NOVAFRICA
Working Paper Series

# Can Technology Improve the Classroom Experience in Primary Education? An African Experiment on a Worldwide Program

**Joana Cardim**
Universidade Nova de Lisboa and NOVAFRICA

**Teresa Molina-Millán**
Universidad de Alicante, NOVAFRICA and IZA

**Pedro C. Vicente**
Universidade Nova de Lisboa, NOVAFRICA, and BREAD

ISSN 2183-0843
Working Paper No 2101
January 2021

# Can Technology Improve the Classroom Experience in Primary Education? An African Experiment on a Worldwide Program *

Joana Cardim†, Teresa Molina-Millán‡, and Pedro C. Vicente§

January 2021

**Abstract**

Primary school coverage has been increasing in most developing countries. Yet, it has not been accompanied by significant improvements in learning indicators. We implemented a randomized experiment in Angola around the introduction of ProFuturo, a worldwide educational program. The program includes a Computer-assisted Learning (CAL) software directed at improving the regular classroom experience. One year after the program started, we find higher familiarity with technology. Teachers miss fewer days of classes and implement better teaching practices. Students become more interested in learning and pro-social. Finally, the program improves students' test scores in the most popular subject in the CAL platform.

**JEL codes:** O12, I21.

**Keywords:** Primary education, computer-assisted learning, CAL, field experiment, RCT,

*We thank seminar and conference participants at NEUDC, IFS, and NOVAFRICA for useful comments. We are grateful to Ana Margarida Araújo, Carine Canonici, Sara Queirós, Xaver Schenker, and Sara Vaz for outstanding field coordination and/or research assistance. We also convey a word of appreciation to all the enumerators we worked with in this project. We acknowledge funding from la Caixa Foundation. Ethics clearance was given by Universidade Nova de Lisboa. All remaining errors are our responsibility.

†Nova School of Business and Economics and NOVAFRICA. Email: joana.cardim@novasbe.pt.

‡Universidad de Alicante, NOVAFRICA, and IZA. Email: teresa.molina@ua.es.

§Nova School of Business and Economics, NOVAFRICA, and BREAD. Email: pedro.vicente@novasbe.pt.

Africa, Angola.

# 1 Introduction

Human capital is widely considered to be vital for growth and human development. This shared belief is illustrated by the fact that, after the Millennium Development Goal of achieving universal primary education, ensuring inclusive and equitable education quality became one of the most prominent Sustainable Development Goals of the United Nations. Indeed, policy makers of developing countries, NGOs, and international institutions have been primarily focused on increasing school enrollment and attendance. However, learning outcomes are still very low in many developing countries. In these settings the main challenge is now to find effective ways to improve education quality, despite the limited availability of skills in the population (Hanushek and Woessmann, 2008). One possibility is to gear the incentives of teachers and parents/students for higher school attendance and better learning. A large literature has devoted attention to incentives schemes, e.g., Duflo, Hanna, and Ryan (2012); Molina-Millán et al. (2019). But the role of school inputs in the classroom should not be disregarded.

Relating to school inputs, it is important to note that recently expanded school systems in developing countries are often focused on elite students. Classical school inputs, such as textbooks, could then be the wrong inputs to help strengthening education quality, e.g., Glewwe, Kremer, and Moulin (2009). Targeting more and better school inputs to those students lagging behind could be the way to go, namely in settings where student heterogeneity is significant (Banerjee et al., 2007). At the same time, computer technology enabled pedagogical methods to be more tailored to the specific needs of individual students. Many programs have been trying to improve education quality in developing countries through the use of technology. However, experimental evidence from various parts of the world reviewed by Escueta et al. (2017) proves that simply providing schools with computers is unlikely to improve children's academic performance. Differently, computer hardware combined with Computer-assisted Learning (CAL) software has demonstrated clear positive impacts on student achievement, in particular when individual customization of contents is possible, e.g., Muralidharan, Singh, and Ganimian (2019).

Important questions do however remain about how to link these CAL programs to regular teachers and classroom dynamics. In this paper, we assess the impact of the introduction of ProFuturo, an innovative CAL program present in 38 countries of three continents, which has already reached more than 450,000 teachers and 12 million children. Importantly, the package provided by the ProFuturo program enables tailoring contents to individual student needs. However, two innovative features of the program stand out. First, its main objective is to help regular primary school teachers in their core teaching activities in the classroom, i.e., after significant teacher training, ProFuturo places teachers at the center of the program's experience. Second, it promotes student interaction with teachers and peers, namely through group work. Apart from the learning software, ProFuturo includes an equipment set composed by individual tablets for all students in a classroom. Since the objective of the program is to familiarize students and teachers with technology while motivating and improving not only teachers' teaching but also students' learning experience in all major contents of regular primary school, we ask in this paper whether the program is effective in that regard. We also investigate the effects of the implementation of ProFuturo on children's cognitive skills.

The context of our study is the city of Luanda, Angola.[1] Angola is a suitable illustration of an expanded but low-performing school system. The Gross Enrollment Rate in Angola's primary education grew from an estimated 71% in 1998 to 113% in 2015. However, in 2014 its youth literacy rate was close to the average of Sub-Saharan African countries, at 77% (World Bank, World Development Indicators, 2019). Sizeable heterogeneity in students' abilities within each classroom is common. Hence, it is clear that the main challenge today in Angola, like in other African countries, is to increase

[1] Angola is a low-income country with 30.8 million inhabitants in 2018. It has maintained political stability since the end of the 27-year civil war in 2002. However, the country still faces substantial development challenges stemming from high dependency on oil and very high poverty rates.

the quality of education.

We implemented a randomized field experiment to test the impact of ProFuturo. From the 42 primary schools in Luanda that were selected to receive ProFuturo, 21 were randomized to receive it in the beginning of 2018 and 21 assigned to start using it in 2019. We employ a comprehensive set of measurements, including detailed school principal, teacher, student, and household surveys, three rounds of class systematic observation, student standardized tests in Mathematics, Portuguese, and Science, as well as administrative data from schools and the program.

