INSTITUTO DE ECONOMÍA

DOCUMENTO de TRABAJO

565

2021

Does Short-Term School Tutoring have Medium-Term Eects? Experimental Evidence from Chile

Verónica Cabezas, José Ignacio Cuesta y Francisco Gallego

www.economia.uc.cl · ISSN (edición impresa) 0716-7334 · ISSN (edición electrónica) 0717-7593

Does Short-Term School Tutoring have Medium-Term Effects? Experimental Evidence from Chile*

Verónica Cabezas
Pontificia Universidad Católica de Chile

José Ignacio Cuesta
Stanford University

Francisco A. Gallego
Pontificia Universidad Católica de Chile and J-PAL

September 2021

Abstract. This paper explores how short-term tutoring affects educational outcomes in the short- and medium-term. We implemented a randomized experiment of a three-month small group tutoring program in Chile that aimed at improving reading outcomes among fourth graders using college student volunteers. We find small short-term effects on reading outcomes. Using administrative data covering up to eight years after the program ended, we find significant decreases in the probability of dropping out, increases in the probability of timely school progression, and increases in attendance, school grades and test scores. These effects are stronger for students who were ex-ante more likely to drop-out from school. The program effects are stronger for students who established stronger personal connections with the tutors. Our results suggest that tutoring programs may have relevant medium-term effects that go beyond short-term impacts on specific subjects with stronger effects on more at-risk children.

Keywords. remedial education, tutoring, short-term programs, medium-term effects
JEL codes. I21, I28, O15

*Email addresses: vcabezag@uc.cl, jicuesta@stanford.edu, and fgallego@uc.cl.

This paper was previous circulated under the title "Effects of short-term tutoring on cognitive and non-cognitive skills: Evidence from a randomized evaluation in Chile". This study is registered in The American Economic Association's registry for randomized controlled trials under RCT ID AEARCTR-0001480. Randomized evaluations like this require the contributions of a large number of people. While it would be impossible to recognize everyone who made a contribution to this project, we would like to thank Martín Canessa, Ryan Cooper, Daniela Guzmán, and several members of the Ministry of Education of Chile and Fundación para la Superación de la Pobreza of Chile. We are also grateful to Gabriel Cruz for research assistance, and Jeanne Lafortune, Mauricio Romero, and seminar participants at Columbia University, PUC-Chile, J-PAL LAC, the LACEA Impact Evaluation Network meetings, and the LACEA-LAMES meetings for comments; and the Ministry of Education of Chile and Fondecyt (Grant # 1100623) for funding support. The usual disclaimer applies.

1 Introduction

Several studies have suggested that non-traditional educational interventions have a positive impact on school effectiveness (e.g., Fryer Jr, 2011; Dobbie and Fryer Jr, 2013). Particularly, tutoring programs have been identified as one of the most versatile and potentially transformative educational tools (e.g., Nickow *et al.*, 2020). In a parallel literature, several studies find medium-term effects of educational interventions (e.g., Resnjanskij *et al.*, 2021; Lavecchia *et al.*, 2020; Lavy, 2020; Oreopoulos *et al.*, 2017; Angrist *et al.*, 2006), motivated by dynamic effects in educational production functions (e.g., Carneiro and Heckman, 2003). This suggests that tutoring programs may have effects not only in the short- but also in the medium-term, a possibility also suggested recently by Guryan *et al.* (2021) but for which there is limited evidence.

Tutoring programs and remedial education more broadly, should have heterogeneous effects on children with different ex-ante learning and educational and social opportunities. This channel might be particularly relevant for children with less attachment to the educational process, as they may have less access to support in and outside school. Tutors may provide not only subject-specific knowledge but also mentoring, support and motivation services that may affect medium-term outcomes. The literature indeed underscores that the relationship between tutors and students is a relevant driver of the effectiveness of tutoring programs (Miller and Connolly, 2013; Fryer Jr and Howard-Noveck, 2020). Still, other papers find that the impacts of tutoring are more consistent with the impact of the academic dimensions provided by tutoring rather than with the mentoring component (Guryan *et al.*, 2021).

In this paper, we use a randomized controlled trial to study a short-term reading tutoring program to shed light on these questions. To explore the medium-term effects of tutoring, we collected administrative data spanning an 8-year period after the program was implemented, covering a comprehensive set of educational outcomes related to both the extensive (enrollment and drop-out decisions), intensive (attendance), and learning (grades and test scores) margins. To study whether the treatment has heterogeneous effects by the attachment of students to the educational process, we report several heterogeneity analyses that consider the ex-ante probability of dropping out from school as a proxy for attachment to the educational process. Finally, to uncover the channels for how tutor support may impact student achievement, we study tutor academic support and tutor-student relationship as two potential mechanisms (Zijlstra *et al.*, 2020). The second mechanism is more related with potential transformative effects of the program on long-run outcomes but not necessarily related to subject learning.

The program we study is called *Servicio País en Educación (SPE),* and was implemented by the Minister of Education and a Chilean NGO called Fundación para la Superación de la Pobreza, from September to December 2010. The target population were fourth graders from relatively low socioeconomic background attending government funded, low-performing schools. One important aspect of the program is the relationship created between the student and the tutor.[1] The program

[1]A related version of the intervention has been implemented in Chile since 1999 by the Fundación de Superación

consisted of 15 weekly-sessions of 90-minutes in which tutors would read suitable materials with 4th graders, following a structured methodology of shared reading. The tutors were volunteer college students, mostly without teaching experience and preparation, who received small stipends to cover transportation costs. The tutoring sessions took place during school hours. The cost of the program was \$99.0 per student (\$123.6 if using PPP rates).

The research design consists of a randomized controlled trial. We randomly allocated the program among 85 schools, with 6,129 students using a stratified randomization, by which 45 schools were assigned to the treatment group—of which 87% accepted to receive the treatment—and 40 schools were assigned to the control group. We collected data on the actual implementation of the program. In practice, the average school in our sample received 12 sessions of the program—with a range varying from 9 to 15 sessions—. In turn, the average student received 7.9 sessions. No children in the control group group received any sessions. The implied take-up of the program at the student level was thus 52.6%.

We report five sets of results. First, the program had short-term effects in a language test we applied by the end of the program implementation. Treated students increased test scores in language, especially in the reading comprehension part of the test. In contrast, we do not identify significant increases in a non-cognitive test that measures attitudes towards reading. Thus, the program seems to have slightly increased outcomes related directly to the program. The increase in the language component of the test is equivalent to 0.06 standard deviations (σ). And the magnitude of the increase in the reading comprehension component was 0.11σ. The effects are not different for students with different ex-ante probabilities of dropping out from school.

Second, the program significantly decrease the likelihood of dropping out from school and increased the probability of graduating from primary and secondary school on time. Our intention-to-treat (ITT) estimates imply a decrease in the probability of dropping out of -1.76 percentage points (p.p.) by 2017, the last year for which we have data and equivalent to $11th$ grade. This is entirely driven by an ITT effect of -3.73 p.p. for students with a high ex-ante probability of dropping out from school. We see smaller effects for the probability of dropping out by the end of 2011 and 2014, one and four years after the program ended. We also find significant ITT effects on the probability of graduating on time from primary school and advancing grades on-time by 2017—this variable combines drop-out rates with repetition rates—. Our ITT estimates imply an increase in the probability of graduating on time from primary school and $11th$ grade of 4.53 p.p. and 3.19 p.p., respectively. Again, the effects are entirely driven by students with higher ex-ante probability of dropping out, for whom ITT effects are of 6.54 p.p. for primary education and 5.44 p.p. for secondary education. These results suggest that the intervention had effects beyond the short-term, and affected the medium-term educational trajectories of students. The effects are entirely driven for relatively at-risk students.

Third, the program had effects on the intensive margin of educational outcomes in the medium-

de la Pobreza, which is motivated by the Perach program that has been implemented in Israel since 1974, and had about 30,000 tutors about 60,000 students in 2008. See Carmeli (2000) for a more detailed description of the program.

term, including attendance and grades (conditional on not dropping out from school). Our ITT estimates imply increases in attendance and grades in primary school, which are again entirely driven by students with higher ex-ante risk of dropping out. The program increased attendance by 0.81 p.p. for the full sample and 1.55 p.p. for students with higher ex-ante risk of dropping out. A similar pattern emerges for grades in primary education, which increased by 0.11σ for the full sample and by 0.15σ for students at risk of dropping out from school. Estimating these effects for secondary education is more complicated because the treatment affects the extensive margin, as previously discussed. Thus, we use three methods to estimate treatment effects with selective attrition: we re-weight the sample using inverse probability weighting (IPW), and we estimate bounds of the ITTs jointly with Tobit estimates following Angrist *et al.* (2006). Most results are robust across the three methods. We estimate significant ITT effects on secondary education attendance and grades. The program increases attendance during secondary school by 1.6 p.p. using IPW, with a statistically significant upper bound of 3.7 p.p.. Again, the effects are much stronger for children with high risk of dropping out. In their case, the IPW estimate is 3.4 p.p. with an upper bound of 5.4 p.p.. In contrast, the effects on average grades are not statistically significant for most of the estimates for the full sample. However, we provide some some evidence of stronger effects for children with high risk of dropping out, although with wide confidence intervals when considering Tobit estimates.

Fourth, the program also has effects on test scores in language and math taken by students in $8th$ and $10th$ grades, which is four and six years after the program ended. The effects are significant only for children at risk of dropping out. Using our methods to deal with selective attrition, we find effects of 0.12σ and 0.20σ for Language and Math in $8th$ grade. The results for test scores in $10th$ grade also suggest stronger effects for student at risk of dropping out but they are less precise. In all, these results suggest that the program had significant effects on measures of the intensive and learning margins of schooling that go beyond short-term effects, especially for students at risk of dropping out. Interestingly, our results show significant impacts also in Math, suggesting that the program provides foundations for human capital also in other subjects.

