esade

EsadeEcPol - Center
for Economic Policy

Online tutoring works: Experimental evidence from a program with vulnerable children

Working paper **#2**

AUTHORS

Lucas Gortazar
EsadeEcPol and World Bank

Claudia Hupkau
CUNEF and LSE-CEP

Antonio Roldán
EsadeEcPol and LSE-SPP

March, 2022

Online tutoring works: Experimental evidence from a program with vulnerable children*

Lucas Gortazar† Claudia Hupkau‡ Antonio Roldán§

March 24, 2022

Abstract

We provide evidence from a randomized controlled trial on the effectiveness of a novel, 100-percent online math tutoring program, targeted at secondary school students from highly disadvantaged neighborhoods in Spain. The intensive, eight-week-long program was delivered by qualified math teachers in groups of two students during after-school hours, six months after Covid-19 related school closures had ended. The intervention significantly increased standardized test scores (+0.26 SD) and end-of-year math grades (+0.48 SD), while reducing the probability of repeating the school year. The intervention also raised aspirations, as well as self-reported effort at school.

Keywords: Tutoring, Covid-19, mentoring, RCT, mathematics, child outcomes

JEL Classification: C93, I24, I28, H75.

*We thank seminar participants at LSE-CEP, UC3M, UPO and UAM. Pablo García-Guzmán, Natalia Collado, Ana Herrero and Angel Martinez provided excellent research assistance. We thank Antonio Cabrales, Jonathan de Quidt, Luis Garicano, Nagore Iriberri, Mónica Martinez-Bravo, Pedro Rey-Biel and Jenifer Ruiz-Valenzuela for the their comments. We are grateful to Empieza por Educar for the implementation of the programme, including Miriam Arriola, Zaida Pillado, Beatriz Morilla, Lara Crespo, Oscar Ferri, Carla Carmona, Paula Royo, Miguel Costa, Carles Lopez, Luis Oliver, and Justo Quintanar. We are also grateful to Iria Mata, Dirk Rosquillas, Miguel Ujeda, and Gonzalo Romero for their support in various layers of implementation. We are grateful to the schools that participated in the intervention. We acknowledge financial support from Esade, Fundación COTEC, Fundación CEOE, United Way Spain, Banco Santander, UNIR, Huawei, Fundación Atlantic Copper, Edvolution and Google. AER RCT Registry ID: AEARCTR - 0007138. The project has obtained IRB approval from University Ramon Llull.

†ESADE Center for Economic Policy and World Bank. Email: lucas.gortazar@esade.edu

‡Department of Economics, CUNEF Universidad; and Centre for Economic Performance, London School of Economics. Email: claudia.hupkau@cunef.edu.

§ESADE Center for Economic Policy and School of Public Policy, London School of Economics. Email: antonio.roldan@esade.edu

1 Introduction

The Covid-19 pandemic has disrupted education in over 150 countries and affected 1.6 billion students worldwide (Azevedo et al., 2021). Mounting evidence shows that lockdowns and school closures had a profound negative impact on the learning, socio-emotional development and mental health of pupils around the world (Schult and Lindner, 2021; Engzell et al., 2021; Tomasik et al., 2021; Contini et al., 2021; Lichand et al., 2021; Newlove-Delgado et al., 2021; Pizarro-Ruiz and Ordóñez-Camblor, 2021; Arenas and Gortazar, 2022). Moreover, there is growing evidence that pupils from disadvantaged families were hit harder than those from more comfortable backgrounds (Maldonado and De Witte, 2021; Engzell et al., 2021; Contini et al., 2021; Blainey and Hannay, 2021; Haelermans et al., 2021; Elliot Major et al., 2020), driven by differences in parental support, access to technology and participation in remote instruction throughout the closures (Agostinelli et al., 2020; Chetty et al., 2020). While these learning losses seem to have been partially or fully recovered in some countries since schools reopened (Birkelund and Karlson, 2021; Department for Education, 2021), these findings raise serious concerns that the pandemic will increase inequality in the long-run.

Policy makers worldwide are scrambling to come up with an effective policy response in order to reverse the widening of this gap. Parents share the concerns about learning losses and strongly support a comprehensive policy response, in particular via programs such as tutoring (Farquharson et al., 2021). Pandemic restrictions and social-distancing rules have, however, made the policy response more difficult, as traditional catch up programs, like face-to-face tutoring or after school activities, are now more difficult to implement.

In this paper, we study the effectiveness of an intensive, eight-week online tutoring program on academic and socio-emotional outcomes of secondary school children in Spain. The program offered free, 100 percent online after school tutoring to pupils aged 12 to 15 from very disadvantaged backgrounds. Tutoring sessions were delivered by professional teachers in groups of two students, focused on math and social-emotional support (motivation, well-being and work routines) and took place from mid-April to early June 2021, with three 50-minutes meetings per week.

The program, called *Menπores*, has four novel features. First, the whole organization of the program and the tutoring sessions were implemented online. Second, tutoring was carried out by paid-for, qualified math teachers who passed a rigorous selection and training process. Third, the tutoring sessions were done in groups of two students per tutor. Fourth, the program took place during a time period when schools were largely back to normal.

The program was implemented in partnership with *Empieza por Educar* (the Spanish branch of *Teach for All*), an NGO specialized in training young teachers working in schools attended by vulnerable and low-income students. The recruitment of program participants was done in two steps. First, we identified a number of schools that showed interest in the program. Second, we asked principals in participating schools to identify students most in need for support in math and disseminate the program among them and their families. Among all students who signed up, we randomly assigned slightly more than half to the program. Randomization was blocked by classes, to ensure that students who ended up in the same group knew each other, and to increase the power of our experimental design. Within blocks, treatment students were randomly divided into groups of two, which were subsequently randomly assigned to a tutor.

We collected a rich array of child and family characteristics at the stage of online registration. We ran base- and endline surveys of pupils, which included a standardized math test and questions on socio-emotional well-being, aspirations and past performance. At the end of the program, we administered an endline survey of families to collect information on the academic results at the end of the year. Finally, we collected very rich real-time data throughout the duration of the program regarding participation, connection time, and quality of the connection.

We find a positive and significant effect of program assignment on end-of-year math grades (+0.48 SD), equivalent to about six months of learning. Further, we find a significant increase of about 32 percent with respect to the control group mean in the likelihood of passing the subject (+22 percentage points). Using our standardized math test, we find an increase in the test score by 26.4 percent of a standard deviation (equivalent to about three months of learning), which is significant at the 10 percent level. Further, we find a large and significant effect on grade retention: the program decreased the likelihood of repeating the school year by 9.4 percentage points, equivalent to a 78 percent decrease with respect to the control group, which had a repetition rate of 12 percent.

In terms of non-cognitive outcomes, we find that the program raised students' aspirations: Students in the treatment group are 13.6 percentage points more likely to state that they would like to go onto the academic track after compulsory schooling (i.e. *Bachillerato*), equivalent to a 33.2 percent increase compared to the control group mean. Students assigned to the treatment were also 12 percentage points more likely to state that they exerted high effort always or most of the time at school, which corresponds to an increase by 22 percent when compared to the control group mean. We do not find an impact on student's self-perceived math competencies or the likelihood of stating they like mathematics. We neither find an effect on locus of control, grit, or on overall well-being.

To put these numbers into context, Nickow et al. (2020)'s meta analysis of the effectiveness of in-person tutoring finds an overall pooled effect size of 0.37 SD. For tutoring delivered by professional teachers, this effect is about 0.5 SD, but this evidence is based on programs delivered face-to-face during school hours only.[1] The pooled effect size estimate for after school programs delivered in person by paraprofessionals is 0.15 SD, but this figure is based on a very limited number of studies.[2] Closest to our setting is the study by Carlana and La Ferrara (2021), a 100 percent online tutoring program implemented in Italy during the onset of the pandemic in Spring 2020. Our measured effect sizes in the standardized test are very similar to theirs.

We find much larger effects of our program on teacher-assessed outcomes than for the standardized test. A possible explanation for this might be the positive effect of the program on students' socio-emotional skills and effort. As suggested by Eisner et al. (2020), this might have improved teacher-student interactions, which are important for teacher-assessed outcomes (Borghans et al., 2016). Evidence from teachers surveyed after the end of the program suggests that students who participated in the program were more focused in class, more likely to have completed their homework and better able to focus on class content. The high *de facto* intensity of the treatment might also have contributed to the comparatively large positive effects of the program. The program was completed by 96.6 percent of pupils of the treatment group, with the median student attending 20 sessions (83 percent of the target of 24) and 960 minutes (80 percent of the target of 1200). These are high numbers, considering the online nature of the program

[1] As noted by Nickow et al. (2020), there are no rigorous studies assessing the effectiveness of after school programs delivered by teachers.

[2] Paraprofessional tutors include mainly tutors employed by the school or community, undergraduate and graduate students and trainees in education-related fields (Nickow et al., 2020).

and the highly complex backgrounds of the participants: In our sample, 42 percent of the pupils live in households with an income of less than €1,000 per month, 46 per have an immigrant background, 52 percent of parents have, at most, lower secondary education, and 28 percent live in single-parent families.

We also look at treatment effect heterogeneity in several dimensions. The program was more effective among pupils who did not experience internet connection problems and who were in a group with someone of the same gender, although effect size differences are imprecisely estimated. We do not find significant differences between boys and girls. A greater impact on performance in the standardized test was observed among pupils in the second year of secondary school, but those from first year of secondary school experienced higher effects on teacher-assessed outcomes. Finally, children with immigrant background and from single-parent families experienced higher increases in end-of-year grades and greater reductions in grade-retention. However, the experiment was not powered to detect modest effect size differences across subgroups and most coefficients are not significantly different from another.

Our study contributes to the understanding of whether online tutoring can work as an effective tool for closing learning gaps for disadvantaged students. While there is extensive evidence on the effectiveness of one-to-one and small group in-person tutoring programs (Nickow et al., 2020), very little evidence exists as regards to the effectiveness of online tutoring programs. Before the pandemic, a sizable amount of research was dedicated to understanding the effectiveness of educational software tools and online learning for university students (Escueta et al., 2020). During the pandemic, some authors have explored the effectiveness of different remote learning methods, such as online peer mentoring to support university students (Hardt et al., 2022) or parental educational support through phone calls and text messages (Angrist et al., 2020). However, none of these studies analyzed the effects of online real-time tutoring between teachers and secondary school students.

To the best of our knowledge, the only other rigorously evaluated online tutoring program was implemented in Spring 2020 by Carlana and La Ferrara (2021) in Italy. Our program departs from theirs in three fundamental ways. First, their program was delivered by volunteer university students, while *Menπores* was implemented by paid-for, qualified secondary school teachers. Secondly, the tutoring was implemented in groups of two students, instead of one-to-one. Third, and perhaps more importantly, their program was implemented during the harshest lockdown period from April to June 2020 in Italy (when all kids were at home and schools where closed). Ours was implemented one year after the onset of the pandemic, several months after schools were fully re-opened in Spain. In that sense, we believe our results show the effectiveness of online tutoring in normal times, when tutoring can be considered a complement rather than a substitute for in-class, regular teaching.[3]

Our contribution is relevant both in terms of policy and for further academic research. As a response to the pandemic, governments are investing large amounts of money in tutoring programs (both in face-to-face and online formats). For instance, the UK has committed £3bn for their publicly funded National Tutoring Program. Our evidence suggests that this money would be well spent. Using the standard deviation increase in the academic outcome per €100 spent on the intervention as a measure of cost-effectiveness, our program compares favorably in terms with alternative programs: Our intervention costs approximately €300 per student, and has a positive impact of 0.26 SD on learning outcomes (in math), translating into a 0.087 SD increase per €100 spent in our cost-effectiveness measure. This compares slightly favorably with summer schools (Cooper et al., 2000), with a cost-effectiveness of 0.066 SD per €100 spent (based on an impact of 0.23 SD and a cost of €350, estimated for the case of Spain). It also compares favorably with increasing instruction time one hour per day, which according to Higgins et al. (2012) costs €1,020 for an increase of 0.24 SD in test scores, resulting in a cost-effectiveness rate of 0.0235 SD per €100 spent.

Regarding potential future scaling up, we would expect our results to be replicable while schools are open. In terms of cost, programs with volunteers are always going to be cheaper. However, at large scales, they are likely to face more practical and political economy limitations than programs with paid professional teachers. The innovative two-to-one online design offers additional cost savings in relation to in-person programs and one-to-one online programs, while showing very similar results. Finally, the two-to-one online design may have other advantages, such as creating positive motivational dynamics and peer pressure not to abandon the program, which might help explaining our successful completion rates. If this was true, it would help to address some of the key shortcomings

[3] The Italian Statistical Institute estimates that around 3 million Italian students aged 6-17 may not have been reached by remote learning during the lockdown (Instituto Nazionale di Statistica, 2020).

found in the literature in online education, such as a lack of perseverance and motivation (Escueta et al., 2020). More research on this will be needed to better understand group dynamics in remote learning.

The rest of the paper is organized as follows. Section 2 describes the context of the intervention. In Section 3, we present the study design and in Section 4 we describe the data. The empirical strategy is presented in Section 5, and results and robustness checks are shown in Section 6. Section 7 concludes.

2 Context of the intervention

Like most education systems across the world, Spain faced unprecedented challenges in March 2020 due to the Covid-19 pandemic. Although education policy in Spain is largely decentralized and devolved to the regions, many coordinated measures were implemented throughout the pandemic. Schools were closed on 11th March and did not reopen until September 2020, when the 2020/21 school year began.[4] The school year 2020/21 can be considered to be close to a normal school year, with no coordinated school closure measures in any of the Spanish regions. Between September 2020 and March 2021, the proportion of confined classrooms remained below two percent at all times (Ministerio de Educación y Formación Profesional, 2021). On March 9th 2021, the latest date for which data is available (and closest to our intervention), only 0.5 percent of classes in Spain were closed. Spain was among the most successful countries among advanced economies regarding the school reopening campaign in 2020/2021. According to the OECD (2021), by May 31st 2021 Spain was in the fourth position in the OECD ranking on the number of days schools were open in the 2020/21 school year. While the vast majority of classes were open, the epidemiological constraints made the school year 2020/21 still complex in pedagogic and logistic terms.