Approximately a year after the program was introduced, We find that it increases the levels of technology usage for both teachers and students. Importantly, teachers' motivation improves with ProFuturo. Specifically, the program led to a reduction in teachers' absenteeism. The point estimate is large when employing administrative data: less 0.69 standard deviations in days missed by teachers. Students also become more motivated, namely towards Mathematics. Central to the program and to our analysis are effects on teacher and student behaviors. Teachers improve the quality of their class preparation and engage more often in active teaching in the classroom. Students devote more time to reading, and to shared time with their guardians using technology. Some evidence also suggests more pro-social interaction between students. Finally, we observe in classrooms higher standards of teacher knowledge. These effects of ProFuturo translate into improved students' test scores in Science, which was the subject most frequently selected to be taught under ProFuturo. We do not find any treatment effects in other subjects.

Our paper relates to the literature on the use of computer technology for educational purposes. The literature has shown that delivering computer hardware to schools has not led to clear improvements in test scores. This literature includes non-experimental but plausibly-identified causal effects: Angrist and Lavy (2002) show impacts of a lottery program in Israel, and Machin, McNally, and Silva (2007) analyze the impact of ICT funding for English schools. Barrera-Osorio and Linden (2009) show null results for a two-year RCT in Colombia. A critical determinant of the extent to which these inputs give rise to better learning outcomes could be how they interact with teachers and pedagogy in the classroom.

Similar null results are found for the One Laptop per Child program (OLPC), which allocates computers to students one-to-one, with an emphasis on home use. Beuermann et al. (2015) and Cristia et al. (2017) provide comprehensive supporting evidence from OLPC experiments in Peru in both the short and longer run, as well as rural and urban settings. A large RCT in the US also finds precise but null effects of access to computers at home on educational outcomes (Fairlie and Robinson, 2013). Employing a regression-discontinuity design, Malamud and Pop-Eleches (2011) find negative effects on school grades of a voucher program in Romania to purchase computers, despite improvements in computer skills.[2]

One pedagogical approach that seems to have a role in student achievement is Computer Assisted Learning (CAL). This includes delivering hardware to students along with a specific software designed to develop particular skills (e.g., improving Mathematics). In some cases, this software may adapt and respond to students' learning needs. This approach may be particularly relevant as schools in developing countries typically have more than a grade per class, a large student-teacher ratio, and highly heterogeneous students. There are promising results of this type of interventions.

Banerjee et al. (2007) evaluate a CAL program in India that consisted of Math games whose level of difficulty responded to students' ability to solve them. These authors find large gains in test scores at the end of one and two years of the program for all types of students. Consistently, for a sample of American schools, Barrow, Markman, and Rouse (2009) find significant improvements in pre-algebra and algebra skills after a targeted CAL program was implemented. These authors hypothesize that these effects arise from increased individualized instruction as impacts appear larger for students in larger classes. Indeed, it could be that the individual tailoring of CAL programs is particularly important for their effectiveness. A series of experimental evaluations of CAL in China have found modest positive impacts on learning. A likely explanation is that the evaluated program

[2]The exception in this literature is the experimental evaluation of Mo et al. (2013) which finds improvements in Maths scores of migrant students in Beijing.

employed technology that did not feature extensive individual customization (Yang et al., 2013; Lai et al., 2013; Mo et al., 2014; Lai et al., 2015).

Muralidharan, Singh, and Ganimian (2019) evaluate a CAL program (Mindspark) explicitly designed to customize pedagogy to the right level of students in grades 6 to 9 in New Delhi, India. The program included an initial screening test and an algorithm that constantly updated its information about the student. It used a CAL system able to finely calibrate students' competencies and tailor academic content to the corresponding level. The study finds high impact in Maths and Language skills, with larger gains for the academically weaker students.

On a comprehensive evidence-based review of the impact of education technology, Escueta et al. (2017) identify that the majority of programs that have positive results focused on improving Maths outcomes instead of Language. Some specific examples are Rouse and Krueger (2004) for an early American program focusing on Language, and Carrillo, Onofa, and Ponce (2011) for a more recent program in Ecuador with a large contrast between Maths and Language. The authors of the review study also claim that the channel through which CAL is expected to improve learning the most is by tailoring education to the needs of the students and by providing immediate feedback to students.

One can argue that programs such as Mindspark may act as substitutes and not complements to teachers, hindering the potential benefits that students may attain from student-teacher interactions. Escueta et al. (2017) claims that little is known about how CAL programs interact with teachers' efforts. Beg et al. (2019) studied a program in Pakistan that consisted of video lectures together with some suggestions of activities for the teacher to implement in the classroom. The authors found that student achievement in Maths and Science tests increased after 4 months of exposure to the program. They found some evidence of complementarity between school inputs: teachers using the program increased attendance and spent more time preparing the lessons. In this case, however, the program did not include individual student customization. Another critique to CAL programs is related to the possibility of decreasing interaction among peers. Araya et al. (2019) experimentally evaluate an innovative technology program that uses gamification to increase Maths learning in low-performing primary schools in Chile. Students improved learning outcomes but the program increased Maths anxiety and reduced students' preferences towards teamwork.

We contribute to this literature by studying the impact of a highly innovative program that combines three important components of successful CAL programs: instruction tailored to students' needs, the inclusion of teachers as the main intermediaries in the implementation of the CAL - i.e., securing substantial interaction between teachers and students -, and the possibility of undertaking activities in student groups through the CAL package. We also focus on a developing setting with implications for where in the world the most acute shortcomings in education quality are present.

The remainder of this paper is organized as follows. Section 2 provides the context of our experiment and the details of the program. Section 3 describes the design of the experiment, including of the measurement. Section 4 explains the hypotheses and the identification strategy. In Section 5 we present and discuss the results. Finally, we conclude.

## 2 The ProFuturo program

The CAL program that we study in this paper, ProFuturo, is an international program sponsored by Fundación la Caixa and Fundación Telefónica in Spain. It is one of the largest digital education initiatives in the world: it aims to target 25 million children in vulnerable areas by 2030. By the end of 2017, 5.8 million children were beneficiaries of the program. Today, ProFuturo is present in 38 countries in Latin America, the Caribbean, Africa, and Asia, having trained more than 450,000 teachers and benefited 12 million children.

The main objective of ProFuturo is to 'bridge the education gap in the world by providing quality, digital education to children from vulnerable environments.' To reach this goal, ProFuturo aims to improve teachers' expertise both at the technological and pedagogical levels, and to promote learning by students through increased motivation, as well as improved quality in classroom dynamics. At the center of ProFuturo is the idea of a CAL tool that intends to be a complement, and not a substitute for teachers in the classroom. ProFuturo takes teachers as 'the main learning activator.'