Fifth, we explore heterogeneous effects of tutoring on educational outcomes by the type of relationship established between the tutor and the student. Using survey data, we construct indices to measure two dimensions (Zijlstra *et al.*, 2020): tutor academic support and tutor-student relationship. We conjecture that the second dimension is more important to explain the previous results because the initial intervention was focused on language and, while we find only moderate effects in the short-term, we do find significantly stronger effects in the medium-run that go beyond language. We find that the tutor-student dimension of the relationship plays a stronger role in explaining the medium-term effects. These effects are concentrated among students with high ex-ante risk of dropping out. In contrast, the academic support component of the tutoring relationship has a positive effect only for short-term results, and for both types of students. Thus, this evidence supports the conjecture that the human connection—following the wording of Nickow *et al.*, 2020—between tutors and students is an important mechanism that explains the medium-term effects we find.

The paper makes several contributions to previous research. By estimating medium-term effects on a broad set of educational outcomes up to 8 years after the program was implemented, we add to the literature on the effects of short-lived educational interventions (e.g., Rockoff and Turner, 2010; Abeberese *et al.*, 2011; Banerjee *et al.*, 2010), to the literature on the the medium-term effects of tutoring programs (e.g., Guryan *et al.*, 2021; Zijlstra *et al.*, 2020; Blachman *et al.*, 2014) and to the literature on medium-term effects of remedial education programs (e.g., Resnjanskij *et al.*, 2021; Lavy, 2020; Zijlstra *et al.*, 2020; Blachman *et al.*, 2014; Lavecchia *et al.*, 2020). Second, by providing support to a mechanism related to the human connection between the student and the tutor to explain our short- and medium-term results, we contribute empirical estimates, especially for medium-term impacts, to the previous literature(Fryer Jr and Howard-Noveck, 2020; Miller and Connolly, 2013; Guryan *et al.*, 2021). Guryan *et al.* (2021) argue that the impact of the mentoring part of the program they study seems to be less relevant than the academic portion by analyzing survey data. We also use perceptions of students but interact them with the treatment. One difference is that they focus on teenagers, while we focus on an intervention among fourth-graders. Future research could study this aspect in more detail. Finally, by estimating the short-term effects of our program we add an additional data point to the literature on the subject (Nickow *et al.*, 2020), with a short-term and structured reading program (e.g., Fryer Jr and Howard-Noveck, 2020; Lindo *et al.*, 2018; Jacob *et al.*, 2016; Miller and Connolly, 2013) developed during school hours (Jayachandran, 2014; Baker *et al.*, 2000; Wasik, 1997; Morris *et al.*, 1990) developed by nonprofessional (Nickow *et al.*, 2020) volunteers tutors (Ritter *et al.*, 2009). Actually, our short-term effects are in the range of effects reported for programs with the same characteristics among the programs reported in Nickow *et al.* (2020).

The remainder of the paper is organized as follows. Section 2 briefly describes the SPE program in detail. Section 3 describes the research design and methods used in the analysis. Section 4 provides the main results and further evidence on the mechanisms underlying the treatment effects. Finally, Section 5 concludes.

2 The Program

The SPE program started from a partnership between the Chilean Ministry of Education (MINEDUC) and the Fundación para la Superación de la Pobreza (FSP) and was first implemented during the period between September and December of 2010. The main goal of the program was to improve attitudes toward reading and reading comprehension (RC) among 4th graders.[2] The program design includes 15 weekly 90-minute sessions delivered during school hours, adding up to 22.5 hours of tutoring over three months. The class was split in small groups of between 5 and 6 students assigned to a tutor, which was a volunteer recruited by FSP. Tutors were college students coming from the same communities than the students, who were selected by the FSP from a pool of applicants on the basis of interviews and tests, and then received a one-day training before imple-

[2] Additionally, the program aimed at affecting the college students who were tutors. We only address the questions related to the program impacts on student educational outcomes and not on the tutors.

menting the program. The sessions include a set of activities regarding group reading following a shared-reading instructional methodology of traditional stories and informative texts, which are age and interest appropriate for students.[3] The actual implementation of the program was managed by a paid professional of the FSP which was inserted permanently in the intervened school. The idea was that this professional would verify the accurate implementation of the program and assist pedagogically the volunteers. Thus, following the classification of Nickow *et al.* (2020), SPE is a reading tutoring program implemented by non-professional volunteers in small groups of students using a structured curriculum.

The program targeted disadvantaged schools. This study focuses on schools from 10 counties,[4] in which the families were classified as low to middle income, and with below average results in the language section of a standardized test called SIMCE in the previous year.[5]

Table 1-A presents information about the implementation of the program. The average number of sessions per school was 12.04, with a minimum of nine sessions per school. In turn, an average student attended 7.9 sessions, equivalent to a 52.8% take-up rate. We asked students to evaluate several dimensions of the tutoring sessions in the follow-up. Students tend to evaluate relatively well their improvements in reading and writing, the quality of the tutors and the relationship with them, with average ratings between 2.3 and 2.5 in a 0–3 scale.

3 Experimental Design

3.1 Sample

On top of being a vulnerable school, a number of logistic restrictions were put by FSP to select schools for the program. In particular, we excluded all schools from counties in which the FSP either was not able to work because they had no human resources in them or had already committed with some schools in it, which made randomization impossible. This reduced the number of counties from which schools were included in the evaluation to 10.[6] In some of these counties, an additional restriction was set in terms of the administrative dependence of the schools, restricting us to include either only public schools (P) in some counties or only private subsidized schools (PS) in others. In addition, to fit the operational model designed by the FSP, each of the schools had to have at

[3]See Holdaway (1979) for a discussion on the motivation for using shared-reading.

[4]The 2010 version of the program also considered schools of two additional Chilean regions that are not included in this paper because the allocation of schools to the program was not random.

[5]The SIMCE test is applied nationwide since 1988 to more than 90% of students in a different grade each year (4th, 8th or 10th graders). The test includes language, mathematics, science, and social science sections.

[6]The counties in the study sample are Santiago, Estación Central, Lo Espejo, Maipú, La Florida and San Bernardo from the Great Santiago region (Metropolitan Region, RM), and Concepción, Coronel, Hualpén and Talcahuano from the Biobio Region (VIII).

least 90 students in fourth grade.[7] Table A1 summarizes both the eligibility restrictions and the eligible number of schools in each of the counties included in the sample.

Using this sample, schools were randomly assigned to treatment and control groups, stratifying by county, socioeconomic group, and previous test scores. As the set of eligible schools was larger than the number of required schools, only some of the schools assigned to each group were included in the evaluation, decision that was random too. The remaining schools were kept as replacement lists for the eventual rejection of schools to take part of the evaluation. With the results of this assignment, schools were contacted and invited to take part of the evaluation in their corresponding group. Five of them rejected the program but, except in two cases, all of them accepted to be evaluated anyway. Additionally, two schools in the control group rejected to be evaluated. Each of these schools was randomly replaced by another schools coming from the replacement lists.

The final composition of the study sample is displayed by Table A1. The treatment and control groups finally included 45 and 40 schools respectively, grouped in 25 and 24 units. In section 3.3, we provide information regarding balance between groups to validate the randomization procedure.

3.2 Data Collection

We combine several sources of data for this study. First, we collect data before and after the implementation of the program, in August, 2010 and December, 2010, respectively. We complement that data with administrative data collected by MINEDUC for the 2009-2017 period.

In terms of the baseline and follow-up, we included two instruments to measure skills: one measuring formal reading skills and the second measuring attitudes towards reading. The reading instrument was *Prueba de Comprensión Lectora y Producción de Textos* (Reading Comprehension and Texts Production Test, CLPT), which measures Reading Comprehension (RC), Texts Production (TP), and use of Language (UL).[8] To measure attitudes towards reading, we use a short questionnaire called *Gusto por la Lectura* (Taste for Reading, GPL) where we ask students several questions on four dimensions: Interest for Reading (IR), Self-perception as a Reader (SPR), Enjoyableness for Reading (ER) and Perception of Reading at School (PRS). These indices vary discretely between 0 and 3, where 0 is the most negative and 3 is the most positive of the alternatives.[9]

We also collected information on the program implementation, to understand the actual treatment and some of the mechanisms though which the program produces effects. First, we monitored the program implementation with random visits to observe the actual tutoring sessions. Second, we gathered administrative information about student, tutor, and professional assistance to tutoring

[7]As the number of schools that fit these size criteria was insufficient in some counties, we set an additional eligibility criteria that implied that if two schools were less than one kilometer away between them, and the sum of their fourth-graders was higher than 90, then that pair of schools could be included in the eligible set of schools too.

[8]Medina and Gajardo (2010) describes the test in detail.

[9]We constructed the instrument motivated by previous research by Solange (2004), McKenna *et al.* (1995) and McKenna and Kear (1990).

sessions and about the numbers of sessions received by each student in the program. Third, we asked the students about tutoring process in the follow-up survey.

Finally, we collected data from MINEDUC on school and student characteristics, and on several short- and medium-term outcomes. Regarding schools, we collected data on language and math SIMCE test scores, average mother years of schooling, average household income, socioeconomic level, a school vulnerability index (IVE), and administrative dependence (public or private). Regarding students, we collected information on gender, school enrollment, attendance, progression and grades for 2009-2017, and SIMCE test scores for students in the sample in their $8th$ and $10th$ grades.

3.3 Summary Statistics, Balance, and Attrition

Table 1-B presents descriptive statistics at the student level before the treatment. Students in the sample have average grades of 5.8 in a 1–7 scale, and 53% of them are male. Attendance in the year before treatment was 91%. On average, students have 52% of correct answers in the CLPT test at baseline, with relatively better outcomes in its language component (67%). In terms of the baseline GPL test, students have better outcomes in the enjoyableness for reading and the perception for reading at school dimensions than in self-perception as a reader and interest for reading.

Table 1-C describes the sample after treatment. The drop out rate in 2011 was 2%, and it increases to 5% in 2014 and to 16% in 2017. The same pattern appears in terms of student progression, which combines not dropping out with not repeating any grade. While in 2011, 91% of students were on time, this number decreased to 58% in 2017. School attendance was 89% in primary education and 82% in secondary education, and average grades were 5.32 and 5.03 in primary and secondary education, respectively. All these dimensions imply several margins of potential effects of the tutoring program in the medium-term. In terms of learning outcomes we also have data for our CLPT and GPL instruments right after the treatment was implemented, the patterns are very similar to those at baseline. In terms of medium-term outcomes, we observe SIMCE test scores in primary and secondary school. Students tend to have better outcomes in math than in language both in primary and secondary education.