Learning losses in Spain are estimated to have been lower than in other countries, likely due to the comparatively short duration of school closures. Using data from the Basque Country, a region in the north of Spain, Arenas and Gortazar (2022) estimate a learning loss of 0.075 SD for math and 0.05 SD for language one year into the pandemic. This is smaller than for the case of the Netherlands, for instance, where Haelermans et al. (2021) estimate a learning loss after a full year of Covid-19 of 0.14 SD for math and

[4]The first region to close all schools, nurseries and universities was Madrid. The remaining communities followed suit on March 12th or 13th 2020.

0.17 SD for reading. Arenas and Gortazar (2022) also find that students with higher learning losses self-report significantly worse levels of socio-emotional well-being due to the pandemic. The fact that losses are larger in mathematics and that the deterioration of socio-emotional wellbeing is correlated with learning losses provides suggestive evidence of the importance of focusing on both academic as well as socio-emotional aspects in catch-up programs.

Our intervention took place in two large regions of Spain, Madrid and Catalonia, but re-opening policies were very similar across regions. In primary school and the first two grades of lower secondary school (Grades 1 to 8, ages six to 13), the relevant years for our study, classes have been operating under a face-to-face model since September 2020. In order to guarantee social distancing, class sizes were slightly reduced. Some schools concentrated hours of instruction between 9 a.m. and 2 p.m., rather than the usual model of classes happening between 9 a.m. and 4 p.m., with a lunch break of 1.5 hours in between, in order to avoid lunch happening at school (a potential source of virus spread). While this did not lead to a reduction in instruction time, it meant that some of the students in our study potentially had one to two hours more free time in the afternoons compared to the pre-pandemic scenario.[5] In the upper grades (grades 9 to 12, ages 14 to 18) of secondary school, schools operated in hybrid models or pure remote models.

To cope with the various learning models and anticipate potential future school closures, the Ministry of Education and Vocational Training and regional ministries made large efforts to provide schools with tablets and computers for the school year 2020/21. The Autonomous Community of Madrid, for instance, invested more than €6.1 million (or $6.9 million) in 36,100 tablets for their schools (Comunidad de Madrid, 2020). Because schools lent these devices to students who did not have access to a computer or tablet, only a very small share of students who enrolled in our program did not have the technology at home to attend online tutoring sessions.

3 Study Design

In this section we describe the intervention design, recruitment of participants and tutors and the timeline of implementation.

[5]In Catalonia, 43 percent of schools ended up operating in morning shifts while in Madrid 72.9 percent of schools did so. This meant an increase of 3 percentage points in Catalonia and an increase of 19.5 percentage points compared to pre-pandemic times in Madrid (Plataforma de Infancia, 2021).

3.1 The program *Menπores*

Our online tutoring program, which we called *Menπores*, was an intensive intervention consisting of three 50-minute sessions per week over a period of eight weeks. The target population were students in Grades 7 and 8 (grades 1 and 2 of secondary school, aged 12 to 15, attending schools in highly disadvantaged neighborhoods. We chose this target for two reasons: First, disadvantaged students were disproportionately affected by learning loss during the pandemic (Haelermans et al., 2021; Blainey and Hannay, 2021) and most likely to benefit from the intervention. The need to invest and experiment with remedial programs which could facilitate catch up for the learning loss of these students was and still is a priority in education policy in many countries (World Bank, 2021a,b). Second, evidence suggests that tutoring in mathematics tends to be more effective for students in higher grades (Nickow et al., 2020), and budget, logistical and time constraints meant that we could deliver tutoring only in one subject area and only in secondary schools.

Tutoring sessions were delivered online by qualified math teachers in groups of two students per tutor. We decided to hire qualified math teachers for several reasons: First, existing evidence on face-to-face tutoring shows that they are significantly more effective than non-professionals or volunteer tutors (Nickow et al., 2020). Second, while we had initially planned a treatment arm delivered by university student tutors as volunteers as in Carlana and La Ferrara (2021), we were neither able to recruit sufficient participating students nor sufficient volunteer tutors in the short time-frame we were operating in.[6] The timing of our intervention (towards the end of the academic year, when university students tend to be more busy because of final examinations) and the fact that life in Spain had largely gone back to normal by March 2021 (students were no longer locked inside their homes as they had been between March 2020 to May 2020) are possible explanations for the low response to our call.

The group composition was fixed throughout the program, with the same students attending meetings with the same tutor in each session. The students in each tutoring group of two were from the same class or grade from the same school to guarantee that students knew each other and would find it easier to connect and accommodate. We decided to go for a two-to-one student-tutor ratio for three reasons. First, the pedagogic team in charge of implementation suggested that being in a group with another child

[6] Like Carlana and La Ferrara (2021), we launched a call searching for volunteer math tutors at five large public and private universities in Barcelona and Madrid, but received less than 50 applications.

had the potential to generate mutual motivation and peer pressure not to abandon the program. Second, existing evidence for face-to-face programs in Nickow et al. (2020) shows that two-to-one tutoring is nearly as effective as one-to-one tutoring. Third, this design meant that we could deliver the tutoring to twice as many students as in a one-to-one setting, given our budget.[7]

A key element of the program was its online nature. The fact that face-to-face interactions were severely constrained by social distancing rules made this the only viable option. Additionally, the demand and interest in online tutoring has surged rapidly since 2020, while to date very limited evidence on its effectiveness exists.

3.2 Content and methodology of the tutoring sessions

We designed the academic and pedagogic content of the intervention together with *Empieza por Educar* (ExE), the Spanish partner of the US based network *Teach for All*. ExE is an NGO specialized in training young teachers working in schools attended by highly vulnerable and low-income students in the regions of Madrid and Catalonia. Its core activity is based on a highly selective model of teacher training, identifying teacher candidates with top academic and socio-emotional skills that are relevant for the teaching profession, as well as an interest in the profession and in social change.[8]

The academic content of the tutoring program was based on the national mathematics curriculum and covered the expected knowledge from 1st and 2nd graders in secondary schools in Spain. Additionally, the program aimed at providing psycho-social and socio-emotional support to students. This was done for several reasons. First, to potentially mitigate the detrimental effects of the pandemic and associated school closures on children's mental health (Newlove-Delgado et al., 2021). Second, there is growing evidence as to the importance of socio-emotional skills in educational attainment and future labor market outcomes (Heckman et al., 2006; Kosse et al., 2020; Kosse and Tincani, 2020; Eisner et al., 2020). There was therefore an explicit mandate for tutors to spend time in the sessions providing such support and reserve at least ten out of the 50 minutes to discuss any issues, fears or concerns the children might be facing at home or at school.

[7] We decided not to go for a three-to-one ratio as we thought it would have been exceedingly challenging from a logistic point of view to coordinate four people to be available at the same time three times a week.

[8] Every year, ExE selects around 80 candidates out of an applicant pool of between 2000 and 3000 to receive training and support during the two years they work in such schools in pedagogy, classroom management, school and community transformation and leadership skills.

The pedagogical approach of the sessions was inspired by the *No Excuses* methodology, which has been shown to be effective in raising academic and non-cognitive outcomes in the context of urban US charter schools for vulnerable children (Dobbie and Fryer, 2013). This methodology emphasizes high expectations, increased instructional time and support, individualized support, continuous feedback and intensive data collection on student progress to guide instruction. We provide details on how tutors were trained in these aspects in Section 3.4.

3.3 Recruitment of schools and participants

The recruitment of program participants was done in two steps. First, we identified a number of schools that showed interest in the program. Second, we asked schools who had agreed to participate to identify potential participants from their pool of students and disseminate the program among them.

For recruitment of participant schools, we leveraged ExE's large school network in the regions of Catalonia and Madrid. School principals were initially contacted by ExE and informed about the program and its characteristics, its target population (disadvantaged students in the 1st and 2nd grade of secondary school lagging behind in mathematics), and the fact that the program was to be evaluated scientifically through a randomized controlled trial. Principals signed an agreement detailing the school's role in the study, including: (i) the identification of a group of students that would benefit most from the program; (ii) dissemination of the application material among these students and their families; (iii) ensuring the administration of baseline and endline surveys during school hours; and (iv) participating in a final survey themselves.

In the second step, parents of children identified by the school as in need were directed to an online registration form, which included an information sheet for parents and children informing them of the fact that the program was to be evaluated and that not all students that registered would eventually be selected. In the registration process, parents were asked to give consent for their children's participation and the usage of data for research purposes. In the registration form we asked for detailed household and student characteristics and whether the student that was being registered had access to a tablet or other device to participate in the online sessions.

3.4 Selection and training of tutors

ExE designed and implemented the selection and training for tutors based on their long-standing experience with teacher selection. A key criterion for selection was to hold a post-graduate (Master's) degree in Teacher Training in a scientific specialization (math, physics, chemistry or biology), which is a formal requirement to teach mathematics in secondary education in Spain. Other skills, such as motivation for the program, having taught in low-income schools, and prior teaching experience, were also considered. Advertisement of the positions was done through various channels, including online hiring portals, ExE's own network of current teachers and alumni, and other teachers whom they work with. A total of 199 applicants which met the minimum pre-requisites were sent a formal application form, and applied. Out of these, 110 candidates were sent a link for an online interview. Finally, 46 tutors which were offered to participate in the program started the tutoring sessions after the training.[9]

Before the start of the program, tutors received between 15 to 20 hours of online training through ExE's teacher training platform. Training included two remote training modules and two online webinars with expert teachers. Training focused on the following key areas: how to establish strong ties with students, student motivation, lesson planning, learning verification and formative assessments, math academic content knowledge and tutoring methodology.

3.5 Timeline

Figure 1 shows the timeline of the intervention. Planning and design took place between January and March 2021, and the registration period for parents and children started in early March 2021, lasting for about two weeks. A total of 356 complete registrations with valid consents were received during this time window.[10] After registrations were closed,

[9]As mentioned previously, we had initially planned a third treatment arm with volunteer mentors. Although we advertised the program at four big public and private universities, we only received 50 applications. We attribute the small number of applications to the fact that the program required a high level of time commitment and coincided with end of term examinations at university. This was also the reason why most of the applicants finally decided to drop out of the process before the start of tutoring sessions. After initial screening and interviews, we were able to include only eight volunteer tutors, who completed the entire application process. We decided to keep these tutors in our pool and included them in the randomization process. This allowed us to fulfill the initial commitment to schools to provide tutoring to around 200 students. Results when including volunteer tutors are basically unchanged compared to the main results presented in this paper.

[10]The total number of valid registrations was 378, which includes those students that were later randomly assigned to a volunteer tutor.

baseline tests and surveys were administered in all participating classrooms, meaning in all classes where at least one student had registered for the program.

Students were randomly assigned to treatment and control group, and in case they had been selected to be in the treatment group, to a partner and tutor, during the Easter break (early April 2021). The tutoring sessions started in the second week of April 2021 and ended in early June, at the time where the final grade evaluation takes place.

Endline tests and questionnaires to students were administered after the end of the intervention and before the end of the academic year (second week of June 2021). We also asked tutors, principals and math teachers to complete brief online surveys at the end of the program. Finally, we administered an online and phone survey to parents during the month of July.

3.6 Experimental design and randomization

The experimental strategy relied on over-subscription. No compensation for students not assigned to the treatment was offered, as at the time of the randomization we did not have funds available that could have covered the cost of a second round of the program at a later stage.

Randomization was done in various steps. First, we assigned the initially 356 students who enrolled in the program randomly into treatment (186 students) and control group (170 students). Randomization was at the person level in blocks, where a block consisted of all students of a class at a school that had signed up for the program. When the number of students from the same class who enrolled was two or less, we combined classrooms of the same grade level within the same school into one block. We did this in order to get blocks of sufficient sizes to assign an even number of students within each block to the treatment group. The total number of blocks was 68, distributed across 18 schools. In a second step, we randomly ordered treatment students within each block and assigned them sequentially into groups of two. For instance, if a given block had four treatment students, students one and two in the random order were assigned to the same group, and students three and four to another group.

In the last step, we randomly assigned tutors to groups of two students. In general, all tutors were assigned to three tutoring groups. Randomization of tutors was stratified by geographic area, where those tutors who indicated they spoke Catalan were assigned to students based in Catalonia, and those who did not speak Catalan to those who were based in the region of Madrid.

3.7 Implementation

Students and tutors were able to organize their own schedule and agree on weekly meeting times.[11] Each student and mentor received personal and unique credentials for accessing a specifically created domain within an online platform from a large, US-based technology firm, consisting of a tool to organize emails, calendars, files and most importantly, hold online meetings. Tutors had to hold sessions through the platform and could only communicate with students through this channel.[12] Students who registered and stated they did not have access to a computer or tablet and/or internet were provided with a tablet with internet access for the duration of the program. In total, 13 students were given tablets, which were donated to their schools at the end of the program.

A key advantage of the online format was that student attendance could be monitored in real time. Throughout the program we collected data for each tutoring session via a management and monitoring dashboard that was fed with data from the technological platform where the virtual sessions were taking place. This data allowed us to immediately identify issues with the connection and quality of video calls and pupils who did not attend their sessions. With this information we could draw up plans of action with tutors, families, and schools to help get them back into the program. Figure 2 shows the distribution of total minutes and total number of tutoring sessions attended by students. Only seven students (3.4 percent of the students assigned to the treatment group) dropped out of the program before it began. Among those that did start the program, the median number of minutes of tutoring received was 960, representing 80 percent of the maximum envisaged number of minutes (1200). The median number of sessions attended was 20, corresponding to 83 percent of the maximum envisaged number of sessions (24).