The program includes the distribution of suitcases which include tablets, a computer for the teacher, and a projector. Each suitcase suits roughly one classroom, with enough tablets for all students. This technology may be easily implemented in the context of a developing country, where there is limited access to electricity, as batteries last for nine hours after being charged for four hours and its software runs only offline. Each tablet is equipped with a software, produced on purpose for ProFuturo. The software contents are engaging and interactive and were approved by several educational partners, such as UNESCO and Instituto Cervantes (Spain). They are adapted for different countries in terms of language and cultural references.

While focusing on core educational contents at the primary level employing international standards, the contents of the software package include lectures on Language (Portuguese), Mathematics, and Science. Other more specific types of contents are also included, namely on Technology and on Ethics and Citizenship (including contents on social cohesion, the relationship with others and the community, rules of conduct in school, among other topics). There are activities available at the end of each module to test students on what they learned. These activities give immediate feedback to students – praising them when they give a correct answer and telling them to try again if they make mistakes. The platform also allows teachers to have access to the progress of students by giving them reports about students' performance in activities performed within ProFuturo. Teachers can then customize contents to be used according to the needs of individual students. They can also create their own contents, which become available together with the full repository of ProFuturo for didactic contents.

Before implementation, ProFuturo trains school principals and teachers in the schools where the program is introduced. Teachers are trained according to their perceived level of computer proficiency by ProFuturo coordinators. Each training module directed to principals typically lasts 5 hours; modules directed to teachers have the duration of around 20 hours.

ProFuturo has been present in Angola since the end of 2015, in close link with the Catholic Church and the Ministry of Education of Angola. It was first introduced in Luena, Moxico province. At the end of our project, it was present in 124 schools, involving 1,208 teachers and around 100,000 children in primary schools.[3] These schools typically serve children from disadvantaged socioeconomic neighborhoods. ProFuturo included at this point in time 22 full-time coordinators in Angola.

We present more details on the ProFuturo program and its adoption in Luanda at the time of our study in Section A of the Appendix. It is important to note that Science contents were the ones most frequently selected by teachers using the ProFuturo platform. Administrative data provided by the ProFuturo platform reports that 40-44 percent of all activities performed using ProFuturo in 2018 and 2019 were in Science, followed by Portuguese (23-24 percent) and Mathematics (16 percent).

# 3 Experimental design

In the end of 2017, ProFuturo selected 42 Catholic schools in Luanda to be included in this study.[4] The randomization procedure for the allocation of schools to treatment was implemented following a stratified clustered design. After schools were paired based on region and school characteristics,[5] half of them were randomly allocated to receive the ProFuturo program immediately (21 schools), with the other half assigned to a control group which was promised to receive the program following the end of the impact evaluation project (approximately a year after the treatment group).

[3]The schools included in the program were at this point located in some of the main cities of Angola: Luanda, Caxito, Viana, Benguela, Huambo, Moxico, Malange, Uige, Dundo, and Saurimo.

[4]See figure A1 for a map with the geographic distribution of selected schools in Luanda.

[5]The employed school characteristics included: number of students, number of teachers, number of classrooms, average number of students per class, maximum number of students per class, school access to electricity, safety from crime within the school, school access to internet, indicators of school infrastructure, and school staff knowledge of information technology.

### 3.1 Measurement

The structure of measurements in this project included: (i) baseline and endline surveys at the school principal, teacher, student, and student's guardian levels; (ii) student cognitive tests; (iii) classroom observation activities; (iv) administrative data from schools as well as from Profuturo on students' and teachers' use of ProFuturo's software package. Figure A7 in Appendix B depicts the timeline of the measurements.

*Surveys:* The surveys we designed and conducted included face-to-face submission of questionnaires to all school principals and to all teachers working in each school at the time of the interview, as well as to a random sample of students and their caregivers. This data collection effort started before the beginning of the intervention for the baseline surveys, from December 2017 to March 2018, and for the endline surveys, 11 months later, from November 2018 to May 2019.

The survey questionnaires targeting school principals included questions on their demographic and socioeconomic characteristics. They also included a module on school management. The survey questionnaires targeting teachers were analogous. Beyond demographic and socioeconomic questions, and importantly for our analysis, they included questions on use of technology, motivation and attitudes towards teaching, as well as time allocation. We also collected self-reported information on teacher absenteeism.

We randomly sampled 60 students in the 4th, 5th, and 6th grades in each school, stratified by class, for the students' and guardians' surveys. The survey questionnaires targeting students included questions on their ability to use technology, their motivation towards the school, as well towards learning in general and in specific subjects. Students were also asked to report their absenteeism, and they were subject to a test on their executive function, in particular a forward and backward digit span test, in which the respondent is asked to repeat a series of numbers read to him/her (Engle, 2002). The questionnaires directed to students' guardians included questions about demographic and socioeconomic characteristics of the student's household. They also included questions regarding the guardian's perception of his/her child's motivation, satisfaction about the school, learning, as well as time allocation when he/she is not in school.

*Classroom observation:* Our systematic classroom observation involved the random selection of five classes within each school from the 4th, 5th, and 6th grades to be observed. We implemented a classroom observation questionnaire, to be answered by enumerators observing the delivery of classes by teachers in the classroom. This was constructed using the 'Stallings Classroom Snapshot Instrument.' The Stallings instrument generates quantitative data describing the activities performed by the teacher during the class and the type of interaction between the teacher and students in the classroom. The classroom observation effort entails the enumerator coding ten different snapshots, using regular intervals of time, during each observed class. During the submission of the questionnaire, the enumerators were asked to write down a brief description of the activity before coding each activity according to the instrument. To reduce possible subjectivity on observations, we deployed two enumerators per classroom. At the end of each class observation, enumerators reported their overall perceptions about the class, registering aspects such as perceived teachers' mastery of the contents being taught. We included three rounds of classroom observation in our measurement design, one at the baseline and two after approximately a year had passed from the introduction of ProFuturo, with a few months in between.