Table 1-D presents variables at the school level before treatment. The average school has students with mothers with 10.51 years of schooling and a household monthly income of CLP 255,175 or around $500 dollars. These patterns indicate that these schools serve low-income families. Actually, an average 62.25% of the students are classified as vulnerable. The average number of students in $4th$ grade is about 73 and two thirds of the schools are public schools, which the remainder being voucher schools.

Table 2 provides balance tests. When comparing students and schools in the control and treatment groups, we observe that they are extremely similar in terms of socio-economic variables, size, type of school, and students outcomes. We do not observe statistical differences in all variables. In addition, there were short-term attriters in both treatment and control groups, as students did not attend to school the day in which the follow-up tests were applied. In order to assure the integrity of the experiment, Table A2 presents the results of comparing the characteristics of attriters and non-attriters in the treatment and control groups. Results imply again no statistical differences between students into each of these groups, and indeed that attrition seem to be balanced across groups. All in all, our reading of these results is that there are no systematic differences between treatment and control groups in most of the relevant variables. These results imply that the main cost of short-term attrition relates to sample sizes and having a lower-powered experiment without significant changes in observable characteristics.

3.4 Methods

The random assignment of the program across schools allows us to estimate the effect of offering the program by comparing average outcomes between the treatment and control groups. We estimate the effect of offering the program—this is the intention to treat (ITT) effect—by running the following OLS regression:

$$Y_i = \alpha + \beta T_{s(i)} + \gamma X_i + \epsilon_i \tag{1}$$

where Y_i is the outcome of interest for student i, $T_{s(i)}$ is a dummy variable that equals 1 is the school of student i was assigned to be treated, and β is the ITT effect of the program. X_i is a set of control variables at the student level that are included in the regression in order to increase the precision of the estimates, including strata fixed effects, gender, baseline values of the CLPT and GPL tests, student average grades and attendance in 2009. Finally, ϵ_i is an error term. In all regression we cluster standard errors at the classroom level.

We study heterogeneous treatment effects considering the probability of dropping out from school. This dimension is important to understand the potential effects of the program in medium-term. Drop-out rates are higher for male students than for female students in Chile, capturing both supply- and demand-side factors (MINEDUC, 2020). In turn, drop-out rates in Chile and abroad depend on school effects and student education outcomes, including grades, attendance, among others (MINEDUC, 2020). Thus we follow the model used by MINEDUC (2020) to predict the probability of dropping out from school by 2017 using the following independent variables: drop-out rates of the school attended in 2010, the share of poor students in the school attended in 2010, student grades in 2009, student attendance in the first half of 2010 (before the treatment started), and student gender. We estimate this model only for the control students. Results are presented in Table A3. All variables have the expected signs and most are are statistically significant. We use the predicted probabilities to split the full sample among students with high probability and low probability of dropping out of school by 2017 using the median of the predicted probability.

Additionally, we report treatment-on-the-treated (ToT) effects in the appendix. This follows from the fact that the number of sessions students actually received varies substantially. To estimate ToT effects, we run the following instrumental variables regression:

$$Y_i = \alpha + \beta N_i + \gamma X_i + \epsilon_i \tag{2}$$

where all the variables are the same as in equation (1), except for N_{is}, which is the share of sessions of the program received by student i, which stands as a measure of the intensity of the program, which we instrument using the intention-to-treat dummy $T_{s(i)}$. The estimate of the cost β measures is the ToT effect.

In estimating medium-term effects, a challenge is that the program may affect the extensive educational margins, including the probability of being enrolled in school and the probability of not repeating. Thus, there is a selection bias in which conventional treatment effects are probably downward biased—this is the case if the program helped some of the weakest students to change their extensive margin decisions—. We use a set of techniques to deal with this challenge. First, we re-weight the sample using inverse probability weighting (IPW), which assumes we can correct the estimates by re-weighting the effects using observable characteristics at the baseline. Second, we estimate parametric Tobit estimates of the ITT effects, following Angrist *et al.* (2006). This procedure provides parametric estimates assuming a normal distribution of the latent variable that measures the outcomes and assumes that any untested student would have scored at or below a threshold, i.e. different quantiles. Given these assumptions, the Tobit procedure estimates using data censored at different thresholds for people with information. We focus on percentiles 1, 5, 10, 25, 50, 75, and 90. The estimated treatment effects capture causal impacts under these assumptions, and comparison of different thresholds helps to study the part of the distribution where the impacts are bigger. Third, we compute non-parametric bounds of the ITTs. This procedure computes non-parametric bounds of the effects assuming that the selection bias is probably negative if treatment effects on the intensive margin are positive, as in Angrist *et al.* (2006). The idea is the following: take the percentile θ of the distribution of the observed outcomes for the control group, construct the upper bound by dropping the lower θ% of the distribution of outcomes for the treatment group to construct the upper bound of the ITT effect, and construct the lower bound by dropping the θ% of the distribution of outcomes for both groups.[10] We report the effects for θ_0% for each variable, i.e. the difference in the share of observations available for the treatment minus the control group. These three procedures present estimates that rely on different assumptions to solve the potential selection bias associated with negative selection for variables in which the treatment affects the extensive margin.

4 Main Results

In this section we present our main results. We first report estimates of short-term effects. Next, we present treatment effects on school enrollment and progression in 2011, 2014 and 2017, which is one year after the program, after primary school and after $11th$ grade, respectively. We then move to the estimates of the effects on school attendance and grades in primary and secondary education. Finally, we present estimates on standardized tests in $8th$ and $10th$ grades.

[10] Notice that given there is negative selective attrition based on the treatment, the lower bound is downward biased. We still report these results as a reference.

4.1 Short-Term Effects

Table 3 presents our findings of ITT effects on CLPT and GPL scores at the time when the program ended. We consider both the aggregate scores and, in the case of CLPT, also effects on its three components. We first report the impact on the CLPT score. This ITT effect is equal to about 0.06σ. The impacts are slightly stronger for students with a high probability of dropping out but the differences are not economically or statistically significant. When looking at ITT effects on the components of CLPT, results in Table 3 suggest that the impacts are stronger for reading comprehension—with an ITT effect of about 0.11σ—and smaller effects for use of language and text production. We do not see significant heterogeneous effects in most cases, maybe with the exception of significantly stronger effects on the use of language component, for which we estimate large effects for students with high probability of dropping out from school. It is worth noting that reading comprehension is the dimension of the CLPT test most related to the design of the program. In contrast, effects on attitudes towards reading are close to 0 for the full sample and for the two groups.

In all, the short-term effects of the program are small to moderate on dimensions related to the actual implementation of the program. If we compare these results with other tutoring interventions, they are smaller than the average impacts found in the meta-analysis on tutoring programs by Nickow *et al.* (2020), but not different than several of the studies in that review that have characteristics similar to SPE, as previously discussed. If we compare with other studies aimed at improving reading scores, our results are smaller than those from the *balsakhis* program in India (Banerjee *et al.*, 2007), and slightly smaller than the *read-a-thon* program in the Philippines (Abeberese *et al.*, 2011). Moreover, SPE effects are similar to the ones found in the *Literacy Hour* in the UK (Machin and McNally, 2008), which is remarkable as the duration of the last program is two years, while SPE lasted just three months.

4.2 Effects on Drop-out Rates and School Progression

The results in the previous section suggest small impacts of the program. However, they do not consider potential impacts on several dimensions that take place after the program was implemented. In this section, we study effects on medium-term drop our rates and schooling progression.

Table 4 presents the results. We find a significant ITT effect of -0.8 p.p. on the probability of dropping out from school in 2011. This is a large effect as the probability of dropping out from school from school in 2011 in our sample is 2%. There is heterogeneity in the effects. The ITT impact is much bigger in absolute value (although imprecise) for students with high probability of dropping out from school. Results for 2014 are similar but slightly smaller and less precise but results for 2017 imply a large negative effect of -1.76 p.p.. This effect is entirely explained by effects for students with high ex-ante probability of dropping out from school at -3.73 p.p.

Next, we analyze the effects of the program on school progression. We define a variable that indicates if the student is enrolled in school and has not repeated any grade up to year t. We present results for 2011, 2014, and 2017. Our results imply significant ITT effects for 2014 and 2017 of 4.53 p.p. and 3.19 p.p., respectively. Comparing these effects with the ones for drop out decisions, they imply a non trivial effect of the program on repetition rates. Again, the effects are bigger for students with a high probability of dropping out from school.

These results suggest non trivial effects of the treatment on the extensive margin, more so for students with higher ex-ante probability of dropping out from school. This confirms the hypothesis that the SPE program may have also affected extensive margin decisions. If we compare these effects on the probability of dropping out with other research, we find that they are comparable to those in Gallego *et al.* (2016) for an information intervention in Peru, and those for programs that have significant effects reported in J-PAL (2017). Moreover, our estimates tend to be stronger and more significant than the effects on dropout rates and on school progression reported in Guryan *et al.* (2021), and smaller than the effects on graduation rates in secondary education in Lavecchia *et al.* (2020). We leave a discussion of cost effectiveness for the final part of the paper.

4.3 Effects on Grades and Attendance

An additional margin of effects is related to school attendance and grades. Table 5 presents the results. We start by analyzing student attendance in primary and secondary school, for students enrolled in school. We find significant effects for attendance in both primary and secondary school. The effects are slightly larger for secondary school attendance than for primary school attendance, at 1.13 p.p. versus 0.81 p.p.. This is relevant as secondary school attendance is lower in our sample than primary school attendance, at 82% versus 89%. As in previous outcomes, the effects for the complete sample are driven by students with high ex-ante probability of dropping out, at 1.55 and 2.22 p.p. for primary and secondary education, respectively. Note that in the case of secondary attendance, our estimates may be biased because of selection into school enrollment based on the treatment. To deal with this concern, we first present IPW estimates in Table 7, which imply larger impacts at 1.64 p.p. for the complete sample and 3.41 for student with high probability of dropping out. Moreover, the upper bound of the effects in Table 9 imply even larger impacts. Finally, Tobit estimates in Figures 1 and 3 confirm these results.