4 Data

In this section we describe the data collection process, the kind of information we collected at base- and endline and the outcome measures we constructed.

[11]Students and tutors were asked and had to confirm at the registration and application stage, respectively, that they were available at least three days a week between 4 p.m. and 7 p.m.

[12]This was done both for organizational as well as for legal reasons of child protection: All communication through these channels could be monitored by us and the implementation team.

4.1 Baseline information

We collected a rich array of family and household characteristics at the stage of online registration for the program, where parents had to fill out a detailed survey. This survey included questions on household composition, civil status of the respondent (the mother or father), education level and household income, as well as the origin of the respondent and the child that was being registered for the program, and the language typically spoken at home. We also asked whether the child was currently receiving tutoring support of any kind and whether the child had a device (computer or tablet) and internet connection available at home with which to connect to the sessions.[13]

After the completion of the registration period and before randomization, we ran a baseline student survey that included a math test and a questionnaire on prior attainment, well-being and other socio-emotional outcomes. The baseline test was completed by all students in classrooms where there was at least one student registered for the program, thereby avoiding stigmatization or association of the test with the program. The tests were paper-based and administered by the children's math teachers or their main classroom teachers (also called *tutor* in Spanish) during a regular math class or the weekly lesson reserved for general matters.[14] Because of the timing of the baseline test - right before the Easter holiday and after grading for the first term had finished - students were not missing regular math content to do the test. We explicitly instructed teachers not to mention the program *Menπores* while they ran the tests, so that students who had registered would not associate this assessment to their likelihood of being selected.

Because there is no standardized test for the age groups included in the program (grade 7 and 8), we had to create our own assessment tool. Together with *ExE* experts with experience as secondary math teachers, we designed two math tests that differed by grade level and was based on the national curriculum for the respective grade levels. Sample questions are shown in Appendix A.1. The test for 7th graders included seven questions, while the test for 8th graders included six questions.

The second part of the survey, which covered well-being, socio-emotional skills and prior attainment was identical for both grade levels. Well-being questions were based on

[13] As noted in Section 3.7, having a device and internet connection was not a pre-requisite for participation, as we provided internet-enabled devices to students who did not have one at home.

[14] In Spanish secondary schools, all classes have one hour per week reserved for a class with their *tutor* in which they discuss general matters.

the well-being module in the age 14 survey of the Millennium Cohort Study (University College London et al., 2020). Students were asked six questions on how they felt about different aspects of their lives, which they had to rate on a scale of 1 to 7, where '1' meant not at all happy and '7' meant completely happy. The exact questions can be found in Appendix A.2. We calculate the average scores across the six items to create a Likert-type well-being scale.

The second set of questions comprised three items from the CARALOC Pupil Questionnaire (University College London et al., 2021), which assesses locus of control, and are answered with 'yes' or 'no', which we assign value zero and one, respectively. We calculate the average across the answers to these three questions, where a number closer to one indicates a more internal locus of control. An internal locus of control indicates that students believe they are more in control of the results of their actions in their daily lives. Students with a more external locus of control believe that fate or others are responsible for what happens in their lives. A more internal locus of control has been found to be associated with better academic outcomes (Shepherd et al., 2006). Additionally, we asked students to self-assess their ability in Spanish language, math and English. Finally, we asked students whether and how often they had attended online classes during the school closures from mid-March to June 2020, to be able to control for potential learning losses experienced during the onset of the pandemic.

4.2 Outcome Measures

During the second week of June, when tutoring sessions had finalized, we administered an endline survey. The endline survey again contained a standardized math test and also included the questions regarding well-being, locus of control and self-rated ability discussed in Section 4.1.

We added new questions on socio-emotional skills and aspirations. Socio-emotional skills were captured in several dimensions. First, as in Carlana and La Ferrara (2021), we measured grit using the Short Grit Scale developed by Duckworth and Quinn (2009). This includes eight questions with a 5-point scale, which are then aggregated into an overall Likert-scale by averaging the valuations across all questions. The exact questions can be found in Appendix A.2. Second, we measured school motivation using three items from the school motivation grid of the sixth wave (age 14 survey) of the Millennium Cohort

Study (University College London et al., 2020). These covered the frequency with which students (1) exerted high effort at school, (2) thought school was interesting, and (3) they found school a waste of time, with answers ranging from 1 (never) to 4 (always). We look at these outcomes individually and also aggregate them into a school motivation index by adding the values for each question and dividing it by the maximum sum (12). We also ask questions on homework, and interest in language and math. To measure aspirations, we asked about the plans students had for after completion of compulsory schooling at age 16 (vocational track, academic track or dropping out of school), as well as their intentions to go to college.

We also conducted an online and phone survey of parents of study participants in early July, when the school year was over. Parents were asked two key questions about their child's academic outcomes: The final math grade obtained by their children at the end of the year and whether the child will have to repeat the school year. We also asked about whether their children had received any other remedial education support program. Finally, for the case of treated children, parents responded to various questions on a scale from 1 to 5 regarding their perception of the program and its impact on their children's educational progress.

Qualitative surveys were also administered to tutors, school principals, and math teachers of treated students. All of them were asked questions on a scale from 1 to 5 regarding their general views on the program, its effect on beneficiaries, as well as its appropriateness for scaling up in the future. Besides this, tutors were also asked about the quality of the training they received and about the development of the tutoring sessions during the program.

5 Empirical Strategy

The estimation of the effect of the intervention is done using two empirical specifications. For outcomes that are measured both at base- and at endline, we estimate the following difference-in-difference regressions by OLS:

$$Y_{ibt} = \alpha_b + \beta Treat_{it} + \gamma Post_t + \delta Treat_i \times Post_t + \lambda X_i + \epsilon_{ib} \tag{1}$$

where Y denotes the outcome for student i, in block b at time $t \in \{pre, post\}$, i.e. either before or after the intervention. The α_b's are block fixed-effect (indicating the classroom of the student). $Treat_i$ is a dummy variable equal to one if the individual is in the treatment group, and $Post_t$ is a dummy variable equal to one for the period after the end of the program. The vector X represents a set of student and parent-level control variables that include student age, grade, gender, region, a dummy indicating school meal eligibility, last math grade, a set of dummy variables indicating the frequency of online lessons during school closures between April and June 2020, a dummy indicating whether the student had a tablet or computer at home before the program, a dummy indicating whether the student was receiving other tutoring before the program, categorical variables indicating the language spoken at home, parental education, household income, and household composition, an indicator for whether the responding parent is a single parent, and a dummy variable indicating whether the parent is of Spanish origin. Finally, ϵ_{ib} is an error term. The coefficient of interest, δ, corresponds to the intention-to-treat (ITT) estimate, which measures the effect of being assigned to participate in Menπores. Standard errors in this specification are clustered at the individual level.

For outcomes measured only at endline we estimate the impact of being assigned to the program with the following OLS regression:

$$Y_{ib} = \alpha_b + \delta Treat_i + \lambda X_i + \epsilon_{ib} \tag{2}$$

where the variables are defined in the exact same way as in the difference-in-difference specification above. Again, the coefficient of interest measuring the ITT effect is captured by δ. We report robust standard errors for this specification.

6 Results

In this section we present our main results, which are summarized in Figure 3. We also perform a set of robustness checks and analyze heterogeneous effects to determine potential mechanisms and groups of students that benefited more (or less) from the program.

6.1 Academic outcomes

Table 3 summarizes the results for the impact of the intervention on academic outcomes. The dependent variable in Column 1 is the score on the standardized math test, standardized by grade level.[15] Treatment students improved their score by 0.264 SD more than control students, which is significant at the 10 percent level.

Columns 2 to 4 show treatment effect estimates for teacher-assessed outcomes. We find that treatment group students have a 0.85 points higher end-of-year math grade than control students, corresponding to an increase by about 0.48 SD (Column 3) compared to the control group. Treatment students are also 22 percentage points (Column 4) more likely to have passed the subject (math), corresponding to a 32 percent increase in the likelihood of passing compared to the control group mean. Further, we find a large, negative and significant effect on grade retention. Treatment students were 9.4 percentage points less likely to have to repeat the school year (Column 5), corresponding to a 78 percent drop in the repetition probability compared to the control group.

In Table 4 we present results on outcomes measuring self-perceived ability and affinity towards math. Columns 1 and 3 show, respectively, that the program did not increase the likelihood of pupils stating that they thought they were good at math or liked math. For comparison, we also asked the same questions about Spanish language, to check whether there was some sort of crowding out (i.e. students shifting preferences toward math or away from math in favor of other subjects). Again, we do not find evidence that this happened (Columns 2 and 4).

6.2 Aspirations, perseverance, effort and motivation

We now look at the impact of the program on aspirations, perseverance and motivation. Column 1 of Table 5 shows that treatment group students were 13.6 percentage points more likely than control group students to state that they would like to go on to complete a *bachillerato* (academic high school track), the pre-requisite for entering university in Spain, after completing compulsory education. This corresponds to a 33.2 percent higher probability than the control group.

We do not find an increase in the likelihood of stating that students plan to go on to university after school (Column 2). While choosing the academic track at upper secondary school tends to be highly correlated with planning to go to university, the fact that we do not find an impact here might be because this is a decision that lies very far in the future for the students in our intervention, who were on average just 13 years old.

[15]The test score is standardized at the grade level (for grade 7 and 8, respectively) and using the mean and standard deviation of the control group at baseline.

In Column 3 of Table 5, we assess whether program assignment had any impact on grit, a measure of perseverance and conscientiousness. We do not find evidence that this was the case. It is likely that the duration of the program was too short to be able to change these personality traits.

Column 4 shows the impact of program assignment on self-perceived effort at school. Students in the treatment group were 11.6 percentage points more likely than the control group to state that they exerted high effort at school always or most of the time, corresponding to a 21.5 percent higher probability than in the control group. Column 5 shows the effect of program assignment on our school motivation index. We do not find a significant impact on this outcome.

6.3 Well-being and socio-emotional outcomes

Table 6 shows the difference-in-difference estimates of the impact of the program on measures of well-being and locus of control (Columns 1-4). We find no impact of the program on overall subjective well-being measured by the well-being index (Column 1). When looking at one of the questions included in the well-being index separately - satisfaction with school - we find a relatively large coefficient estimate equivalent to a 0.20 SD increase (Column 2), which is however not significant at conventional levels.

Column 3 shows that the intervention had no significant impact on our measure of locus of control. However, when looking at one of the variables composing the locus of control index separately - whether students agreed with the statement that when something bad happened to them it tended to be the fault of others (Column 4) - we find that this increases significantly more for treatment students than for control students, corresponding to a more external locus of control. While this result is counter-intuitive, as we would have expected the intervention, if anything, to increase internal locus of control, we interpret this as evidence for a reduction in self-blame among treatment students.

6.4 Tutor, teacher and parent feedback

At endline, we collected feedback from parents, tutors and schools, asking them to evaluate their experience with the program. While these evaluations are of a purely subjective character and have no causal interpretation, we nevertheless believe they are important to analyze potential obstacles or lessons for a potential scale up of the program in the future.

In the final surveys of the families of the pupils participating in the program, we found a general satisfaction with *Menπores*. More than 80 percent of the families agree or strongly agree with the statement 'My mentored child is more confident in the subject of mathematics'. Some 80 percent of the families agree or strongly agree with the statement: 'Tutoring has improved my child's results in mathematics at school'. Finally, 85 percent agree with the statement: 'The mathematics reinforcement program has been useful for my child'.

Tutors were satisfied with the training received. When asked about the training and support received before and during the program, 89 percent of the tutors agreed or strongly agreed that the training plan was useful. Some 89 percent also agreed or strongly agreed that the training on the platform was adequate and sufficient. Some 68 percent agreed or strongly agreed that the webinars were useful and sufficient. Moreover, 86 percent of tutors consider the combination of online training and webinars to be adequate.

Tutors also answered several questions about their satisfaction with the project. Some 82 percent agreed or strongly agreed that participating in the project has helped them better understand the social reality facing students. Some 95.4 percent of tutors agreed or strongly agreed that the program has enriched them personally and professionally.

Mathematics teachers and headteachers of the participating schools rated the impact of the program positively. More than 70 percent of the teachers and 57 percent of the headteachers surveyed agreed or strongly agreed that the program has been useful for their pupils. Some 69 percent of the teachers believe that the program is a good support for their teaching. Finally, 71 percent think that the program should continue, which is also shared by 100 percent of the headteachers surveyed. Finally, 40 percent of surveyed mathematics teachers believe that the fact that pupils participated in the program helped them to work better, and another 42 percent believe that the coordination meetings with the mentors were useful. Some teachers said that they were overwhelmed by the additional workload during the program (due to coordination with tutors and administering base and endline tests), which might explain the lower proportion of positive responses. Finally, in the open-ended responses, several teachers and headteachers mentioned the need to start these programs before April and make it longer.

In terms of lessons for a potential scale-up, in order for the program to be positively regarded by teachers it should be ensured that it does not mean additional workload for them. It would also be valuable to test whether a different timing (more towards the beginning or middle of the academic year) and a longer duration would make the program even more effective.

6.5 Discussion of results and robustness checks

We now discuss several issues around the internal and external validity of our main results. The choice of specification or selective attrition might also cause bias in our estimates. We therefore perform several robustness checks to see whether our results remain robust to different specifications, and taking into account selective attrition.

Contamination of control group In response to the pandemic school closures, many governments launched additional support programs to close learning gaps that emerged during lockdowns. Such competing programs were also launched in Spain around the same time as ours, which constituted a risk of contamination of the control group. To check whether this was likely a problem, during the endline survey we asked parents whether students received any other tutoring or academic support program in math or other subjects during the period while *Menπores* was implemented. Indeed, as Figure 4 shows, nearly 40 percent of control group students received some other tutoring or academic support in math, compared to only around 12 percent of the treatment group. The control group was also more likely to have received additional support in another subject, indicating that schools and/or parents might have compensated control group students with other offers. Overall, these findings suggest that our impact estimates could be interpreted as lower bounds.