*Cognitive tests:* As part of the implementation of the teachers' surveys, we included an assessment of teachers' cognitive skills based on the Survey of Adult Skills (PIAAC) - an international test developed by the OECD, and adapted to the context of a developing country. The test consisted of a reading comprehension question about a text in Portuguese and three Maths questions. In addition, we submitted cognitive tests to students. We randomly selected six classes from grades 4, 5, and 6 within each school to be submitted these tests. All tests were constructed based on the materials of the learning initiative Uwezo. This is a platform that conducts annual, large-scale, citizen-led, household-based assessments that measure actual levels of children's literacy and numeracy. The platform is dedicated to Kenya, Tanzania, and Uganda. The curricula of Portuguese, Mathematics, and Science for each grade were also taken into account. The tests begin with relatively easy questions and gradually become more difficult.

*Administrative data:* We collected administrative data from each school on various dimensions of school operation.[6] Since some of the data we accessed were incomplete, we will dedicate particular attention in the analysis that follows to the most complete dimensions of school operation, which included the data on teacher absenteeism. Data from ProFuturo's software platform are also available to complement data collected in the field. However, it is not possible to link these data to individual data collected for the students (surveys and test scores) due to no name identifiers being available in the ProFuturo platform for data protection reasons. However, we use these data to formulate a better understanding of the main activities and subjects studied through the platform.[7]

All main outcome questions employed in our study are fully described in Section C of the Appendix. We organize them by families of outcomes. We begin with use of technology. We then list the outcomes we employ to describe the levels of motivation by teachers, students, and guardians. The time allocation of teachers, and the observed class delivery are next. Subsequently, we take students' time allocation, behavior, and interactions at school. We finally depict the variables we employ for cognitive skills of both teachers and students.

# 4 Hypotheses and estimation strategy

Our main hypothesis is that ProFuturo improves the classroom experience of students in primary education leading to an improvement in their skills.[8] Various channels for these effects are at stake. These are described as follows.

**Hypothesis 1:** The program increases the use of technology for teachers and

[6] These included the number of teachers and students in each school, the number of classes and number of children per class, attendance of teachers and students, dropout and enrollment rates of students, national exam grades (at the time of the 6th grade), internal school grades, teacher evaluations, and students' discipline.

[7] In addition, we implemented three behavioral activities to assess non-cognitive behaviors. These measured: (i) children's motivation with school and learning; (ii) altruism and pro-social behavior of children; and (iii) teachers' motivation towards the school and teaching. However, we encountered many problems when implementing these activities, and we found out that participation of students and teachers was not completely voluntary in some schools. Therefore, we do not include these measures in the analysis.

[8] For details on the proposed research design prior to data access, see the research proposal available at Research proposal, June 2017.

students.

**Hypothesis 2:** The program leads to an increase in the motivation levels of teachers and students, which includes reducing absenteeism.

**Hypothesis 3:** The program leads to higher quality teaching by teachers, including better preparation and delivery of classes.

**Hypothesis 4:** The program leads to a better learning experience by students, including more time devoted to learning at home and more effective social interactions at school.

**Hypothesis 5:** The program improves the cognitive skills of teachers and students.

We estimate the Intent-to-treat (ITT) effect of ProFuturo on our broad set of outcome variables. The basic specification of the model is:

$$y_{is} = \alpha + \beta \text{ProFuturo}_s + \delta y_{is0} + X_{is}'\gamma + \epsilon_{is} \tag{1}$$

where $y_{is}$ is the outcome of interest for individual $i$ in school $s$, measured at the endline. Note that individual $i$ can be a teacher, a student, or a student's guardian. The variable $\text{ProFuturo}_s$ is a treatment indicator taking value 1 for schools which were assigned to receive ProFuturo and 0 otherwise. $X_{is}$ is a set of individual characteristics including strata fixed effects, for either teachers, students, and/or students' guardians depending on the outcome at stake.[9] $y_{is0}$ is the baseline value of the dependent variable, and $\epsilon_i s$ is an idiosyncratic error term. To account for possible correlation in outcomes within schools, the error term is clustered at the school level. If the auto-correlation of the outcome variable is low, which is the case for most survey outcomes, this specification maximizes statistical power in field experiments (McKenzie, 2012).

Given that baseline values of the outcome variable are not available for classroom

[9]Control variables are as follows. When analyzing outcome variables at the level of the teacher: gender, age, age squared, education, house ownership, index of assets, an indicator variable on whether the house has piped water, and an index on the use of technology. When analyzing outcomes at the level of the student and the guardian: students' gender, age, and grade; an indicator variable on whether the student has attended kindergarten, an indicator variable on whether the student has failed at least one course in the past, and respondent's month of interview. When analyzing outcome variables at the level of the guardian, in addition to the above: guardian's gender, age, age squared, and education; number of members of the household, house ownership, index of assets, and an indicator variable on whether the house has piped water.

observation data, we employ the following specification when employing the referred type of data:

$$y_{cs} = \alpha + \beta \text{ProFuturo}_s + X'_{cs}\gamma + \epsilon_{cs} \tag{2}$$

where $y_{cs}$ is the outcome of interest for classroom $c$ in school $s$. $X_{cs}$ is a set of classroom characteristics including strata fixed effects, grade, and month of observation.

To assess whether the relatively small number of clusters (schools) biases any of the results in terms of statistical confidence, we follow Athey and Imbens (2017), as well as Young (2019), and replicate all our hypothesis tests using randomization-based inference tests. In randomization-based inference, uncertainty in estimates arises naturally from the random assignment of the treatment, rather than from hypothesized sampling from a large population. This method allows estimating the exact p-value under the sharp null hypothesis that the treatment effect is null, by calculating all possible realizations of a test statistic and rejecting if the observed realization in the experiment itself is extreme enough (see Heß (2017)).

In the results section we also check whether the main results of the paper are robust to using the Post-Double Selection LASSO procedure to select control variables.

# 5 Results

## 5.1 Descriptive statistics, balance, and attrition

We now turn to describing the baseline characteristics of teachers, students, and students' guardians. In the process, we also provide an assessment of balance between treatment and control groups for the referred traits.

The top panel of Table D1 in Appendix D shows descriptive statistics on school teachers' baseline characteristics. Forty percent of teachers in the control group are women They are on average 37 years old, around half of them are married, and 74 percent have children. There are no significant differences between the control and treatment group in terms of teachers' characteristics, including teaching experience and asset ownership. The exception is that teachers in the treatment group are 11 percentage points more likely to have completed university studies, which is significant at the 1 percent level.

