisfaction:** we included few questions only for treated tutors.

C Heterogeneous Treatment Effects using Causal Foest

In different empirical settings, treatment effects may vary widely between subgroups of the population. It is important then to evaluate if there are significant heterogeneities in treatment effects across observable variables. There are a few limitations to the traditional approach in which economists include interactions between different variables to estimate treatment effects. Including too many variables might lead to computational challenges and increase the risk of overfiting. Choosing one particular set of variables to estimate heterogeneous treatment effect can lead to arbitrary decision making and to the loss of information on possibly important heterogeneities. To deal with these challenges, we compute heterogeneous treatment effect using the honest causal forest algorithm of Wager and Athey (2018).

In our estimation procedure, we follow Davis and Heller (2017) and average treatment effects only across predictions made on observations that were neither part of the tree growing subsample nor the subsample used to estimate the conditional treatment effects. Accordingly, before estimating the causal forest we split the data into a subsample that will be used in the implementation of the causal forest algorithm, the *training sample* and a subsample that will be used to compute average treatment effects, the *test sample*. To do so, we implement the following procedure:

1. From our final sample, we obtain a random subsample—without replacement—consisting of 50% of the observations in the original sample. This subsample is our training sample and the remaining data is our test sample.

2. We use the training sample to estimate the causal forest (using R's command *causal_forest* from the *grf* package). We implement this command building a forest with 2500 trees. To build each tree, we divide the training sample in half and then use 70% of this new sample to determine splits. The other 30% is used to estimate the conditional treatment effect.

3. We use the conditional treatment effect obtained in step 2 and the test sample to estimate average treatment effects.

4. We implement 1000 replications of steps 1, 2, and 3.

5. The final estimate of average treatment effect is given by average the result across all simulations.