External validity Our program was specifically targeted at schools in disadvantaged areas, which means that our sample is not representative of the population of Spanish 7th and 8th grade students as a whole. To get a sense of how students at participating schools compare to our schools, Table B1 shows summary statistics of learning outcome indicators and socio-economic characteristics spereately for the entire population of schools in the region of Madrid and for those schools that participated in our experiment.[16] In Column

[16]We cannot do the same exercise for the schools in Catalonia as we do not have access to school level statistics for that region.

1 we can see that the ESCS index, measuring student socio-economic status, of schools participating in the program is half a standard deviation below the regional average, approximately placed in the 25th percentile in the overall socio-economic distribution. Columns 2 to 5 show that students in participating schools were on average much lower performing, by between 73 (Spanish), 16 (Math), 90 (English) and 74 (Social subjects) percent of a standard deviation with respect to the regional average. Participating schools also have a lower overall share of children born to Spanish parents. Overall, students at our participating schools are on average lower performing and more disadvantaged than the average population of students in Madrid. This should be kept in mind when interpreting our results.

Volunteer tutors We had initially planned a third treatment arm, where we wanted to compare the effectiveness of volunteer tutors with that of our professional tutors. While we did not achieve enough volunteer applications in order to fully implement this third treatment arm, we included them in the randomization and eventually 19 students were taught by such volunteers. All results presented so far exclude these students, but in Tables B2 to B5 we present our main results including all students that were part of the randomization. As can be seen, results are essentially identical to our main specifications, apart from slightly higher point estimates in our main effects when volunteer tutors are excluded. This is consistent with evidence showing that professional tutors are more effective than volunteers (Nickow et al., 2020), but we cannot draw any strong conclusions from this evidence given the extremely small sample sizes.

Alternative specifications The results shown thus far correspond to those using our preferred, full specification. In Columns 1-3 of Tables 7 to 10, we additionally show regressions using alternative specifications. Column 1 shows the most basic specifications, including only the treatment dummies and block fixed effects. In Column 2 we add demographic characteristics (age, gender, grade, autonomous community, whether eligible for school meal subsidy, baseline math grade, whether had online classes during lockdown, whether had a device to connect to tutoring sessions available, whether received some for of academic tutoring at baseline) and in Column 3 we add variables relating to socio-economic status (whether speaks Spanish at home, household income, number of household members below age 18, parental education, whether living in a single-parent household, and whether the parent is of Spanish origin), corresponding to our main specification shown so far.

When looking at academic outcomes in Table 7, results are very stable across specifications, with effect size estimates mostly increasing as we add more controls to take into account heterogeneity at baseline across treatment and control group. For non-academic outcomes reported in the remaining Tables 8 to 10 the same holds: Estimates are remarkably stable across specifications and so are their statistical significance levels. We therefore conclude that our main results are robust to the inclusion or exclusion of specific control variables.

Attrition Out of the 356 children who initially signed up for the program, 93.3 percent took part in the baseline survey and math test, and 87.7 percent did so in the endline survey and math test.[17] To check whether there is selective attrition at endline, in Table 11 we run regressions where the outcome variable is a dummy equal to one if the child participated in the endline survey and math test, and zero otherwise. Column 1 shows that children in the treatment group were 4.5 percentage points more likely to respond, but the coefficient is not significant. In Column 2, we add control variables measured at baseline to see whether they correlate with attrition. Students in grade eight (versus those in grade seven) were more likely to respond at endline, but keeping grade constant, the association between age and responding at endline is negative. Other child characteristics, such as gender or whether the child is receiving a school meal subsidy, show no significant association with responding at endline.

In terms of parental and household characteristics, there is no significant association between parental education and responding at endline, even though coefficients tend to be negative, meaning that compared to the omitted category (primary education or no education), children of more highly educated parents were less likely to respond at endline. We observe a positive correlation between household earnings and the probability of responding at endline, but none of these coefficients is significant at standard levels.

Among the group assigned to treatment, we observe very similar associations between the likelihood of responding and child and family characteristics as for the whole sample (Column 3). The same holds for control group students (Column 3). Given these results,

[17]Recall that tests were done during class hours and the main reason why children might have been absent is because they happened not to be in school that day, which explains the low overall rate of attrition.

it is not clear in which direction selection bias for the child survey at endline might go, as on the one hand we see a negative association between parental education and response probability and on the other hand a positive association between household income and responding at endline. We conclude that it is likely that, if anything, we have a slight positive selection bias in terms of child responses, which would mean that our effect size estimates would be downwardly biased.

In Table 12 we perform the same exercise as above, but now looking at the the likelihood of parents responding to the endline phone survey, which we undertook in order to obtain teacher-assessed outcomes, such as the final grade in math, and whether the child had to repeat the school year. Parents of children in the treatment group were 13.6 percentage points more likely to respond than those in the control group (the overall response rate was only 62 percent, much lower than the response rate for the child endline test and survey at school). More highly educated parents were more likely to respond, but this was not true for parents of children in the treatment group, meaning we might have some negative selection in responses within this group. We see a clear negative association between household income and responding at endline for the parent survey. This indicates that overall, selection of parents into responding at endline might be slightly negative, especially for the treatment group. This would potentially bias our impact estimates based on parent reported outcomes, like the end-of-year math grades and passing the school year, downwards.

To quantify how much this might matter for our results, Column 4 in Table 7 shows the estimated impact of program assignment on academic outcomes using inverse-probability weights to account for selective attrition. We can see that the effects size on our standardized math test (Panel A) is virtually identical when using these weights compared to our main result (reported in Column 3). Panels B to C are outcomes that rely on parental responses to the endline survey. If anything, effect size estimates here are higher for all three outcomes when using inverse-probability weights, which is in line with the potential bias (downward) predicted from our analysis of selective attrition.

In Tables 8 to 10 we perform the same exercise as above for non-academic outcomes. Column 4 of each table shows the effect size estimates using inverse-probability weights, and Column 3 reports our main results for comparison. Effect size estimates are not substantially different from one another. To conclude, our estimates are robust to accounting for attrition and if anything, are slightly downwardly biased in the case of academic outcomes reported by parents.

6.6 Heterogeneous effects

We now turn to analyzing potential mechanisms and heterogeneous effects in order to see what might be driving the impact of the program and to see whether certain groups of students benefited more from the program than others. Due to the relatively small sample size, we are not generally powered to detect small effect size differences between different groups. Nevertheless, the evidence presented gives some indications as to the direction of heterogeneous effects. We focus on academic outcomes and outcomes measuring aspirations and motivation for this analysis, given that we do not find much of an effect of the program on either self-perceived ability and affinity (toward math) or well-being.

Internet connection quality In Table 13 we report heterogeneous effects by connection quality (Panel A). We define the internet connection quality based on data collected during all sessions a child connected to. A connection is classified as good if the average ping (or latency) across all online sessions is lower than 100.[18] Having experienced a good connection quality is not associated with a higher impact of program assignment on the score in the standardized test (Column 1). For the final math grade (Column 2) and passing math (Column 3), the positive effect of the intervention seems to be concentrated entirely on those who experienced a good internet connection. These results are indicative of the importance of technology and high quality internet connections if online tutoring is meant to be used as a public policy tool to combat learning gaps. However, we might also be picking up unobserved heterogeneity here, if having a bad internet connection is associated with other unobserved characteristics that affect treatment impact. We do not see any significant associations between our main non-academic outcomes and connection quality (Columns 5-8).

Device used to connect We next check whether it mattered on what kind of device the tutoring sessions were attended; tablet or phone versus a laptop or computer. Results

[18] A ping or latency is the time (measured in milliseconds) it takes to send information from your local computer to the internet server. A value of above 100 is considered a "slow" speed.

are shown in Panel B of Table 13. Having used a tablet (or phone) rather than a computer to connect to tutoring sessions is generally associated with a smaller effect of treatment assignment, but none of the reported coefficients is significant. We cannot draw strong conclusions on whether the type of device used matters. However, these results are in line with the findings in Carlana and La Ferrara (2021).

Tutor gender Because tutors were randomly assigned to groups, we can assess the relevance of tutor gender for program impact. About 77 percent of students were tutored by females. Panel C of Table 13 shows no clear pattern of whether male or female tutors were more effective. In terms of the standardized test, the effect size is smaller for females, even though the difference is not statistically significantly different from zero. For teacher-assessed outcomes, we find a higher effect on the final math grade, but a lower effect on passing math (neither being significant). This suggests that possibly female tutors were more effective with students that were not at the margin of failing math, but those slightly higher up the grade distribution.

Gender composition of group In Panel D of Table 13 we show heterogeneous effects by whether the students in the group were of the same gender. In our sample, about half of treatment groups students were in groups where both students were either both girls or both boys. The signs of the coefficients suggest that students in such groups experienced higher improvements in their final math grade (not significant), the likelihood to pass the course (significant), and greater reductions in their likelihood of having to repeat the year (not significant). We also find higher effect size estimates for groups where both students were of the same gender when it comes to non-academic outcomes: Wanting to do a *Bachillerato* (academic upper secondary track) and exerting high effort at school often or always. These results suggest that same-gender groups were more motivating and had a bigger effect on non-cognitive development, which in turn might have positively affected behavior and interactions in class, leading to positive effects on teacher-assessed outcomes.

Group ability composition In Panel E of Table 13 we check whether the ability composition of the group mattered for program assignment impact. We create a dummy variable equal to one whenever students in the same tutoring group scored either both above or both below the median at baseline. None of the interaction terms between this dummy and and the treatment dummy are significant, but the magnitudes of the coefficients and their signs suggest that improvements in teacher-assessed outcomes might have been smaller for groups with more homogeneous abilities.

Child gender Panel A of Table 14 shows heterogeneous effects by student gender. Girls seem to have benefited more from the program in academic terms when using our standardized test score (Column 1) as the outcome of interest, but boys experienced larger increases in their final math grade and in their probability to pass the subject (Columns 2 and 3). When it comes to non-academic outcomes (Columns 5-8), the interaction term between treatment dummies and girl dummies tend to be negative for all outcomes, except for the probability to agree that when bad things happen to them, it tends to be the fault of others (Column 8), suggesting a more external locus of control for girls assigned to the treatment group. This suggests that overall, improvements in these outcomes where concentrated among boys. However, most of these interaction terms are not statistically significantly different from zero.

Grade We also look at whether the effect differed between children of different grade levels. Panel B of Table 14 shows that while those in second year of secondary school seem to have higher improvements on the standardized test (Column 1), the opposite is true for final math grade (Column 2), where the students in first grade of secondary school experienced bigger improvements, which also translated into a greater increase in the probability of passing the subject (Column 3). However, none of these interaction terms is significant at conventional levels. School satisfaction overall seems to have improved more for the older students (Column 7).

Household characteristics Lastly, we look at characteristics of the family. Panel C of Table 14 looks at heterogeneous effects by whether the child lives in a single-parent household. We find that improvements on the standardized test are lower for these children (Column 1, not significant), but improvements on teacher-assessed outcomes are generally higher (but not significant). Also aspirations to do the academic upper secondary track increase more for those from single-parent households (Column 5, not significant), as does the likelihood of stating that bad things tend to be the fault of others (Column 8, significant). Finally, children with an immigrant background (Panel D) show a similar pattern as those from single-parent households: They benefit less when measuring academic success through the score on the standardized test (Column 1) , but more when measuring it through final math grade and passing the subject (Columns 2 and 3). However, it raised their aspirations to do the academic upper secondary track by less then for children without immigrant background (Column 5, not significant), and assignment to the intervention does not seem to have increased their likelihood of making high effort at school (Column 6, significant).

7 Conclusion

The pandemic crisis and the school closures that followed resulted in significant learning losses and widening educational inequalities for millions of pupils around the world. Catch-up interventions to reverse such losses should be a priority for policy makers. While face-to-face tutoring has been widely evaluated as an effective policy response, very little experimental evidence exists on the effectiveness of online tutoring programs for secondary school students.

In this study we show that in a nearly normal schooling environment, our novel 100 percent online intensive tutoring program in small groups of two students improved academic outcomes, effort and aspirations of socially disadvantaged students. The 8-week program significantly increased standardized test scores (+ 0.26 SD), end of year math grades (+0.48 SD) and the probability of passing the subject (by about 22 percent with respect to the control group mean), while reducing the probability of repeating the school year (by about 78 percent with respect to the control group). In terms of non-academic outcomes, the intervention contributed to improve self-reported effort and aspirations. The students that were selected to receive tutoring were 14 percentage points more likely to state that they would go to the academic track after compulsory schooling, and 12 percentage points more likely than the control group to state that they exerted high effort at school always or most of the time.

Governments and international organizations are spending large amounts of money and resources to respond to the learning loss caused by the pandemic worldwide. Our results have immediate policy relevance to inform on how to design effective policy responses to recover these learning losses. Online tutoring programs have the advantage of reaching children at a lower cost and can be provided to any child with an internet connection, including those in remote places were traditional tutoring programs are harder to deliver. Moreover, our two-students-per-tutor format has the benefit of being more cost-effective than other alternatives with professional teachers, such as face-to-face small groups or one-to-one online programs.

In terms of implementation, a key advantage of the online format is that attendance can be tracked in real time and one can react with action plans immediately when students are starting to lag behind or be absent. In our experiment, this ability might have been one of the reasons explaining the high completion rate of the program, despite the intensity (three sessions per week) and the fact that most participants came from highly disadvantaged backgrounds. When thinking about implementing such a program at bigger scale, these considerations are important, as data driven monitoring is a key advantage of online programs.

As regards to external validity, one of the main contributions of our study is that the program was implemented while schools were open thus providing a complement to formal schooling, which is closer to a normal setting. A potential limitation of our design for large scale programs might be the secular shortage of qualified math teachers (Santiago, 2002).