The middle panel of Table D1 shows descriptive statistics on students' baseline characteristics. 53 percent of the sample is female and the average student in the control group is 11 years old. Both age and gender are balanced across comparison groups. However, we observe differences across students in two dimensions, pointing in the same direction. First, students from the treatment group are 3 percentage points less likely to have attended kindergarten. They are also 4 percentage points more likely to have failed at least one course in the past. Both differences are significant at the 10 percent level. There are not significant differences with respect to students' baseline test scores.[10]

The bottom panel of Table D1 shows descriptive statistics of students' guardians at the baseline. In both treatment and control groups, around 56 percent of the interviewed caregivers are women. On average guardians in the control group are 39 years old, while guardians in the treatment group are on average 1.2 years younger. This difference is significant at the 5 percent level. Guardians from the treatment group are 5 percentage points less likely to have completed university studies, which is significant at the 5 percent level. In term of households' wealth, on average 60 percent of the caregivers own a house and 38 percent of them have piped water, with no significant differences between groups. There are no significant differences with respect to IT goods ownership.[11]

Overall, we can conclude that our randomization procedure was able to identify comparable groups, namely in terms of demographic characteristics. Nevertheless, we found some statistically significant differences between individuals in treatment and con-

[10]Figure A1 in the Appendix shows the distribution of Mathematics and Portuguese test scores by grade. The figure confirms that the tests in both fields of assessment performed well in capturing a wide range of achievement within grades and across grades. As expected, the distributions move from the right to the left as students' grades increase.

[11]Note that sample averages for the guardians interviewed in the baseline survey should be taken with caution. We managed to interview 1,087 caregivers from a targeted sample of 2,520. Thus, it is possible that the averages presented are not representative of the study population: guardians that were interviewed are likely to be more involved with the school. However, our analysis suggests that this sample selection did not translate into systematic differences between treatment and control.

trol schools. To take these imbalances into account we apply Inverse Probability of Treatment Weighting (IPTW) to the main results (see Section H in the Appendix).

A note is also due on survey attrition. Table E1 in Appendix E presents response rates in the teachers', students', and guardians' surveys, when comparing baseline to endline. In the endline, attrition in the teachers' survey was 13 percent when considering the total number of interviews and 38 percent when considering the specific teachers interviewed at the baseline (control group). In the students' survey, it was 7 percent when considering either the total or the panel (control group). Attrition in the guardians' survey was larger. We interviewed 74 percent of the guardians in the endline, and we have panel data reported by the guardians for 44 percent of the baseline sample. Overall, attrition rates are not significantly different between the treatment and the control groups. Appendix Table A1 shows balance tests on baseline characteristics for respondents surveyed at the endline. Overall, the endline samples of teachers, students, and guardians are very similar to the corresponding samples at baseline.

## 5.2 Treatment effects

### 5.2.1 Use of technology

We now turn to our analysis of treatment effects. We begin our analysis of outcome variables with measures of familiarity with technology for teachers and students. These are shown in Table 1. Column 1 is dedicated to teachers' use of technology during the lectures. Columns (2) to (4) are dedicated to students' outcomes. In particular, we analyze students' self-reported ability in performing sets of basic and advanced activities employing a computer, and their desire to use more technology at school. Column 5 is devoted to students' use of technology at home, as reported by their guardians.

Table 1 shows that teachers in treated schools are more likely to employ computers during their lectures than teachers in the control group. The magnitude of the effect is 0.09 standard deviations, significant at the 1 percent level. The exact p-value from randomization inference is lower than 1 percent. Students in treated schools report being able to perform more basic activities (turn on/off a computer, write, and open/close programs and applications) than students in the control group. The size of this effect is 0.09 standard deviations, significant at the 10 percent level and not significant using randomization inference. We do not find any effect of ProFuturo on the advanced use of technology, i.e., including performing searches in internet, saving a file, and printing of documents. Students in treated schools are more likely to report that they would like to use more technology at school. The size of the effect is 0.03 standard deviations and it is significant at the 1 percent level. The exact p-value is 0.01. Guardians also report an increase on the use of technology at home by their children. This effect is statistically significant at the 1 percent level and its size is 0.23 standard deviations. The randomization inference p-value is 0.04.

We validate Hypothesis 1, as we find evidence that both students and teachers in ProFuturo schools are more likely to use technology, not only at school, as is the case for teachers, but also at home, as is the case for students.

#### 5.2.2 Motivation and absenteeism

We now turn to measures of teacher and student motivation towards the school, including absenteeism. Table 2 depicts results related to teachers. We employ outcomes from the schools' administrative data, from the teachers' survey, from the guardians' survey, and from classroom observations. Specifically, we analyze the number of days the teacher was absent (from administrative school records and self-reported), the number of days the teacher arrived late to school (self-reported), and teacher motivation (from the teachers' survey, the guardians' survey, and the classroom observations' data).

We find clear effects of the treatment on absenteeism from the administrative data both for the month prior to the survey and for the full academic year. Both self-reported survey data and school administrative data yield treatment effects on missing less days of classes. There are however no significant effects on being late when arriving at school as reported by teachers. The effect of ProFuturo on teachers' absenteeism derived when employing the administrative data is large: the program reduces the number of days the teachers missed in the previous month by 0.69 standard deviations. This result is statistically significant at the 1 percent level. When employing randomization inference, it becomes significant at the 5 percent level. We also observe that ProFuturo reduces the number of days that teachers were absent during the complete school year by 0.4 standard deviations, which is significant at the 1 percent level (also when employing randomization inference). Turning to the teachers' surveys, ProFuturo reduces the number of days that teachers reported to be absent by 0.1 standard deviations, which is significant at the 5 percent level and marginally insignificant using randomization inference. There are no significant differences in self-reported levels of motivation or on the levels of teacher motivation reported by guardians. There is a positive treatment effect of 0.23 standard deviations on the motivation of teachers reported by the enumerators that observed classroom teaching in the final round. However, this is significant at the 10 percent level, and not significant using randomization inference p-values.