Table 5 also presents results for ITT effects on primary and secondary school grades. Results for primary school grades are positive and statistically significant at 0.09, equivalent to 0.11σ. This effect is again driven by students at risk of dropping out. In the case of secondary school grades, the effects are not statistically significant using our main specification. When using estimators robust to potential selection bias, we tend to find bigger and statistically significant effects, particularly for students at risk of dropping out. IPW estimates in Table 7 almost double OLS estimates for both the full sample and for students at risk of dropping out, but they are only statistically significant for this subgroup. The upper bounds in Table 9 are even larger. Finally, Tobit estimates in Figures 1 and 3 present estimates that are only marginally significant. Thus, the evidence for secondary school grades suggests treatment effects on this margin, but they are less precisely estimated than for attendance.

4.4 Effects on Test Scores in the Medium-Term

We now estimate ITT effects on language and math SIMCE test scores in $8th$ and $10th$ grades. Concerns about selection bias in our OLS specification are likely more relevant for this outcome, although we still present those results in Table 6. We estimate effects of 0.07σ and 0.06σ on math and language in $8th$ grade, but neither of them is statistically significant. However, we find statistically significant and larger impacts for students at risk of dropping out, at 0.17σ and 0.09σ for math and language, respectively. Our IPW estimates increase slightly in magnitude but again are only statistically significant for children at higher risk of dropping out, with estimates of 0.20σ and 0.10σ for math and language, respectively. Moreover, Tobit estimates also imply larger and statistically significant effects for both the full sample and for students at risk of dropping out, as displayed by Figures 2 and 4.

Results for test scores in $10th$ grade display a similar pattern, particularly for estimators that correct for potential selection bias. IPW estimates in Table 8 imply significant effects for the full sample, with coefficients of 0.13σ and 0.11σ for math and language, respectively. Again we observe larger impacts for students at risk of dropping out, but the differences are not as large as for other outcomes. Estimates for bounds in Table 9 again suggest statistically significant upper bounds, which are larger for students at risk of dropping out from school. Tobit estimates also point out at statistically significant effects in both tests for the complete population in Figure 1, and for students with high probability of dropping out in Figure 4, although the differences with students with lower probability of dropping out are smaller.

The results in this section suggest that there are significant treatment effects in both primary and secondary school test scores, and that the impacts are larger for students at risk of dropping out. Interestingly, the effects seem to be at least of the same order of magnitude of short-term effects for language estimated at the end of the implementation of the program. Moreover, we estimate significant effects on math, suggesting the program affects outcomes beyond its main area of focus.

4.5 Treatment on the Treated Estimates

We now discuss our results for ToT effects. We present these results in Tables A5 and A6. We define the intensity of the treatment at the student level and as the share of sessions attended by each student, out of a maximum of 15. Table A4 presents the first stage. Interestingly, the first stage results suggest that children at risk of dropping out are slightly more likely to attend the tutoring sessions. The coefficient is 4.1 p.p. higher for students with high risk of dropping out versus students with low risk of dropping out, with a p-value of 0.11 for a test of the difference. This is equivalent to about 0.5 additional sessions, relative to an average of 7.7 sessions for a student with low risk of dropping out. While this result is informative of the program, it cannot explain the difference in ITT effects we find. One way of looking at this is to analyze the ToT estimates.

The pattern of estimates is mechanically consistent with our previous results in the sense that ToT estimates are scaled-up versions of ITT effects because of imperfect attendance to the tutoring sessions. Thus, ToT effects are informative because they provide an estimate of the size of treatment effects on the treated students. For instance, short-term effects on CLPT increase to 0.11σ for the full sample, without any statistical difference for students with low and high probability of dropping out. As with ITT estimates, the only relevant difference in this dimension is that effects are stronger among kids at risk of dropping out for the use of language component of CLPT, at 0.19σ. Results for medium-term effects also increase in magnitude. For instance, ToT effects on the probability of dropping out from school by 2017 are -3.34 p.p., explained by a huge impact for kids with higher risk of dropping out. The same pattern emerges for the other outcomes discussed above. In all, ToT estimates confirm the previous patterns of stronger effects for children with high probability of dropping out, and thus discard the possibility that these are explained by small differences in attendance to the tutoring sessions.

4.6 Heterogeneous effects: Human Connection or Academic Contents?

Results in previous sections suggest that the program affected several outcomes that go beyond dimensions directly targeted by the program. In this section, we study the potential mechanisms behind this by exploiting some questions in the follow-up survey in which treated students respond questions on the academic quality of the tutoring program and on their personal relationships with the tutors. The questions are answered in a scale that goes from 0 to 3, from very bad to very good. As previously discussed, the average for the variable measuring quality of instruction is 2.49 and the average for the variable measuring the quality of the relationship with the tutor is 2.25. For this analysis, we create dummies that take a value of one if the student answers 3 (very good) in a question and interact this variable with the treatment dummy in our main specification.

Results for the complete sample are presented in Table 10-A.We start with effects on short-term outcomes. The evidence suggests that the interaction effects with the dummy for tutor-student relationship has a positive impact for both the CLPT and GPL tests. In the case of the interaction with the quality of academic instruction, the interaction is significant only for the GPL test. Next, moving to the medium-term effects, the pattern that emerges suggest relevant interaction effects with the dummy for the tutor-student relationship on most variables. Notably, several of them are statistically significant, especially for variables for which we find larger medium-term impacts, such as on-time school progression in 2014 and 2017, and attendance and grades in secondary school. The interaction with the quality of the tutoring section is in general not significant and, if anything, has negative effects on a couple of outcomes.

Tables 10-B and 10-C present the results splitting the sample between students with high and low ex-ante probability of dropping out. The evidence suggests that the medium-term impacts are driven by students with high probability of dropping out, and that tutor-student relationships matter. The latter is consistent with the notion that medium-term effects are mostly driven by the influence of the human connection rather than by the academic contents of the program. Consistent with this view and with previous results, the interaction effects for the short-term outcomes are relevant for both groups of students and for both tutor-student relationships and tutor quality. This suggests that short-term outcomes are affected by both dimensions but that the medium-term effects we identify are mostly affected by the personal influence of tutors on tutees.

5 Conclusions

The process of human capital accumulation poses important challenges to actual educational systems in terms of the implementation of public policies that are cost-effective (World Bank, 2020). A particularly challenging dimension relates to the dynamic dimension of this process, in that current investments affect future investments and outcomes. Therefore, the study of the medium-term effects of educational interventions is crucial to complement research on short-term effects of interventions. This paper examines both the short and the medium-term effects of access to a tutoring intervention that took place over a three-month period in 2010.

We designed a randomized experiment to test how several short-term and medium-term educational outcomes responded to a short-lived tutoring program with a focus on reading. The program considered small group tutoring using college student volunteers as tutors, which belonged to the same communities as students. Actual attendance to the program was roughly 50% and students evaluated the program well. We exploit variation in ex-ante risk of dropping out across students to study heterogeneous effects of the treatment along this dimension. We hypothesize that the potential medium-term effects of the program on students with lower attachment to the educational process should be stronger. We applied a baseline and a short-term follow-up surveys and test to characterize the population, and measure learning outcomes and self-reported evaluation of the program by students. We also collected administrative information on educational outcomes to measure the medium-term effects of the program.

Our results show that the tutoring program leads to statistically significant but small improvements in short-term educational outcomes related to the program learning objectives. We find no heterogeneous treatment effects for children with different ex-ante probabilities of dropping out from school. In contrast, we find statistically and economically relevant effects on a range of outcomes several years after the program ended, including the probability of dropping out from school and of school progression on time, attendance and grades in primary and secondary education, and test scores in $8th$ and $10th$ grade. These effects are mostly driven by impacts on students at risk of dropping out from school. This suggests that the program provided inputs that go beyond its specific focus on improving reading. This conclusion is reinforced when analyzing heterogeneous treatment effect by student self-reported measures of personal relationships with tutors: the medium-term impacts are stronger for students who report close personal links with tutors. In contrast, the interaction of the treatment with a measure of the instructional quality of the tutor does not increase positive impacts on medium-term outcomes. Taken together, our findings indicate that short-lived educational interventions can enhance medium-term educational outcomes, emphasizing the dynamic effects of these interventions and the benefits to protecting students with low attachment to the educational process.

Finally, we discuss the size and cost-effectiveness of the program relative to other interventions. As previously mentioned, the program costs about \$100 per student (about \$125 if using the 2010 IMF's PPP conversion factor). Focusing on short-term effects, the program is less cost-effective than several other educational interventions reported in the literature (Kremer and Holla 2009; World Bank 2020), but more cost-effective than others interventions like the *Literacy Hour* in the UK (Machin and McNally, 2008) and remarkably more cost effective than JEC, a full day school program implemented in Chile (Bellei, 2009). In terms of other tutoring programs, our estimates of medium-term effects for school enrollment (test scores) are larger (smaller) than the ones reported in Guryan *et al.* (2021) for the SAGA tutoring program, which has much higher costs per student, at \$3,500–\$4,500 per student-year. Given all the margins that the SPE program affects, we can also use the concept of learning-adjusted years of schooling (LAYS) suggested by Angrist *et al.* (2020). This concept is useful because it combines all the margins we study—enrollment, attendance, and learning outcomes—in one metric. Moreover, it also captures the dynamic effects we estimate. Our results imply that for the full sample, the program produces 1.39 LAYS per \$100 spent per student.[11] This places the program in the mid range of estimates of cost-effectiveness well below the most effective ones but in a range comparable with the the *read-a-thon* program in the Phillipines (Abeberese *et al.*, 2011). This is an important point because if we just consider the short-term effects, the *read-a-thon* dominates the program we study (mostly due to its low cost), but when we consider the medium-term effects we estimate, that conclusion changes. This result emphasizes that the medium-term effects of educational policies are quite relevant for policy analysis.

References

Abeberese, A. B., Kumler, T. J. and Linden, L. L. (2011). *Improving reading skills by encouraging children to read: A randomized evaluation of the Sa Aklat Sisikat reading program in the Philippines*. Tech. rep., IZA Discussion Papers.

Angrist, J., Bettinger, E. and Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in colombia. *American Economic Review*, **96** (3), 847–862.

[11] We do the following: LAYS is years attended times average attendance times learning adjusted to an international benchmark, which is 0.75 in the baseline accordingly to Angrist *et al.* 2020. This produces 7.52 LAYS for the control group (11.52 years of actual education times average attendance of 0.87 times 0.75). Then, we use our estimated treatment effects on drop-out rates, attendance, and learning to compute the same value for the treatment group. We consider learning effects of 0.10σ, consistent with our estimates on grades and test scores. This produces an estimate of 8.91 LAYS for the treatment group.