For future research, it will be relevant to explore whether the positive results of online tutoring shown here hold in different contexts: with primary school students, with variations in socio-emotional support or teacher training, focusing on other subjects, such as reading, or increasing the number of students per tutor (three-to-one, for instance). Also, it would be interesting to explore in more detail the potential benefits of small-group positive peer dynamics in online teaching. The remarkable academic effect of the program as well as the success in attendance and completion rates indicate that our two-to-one design might have helped to mitigate some of the shortcomings found in the literature in online education, such as a lack of perseverance and motivation (Escueta et al., 2020).

References

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti, "When the Great Equalizer Shuts Down: Schools, Peers, and Parents in Pandemic Times," Working Paper 28264, National Bureau of Economic Research December 2020.

Angrist, Noam, Peter Bergman, and Moitshepi Matsheng, "School's Out: Experimental Evidence on Limiting Learning Loss Using "Low-Tech" in a Pandemic," Working Paper 28205, National Bureau of Economic Research December 2020.

Arenas, A. and L. Gortazar, "Learning Loss one Year after School Closures: Evidence from the Basque Country," *IEB Working Paper 2022/03*, 2022.

Azevedo, Joao Pedro, Halsey Rogers, Sanna Ellinore Ahlgren, Marie-Helene Cloutier, Borhene Chakroun, Gwang-Cho Changl, Suguru Mizunoya, Jean Nicolas Reuge, Matt Brossard, and Jessica Lynn., "The State of the Global Education Crisis : A Path to Recovery," Technical Report 2021.

Birkelund, Jesper Fels and Kristian Bernt Karlson, "No Evidence of a Major Learning Slide 14 Months into the COVID-19 Pandemic in Denmark," 2021.

Blainey, Katie and Timo Hannay, "The impact of school closures on autumn 2020 attainment," White papers, RS Assessment from Hodder Education 2021.

Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman, and John Eric Humphries, "What grades and achievement tests measure," *Proceedings of the National Academy of Sciences*, 2016, *113* (47), 13354–13359.

Carlana, Michela and Eliana La Ferrara, "Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic," Discussion Paper IZA DP No.14094, IZA Institute of Labor Economics 2021.

Chetty, Raj, John N Friedman, Nathaniel Hendren, Michael Stepner, and The Opportunity Insights Team, "How Did COVID-19 and Stabilization Policies Affect Spending and Employment? A New Real-Time Economic Tracker Based on Private Sector Data," Working Paper 27431, National Bureau of Economic Research June 2020.

Comunidad de Madrid, "Invertimos más de 6,1 M en adquirir 36.100 tablets para los centros educativos," 2020. Accessed on 27/11/2021 from https://www.comunidad.madrid/noticias/2020/11/18/invertimos-61-m-adquirir-36100-tablets-centros-educativos.

Contini, Dalit, Maria Laura Di Tommaso, Caterina Muratori, Daniela Piazzalunga, and Lucia Schiavon, "The COVID-19 Pandemic and School Closure: Learning Loss in Mathematics in Primary Education," Discussion Paper IZA DP No.14785, IZA Institute of Labor Economics 2021.

Cooper, Harris, Kelly Charlton, Jeff C Valentine, Laura Muhlenbruck, and Geoffrey D Borman, "Making the most of summer school: A meta-analytic and narrative review," *Monographs of the society for research in child development*, 2000, pp. i–127.

Department for Education, *Understanding Progress in the 2020/21 Academic Year. Findings from the summer term and summary of all previous findings. October 2021*, Department for Education, 2021.

Dobbie, Will and Jr. Fryer Roland G., "Getting beneath the Veil of Effective Schools: Evidence from New York City," *American Economic Journal: Applied Economics*, October 2013, *5* (4), 28–60.

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

Eisner, Manuel, Denis Ribeaud, Giuseppe Sorrenti, and Ulf Zölitz, "The Causal Impact of Socio-Emotional Skills Training on Educational Success," *Mimeo*, 2020.

Elliot Major, Lee, Stephen Machin, and Andrew Eyles, "Generation COVID: Emerging work and education inequalities," *CEP Covid-19 analysis No. 011*, 2020.

Engzell, Per, Arun Frey, and Mark D. Verhagen, "Learning loss due to school closures during the COVID-19 pandemic," *Proceedings of the National Academy of Sciences*, 2021, *118* (17).

Escueta, Maya, Andre Joshua Nickow, Philip Oreopoulos, and Vincent Quan, "Upgrading Education with Technology: Insights from Experimental Research," *Journal of Economic Literature*, December 2020, *58* (4), 897–996.

Farquharson, Christine, Sonya Krutikova, Angus Phimister, Adam Salisbury, and Almudena Sevilla, "The return to school and catch-up policies," Briefing note BN318, Institute for Fiscal Studies March 2021.

Haelermans, Carla, Madelon Jacobs, Lynn van Vugt, Bas Aarts, Henry Abbink, Chayenne Smeets, Rolf van der Velden, and Sanne van Wetten, "A full year COVID-19 crisis with interrupted learning and two school closures: The effects on learning growth and inequality in primary education," 2021.

Hardt, David, Markus Nagler, and Johannes Rincke, "Tutoring in (Online) Higher Education: Experimental Evidence," *CESifo Working Paper No. 9555*, 2022.

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

Higgins, Steve, Dimitra Kokotsaki, and Robert Coe, "The teaching and learning toolkit," Technical Report 2012.

Instituto Nazionale di Statistica, "Rapporto annuale 2020. La situazione del paese.," 2020. Accessed on 21/12/2020 from https://www.istat.it/storage/rapporto-annuale/2020/Sintesi2020.pdf.

Kosse, Fabian and Michela Tincani, "Prosociality Predicts Labor Market Success Around the World," *Nature Communications*, 10 2020, *11*, 5298.

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

Lichand, Guilherme, Carlos A. Dória, Onicio Leal Neto, and João Cossi, "The Impacts of Remote Learning in Secondary Education: Evidence from Brazil during the Pandemic," Technical Report IDB-TN-02214, Inter-American Development Bank 2021.

Maldonado, Joana Elisa and Kristof De Witte, "The effect of school closures on standardised student test outcomes," *British Educational Research Journal*, 2021.

Ministerio de Educación y Formación Profesional, "Un año de pandemia: impacto en la educación," 2021. Accessed on 11/7/2021 from https://www.educacionyfp.gob.es/prensa/actualidad/2021/03/120321-balanceestadodealarma.html.

Newlove-Delgado, T., S. McManus, K. Sadler, S. Thandi, T. Vizard, C. Cartwright, T. Ford, Mental Health of Children, and Young People group, "Child mental health in England before and during the COVID-19 lockdown," *Lancet Psychiatry*, May 2021, *8* (5), 353–354. © 2021. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/.

Nickow, Andre, Philip Oreopoulos, and Vincent Quan, "The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence," Working Paper 27476, National Bureau of Economic Research July 2020.

OECD, "The State of Global Education: 18 Months into the Pandemic," Technical Report 2021.

Pizarro-Ruiz, Juan Pablo and Nuria Ordóñez-Camblor, "Effects of Covid-19 confinement on the mental health of children and adolescents in Spain," *Scientific Reports*, 2021, *11* (1), 1–10.

Plataforma de Infancia, "Un curso escolar atravesado por la pandemia: Análisis de la inversión en educación para la adaptación al contexto COVID-19," 2021. Accessed on 21/12/2021 from https://coleseguro.es/resultados/#downloads.

Santiago, Paulo, "Teacher demand and supply: Improving teaching quality and addressing teacher shortages," 2002.

Schult, Johannes and Marlit Annalena Lindner, "Did students learn less during the COVID-19 pandemic? Reading and mathematics competencies before and after the first pandemic wave," 2021.

Shepherd, Stephanie, Dean Owen, Trey J Fitch, and Jennifer L Marshall, "Locus of control and academic achievement in high school students," *Psychological reports*, 2006, *98* (2), 318–322.

Tomasik, Martin J., Laura A. Helbling, and Urs Moser, "Educational gains of in-person vs. distance learning in primary and secondary schools: A natural experiment during the COVID-19 pandemic school closures in Switzerland," *International Journal of Psychology*, 2021, *56* (4), 566–576.

University College London, UCL Institute of Education, and Centre for Longitudinal Studies, *Millennium Cohort Study: Sixth Survey, 2015, SN-8156*, [data collection]. 7th Edition. UK Data Service, 2020.

_ , _ , and _ , *1970 British Cohort Study Response Dataset, 1970-2016, SN: 5641*, [data collection]. 4th Edition. UK Data Service, 2021.

World Bank, "Accelerate Learning Recovery," Technical Report 2021.

_ , "Remediating Learning Loss," Technical Report 2021.

Figures

Figure 1: Timeline of the Menπores program implementation

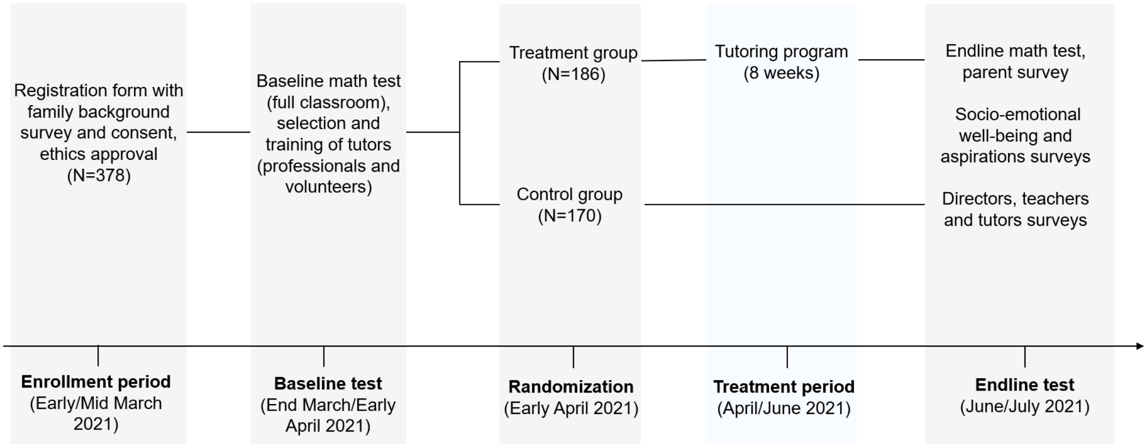
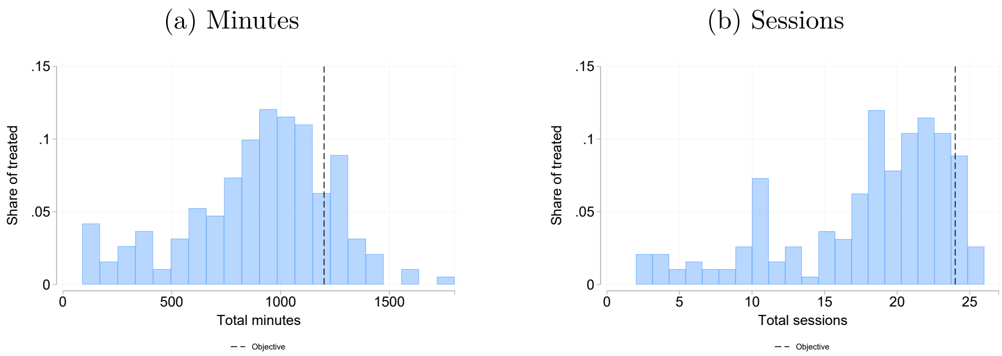


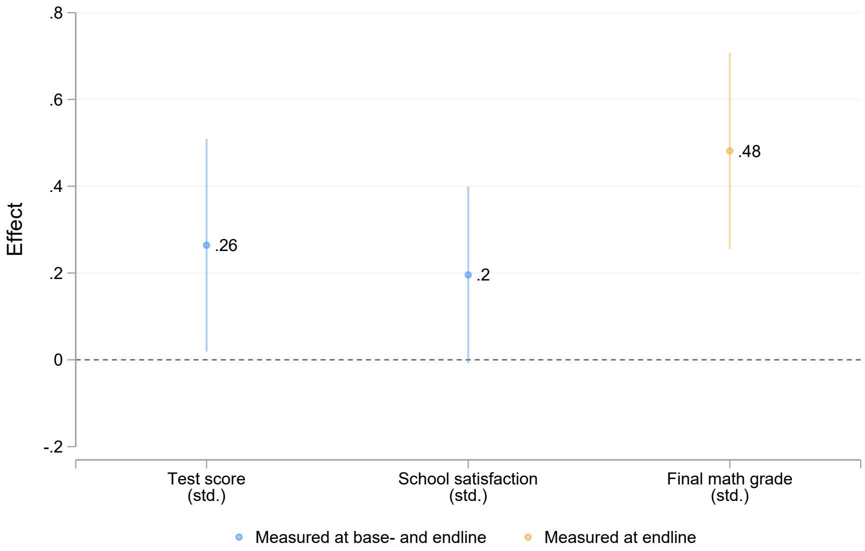
Figure 2: Distribution of total minutes and number of sessions of tutoring attended



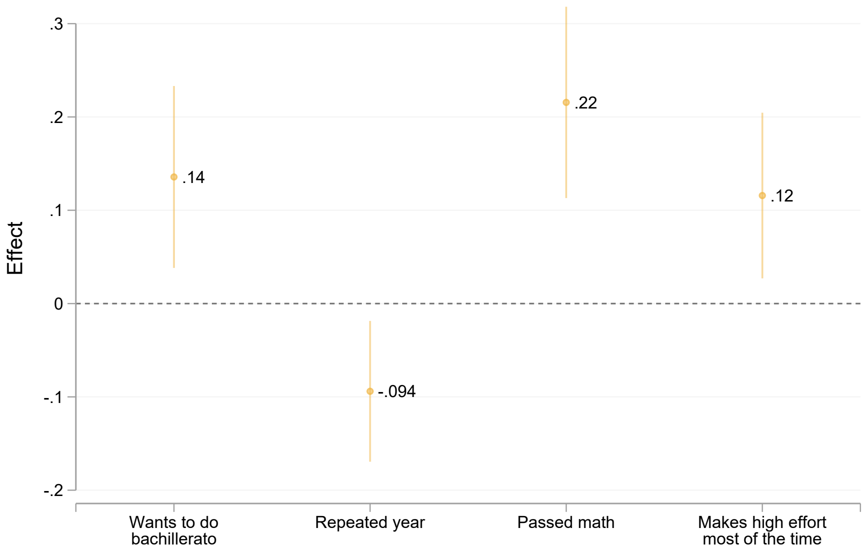
Note: This figure shows histograms of the total number of tutoring minutes attended (left panel) and the number of sessions attended (right panel). The data comes from electronic records of connection times to online meetings.