Table 3 presents results related to students' and guardians' motivation and their attitudes towards the school using data from the students' and the guardians' surveys, as well as classroom observations. Columns (1) to (3) are dedicated to students' survey data. Specifically, we analyze whether students report a positive attitude towards the school, whether they like Mathematics, and whether they like reading. Column (4) is dedicated to guardians' survey data, namely on guardians' reports about school satisfaction. Column (5) is dedicated to students' survey data, in particular to the number of days the student missed school. Columns (6) and (7) are dedicated to data collected through classroom observations.

Students in treatment schools report they like Mathematics more than in control schools. ProFuturo increased the likelihood students report liking Mathematics by 0.02 standard deviations, which is statistically significant at the 5 percent level (at the 10 percent level using exact p-values from randomization inference). We do not find significant treatment effects on attitudes of students towards school or reading. Turning to the guardians' survey, we find a positive impact, of 0.13 standard deviations, on reported overall school satisfaction. This effect is significant at the 10 percent level but insignificant when employing randomization inference. We do not find any treatment effect on students' absenteeism, i.e., on the number of school days missed. Finally, when analyzing classroom observation data, namely from enumerators assessments of students' motivation, we do not report any significant effects of ProFuturo.

We conclude that ProFuturo was effective at decreasing the absenteeism of teachers and increasing their motivation, consistently with Hypothesis 2. The effects on absenteeism are particularly clear. Regarding students, we find that the program has some effects on increasing their motivation, namely on having a more positive attitude towards Mathematics.

#### 5.2.3 Teaching

We turn now to analyzing the quality of teachers' preparation of the classes they teach, their broader time allocation, as well as their behavior in the classroom. Table 4 depicts treatment effects on a self-reported teachers' classroom preparation index as well as teachers' self-reported allocation of time in a regular working week by activity. Specifically, the classroom preparation index includes information about whether the teacher has a plan of the subjects to teach during the academic year, a book of class registries and summaries, and a notebook in which they prepare classes. We also analyze teachers' allocation of hours to teaching, planning teaching activities in school, planning teaching activities at home, and undertaking administrative tasks.

Related to teachers' planning, we find that teachers in ProFuturo schools become more careful about preparing their classes. The magnitude of this effect is 0.12 standard deviation units, significant at the 5 percent level. The exact p-value is, however, marginally above standard levels of statistical significance. For a regular week, teachers report to spend more time teaching and less time planning teaching activities at school. The effect magnitudes are 0.04 standard deviations and -0.06 standard deviations, respectively, both significant at the 10 percent level, although not significant with randomization inference. There are no statistical significant differences in time planning at home and undertaking administrative activities.

Table 5 shows results from the second and third rounds of the classroom observation data. Specifically, columns (1) to (3) show results on teacher allocation of time during class. We analyze teachers' allocation of time in the classroom to reading, instruction, and discussion, to practice and drill, and to monitoring. Column 4 is dedicated to knowledge of the subject shown by teachers during the classes as observed by enumerators.

In the second round of observations, there are no significant treatment effects except for a positive effect on time allocated to monitoring. Teachers in the treatment group allocate 6.5 percentage points more time to that activity, which is significant at the 5 percent level (the exact p-value from randomization inference is 0.108). In the third round of classroom observations, we find clear treatment effects on expanding time allocated to practice and drill, and on increasing knowledge shown during the classroom observation. The magnitude of the effects are 1.2 percentage points and 0.35 standard deviations respectively, both significant at the 5 percent level. The corresponding exact p-values are 0.117 and 0.058.

We conclude that ProFuturo induced some changes in line with Hypothesis 3. Specifically, we find higher quality in teachers' preparation of their classes, which translates into more time devoted to teaching as reported by teachers. Looking at observed classroom activities, we find short-term effects on increasing passive teaching (e.g., monitoring), perhaps as compensation for classes employing ProFuturo teaching.[12] However, these short-term effects do not last and are substituted by positive effects on active teaching, namely through practice and drill, and on improved knowledge of the subject taught in the last round of class observations. [13]

[12]We could observe in ProFuturo classes that there was no time devoted to copying activities (see Table A2 in Appendix A). Indeed, outside ProFuturo classes, teachers could be giving students some material for them to take home, given that they do not take anything home after ProFuturo classes.

[13]Table F1 in Appendix F shows results on teacher allocation of time during class observation broken down by activity. Specifically, we analyze teachers' allocation of time in the classroom to reading, instruction, discussion, practice and drill, monitoring seatwork, monitoring copying, disciplining, managing the classroom, being off-task, and being absent. In addition to the results shown in Table 5, the only treatment effect is on increasing time allocated to managing the classroom in the third round. The magnitude of the effects is 3.8 percentage points, significant at the 5 percent level. The exact p-values is 0.064.

### 5.2.4 Students' time, behavior, and interactions

We now consider students' allocation of time to various activities at home as reported by guardians interviewed in the corresponding endline survey, as well as students' behavior and interactions with their teachers and peers at school. Table 6 shows results on students' allocation of time. We analyze in particular the time allocated by students to reading, studying, and playing in a regular week. We also devote attention to how the shared time between guardians and their children is spent, namely on activities using technology and on studying.

Students in the ProFuturo schools spend 0.23 standard deviations more time reading and 0.15 standard deviations more time playing. Results are significant at the 1 percent and 5 percent levels, respectively. The corresponding randomization inference p-values are 0.03 and 0.05. Regarding the time treated guardians and their children spend together, it is more likely to be spent using technology, by additional 0.11 standard deviations, when compared to control individuals. This result is significant at the 10 percent level but not significant using exact p-values. There are no differences regarding time devoted to studying (by students alone or together with their guardians).

In Table 7 we report our results on students' behavior and interactions with their teachers and peers at school, while employing data from students' surveys. Specifically, we analyze a measure of students' self-reported altruism, students' perceptions about the level of collaboration among themselves, and their beliefs about whether they have many friends at school.

We observe a positive treatment effect of 0.03 standard deviations, significant at the 1 percent level, on the measure of students' altruism. This result is significant at the 10 percent level using randomization inference. Students from treatment schools are 0.02 percentage points more likely to have received help from other students, which is statistically significant at the 10 percent level, but not significant when looking at exact p-values from randomization inference. This effect does not translate to having more friends at school.