Angrist, N., Evans, D. K., Filmer, D., Glennerster, R., Rogers, F. H. and Sabarwal, S. (2020). How to improve education outcomes most efficiently? *Development Research*.

Baker, S., Gersten, R. and Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly*, **35** (4), 494–519.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R. and Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in india. *American Economic Journal: Economic Policy*, **2** (1), 1–30.

—, Cole, S., Duflo, E. and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, **122** (3), 1235–1264.

Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? results from a natural experiment in chile. *Economics of Education Review*, **28** (5), 629–640.

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A. and Vaughn, M. G. (2014). Intensive reading remediation in grade 2 or 3: Are there effects a decade later? *Journal of educational psychology*, **106** (1), 46.

Carmeli, A. (2000). Perach: A countrywide tutoring and mentoring scheme from israel. *Widening Participation and Lifelong Learning*, **2** (1), 46–48.

Carneiro, P. M. and Heckman, J. J. (2003). Human capital policy.

Dobbie, W. and Fryer Jr, R. G. (2013). Getting beneath the veil of effective schools: Evidence from new york city. *American Economic Journal: Applied Economics*, **5** (4), 28–60.

Fryer Jr, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, **126** (4), 1755–1798.

— and Howard-Noveck, M. (2020). High-dosage tutoring and reading achievement: evidence from new york city. *Journal of Labor Economics*, **38** (2), 421–452.

Gallego, F., Molina, O. and Neilson, C. (2016). The impact of information provision on human capital accumulation and child labor in peru. Manuscript.

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M., Dodge, K., Farkas, G., Fryer Jr, R. G., Mayer, S., Pollack, H. *et al.* (2021). *Not Too Late: Improving Academic Outcomes Among Adolescents*. Tech. rep., National Bureau of Economic Research.

Holdaway, D. (1979). *The foundations of literacy*, vol. 138. Ashton Scholastic Sydney.

J-PAL (2017). Roll call: Getting children into school. J-PAL Policy Bulletin.

Jacob, R., Armstrong, C., Bowden, A. B. and Pan, Y. (2016). Leveraging volunteers: An experimental evaluation of a tutoring program for struggling readers. *Journal of Research on Educational Effectiveness*, **9** (sup1), 67–92.

Jayachandran, S. (2014). Incentives to teach badly: After-school tutoring in developing countries. *Journal of Development Economics*, **108**, 190–205.

Kremer, M. and Holla, A. (2009). Improving education in the developing world: What have we learned from randomized evaluations? *Annual Review of Economics*, **1** (1), 513–542, pMID: 23946865.

Lavecchia, A. M., Oreopoulos, P. and Brown, R. S. (2020). Long-run effects from comprehensive student support: Evidence from pathways to education. *American Economic Review: Insights*, **2** (2), 209–24.

Lavy, V. (2020). Teachers' pay for performance in the long-run: The dynamic pattern of treatment effects on students' educational and labour market outcomes in adulthood. *The Review of Economic Studies*, **87** (5), 2322–2355.

Lindo, E. J., Weiser, B., Cheatham, J. P. and Allor, J. H. (2018). Benefits of structured after-school literacy tutoring by university students for struggling elementary readers. *Reading & Writing Quarterly*, **34** (2), 117–131.

Machin, S. and McNally, S. (2008). The literacy hour. *Journal of Public Economics*, **92** (5-6), 1441–1462.

McKenna, M. C. and Kear, D. J. (1990). Measuring attitude toward reading: A new tool for teachers. *The reading teacher*, **43** (9), 626–639.

—, — and Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading research quarterly*, pp. 934–956.

Medina, A. and Gajardo, A. M. (2010). *Pruebas de comprensión lectora y producción de textos (CL-PT): 5o a 8o año básico: marco conceptual y manual de aplicación y corrección*. Eds. Universidad Católica de Chile.

Miller, S. and Connolly, P. (2013). A randomized controlled trial evaluation of time to read, a volunteer tutoring program for 8-to 9-year-olds. *Educational Evaluation and Policy Analysis*, **35** (1), 23–37.

MINEDUC (2020). Deserción escolar: Diagnóstico y proyección en tiempos de pandemia. Mimeo, Centro de Estudios, Ministerio de Educación de Chile.

Morris, D., Shaw, B. and Perney, J. (1990). Helping low readers in grades 2 and 3: An after-school volunteer tutoring program. *The Elementary School Journal*, **91** (2), 133–150.

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

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

Resnjanskij, S., Ruhose, J., Wiederhold, S. and Woessmann, L. (2021). Can mentoring alleviate family disadvantage in adolscence? a field experiment to improve labor-market prospects.

Ritter, G. W., Barnett, J. H., Denny, G. S. and Albin, G. R. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research*, **79** (1), 3–38.

Rockoff, J. and Turner, L. J. (2010). Short-run impacts of accountability on school quality. *American Economic Journal: Economic Policy*, **2** (4), 119–47.

Solange, O. G. M. (2004). *El tratamiento didáctico de las lecturas literarias en el primer nivel de enseñanza media en chile: una propuesta de formación en didáctica de la literatura*. Ph.D. thesis, Universidad Complutense de Madrid.

Wasik, B. A. (1997). Volunteer tutoring programs. a review of research on achievement outcomes. report no. 14.

World Bank (2020). Cost-effective approaches to improve global learning.

Zijlstra, H., Van Bergen, E., Regtvoort, A., De Jong, P. F. and Van Der Leij, A. (2020). Prevention of reading difficulties in children with and without familial risk: Short-and long-term effects of an early intervention. *Journal of Educational Psychology*.

Table 1: Summary statistics

	N	Mean	SD	Min	Max
Panel A: Process Information					
Students' attendance	2,752	0.76	0.30	0.00	1.00
Sessions per student	2,752	9.13	3.73	0.00	15.00
Sessions per school	2,752	12.04	1.35	9.00	15.00
GPL-Feeling about reading improvements	2,700	2.4000	0.8337	0	3
GPL-Feeling about writing improvements	2,692	2.4331	0.7533	0	3
GPL-Relationship with tutor	2,686	2.2520	0.8700	0	3
GPL-Tutor's quality instruction	2,688	2.4929	0.7264	0	3
Panel B: Student Variables, before treatment					
Gender (=1 if Male)	6,066	0.53	0.49	0.00	1.00
Grades 2009	5,999	5.77	0.58	2.20	7.00
Attendance 2009	5,999	0.91	0.72	0.17	1.00
Attendance 2010	6,028	0.94	0.68	0.43	1.00
CLPT-Reading Comprehension	5,506	0.50	0.17	0.00	0.99
CLPT-Use of Language	5,506	0.67	0.25	0.00	1.00
CLPT-Texts Production	5,506	0.49	0.25	0.00	0.99
CLPT-Total	5,506	0.52	0.18	0.00	0.95
GPL-Self-Perception as a Reader	5,432	1.82	0.76	0.00	3.00
GPL-Enjoyableness for Reading	5,416	2.44	0.74	0.00	3.00
GPL-Interest for Reading	5,435	2.04	0.77	0.00	3.00
GPL-Perception of Reading at School	5,346	2.27	0.49	0.00	3.00
Panel C: Student Level Variables, after treatment					
Drop 2011	6,129	0.02	0.15	0.00	1.00
Drop 2014	6,129	0.05	0.22	0.00	1.00
Drop 2017	6,129	0.16	0.37	0.00	1.00
On Time 2011	6,129	0.91	0.28	0.00	1.00
On Time 2014	6,129	0.79	0.41	0.00	1.00
On Time 2017	6,129	0.58	0.49	0.00	1.00
Average grades primary	6,090	5.32	0.84	00.0	6.90
Average grades secondary	5,864	5.03	1.22	0.00	7.00
Attendance primary	6,090	0.89	0.11	0.00	1.00
Attendance secondary	5,864	0.82	0.20	0.00	1.00
CLPT-Reading Comprehension	5,236	0.56	0.18	0.00	1.00
CLPT-Use of Language	5,236	0.72	0.23	0.00	1.00
CLPT-Texts Production	5,236	0.47	0.23	0.00	0.96
CLPT-Total	5,236	0.55	0.17	0.00	0.93
GPL-Self-Perception	5,206	1.84	0.75	0.00	3.00
GPL-Enjoyableness for Reading	5,169	2.32	0.81	0.00	3.00
GPL-Interest for Reading	5,196	1.89	0.80	0.00	3.00
GPL-Perception of Reading at School	5,110	2.22	0.52	0.00	3.00
Primary language test score	4,751	234.47	51.74	111.57	373.24
Primary math test score	4,800	257.52	45.24	148.24	402.69
Secondary language test score	3,348	245.18	51.87	134.25	401.64
Secondary math test score	3,326	265.65	61.94	100.95	425.72
Panel D: School Variables, before treatment					
Mothers' years of schooling	84	10.51	1.41	6.18	12.64
Household income	84	255,174.70	75,701.12	102,941.20	423,571.40
Language SIMCE 2009	82	254.30	21.78	195.00	293.00
Math SIMCE 2009	82	246.73	23.04	188.00	292.00
IVE 2010	84	62.25	11.90	42.20	91.72
School size	84	72.94	35.06	10.00	201.00
School dependence (=1 if PS)	84	0.33	0.47	0.00	1.00