Figure 3: Main results

(a) Standardized outcome variables

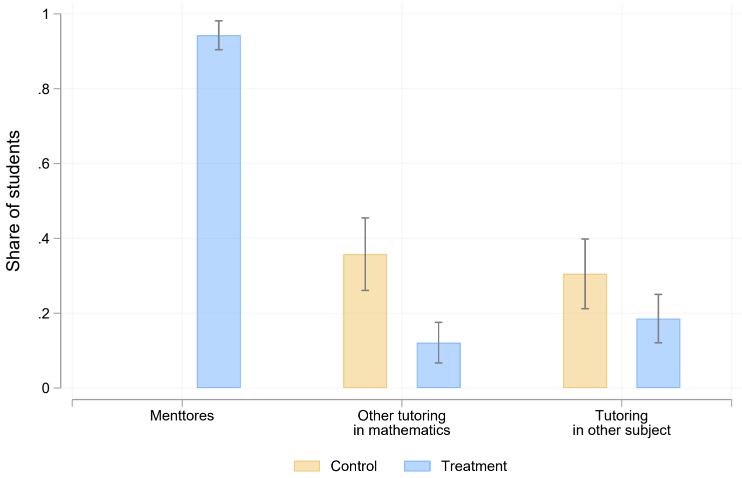


(b) Non-standardized outcome variables



Note: This figure shows the coefficient estimates of the impact of assignment to treatment for the main outcomes of interest and their 90% confidence intervals. For standardized variables, the control group has mean 0 and standard deviation of 1. In the case of variables that are measured both at base- and endline, we standardize using the control group baseline mean and standard deviation.

Figure 4: Counterfactuals



Note: This figure shows the share or treatment and control group students who received i) the *Menπores* program, ii) another tutoring or academic support program in math, and iii) another tutoring or academic support program in another subject. Sample of students whose parents responded to the endline survey.

Tables

Table 1: Balancing table - parental characteristics

	(1) Treat	(2) Control	(3) Difference (2)-(3)	(4) p-value Col. (3)
Mother responded	77.96	76.47	1.49	0.74
Lives as couple	72.04	71.18	0.87	0.86
Spanish origin	55.38	51.76	3.61	0.50
Compulsory schooling or below	52.15	51.76	0.39	0.94
Income < 1000 EUR	44.09	47.65	-3.56	0.50
Catalunya	26.34	31.18	-4.83	0.32
Madrid	73.66	68.82	4.83	0.32
HH size	4.08	4.11	-0.03	0.80
Nb. children age ≤ 18	2.04	1.88	0.16	0.16
Age of youngest child	9.76	9.96	-0.20	0.63
N	186	170		

Table 2: Balancing table - child characteristics

	(1) Treat	(2) Control	(3) Difference (1)-(2)	(4) p-value Col. (3)
Age	13.02	13.04	-0.02	0.83
Girl	47.85	48.82	-0.97	0.85
Public school	22.04	25.29	-3.25	0.47
Born in Spain	84.41	82.94	1.47	0.71
Spanish spoken at home	86.02	81.18	4.85	0.22
School meal stipend	9.14	7.65	1.49	0.61
No tablet or computer	9.68	6.47	3.21	0.27
Has access to internet	99.46	98.82	0.64	0.51
Is receiving some form of academic support	19.89	18.24	1.66	0.69
Grade 7	47.85	51.76	-3.92	0.46
Grade 8	52.15	48.24	3.92	0.46
Baseline child survey outcome				
Partial completion, 1+ maths question (%)	87.63	91.18	-3.54	0.28
Completed survey fully (%)	47.31	47.65	-0.34	0.95
Baseline performance				
Has failed math	52.29	46.04	6.24	0.29
Has failed at least one subject	80.00	77.14	2.86	0.55
Repeated grade at least once	24.73	26.47	-1.74	0.71
Test score (%)	24.69	26.06	-1.37	0.50
N	186	170		

Table 3: Impact on academic outcomes

	(1) Standardized test score	(2) Final math grade	(3) Passed math	(4) Repeated year
Post x Treat	0.264*			
	(0.148)			
Treat		0.846***	0.216***	-0.094**
		(0.241)	(0.062)	(0.046)
Constant	-0.396	9.361***	1.089**	0.374
	(0.598)	(2.451)	(0.482)	(0.350)
Mean dep. var.	-0.01	5.10	0.68	0.12
SD dep. var.	1.00	1.75	0.47	0.32
R^2	0.30	0.63	0.64	0.58
Obs.	644	220	220	218

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SEs clustered at student level (Column 1) and robust SEs (Columns 2-4) in parenthesis. The table shows the δ coefficients from Equation (1) in Column 1 and from Equation (2) for Columns 2-4. All regressions include block FEs and control for student age, grade, gender, region, a dummy indicating school meal eligibility, last math grade, a set of dummy variables indicating the frequency of online lessons during school closures in April and May 2020, a dummy indicating whether the student had a tablet or computer at home before the program, a dummy indicating whether the student was receiving other tutoring before the program, categorical variables indicating the language spoken at home, parental education, household income, and household composition, an indicator for whether the responding parent is a single parent, and a dummy variable indicating whether the parent is of Spanish origin.

Table 4: Impact on self-perceived ability and affinity

	(1) Good at math	(2) Good at Spanish	(3) Likes math	(4) Likes Spanish
Post x Treat	-0.030	0.002		
	(0.048)	(0.067)		
Treat			-0.052	-0.042
			(0.050)	(0.055)
Constant	-0.215	0.184	0.283	0.280
	(0.333)	(0.397)	(0.573)	(0.404)
Mean dep. var.	0.25	0.55	0.36	0.41
SD dep. var.	0.44	0.50	0.48	0.49
R^2	0.38	0.26	0.45	0.39
Obs.	628	629	351	353

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SEs clustered at student level (Columns 1-2) and robust SEs (Columns 3-4) in parenthesis. The table shows the δ coefficients from Equation (1) in Columns 1-2 and from Equation (2) for Columns 3-4. All regressions include block FEs and control for student age, grade, gender, region, a dummy indicating school meal eligibility, last math grade, a set of dummy variables indicating the frequency of online lessons during school closures in April and May 2020, a dummy indicating whether the student had a tablet or computer at home before the program, a dummy indicating whether the student was receiving other tutoring before the program, categorical variables indicating the language spoken at home, parental education, household income, and household composition, an indicator for whether the responding parent is a single parent, and a dummy variable indicating whether the parent is of Spanish origin.

Table 5: Impact on aspirations and motivation

	(1) Bachi llerato	(2) College	(3) Grit	(4) High effort	(5) Motivation school
Treat	0.136**	0.034	0.091	0.116**	0.011
	(0.059)	(0.050)	(0.063)	(0.054)	(0.019)
Constant	0.401	1.375***	2.485***	0.072	0.759***
	(0.530)	(0.325)	(0.511)	(0.404)	(0.192)
Mean dep. var.	0.41	0.81	3.04	0.54	0.75
SD dep. var.	0.49	0.39	0.50	0.50	0.16
R^2	0.46	0.43	0.40	0.47	0.43
Obs.	312	301	311	351	303

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. Robust SEs in parenthesis. The table shows the δ coefficients from Equation (2). All regressions include block FEs and control for student age, grade, gender, region, a dummy indicating school meal eligibility, last math grade, a set of dummy variables indicating the frequency of online lessons during school closures in April and May 2020, a dummy indicating whether the student had a tablet or computer at home before the program, a dummy indicating whether the student was receiving other tutoring before the program, categorical variables indicating the language spoken at home, parental education, household income, and household composition, an indicator for whether the responding parent is a single parent, and a dummy variable indicating whether the parent is of Spanish origin.

Table 6: Impact on socio-emotional outcomes

	(1) Wellbeing index	(2) School satisfaction	(3) Locus of control	(4) Fault of others
Post x Treat	0.032	0.265	-0.058	0.154***
	(0.109)	(0.167)	(0.037)	(0.059)
Constant	5.871***	4.624***	0.324*	-0.050
	(0.740)	(1.078)	(0.188)	(0.276)
Mean dep. var.	6.21	5.47	0.61	0.29
SD dep. var.	1.43	1.36	0.31	0.45
R^2	0.66	0.30	0.29	0.25
Obs.	644	632	641	611

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SEs clustered at student level in parenthesis. The table shows the δ coefficients from Equation (1). All regressions include block FEs and control for student age, grade, gender, region, a dummy indicating school meal eligibility, last math grade, a set of dummy variables indicating the frequency of online lessons during school closures in April and May 2020, a dummy indicating whether the student had a tablet or computer at home before the program, a dummy indicating whether the student was receiving other tutoring before the program, categorical variables indicating the language spoken at home, parental education, household income, and household composition, an indicator for whether the responding parent is a single parent, and a dummy variable indicating whether the parent is of Spanish origin.

Table 7: Robustness - academic outcomes

	(1) block FE	(2) +demog	(3) +SES	(4) IPW
	Panel A: Standardized test score			
Post x Treat	0.232	0.252*	0.264*	0.270*
	(0.143)	(0.145)	(0.148)	(0.145)
Constant	-0.289**	-0.745***	-0.396	-0.309
	(0.127)	(0.284)	(0.598)	(0.639)
R^2	0.22	0.26	0.30	0.31
Obs.	644	644	644	644
	Panel B: Final math grade			
Treat	0.797***	0.932***	0.846***	1.022***
	(0.230)	(0.215)	(0.241)	(0.285)
Constant	5.469***	5.860***	9.361***	9.591***
	(0.635)	(0.867)	(2.451)	(2.722)
R^2	0.36	0.51	0.63	0.84
Obs.	220	220	220	220
	Panel C: Passed math			
Treat	0.162***	0.194***	0.216***	0.235***
	(0.062)	(0.055)	(0.062)	(0.067)
Constant	0.892***	0.792***	1.089**	1.274**
	(0.068)	(0.169)	(0.482)	(0.558)
R^2	0.35	0.57	0.64	0.85
Obs.	220	220	220	220
	Panel D: Repeated year			
Treat	-0.094**	-0.090**	-0.094**	-0.149***
	(0.037)	(0.040)	(0.046)	(0.056)
Constant	0.062	0.177	0.374	0.025
	(0.040)	(0.123)	(0.350)	(0.502)
R^2	0.43	0.51	0.58	0.84
Obs.	218	218	218	218

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SEs clustered at the individual level (Panel A) or robust SEs (Panels B-D) in parenthesis. Column 1 shows OLS regression so the dependent variable in the column heading on a treatment dummy and block fixed effects only. In Column 2, the following additional controls are added: Child age, gender, grade, autonomous community, whether eligible for school meal subsidy, baseline math grade, whether had online classes during lockdown, whether had a device to connect to tutoring sessions available, whether received some for of academic tutoring at baseline. In Column 3 we further add controls relating to socio-economic status and parental characteristics; Whether speaks Spanish at home, dummies for household income, number of household members below age 18, dummies for parental education, a dummy for whether living in a single-parent household, and whether the parent is of Spanish origin). In Column 4 we present the full specification as in Column 3, but using inverse-probability weights to derive estimates.

Table 8: Robustness - self-perceived ability and affinity

	(1) block FE	(2) +demog	(3) +SES	(4) IPW
	Panel A: Good at math			
Post x Treat	-0.017	-0.036	-0.030	-0.029
	(0.047)	(0.047)	(0.048)	(0.050)
Constant	-0.003	0.116	-0.215	-0.181
	(0.027)	(0.131)	(0.333)	(0.380)
R^2	0.26	0.34	0.38	0.38
Obs.	628	628	628	628
	Panel B: Good at Spanish			
Post x Treat	-0.002	0.004	0.002	0.018
	(0.064)	(0.066)	(0.067)	(0.075)
Constant	0.719***	0.637**	0.184	0.144
	(0.251)	(0.277)	(0.397)	(0.326)
R^2	0.17	0.21	0.26	0.27
Obs.	629	629	629	629
	Panel C: Likes math			
Treat	-0.052	-0.052	-0.052	-0.052
	(0.050)	(0.050)	(0.050)	(0.042)
Constant	0.283	0.283	0.283	0.409
	(0.573)	(0.573)	(0.573)	(0.507)
R^2	0.45	0.45	0.45	0.53
Obs.	351	351	351	351
	Panel D: Likes Spanish			
Treat	-0.042	-0.042	-0.042	-0.032
	(0.055)	(0.055)	(0.055)	(0.056)
Constant	0.280	0.280	0.280	0.190
	(0.404)	(0.404)	(0.404)	(0.276)
R^2	0.39	0.39	0.39	0.53
Obs.	353	353	353	353

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SEs clustered at the individual level (Panels A-B) or robust SE (Panels C-D) in parenthesis. Controls for specifications in Columns 1-4 are as described in Table 7.