Overall, we find some consistent patterns with hypothesis 4. Our findings imply positive effects of ProFuturo on time devoted to reading by students and on shared time with their guardians using technology, possibly playing games. They are also suggestive that ProFuturo induces some movement towards pro-social interaction between students.

#### 5.2.5 Cognitive skills

In Tables 8-9, we show treatment effects on outcomes related to teachers and students' cognitive skills. In the case of Table 8, we analyze treatment effects on the performance of teachers in standardized test scores following PIAAC (columns 1 and 2) and on their knowledge self-assessment (columns 3-6).

We find that teachers in the treatment group have lower performance in standardized test scores assessing knowledge of Portuguese. The magnitude of this effect is -0.15 standard deviations, statistically significant at the 5 percent level, and significant at 10 percent level using exact p-values from randomization inference. A possible interpretation for this counter-intuitive result is related to the fact that the ProFuturo platform was set up in Brazilian Portuguese which is quite different from the Portuguese commonly used in Angola. We do not find significant treatment effects on teachers' performance in standardized test scores assessing their knowledge of Mathematics. In terms of self-assessment, we see no effects regarding overall self-assessment or self-assessments on Portuguese and Mathematics, but there is a positive treatment effect of 0.13 standard deviations on the self-assessment of knowledge in Science, significant at the 5 percent level, although marginally insignificant when employing randomization inference.

Table 9 is dedicated to the analysis of treatment effects on outcomes related to students' cognition. Specifically, we study impacts on students' scores in the memory for digit span test and in standardized test scores.

We find treatment effects on the standardized test scores in Science. The size of this effect is 0.07 standard deviations, significant at the 5 percent level. We also encounter marginal statistical significance at standard levels when employing randomization inference. We do not find any significant effects on other subjects.[14]

[14] Table F2 in Appendix F shows treatment effects on students' self-assessment in Portuguese and Mathematics, and on students' ability to estimate their own performance. Overall, there is no evidence

Teachers report higher levels of knowledge in Science, which is consistent with higher test scores by students in that subject. This may be explained by the fact that this is the subject most frequently selected in classes employing Profuturo. See Section A in the appendix for further details of content adoption. We do not observe clear changes in other subjects. Hence, the evidence in favor of Hypothesis 5 is mixed.

#### 5.2.6 Robustness

A further note is needed to account for robustness exercises we conduct on the choice of control variables for teachers, students, and guardians, as well as on a weighting exercise to improve the comparability of the treatment and control schools.

Specifically, in Section G of the Appendix we show the replication of the main results of the paper while employing the Post-double Selection Lasso procedure for selecting the referred control variables.[15] In Section H of the Appendix we describe how we construct the inverse probability weights to correct for sample imbalance at baseline and show the replication of the main results of the paper while applying IPTW.[16] Overall, we do not identify any relevant departure from the benchmark results of the paper.

### 5.3 Aggregation of outcomes

In order to address the risks posed by the analysis of multiple outcomes, we now devote attention to aggregating the outcomes we analyzed in detail in the previous section. We bundle outcomes in indices that are built using the procedure detailed in Kling, Liebman, and Katz (2007). We then calculate within-sample z-scores for each individual outcome, employing the mean and the standard deviation of the control group. Subsequently, we obtain the unweighted average z-score for each set of outcomes. In order to aggregate

that students from ProFuturo schools are more likely to correctly estimate their own performance. In Portuguese, however, students are 5.2 percentage points less likely to overestimate their performance, significant at the 5 percent level and at the 10 percent level using exact p-values. In Mathematics, they are 2.4 percentage points more likely to underestimate their performance. This effect is significant at the 10 percent level, but not at standard levels when using randomization inference.

[15] We do not apply Lasso to the analysis done using the classroom observation questionnaire as there we have very few control variables.

[16] We do not apply IPTW to the analysis done using the classroom observation questionnaire as we have very few control variables to construct weights at the classroom level.

outcomes defined at different units of analysis, i.e., at the level of the teacher, student, guardian, and class, we construct indices at the school and grade level. This means that we take a low number of observations in the regressions that follow, meaning that this is a very conservative exercise. Specifically, we consider indices on: Technology use, built from outcomes of Table 1; Teachers' absenteeism and motivation, built from outcomes of Table 2; Students' absenteeism and motivation, built from outcomes of Table 3; Teachers' time allocation: built from the outcomes in Table 4; Observed class delivery: built from outcomes in Table 5; Students' time allocation: built from the outcomes in Table 6; Students' behavior and interactions at school: built from the outcomes of Table 7; Teachers' cognitive skills: built from the outcomes in Table 8; and Students' cognitive skills: built from the outcomes in Table 9.

Figure 1 shows treatment effects analogous to the ones shown in the previous section on the aggregate indices we described above. Confidence intervals are built using statistical significance at the 10 percent level.[17] In face of the standardization of outcome variables embedded in the procedure we adopted, all treatment effects are expressed in standard deviation units.

We find significant treatment effects of ProFuturo in technology use and teacher's motivation. The magnitudes of both these effects are 0.19 and 0.14 standard deviations, significant at the 1 percent level (also using randomization inference). We find positive effects on teachers' time allocation towards teaching and students' behavior and interaction at school, significant at the 5 percent level (but marginally significant when employing exact p-values). Treatment effects on students' absenteeism and motivation, and time allocation towards reading, studying, playing, and using technology with their guardians are on the aggregate positive but not robustly significant in statistical terms. We do not find significant effects on the remaining aggregates.

[17] Appendix F includes the table corresponding to this graph.

# 6 Concluding remarks

ProFuturo is a technology-assisted learning program, which features both hardware and software enabling the teaching of all contents of primary school. To date, ProFuturo has reached 12 million children in 38 countries. Like other programs studied in the literature, it adapts content delivery to individual-student needs. ProFuturo comes with two clear innovations. First, it places teachers at the center of the learning experience, as they are the ones managing the delivery of the program in the classroom. This implies the delivery of significant teacher training. Second, ProFuturo incentivizes interaction in the classroom, between teacher and students, as well as between students.