Table 2: Balance by treatment status

	Treatment mean	Control mean	Difference
Panel A: School characteristics			
Mothers' years of schooling	10.65	10.35	0.30
	(1.34)	(1.48)	(0.31)
Household income	261,615.80	248,089.60	13,526.20
	(80,009.94)	(70,989.89)	(16,476.93)
Language SIMCE 2009	257.31	251.15	6.16
	(19.24)	(24.00)	(4.82)
Math SIMCE 2009	249.98	243.33	6.65
	(19.98)	(25.68)	(5.10)
IVE 2010	60.69	63.97	-3.28
	(11.24)	(12.50)	(2.60)
School size	71.70	74.30	-2.60
	(35.44)	(35.05)	(7.70)
Panel B: Student characteristics			
Grades 2009	5.79	5.76	0.03
	(0.57)	(0.59)	(0.05)
Attendance 2009	91.56	91.37	0.19
	(6.64)	(7.69)	(0.67)
Attendance 2010	93.15	94.06	-0.90
	(6.68)	(6.81)	(0.87)
CLPT-Reading Comprehension	50.71	49.33	1.39
	(17.14)	(16.79)	(1.57)
CLPT-Use of Language	67.50	65.20	2.31
	(24.44)	(25.04)	(2.11)
CLPT-Texts Production	49.32	48.36	0.97
	(25.18)	(24.87)	(2.38)
CLPT-Total	52.67	51.32	1.35
	(17.78)	(17.22)	(1.84)
GPL-Self-Perception	1.83	1.80	0.03
	(0.76)	(0.76)	(0.03)
GPL-Enjoyableness for Reading	2.46	2.42	0.04
	(0.73)	(0.74)	(0.03)
GPL-Interest for Reading	2.05	2.03	0.02
	(0.77)	(0.77)	(0.04)
GPL-Perception of Reading at School	2.27	2.28	-0.01
	(0.50)	(0.49)	(0.02)

Notes: Means and standard deviations for each variable are presented, along with differences between groups with their corresponding standard errors. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table 3: Short-term effects: ITT estimates

Dependent variable	(1) Full sample	(2) High dropout	(3) Low dropout
CLPT Total	0.0585*	0.0856	0.0649*
	(0.0349)	(0.0533)	(0.0359)
N	5,234	2,183	3,051
GPL Total	0.00652	-0.00758	0.0223
	(0.0193)	(0.0241)	(0.0254)
N	5,091	2,114	2,977
CLPT Reading Comprehension	0.107**	0.103	0.121**
	(0.0459)	(0.0700)	(0.0517)
N	5,234	2,183	3,051
CLPT Use of Language	0.0483	0.112*	0.00501
	(0.0366)	(0.0601)	(0.0371)
N	5,234	2,183	3,051
CLPT Texts Production	0.0369	0.0767	0.0570
	(0.0516)	(0.0728)	(0.0537)
N	5,234	2,183	3,051

Notes: All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. Estimates in column (2) consider students with estimated probability of dropping-out above the median of the sample and estimates in column (3) students with estimated probability of dropping-out below the median of the sample. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table 4: Medium-term effects on drop-outs and school progression: ITT estimates

	(1)	(2)	(3)
Dependent variable	Full sample	High dropout	Low dropout
Dropout 2011	-0.00806**	-0.0142	-0.000412
	(0.00405)	(0.00883)	(0.00211)
N	6,129	2,721	3,408
Dropout 2014	-0.00650	-0.0131	0.00343
	(0.00587)	(0.0120)	(0.00408)
N	6,129	2,721	3,408
Dropout 2017	-0.0176*	-0.0373*	0.00479
	(0.0106)	(0.0196)	(0.00945)
N	6,129	2,721	3,408
On time 2011	0.00787	-0.00667	0.00709
	(0.00813)	(0.0160)	(0.00583)
N	6,129	2,721	3,408
On time 2014	0.0453***	0.0654***	0.0201
	(0.0139)	(0.0233)	(0.0137)
N	6,129	2,721	3,408
On time 2017	0.0319**	0.0544**	0.00865
	(0.0158)	(0.0225)	(0.0200)
N	6,129	2,721	3,408

Notes: All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. Estimates in column (2) consider students with estimated probability of dropping-out above the median of the sample and estimates in column (3) students with estimated probability of dropping-out below the median of the sample. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table 5: Medium-term effects on school performance: ITT estimates

	(1)	(2)	(3)
Dependent variable	Full sample	High dropout	Low dropout
Attendance primary	0.814**	1.547**	0.0727
	(0.373)	(0.654)	(0.324)
N	6,090	2,684	3,406
Attendance secondary	1.139*	2.224**	0.132
	(0.626)	(1.093)	(0.617)
N	5,864	2,490	3,374
Average grades primary	0.0875***	0.133***	0.0499*
	(0.0277)	(0.0357)	(0.0279)
N	6,090	2,684	3,406
Average grades secondary	0.0538	0.0966	0.00859
	(0.0385)	(0.0623)	(0.0394)
N	5,864	2,490	3,374

Notes: All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. Estimates in column (2) consider students with estimated probability of dropping-out above the median of the sample and estimates in column (3) students with estimated probability of dropping-out below the median of the sample. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table 6: Medium-term effects on test scores: ITT estimates

	(1)	(2)	(3)
Dependent variable	Full sample	High dropout	Low dropout
Primary language test score	0.0505	0.0825*	0.0509
	(0.0449)	(0.0463)	(0.0606)
N	4,751	1,845	2,906
Primary math test score	0.0686	0.170***	0.0165
	(0.0503)	(0.0523)	(0.0622)
N	4,800	1,864	2,936
Secondary language test score	0.0642	0.000513	0.0882*
	(0.0408)	(0.0684)	(0.0477)
N	3,348	1,044	2,304
Secondary math test score	0.0854*	0.0711	0.0891*
	(0.0447)	(0.0652)	(0.0512)
N	3,326	1,047	2,279

Notes: All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. Estimates in column (2) consider students with estimated probability of dropping-out above the median of the sample and estimates in column (3) students with estimated probability of dropping-out below the median of the sample. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table 7: Medium-term effects on school performance: Estimates controlling for attrition using IPW

	(1)	(2)	(3)
Dependent variable	Full sample	High dropout	Low dropout
Average grades primary	0.106***	0.172***	0.0541*
	(0.0329)	0.172***	(0.0327)
N	6,090	2,684	3,406
Average grades secondary	0.0892	0.178**	-0.00408
	(0.0567)	(0.0827)	(0.0450)
N	5,864	2,490	3,374
Attendance primary	1.196**	2.412***	-0.0340
	(0.526)	(0.824)	(0.353)
N	6,090	2,684	3,406
Attendance Secondary	1.636*	3.408**	-0.244
	(0.883)	(1.392)	(0.634)
N	5,864	2,490	3,374

Notes: All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. Estimates in column (2) consider students with estimated probability of dropping-out above the median of the sample and estimates in column (3) students with estimated probability of dropping-out below the median of the sample. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table 8: Medium-term effects on test scores: Estimates controlling for attrition using IPW

	(1)	(2)	(3)
Dependent variable	Full sample	High dropout	Low dropout
Primary language test score	0.0908	0.118*	0.0619
	(0.0606)	(0.0623)	(0.0725)
N	4,751	1,845	2,906
Primary math test score	0.0987	0.198***	0.0208
	(0.0662)	(0.0645)	(0.0765)
N	4,800	1,864	2,936
Secondary language test score	0.107*	0.120	0.104*
	(0.0551)	(0.0766)	(0.0632)
N	3,348	1,044	2,304
Secondary math test score	0.128**	0.170**	0.111
	(0.0642)	(0.0789)	((0.0718)
N	3,326	1,047	2,279

Notes: All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. Estimates in column (2) consider students with estimated probability of dropping-out above the median of the sample and estimates in column (3) students with estimated probability of dropping-out below the median of the sample. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table 9: Bounds on ITT treatment effects

	(1)	(2)	(3)	(4)	(5)	(6)
	All Sample		High dropout		Low dropout	
Dependent variable	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
	Percentile 5		Percentile 9		Percentile 1	
Secondary grades	0.0429	0.203***	0.0457	0.274***	0.0407	0.0868**
	(0.0264)	(0.0328)	(0.0405)	(0.0464)	(0.0314)	(0.0362)
	Percentile 5		Percentile 9		Percentile 1	
Secondary attendance	0.883**	3.732***	1.085	5.350***	0.461	1.780***
	(0.404)	(0.524)	(0.738)	(0.823)	(0.471)	(0.579)
	Percentile 26		Percentile 36		Percentile 17	
Primary language test score	0.0353	0.394***	0.0834*	0.471***	0.0240	0.308***
	(0.0358)	(0.0409)	(0.0432)	(0.0494)	(0.0457)	(0.0461)
	Percentile 26		Percentile 36		Percentile 17	
Primary math test score	0.0569	0.404***	0.133***	0.473***	0.0216	0.307***
	(0.0438)	(0.0503)	(0.0424)	(0.0574)	(0.0502)	(0.0542)
	Percentile 49		Percentile 65		Percentile 33	
Secondary language test score	0.0285	0.761***	0.0817	0.974***	0.0168	0.558***
	(0.0295)	(0.0382)	(0.0740)	(0.0639)	(0.0356)	(0.0401)
	Percentile 49		Percentile 65		Percentile 33	
Secondary math test score	0.0253	0.701***	-0.0319	0.865***	0.0481	0.548***
	(0.0317)	(0.0439)	(0.0674)	(0.0625)	(0.0354)	(0.0461)

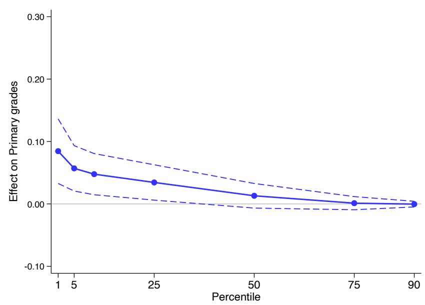
Notes:All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. We use the methodology developed vy Angrist *et al.* (2006). We report the effects for $\theta_0\%$ for each variable, i.e. the difference in the share of observations available for the treatment minus the control group. Estimates in columns (3) and (4) consider students with estimated probability of dropping-out above the median of the sample and estimates in columns (5) and (6) students with estimated probability of dropping-out below the median of the sample. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table 10: Interaction effects with student-tutor relationships and tutor quality