Table 9: Robustness - aspirations and motivation

	(1) block FE	(2) +demog	(3) +SES	(4) IPW
Panel A: Bachillerato				
Treat	0.136**	0.138**	0.136**	0.134**
	(0.058)	(0.057)	(0.059)	(0.065)
Constant	0.242	0.391	0.401	0.439
	(0.354)	(0.375)	(0.530)	(0.376)
R^2	0.27	0.38	0.46	0.46
Obs.	312	312	312	312
Panel B: College				
Treat	0.000	0.009	0.034	0.033
	(0.049)	(0.050)	(0.050)	(0.051)
Constant	1.000***	1.028***	1.375***	1.388***
	(0.032)	(0.165)	(0.325)	(0.383)
R^2	0.26	0.35	0.43	0.43
Obs.	301	301	301	301
Panel C: Grit				
Treat	0.066	0.066	0.091	0.083
	(0.059)	(0.064)	(0.063)	(0.066)
Constant	3.420***	3.211***	2.485***	2.467***
	(0.135)	(0.171)	(0.511)	(0.561)
R^2	0.24	0.31	0.40	0.41
Obs.	311	311	311	311
Panel D: High effort				
Treat	0.099*	0.098*	0.116**	0.093*
	(0.054)	(0.051)	(0.054)	(0.047)
Constant	0.701***	0.892***	0.072	0.091
	(0.230)	(0.160)	(0.404)	(0.373)
R^2	0.20	0.39	0.47	0.69
Obs.	351	351	351	351
Panel E: Motivation school				
Treat	0.000	0.001	0.011	0.010
	(0.017)	(0.019)	(0.019)	(0.022)
Constant	0.861***	0.893***	0.759***	0.767***
	(0.069)	(0.102)	(0.192)	(0.164)
R^2	0.25	0.34	0.43	0.43
Obs.	303	303	303	303

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. Robust SE in parenthesis. Controls for specifications in Columns 1-4 are as described in Table 7.

Table 10: Robustness - socio-emotional outcomes

	(1) block FE	(2) +demog	(3) +SES	(4) IPW
	Panel A: Wellbeing index			
Post x Treat	0.020	0.006	0.032	0.036
	(0.107)	(0.108)	(0.109)	(0.113)
Constant	7.527***	7.433***	5.871***	5.750***
	(0.359)	(0.567)	(0.740)	(0.600)
R^2	0.59	0.62	0.66	0.66
Obs.	644	644	644	644
	Panel B: School satisfaction			
Post x Treat	0.264*	0.290*	0.265	0.288
	(0.160)	(0.162)	(0.167)	(0.182)
Constant	5.699***	5.686***	4.624***	4.875***
	(0.584)	(0.770)	(1.078)	(1.138)
R^2	0.20	0.24	0.30	0.29
Obs.	632	632	632	632
	Panel C: Locus of control			
Post x Treat	-0.051	-0.057	-0.058	-0.061
	(0.035)	(0.036)	(0.037)	(0.039)
Constant	0.587***	0.611***	0.324*	0.306
	(0.072)	(0.078)	(0.188)	(0.204)
R^2	0.18	0.24	0.29	0.30
Obs.	641	641	641	641
	Panel D: Fault of others			
Post x Treat	0.153***	0.154***	0.154***	0.159***
	(0.056)	(0.057)	(0.059)	(0.048)
Constant	0.023	0.092	-0.050	-0.028
	(0.032)	(0.119)	(0.276)	(0.291)
R^2	0.18	0.21	0.25	0.25
Obs.	611	611	611	611

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SEs clustered at the individual level in parenthesis. Controls for specifications in Columns 1-4 are as described in Table 7.

Table 11: Attrition between baseline and endline - child questionnaire and test

	(1) Full	(2) Full	(3) Treat	(4) Control
Treat	0.045	0.036		
	(0.035)	(0.035)		
Child characteristics				
8th grade		0.079*	0.008	0.144**
		(0.046)	(0.066)	(0.073)
Age		-0.102***	-0.081**	-0.102**
		(0.032)	(0.041)	(0.051)
Catalonia		-0.042	-0.072	-0.018
		(0.049)	(0.066)	(0.078)
Girl		-0.044	-0.085	-0.008
		(0.037)	(0.052)	(0.059)
Receives school meal subsidy		0.035	0.100	-0.037
		(0.068)	(0.074)	(0.121)
Had no online classes during lockdown (April-June 2020)		0.021	0.074	-0.125
		(0.060)	(0.066)	(0.118)
Has a laptop		0.125	0.144	0.171
		(0.091)	(0.114)	(0.146)
Has a tablet or touchscreen laptop		0.111	0.130	0.158
		(0.088)	(0.111)	(0.148)
Received some form of academic support at baseline		-0.003	-0.061	0.045
		(0.047)	(0.064)	(0.080)
Speaks Spanish at home		-0.065	-0.058	-0.080
		(0.056)	(0.071)	(0.087)
Immigrant background		-0.009	-0.028	-0.030
		(0.039)	(0.053)	(0.067)
Parental education				
Compulsory schooling		-0.037	-0.070	-0.036
		(0.053)	(0.077)	(0.083)
Vocational degree		-0.047	0.027	-0.129
		(0.058)	(0.078)	(0.095)
Higher education		-0.031	-0.061	-0.023
		(0.053)	(0.077)	(0.083)
Household income				
€500 - €999		-0.016	-0.033	-0.022
		(0.061)	(0.089)	(0.081)
€1000 - €1499		0.042	0.115	-0.062
		(0.060)	(0.085)	(0.091)
€1500 - €1999		0.075	0.077	0.034
		(0.061)	(0.091)	(0.082)
$\geq$ €2000		0.023	0.021	-0.039
		(0.070)	(0.092)	(0.121)
Parental characteristics				
Single		-0.014	0.001	-0.047
		(0.043)	(0.053)	(0.071)
Mean dep. var.	0.88	0.88	0.90	0.85
R^2	0.00	0.09	0.16	0.11
Obs.	356	356	186	170

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. Robust standard errors in parentheses. The table reports coefficients from a linear probability model estimated by OLS. The dependent variable is a dummy equal to one if the child responded to the endline survey and math test, and zero otherwise. The full sample (Columns 1-2) includes all students who registered for the program and whose parents completed the registration form and consents and were included in the randomization sample. Column 3 restricts the sample to those assigned to the treatment, and Column 4 to those that were assigned to the control group.

Table 12: Attrition between baseline and endline - parent questionnaire

	(1) Full	(2) Full	(3) Treat	(4) Control
Treat	0.136***	0.140***		
	(0.051)	(0.052)		
Child characteristics				
8th grade		0.043	0.042	0.019
		(0.066)	(0.098)	(0.096)
Age		-0.023	0.018	-0.035
		(0.041)	(0.057)	(0.059)
Catalonia		-0.068	-0.097	-0.064
		(0.071)	(0.101)	(0.098)
Girl		-0.017	-0.009	-0.026
		(0.053)	(0.070)	(0.082)
Receives school meal subsidy		-0.060	0.101	-0.191
		(0.099)	(0.129)	(0.154)
Had no online classes during lockdown (April-June 2020)		-0.080	-0.023	-0.082
		(0.084)	(0.116)	(0.145)
Has a laptop		-0.144	-0.034	-0.396***
		(0.098)	(0.155)	(0.138)
Has a tablet or touchscreen laptop		-0.130	0.037	-0.419***
		(0.098)	(0.148)	(0.147)
Received some form of academic support at baseline		-0.003	-0.055	0.038
		(0.066)	(0.092)	(0.103)
Speaks Spanish at home		-0.034	-0.125	0.079
		(0.086)	(0.133)	(0.116)
Immigrant background		-0.021	-0.162*	0.099
		(0.060)	(0.082)	(0.091)
Parental education				
Compulsory schooling		0.071	-0.073	0.144
		(0.076)	(0.098)	(0.117)
Vocational degree		0.057	-0.010	0.135
		(0.084)	(0.124)	(0.123)
Higher education		0.119	-0.009	0.198
		(0.087)	(0.106)	(0.131)
Household income				
€500 - €999		-0.086	-0.197*	0.063
		(0.079)	(0.118)	(0.123)
€1000 - €1499		-0.128	-0.034	-0.147
		(0.086)	(0.119)	(0.146)
€1500 - €1999		-0.246**	-0.511***	0.119
		(0.107)	(0.152)	(0.160)
$\geq$ €2000		-0.170	-0.242	-0.060
		(0.119)	(0.171)	(0.205)
Parental characteristics				
Single		-0.198***	-0.240***	-0.195**
		(0.061)	(0.083)	(0.092)
Mean dep. var.	0.62	0.62	0.68	0.55
R^2	0.02	0.08	0.16	0.17
Obs.	356	356	186	170

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. Robust standard errors in parentheses. The table reports coefficients from a linear probability model estimated by OLS. The dependent variable is a dummy equal to one if the parent responded to the endline survey, and zero otherwise. The full sample (Columns 1-2) includes all children who registered for the program and whose parents completed the registration form and consents and were included in the randomization sample. Column 3 restricts the sample to those assigned to the treatment, and Column 4 to those that were assigned to the control group.

Table 13: Heterogeneous effects by tutoring/session characteristics

	(1) Standardized test score	(2) Final math grade	(3) Passed math	(4) Repeated year	(5) Bachi llerato	(6) High effort	(7) School satisfaction	(8) Fault of others
Panel A: Connection quality								
Treat	0.365	-0.707	-0.037	-0.089	0.112	0.148	-0.106	0.056
	(0.272)	(0.478)	(0.120)	(0.078)	(0.121)	(0.106)	(0.305)	(0.103)
Treat x Good connection	-0.120	1.730***	0.281**	-0.006	0.027	-0.038	0.441	0.112
	(0.276)	(0.498)	(0.121)	(0.079)	(0.128)	(0.109)	(0.300)	(0.102)
Constant	-0.405	10.697***	1.306***	0.369	0.403	0.069	4.655***	-0.037
	(0.598)	(2.329)	(0.482)	(0.353)	(0.529)	(0.404)	(1.065)	(0.277)
Panel B: Device type								
Treat	0.280*	0.898***	0.246***	-0.131**	0.159**	0.110*	0.262	0.169**
	(0.166)	(0.278)	(0.076)	(0.057)	(0.068)	(0.065)	(0.185)	(0.069)
Treat x Tablet	-0.047	-0.146	-0.086	0.102	-0.062	0.014	0.008	-0.037
	(0.206)	(0.389)	(0.099)	(0.070)	(0.101)	(0.081)	(0.208)	(0.077)
Constant	-0.360	9.303***	1.055**	0.408	0.390	0.074	4.628***	-0.069
	(0.595)	(2.448)	(0.480)	(0.341)	(0.534)	(0.404)	(1.080)	(0.278)
Panel C: Tutor gender								
Treat	0.515**	0.653	0.289**	-0.115	-0.058	0.109	0.238	0.231**
	(0.237)	(0.498)	(0.135)	(0.096)	(0.144)	(0.101)	(0.207)	(0.108)
Treat x Female tutor	-0.324	0.231	-0.088	0.025	0.245	0.009	0.037	-0.099
	(0.235)	(0.504)	(0.142)	(0.101)	(0.163)	(0.109)	(0.199)	(0.108)
Constant	-0.410	9.401***	1.074**	0.379	0.406	0.071	4.632***	-0.052
	(0.608)	(2.453)	(0.488)	(0.354)	(0.534)	(0.405)	(1.080)	(0.277)
Panel D: Gender composition								
Treat	0.293*	0.716**	0.108	-0.048	0.047	-0.026	0.404**	0.146**
	(0.173)	(0.301)	(0.082)	(0.051)	(0.079)	(0.065)	(0.198)	(0.069)
Treat x Same gender	-0.070	0.272	0.225**	-0.096	0.185*	0.315***	-0.319	0.018
	(0.187)	(0.438)	(0.108)	(0.076)	(0.107)	(0.082)	(0.229)	(0.078)
Constant	-0.334	9.214***	0.967*	0.426	0.269	-0.189	4.812***	-0.052
	(0.625)	(2.452)	(0.489)	(0.352)	(0.555)	(0.407)	(1.060)	(0.273)
Panel E: Group ability composition								
Treat	0.263*	0.888***	0.314***	-0.147**	0.097	0.126*	0.278*	0.153***
	(0.149)	(0.320)	(0.093)	(0.061)	(0.086)	(0.071)	(0.168)	(0.059)
Treat x Similar ability	0.052	-0.090	-0.212	0.116	0.083	-0.022	-0.457	0.030
	(0.181)	(0.482)	(0.140)	(0.095)	(0.129)	(0.109)	(0.300)	(0.086)
Constant	-0.418	9.453***	1.307***	0.261	0.354	0.083	4.826***	-0.063
	(0.609)	(2.497)	(0.481)	(0.352)	(0.551)	(0.414)	(1.060)	(0.276)
Mean dep. var.	-0.01	5.10	0.68	0.12	0.41	0.54	5.47	0.29
SD dep. var.	1.00	1.75	0.47	0.32	0.49	0.50	1.36	0.45
Obs.	644	220	220	218	312	351	632	611

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SEs clustered at student level (Columns 1, 7, and 8) and robust SEs (Columns 2-6) in parenthesis. The table shows the δ coefficients from Equation (1) in Columns 1, 7, and 8, and from Equation (2) for Columns 2-6. Panel A includes an interaction term of the treatment dummy with a dummy for having experienced a good connection quality. The connection quality is defined as good if the average ping (or latency) across all online sessions is lower than 100. Panel B includes includes an interaction term of the treatment dummy with a dummy for having used a tablet or phone to connect to the tutoring sessions. Panel C includes an interaction term of the treatment dummy with a dummy that is equal to one if tutor was female. Panel D includes an interaction term of the treatment dummy with a dummy that is equal to one if both students in the group were of the same gender. Controls included as specified in Table 3.

Table 14: Heterogeneous effects by child characteristics

	(1) Standardized test score	(2) Final math grade	(3) Passed math	(4) Repeated year	(5) Bachi llerato	(6) High effort	(7) School satisfaction	(8) Fault of others
Panel A: Gender								
Treat	0.225	1.050***	0.306***	-0.098	0.198**	0.169**	0.368*	0.081
	(0.179)	(0.331)	(0.087)	(0.072)	(0.082)	(0.073)	(0.188)	(0.073)
Treat x Girl	0.101	-0.497	-0.221	0.009	-0.138	-0.137	-0.207	0.150*
	(0.198)	(0.587)	(0.145)	(0.099)	(0.128)	(0.113)	(0.221)	(0.079)
Constant	-0.322	9.213***	1.023**	0.376	0.371	0.083	4.660***	-0.063
	(0.594)	(2.446)	(0.476)	(0.357)	(0.528)	(0.408)	(1.083)	(0.279)
Panel B: Grade								
Treat	0.113	1.159***	0.263***	-0.088	0.111	0.094	0.049	0.189***
	(0.164)	(0.393)	(0.091)	(0.080)	(0.088)	(0.079)	(0.192)	(0.063)
Treat x 2º ESO	0.300	-0.593	-0.090	-0.011	0.048	0.042	0.423**	-0.067
	(0.195)	(0.484)	(0.129)	(0.097)	(0.121)	(0.105)	(0.204)	(0.078)
Constant	-0.335	8.966***	1.029**	0.366	0.425	0.089	4.799***	-0.117
	(0.608)	(2.516)	(0.490)	(0.369)	(0.536)	(0.411)	(1.069)	(0.280)
Panel C: Single parent household								
Treat	0.340**	0.813***	0.161**	-0.073	0.116	0.138**	0.332*	0.106
	(0.167)	(0.264)	(0.075)	(0.056)	(0.071)	(0.064)	(0.188)	(0.065)
Treat x Single parent household	-0.218	0.155	0.253	-0.098	0.084	-0.083	-0.225	0.131*
	(0.181)	(0.780)	(0.165)	(0.103)	(0.154)	(0.120)	(0.229)	(0.079)
Constant	-0.324	9.445***	1.226**	0.321	0.425	0.025	4.500***	-0.095
	(0.601)	(2.523)	(0.476)	(0.358)	(0.530)	(0.413)	(1.089)	(0.276)
Panel D: Immigrant background								
Treat	0.317*	0.678**	0.084	-0.087	0.171**	0.207***	0.203	0.106*
	(0.169)	(0.330)	(0.087)	(0.067)	(0.079)	(0.071)	(0.194)	(0.062)
Treat x Immigrant background	-0.123	0.381	0.300**	-0.016	-0.086	-0.213*	0.138	0.109
	(0.204)	(0.594)	(0.136)	(0.088)	(0.135)	(0.109)	(0.208)	(0.083)
Constant	-0.389	9.287***	1.031**	0.377	0.425	0.107	4.651***	-0.070
	(0.597)	(2.457)	(0.476)	(0.352)	(0.537)	(0.411)	(1.091)	(0.277)
Mean dep. var.	-0.01	5.10	0.68	0.12	0.41	0.54	5.47	0.29
SD dep. var.	1.00	1.75	0.47	0.32	0.49	0.50	1.36	0.45
Obs.	644	220	220	218	312	351	632	611

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SEs clustered at student level (Columns 1, 7, and 8) and robust SEs (Columns 2-6) in parenthesis. The table shows the δ coefficients from Equation (1) in Columns 1, 7, and 8, and from Equation (2) for Columns 2-6. Panel A includes an interaction term of the treatment dummy with a dummy for tutored child's gender. Panel B includes an interaction term of the treatment dummy with a dummy for whether the kid lives in a single parent household. Panel C includes an interaction term of the treatment dummy with a dummy that is equal to one if the tutored child was in 2nd grade of secondary school (2º ESO, grade 8). Panel D includes an interaction term of the treatment dummy with a dummy that is equal to one if one of their parents was born outside Spain. Controls included as specified in Table 3.

A Sample questions

A.1 Math test

Example for grade 7 Solve the following equation for x and simplify the solution if possible. You must write down the entire procedure.

- $3x + 5(x - 3) = 4x - 2(x - 5)$
  - ☐ $x = \frac{1}{2}$
  - ☐ $x = \frac{5}{6}$
  - ☐ $x = \frac{6}{25}$
  - ☐ $x = \frac{25}{6}$

Example for grade 8 Solve the following equation for x:.

- $x^2 + 2x - 15 = 0)$
  - ☐ $x = -3, x = 5$
  - ☐ $x = 3, x = -5$
  - ☐ x does not belong to the set of real numbers
  - ☐ $x = 31, x = -33$

A.2 Questions on socio-emotional skills, well-being and aspirations

Grit 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. Indicate one of ”Very much like me”, ”Mostly like me”, ”Somewhat like me”, ”Not much like me”, and ”Not like me at all”.

- New ideas and projects sometimes distract me from previous ones.
- Setbacks don’t discourage me. I don’t give up easily.
- I have been obsessed with a certain idea or project for a short time but later lost interest.
- I am a hard worker.
- I often set a goal but later choose to pursue a different one.
- I have difficulty maintaining my focus on projects that take more than a few months to complete.
- I finish whatever I begin.
- I am diligent. I never give up.

Locus of control For each of the following questions, mark ”Yes” or ”No”:

- Do you usually feel that it’s almost useless to try in school because most children are cleverer than you?
- When bad things happen to you, is it usually someone else’s fault?

Well-being On a scale from 1 to 7, where 1 means ”not happy at all” and 7 means ”completely happy”, how do you feel about the following parts of your life?

- Your school work
- The way you look
- The school you go to
- Your friends
- Your life as a whole
- Think about the period of lockdown during Covid-19. How did you feel during that period?

Aspirations What are your plans after you complete compulsory schooling?

- Select one option:
  - ☐ Vocational education
  - ☐ Continue studying (*Bachillerato*)
  - ☐ Find a job
  - ☐ I don’t know
- Mark ”Yes” or ”No”:
  - Would you like to go to college in the future?
  - If so, do you think it would be possible?

Motivation for school How often do you... (indicate one of ”always”, ”most of the time”, ”sometimes”, ”never)

- ...put effort into school?
- ...find school interesting?
- ...feel that school is a waste of time?

Frequency of homework Thinking about last May, how much time did you devote to schoolwork per day on average? Select one option:

☐ Less than 15 minutes

☐ 15-30 minutes

☐ 30-60 minutes

☐ 1-1.5 hours

☐ 1.5-2 hours

☐ 2-2.5 hours

☐ More than 2.5 hours

Interest in math and reading How much do you like the following subjects? Select one option:

- Spanish/catalan language:
  - ☐ A lot
  - ☐ Quite a bit
  - ☐ I somewhat like it
  - ☐ A bit
  - ☐ I don't like it at all
- Math:
  - ☐ A lot
  - ☐ Quite a bit
  - ☐ I somewhat like it
  - ☐ A bit
  - ☐ I don't like it at all

B Additional Tables

Table B1: Socio-economic and scores distribution of schools in Madrid (all schools vs. *Menπores* schools)

	(1) ESCS	(2) Spanish - Score	(3) Maths - Score	(4) English - Score	(5) Social - Score	(6) % of native-born parents (both)
All schools (region of Madrid)						
Mean	0.00	0.00	0.00	0.00	0.00	71.2
SD	1.00	1.00	1.00	1.00	1.00	24.2
p10	-1.15	-1.37	-1.23	-1.45	-1.31	35.7
p25	-0.59	-0.59	-0.67	-0.68	-0.68	62.5
p50	-0.01	0.08	-0.03	0.04	0.00	78.8
p75	0.61	0.65	0.57	0.69	0.61	87.9
p90	1.26	1.19	1.27	1.38	1.31	92.9
Menπores schools						
Mean	-0.51	-0.73	-0.16	-0.90	-0.74	68.8
SD	0.54	0.60	0.71	0.59	0.41	14.1
p10	-1.07	-1.62	-1.01	-1.51	-1.23	50.0
p25	-0.90	-1.29	-0.50	-1.45	-1.10	59.8
p50	-0.58	-0.56	-0.28	-1.01	-0.75	69.1
p75	-0.10	-0.33	0.12	-0.30	-0.39	80.5
p90	0.09	-0.24	0.69	-0.23	-0.35	82.1

Notes: This table shows the socio-economic and scores distribution for all schools in the region of Madrid and the socio-economic and scores distribution for schools participating in the tutoring program. ESCS is an Index of Social, Cultural and Economic Status derived from a student background questionnaire with information from the household. The index as well as the score values are standardized to mean 0 and standard deviation one at the regional level. Source: LOMCE evaluations from academic year 2017/18: students participating. All statistics have been derived assuming equal weights for each school.

Table B2: Impact on academic outcomes - All tutors

	(1) Standardized test score	(2) Final math grade	(3) Passed math	(4) Repeated year
Post x Treat	0.254*			
	(0.146)			
Treat		0.838***	0.201***	-0.091**
		(0.233)	(0.058)	(0.044)
Constant	-0.710	8.058***	0.975***	0.250
	(0.538)	(2.062)	(0.373)	(0.260)
Mean dep. var.	-0.01	5.10	0.68	0.12
SD dep. var.	1.00	1.75	0.47	0.32
R^2	0.29	0.64	0.63	0.58
Obs.	677	233	233	231

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SSEs clustered at student level (Column 1) and robust SEs (Columns 2-4) in parenthesis. The table shows the δ coefficients from Equation (1) in Column 1 and from Equation (2) for Columns 2-4. All regressions include block FEs and control for student age, grade, gender, region, a dummy indicating school meal eligibility, last math grade, a set of dummy variables indicating the frequency of online lessons during school closures in April and May 2020, a dummy indicating whether the student had a tablet or computer at home before the program, a dummy indicating whether the student was receiving other tutoring before the program, categorical variables indicating the language spoken at home, parental education, household income, and household composition, an indicator for whether the responding parent is a single parent, and a dummy variable indicating whether the parent is of Spanish origin. Sample includes both students taught by professional as well as volunteer tutors.

Table B3: Impact on self-perceived ability and affinity - All tutors

	(1) Good at math	(2) Good at Spanish	(3) Likes math	(4) Likes Spanish
Post x Treat	-0.027	0.011		
	(0.049)	(0.066)		
Treat			-0.040	-0.067
			(0.049)	(0.052)
Constant	-0.367	0.155	0.197	0.180
	(0.316)	(0.370)	(0.553)	(0.410)
Mean dep. var.	0.25	0.55	0.36	0.41
SD dep. var.	0.44	0.50	0.48	0.49
R^2	0.36	0.27	0.43	0.39
Obs.	658	659	368	370

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SEs clustered at student level (Columns 1-2) and robust SEs (Columns 3-4) in parenthesis. The table shows the δ coefficients from Equation (1) in Columns 1-2 and from Equation (2) for Columns 3-4. All regressions include block FEs and control for student age, grade, gender, region, a dummy indicating school meal eligibility, last math grade, a set of dummy variables indicating the frequency of online lessons during school closures in April and May 2020, a dummy indicating whether the student had a tablet or computer at home before the program, a dummy indicating whether the student was receiving other tutoring before the program, categorical variables indicating the language spoken at home, parental education, household income, and household composition, an indicator for whether the responding parent is a single parent, and a dummy variable indicating whether the parent is of Spanish origin. Sample includes both students taught by professional as well as volunteer tutors.

Table B4: Impact on aspirations and motivation - All tutors

	(1) Bachi llerato	(2) College	(3) Grit	(4) High effort	(5) Motivation school
Treat	0.123**	0.021	0.075	0.104**	0.005
	(0.056)	(0.050)	(0.061)	(0.052)	(0.018)
Constant	0.563	1.275***	2.586***	0.136	0.739***
	(0.487)	(0.329)	(0.458)	(0.389)	(0.183)
Mean dep. var.	0.41	0.81	3.04	0.54	0.75
SD dep. var.	0.49	0.39	0.50	0.50	0.16
R^2	0.45	0.42	0.36	0.46	0.41
Obs.	328	315	327	368	317

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. Robust SEs in parenthesis. The table shows the δ coefficients from Equation (2). All regressions include block FEs and control for student age, grade, gender, region, a dummy indicating school meal eligibility, last math grade, a set of dummy variables indicating the frequency of online lessons during school closures in April and May 2020, a dummy indicating whether the student had a tablet or computer at home before the program, a dummy indicating whether the student was receiving other tutoring before the program, categorical variables indicating the language spoken at home, parental education, household income, and household composition, an indicator for whether the responding parent is a single parent, and a dummy variable indicating whether the parent is of Spanish origin. Sample includes both students taught by professional as well as volunteer tutors.

Table B5: Impact on socio-emotional outcomes - All tutors

	(1) Wellbeing index	(2) School satisfaction	(3) Locus of control	(4) Fault of others
Post x Treat	0.034	0.286*	-0.062*	0.168***
	(0.106)	(0.164)	(0.036)	(0.058)
Constant	5.913***	5.080***	0.317*	-0.110
	(0.695)	(1.064)	(0.181)	(0.255)
Mean dep. var.	6.21	5.47	0.61	0.29
SD dep. var.	1.43	1.36	0.31	0.45
R^2	0.65	0.29	0.29	0.24
Obs.	677	665	672	641

Notes: Significance levels are indicated by * < .1, ** < .05, *** < .01. SEs clustered at student level in parenthesis. The table shows the δ coefficients from Equation (1). All regressions include block FEs and control for student age, grade, gender, region, a dummy indicating school meal eligibility, last math grade, a set of dummy variables indicating the frequency of online lessons during school closures in April and May 2020, a dummy indicating whether the student had a tablet or computer at home before the program, a dummy indicating whether the student was receiving other tutoring before the program, categorical variables indicating the language spoken at home, parental education, household income, and household composition, an indicator for whether the responding parent is a single parent, and a dummy variable indicating whether the parent is of Spanish origin. Sample includes both students taught by professional as well as volunteer tutors.