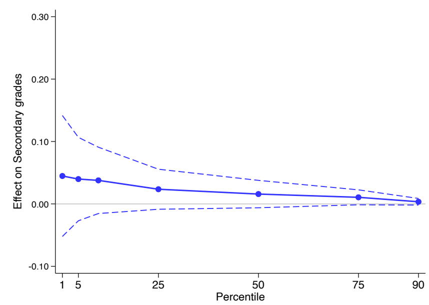
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	CLP	GPL	Dropout		On time		Attendance		Grade		Test score
	total	total	2014	2017	2014	2017	primary	secondary	primary	secondary	average
Panel A: Complete Sample											
Treatment	-0.004	-0.175***	-0.010	-0.020	0.049***	0.033	0.964*	0.775	0.096***	0.043	0.046
	(0.039)	(0.026)	(0.009)	(0.015)	(0.018)	(0.023)	(0.518)	(0.873)	(0.036)	(0.053)	(0.053)
Treatment × tutor-student	0.096***	0.165***	-0.000	-0.021	0.027	0.048**	0.199	1.940**	0.002	0.112**	0.064
	(0.029)	(0.026)	(0.008)	(0.014)	(0.018)	(0.024)	(0.446)	(0.927)	(0.026)	(0.054)	(0.046)
Treatment × tutor quality	0.017	0.229***	-0.002	0.013	-0.009	-0.034	-0.141	-2.285**	0.018	-0.128**	-0.026
	(0.028)	(0.025)	(0.008)	(0.015)	(0.017)	(0.024)	(0.379)	(0.952)	(0.027)	(0.056)	(0.038)
N	4342	4245	4823	4823	4823	4823	4789	4621	4789	4621	4137
Panel B: High dropout											
Treatment	-0.018	-0.176***	-0.027	-0.037	0.074**	0.055	2.119**	2.131	0.173**	0.103	0.039
	(0.058)	(0.030)	(0.017)	(0.025)	(0.036)	(0.037)	(0.873)	(1.455)	(0.066)	(0.081)	(0.070)
Treatment × tutor-student	0.127**	0.142***	-0.009	-0.041	0.059	0.089*	0.426	3.393**	-0.025	0.178**	0.090
	(0.050)	(0.035)	(0.016)	(0.027)	(0.038)	(0.046)	(0.869)	(1.456)	(0.043)	(0.078)	(0.080)
Treatment × tutor quality	0.067	0.280***	0.025	0.018	-0.029	-0.071*	-0.756	-4.190**	-0.005	-0.242**	-0.027
	(0.046)	(0.038)	(0.018)	(0.033)	(0.036)	(0.042)	(0.722)	(2.074)	(0.058)	(0.107)	(0.085)
N	1800	1757	2118	2118	2118	2118	2085	1937	2085	1937	1603
Panel C: Low dropout											
Treatment	0.045	-0.172***	0.009	0.008	0.018	0.011	-0.028	-0.312	0.046	-0.002	0.081
	(0.042)	(0.036)	(0.008)	(0.016)	(0.018)	(0.029)	(0.465)	(0.894)	(0.034)	(0.058)	(0.066)
Treatment × tutor-student	0.071**	0.175***	0.003	-0.018	0.011	0.025	0.263	1.453	0.016	0.082	0.034
	(0.034)	(0.034)	(0.007)	(0.017)	(0.018)	(0.026)	(0.444)	(1.101)	(0.031)	(0.065)	(0.055)
Treatment × tutor quality	-0.030	0.195***	-0.019***	0.006	0.009	-0.019	0.098	-1.438	0.008	-0.092	-0.061
	(0.034)	(0.033)	(0.007)	(0.016)	(0.015)	(0.027)	(0.368)	(0.985)	(0.028)	(0.064)	(0.041)
N	2542	2488	2705	2705	2705	2705	2704	2684	2704	2684	2534

Notes: All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. Estimates in Panel A consider the complete sample, estimates in Panel B consider only students with estimated probability of dropping-out above the median of the sample and estimates in Panel C only students with estimated probability of dropping-out below the median of the sample. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%. Standard errors are clustered at the classroom level and presented in parentheses.

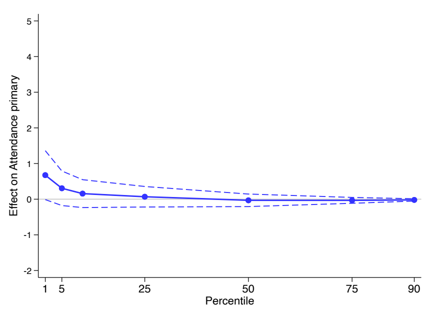
Figure 1: Tobit coefficients, medium-term effects on school performance



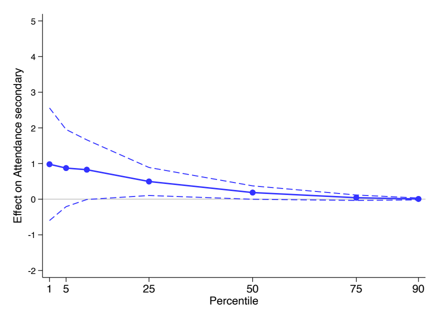
(a) Average grades primary



(b) Average grades secondary



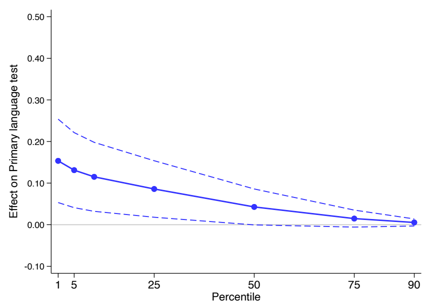
(c) Attendance primary



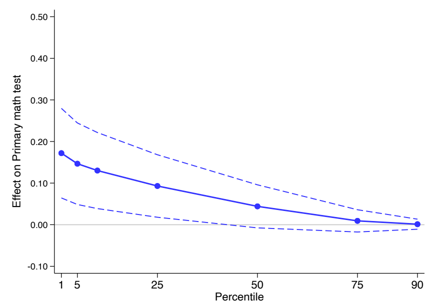
(d) Attendance secondary

Notes: The figure plots Tobit estimates of medium-term effects on school performance, using data censored at the point indicated on the X-axis (i.e., values below the indicated percentile are assigned a value of zero).

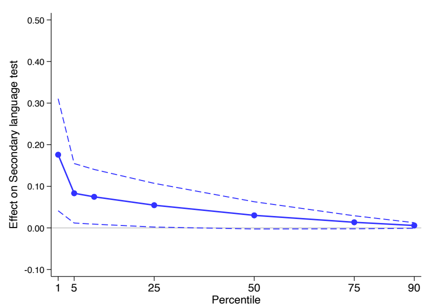
Figure 2: Tobit coefficients, medium-term effects on test scores



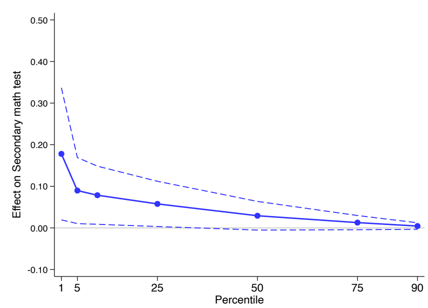
(a) Primary language test score



(b) Primary math test score



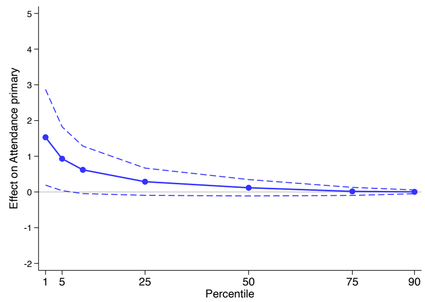
(c) Secondary language test score



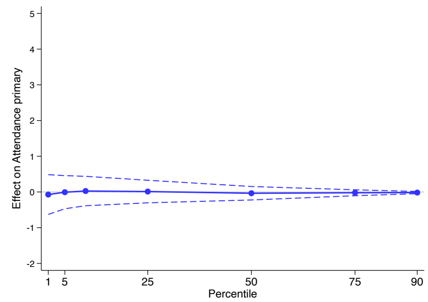
(d) Secondary math test score

Notes: The figure plots Tobit estimates of medium-term effects on test scores, using data censored at the point indicated on the X-axis (i.e., values below the indicated percentile are assigned a value of zero).

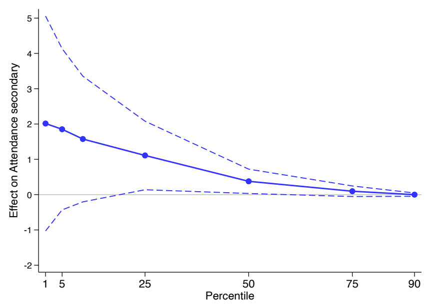
Figure 3: Tobit coefficients, medium-term effects on school performance, by probability of dropping out



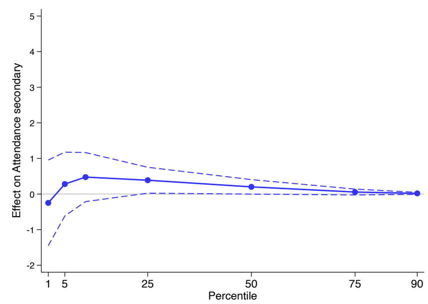
(a) High dropout - Attendance primary



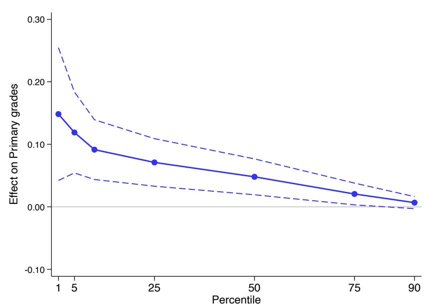
(b) Low dropout - Attendance primary



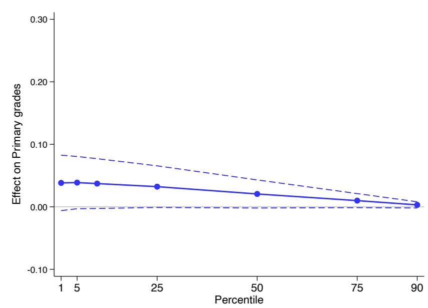
(c) High dropout - Attendance secondary



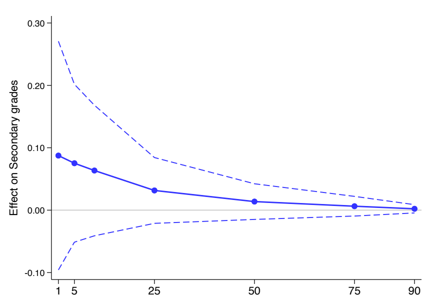
(d) Low dropout - Attendance secondary



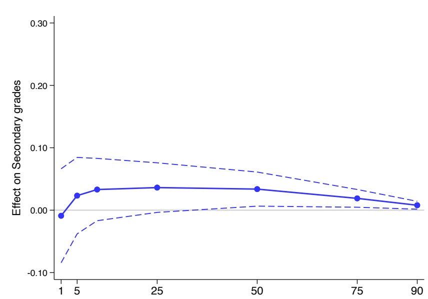
(e) High dropout - Primary grades



(f) Low dropout - Primary grades



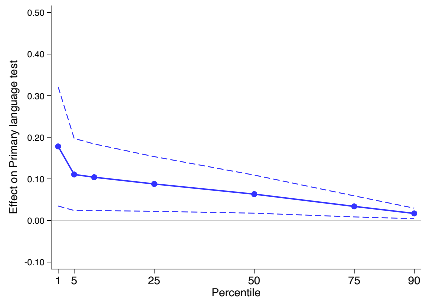
(g) High dropout - Secondary grades



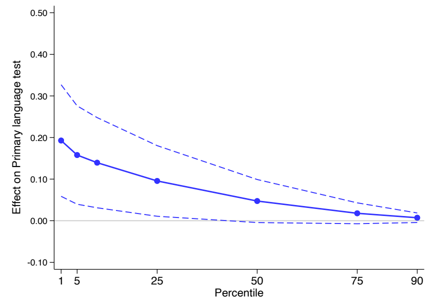
(h) Low dropout - Secondary grades

Notes: The figure plots Tobit estimates of medium-term effects on school performance by probability of dropping out, using data censored at the point indicated on the X-axis (i.e., values below the indicated percentile are assigned a value of zero).

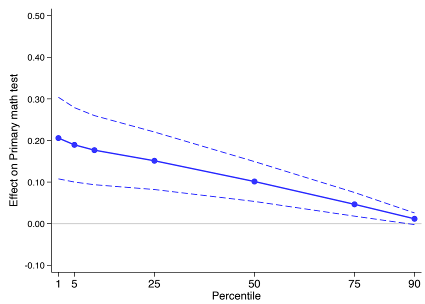
Figure 4: Tobit coefficients, medium-term effects on test scores, by probability of dropping out



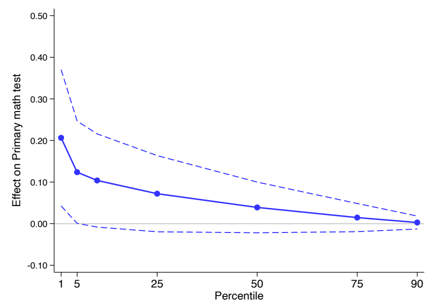
(a) High dropout - Primary language test score



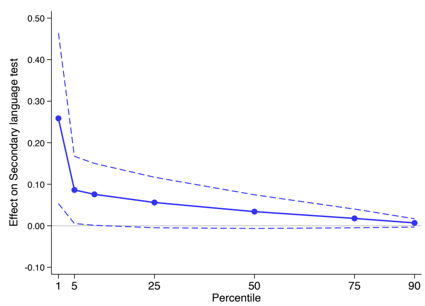
(b) Low dropout - Primary language test score



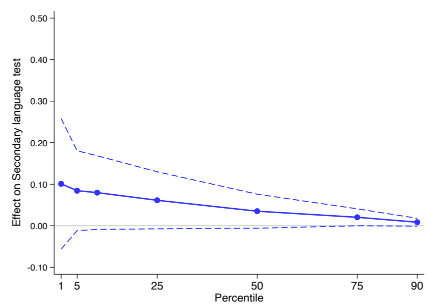
(c) High dropout - Secondary language test score



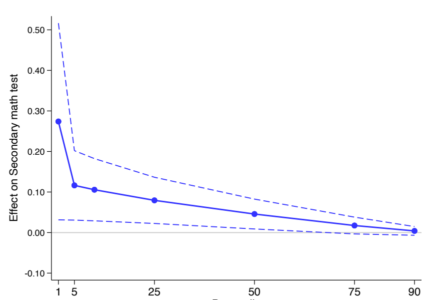
(d) Low dropout - Secondary language test score



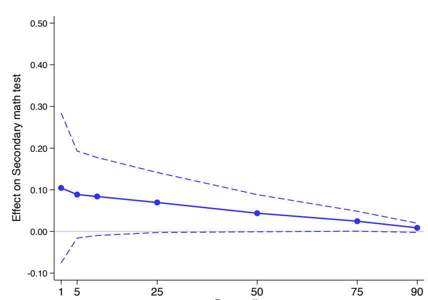
(e) High dropout - Primary math test score



(f) Low dropout - Primary math test score



(g) High dropout - Secondary math test score



(h) Low dropout - Secondary math test score

Notes: The figure plots Tobit estimates of medium-term effects on test scores by probability of dropping out, using data censored at the point indicated on the X-axis (i.e., values below the indicated percentile are assigned a value of zero).

Table A1: Sample restrictions and elegible schools

		FSP Restrictions			Availability	Randomization			
Region	County	Required schools	Required students	Dependence	Eligible schools	Treatment		Control	
						schools	units	schools	units
RM	Santiago	6	600	P or PS	13	9	5	5	3
	Estación Central	4	400	P or PS	14	7	2	7	3
	Lo Espejo	1	100	P	2	2	1	1	1
	Maipú	4	400	P	20	5	5	7	4
	La Florida	2	200	PS	15	4	2	5	3
	San Bernardo	1	100	P	24	2	1	4	2
	Total RM	18	1,800		88	29	16	29	16
VIII	Concepción	3	300	P or PS	8	5	3	4	3
	Coronel	1	100	P or PS	8	2	1	1	1
	Hualpén	2	200	P or PS	8	4	3	3	2
	Talcahuano	2	200	P or PS	5	5	2	3	2
	Total VIII	8	800		29	16	9	11	8

Notes: P: public school; PS: private subsidized school. As the number of schools that fit these size criteria was insufficient in some counties, we set an additional eligibility criteria that implied that if two schools were less than one kilometer away between them, and the sum of their fourth-graders was higher than 90, then that pair of schools could be included in the eligible set of schools too. The variable Units represents the number of groups that were grouped with this criteria.

Table A2: Balance between groups among attriters and non-attriters

	N	Treatment mean	Control mean	Difference
Non Attriters				
Grades 2009		5.80	5.80	0.01
		(0.57)	(0.57)	(0.05)
Attendance 2009		91.92	92.08	-0.16
		(6.36)	(6.84)	(0.62)
Attendance 2010		93.64	94.73	-1.09
		(5.99)	(5.74)	(0.79)
Observations	4,858			
Attriters				
Grades 2009		5.73	5.63	0.10
		(0.58)	(0.64)	(0.06)
Attendance 2009		90.01	88.66	1.35
		(7.59)	(9.84)	(0.98)
Attendance 2010		91.02	91.44	-0.42
		(8.79)	(9.50)	(1.40)
Observations	1,208			

Notes: Means and standard deviations for each variable are presented, along with differences between groups with their corresponding standard errors. Non Attriters are defined as those students that were present the days in which the tests were applied. Inversely, Attriters are defined as those students that were either absent the days of the tests or that had dropped school by then. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table A3: Prediction of probability of dropping out

	(1) 1(Dropout by 2017)
School drop-out rate	-0.913
	(0.970)
Student attendance	-0.00256***
	(0.000624)
Student grade	-0.169***
	(0.0123)
Share poor students	0.00674***
	(0.00158)
Female	0.0316**
	(0.0158)
N	2,961

Notes: The estimation follows the model used by Mineduc (MINEDUC, 2020) to predict the probability of dropping out from school by 2017 using the following independent variables: drop-out rates of the school attended in 2010, the share of poor students in the school attended in 2010, student grades in 2009, student attendance in the first half of 2010 (before the treatment started), and student gender. We estimated this model only for the control students. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table A4: First stage estimates: Effects on the share of sessions of the program attended

	(1)	(2)	(3)
	All Sample	High dropout	Low dropout
Treatment	0.525***	0.550***	0.509***
	(0.0238)	(0.0177)	(0.0304)
Constant	-0.0759	-0.0580	-0.128
	(0.0527)	(0.0587)	(0.179)
N	6,129	2,721	3,408
R-squared	0.638	0.705	0.606

Notes: The dependent variable is the share of sessions of the program attended by each student. All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. Estimates in column (2) consider students with estimated probability of dropping-out above the median of the sample and estimates in column (3) students with estimated probability of dropping-out below the median of the sample. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table A5: Medium-term effects on school performance: ToT estimates

	(1)	(2)	(3)
Dependent variable	Full sample	High dropout	Low dropout
Average grades primary	0.166***	0.240***	0.0981*
	(0.0529)	(0.0866)	(0.0549)
N	6,090	2,684	3,406
Average grades secondary	0.102	0.173	0.0168
	(0.0729)	(0.112)	(0.0773)
N	5,864	2,490	3,374
Attendance primary	1.546**	2.796**	0.143
	(0.707)	(1.182)	(0.637)
N	6,090	2,684	3,406
Attendance secondary	2.152*	3.988**	0.259
	(1.184)	(1.964)	(1.208)
N	5,864	2,490	3,374

Notes: All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. Estimates in column (2) consider students with estimated probability of dropping-out above the median of the sample and estimates in column (3) students with estimated probability of dropping-out below the median of the sample. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.

Table A6: Medium-term effects on test scores: ToT estimates

	(1)	(2)	(3)
Dependent variable	Full sample	High dropout	Low dropout
Primary language test score	0.0990	0.150*	0.102
	(0.0846)	(0.0812)	(0.119)
N	4,751	1,845	2,906
Primary math test score	0.141	0.315***	0.0392
	(0.0959)	(0.0942)	(0.123)
N	4,800	1,864	2,936
Secondary language test score	0.121	0.000539	0.170*
	(0.0777)	(0.123)	(0.0934)
N	3,348	1,044	2,304
Secondary math test score	0.161*	0.126	0.173*
	(0.0855)	(0.119)	(0.101)
N	3,326	1,047	2,279

Notes: All regressions include the student's baseline test score, her gender, average grades and attendance in 2009, and dummies for each stratum among which the program was randomized as controls. Estimates in column (2) consider students with estimated probability of dropping-out above the median of the sample and estimates in column (3) students with estimated probability of dropping-out below the median of the sample. Standard errors are clustered at the classroom level and presented in parentheses. *: Significant at 10%, **: Significant at 5%, ***: Significant at 1%.