We implemented a randomized impact evaluation of the ProFuturo program in Luanda. Despite the short time window of our program evaluation, in some cases of less than one year from the beginning of treatment to endline measurements, and a relatively low intensity of weekly exposure to the program, we are able to identify some encouraging findings. First, we observe direct effects on familiarity with technology by both teachers and students. Second, we report on increased motivation of teachers, illustrated by a clear decrease on the number of days teachers missed school. Some evidence suggests that teachers improved class preparation and active classroom teaching, while students became more interested in reading at home and engaged in more altruistic interactions at school. We only find positive effects of ProFuturo on students' test scores for Science, which was the subject most frequently selected under the ProFuturo platform in our setting.

It will be important to extend this research to check whether the effects we encounter could mediate broader and stronger effects on student cognition in the medium to long term. These could lead the way to a clear agenda on education policy in developing countries towards employing technology side by side with training teachers, enabling effective and wide-ranging skill development of children in primary schools.

## References

Ahrens, Achim, Christian Hansen, and Mark Edwin Schaffer. 2018. "LASSOPACK: Stata Module for LASSO, Square-root LASSO, Elastic Net, Ridge, Adaptive LASSO Estimation and Cross-validation." .

Angrist, Joshua and Victor Lavy. 2002. "New evidence on classroom computers and pupil learning." *The Economic Journal* 112 (482):735–765.

Araya, Roberto, Elena Arias Ortiz, Nicolas L Bottan, and Julian Cristia. 2019. "Does gamification in education work? Experimental evidence from Chile." Tech. rep., IDB Working Paper Series.

Athey, Susan and Guido W Imbens. 2017. "The econometrics of randomized experiments." In *Handbook of economic field experiments*, vol. 1. Elsevier, 73–140.

Banerjee, Abhijit V, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying education: Evidence from two randomized experiments in India." *The Quarterly Journal of Economics* 122 (3):1235–1264.

Barrera-Osorio, Felipe and Leigh L Linden. 2009. "The use and misuse of computers in education: Evidence from a randomized experiment in Colombia." *World Bank Policy Research Working Paper* 4836.

Barrow, Lisa, Lisa Markman, and Cecilia Elena Rouse. 2009. "Technology's edge: The educational benefits of computer-aided instruction." *American Economic Journal: Economic Policy* 1 (1):52–74.

Beg, Sabrin A, Adrienne M Lucas, Waqas Halim, and Umar Saif. 2019. "Beyond the basics: Improving post-primary content delivery through classroom technology." Tech. rep., National Bureau of Economic Research.

Beuermann, Diether W, Julian Cristia, Santiago Cueto, Ofer Malamud, and Yyannu Cruz-Aguayo. 2015. "One laptop per child at home: Short-term impacts from a randomized experiment in Peru." *American Economic Journal: Applied Economics* 7 (2):53–80.

Carrillo, Paul E, Mercedes Onofa, and Juan Ponce. 2011. "Information technology and student achievement: Evidence from a randomized experiment in Ecuador." .

Cristia, Julian, Pablo Ibarrarán, Santiago Cueto, Ana Santiago, and Eugenio Severín. 2017. "Technology and child development: Evidence from the one laptop per child program." *American Economic Journal: Applied Economics* 9 (3):295–320.

Duflo, Esther, Rema Hanna, and Stephen P Ryan. 2012. "Incentives work: Getting teachers to come to school." *American Economic Review* 102 (4):1241–78.

Engle, Randall W. 2002. "Working memory capacity as executive attention." *Current directions in psychological science* 11 (1):19–23.

Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. 2017. "Education technology: An evidence-based review." Tech. rep., National Bureau of Economic Research.

Fairlie, Robert W and Jonathan Robinson. 2013. "Experimental evidence on the effects of home computers on academic achievement among schoolchildren." *American Economic Journal: Applied Economics* 5 (3):211–40.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many children left behind? Textbooks and test scores in Kenya." *American Economic Journal: Applied Economics* 1 (1):112–35.

Hanushek, Eric A and Ludger Woessmann. 2008. "The role of cognitive skills in economic development." *Journal of economic literature* 46 (3):607–68.

Heß, Simon. 2017. "Randomization inference with Stata: A guide and software." *The Stata Journal* 17 (3):630–651.

Joffe, Marshall M, Thomas R Ten Have, Harold I Feldman, and Stephen E Kimmel. 2004. "Model selection, confounder control, and marginal structural models: review and new applications." *The American Statistician* 58 (4):272–279.

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

Lai, Fang, Renfu Luo, Linxiu Zhang, Xinzhe Huang, and Scott Rozelle. 2015. "Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing." *Economics of Education Review* 47:34–48.

Lai, Fang, Linxiu Zhang, Xiao Hu, Qinghe Qu, Yaojiang Shi, Yajie Qiao, Matthew Boswell, and Scott Rozelle. 2013. "Computer assisted learning as extracurricular tutor? Evidence from a randomised experiment in rural boarding schools in Shaanxi." *Journal of Development Effectiveness* 5 (2):208–231.

Machin, Stephen, Sandra McNally, and Olmo Silva. 2007. "New technology in schools: Is there a payoff?" *The Economic Journal* 117 (522):1145–1167.

Malamud, Ofer and Cristian Pop-Eleches. 2011. "Home computer use and the development of human capital." *The Quarterly journal of economics* 126 (2):987–1027.

McKenzie, David. 2012. "Beyond baseline and follow-up: The case for more T in experiments." *Journal of development Economics* 99 (2):210–221.

Mo, Di, Johan Swinnen, Linxiu Zhang, Hongmei Yi, Qinghe Qu, Matthew Boswell, and Scott Rozelle. 2013. "Can one-to-one computing narrow the digital divide and the educational gap in China? The case of Beijing migrant schools." *World development* 46:14–29.

Mo, Di, Linxiu Zhang, Renfu Luo, Qinghe Qu, Weiming Huang, Jiafu Wang, Yajie Qiao, Matthew Boswell, and Scott Rozelle. 2014. "Integrating computer-assisted learning into a regular curriculum: Evidence from a randomised experiment in rural schools in Shaanxi." *Journal of development effectiveness* 6 (3):300–323.

Molina-Millán, Teresa, Tania Barham, Karen Macours, John A Maluccio, and Marco Stampini. 2019. "Long-term impacts of conditional cash transfers: review of the evidence." *The World Bank Research Observer* 34 (1):119–159.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J Ganimian. 2019. "Disrupting education? Experimental evidence on technology-aided instruction in India." *American Economic Review* 109 (4):1426–60.

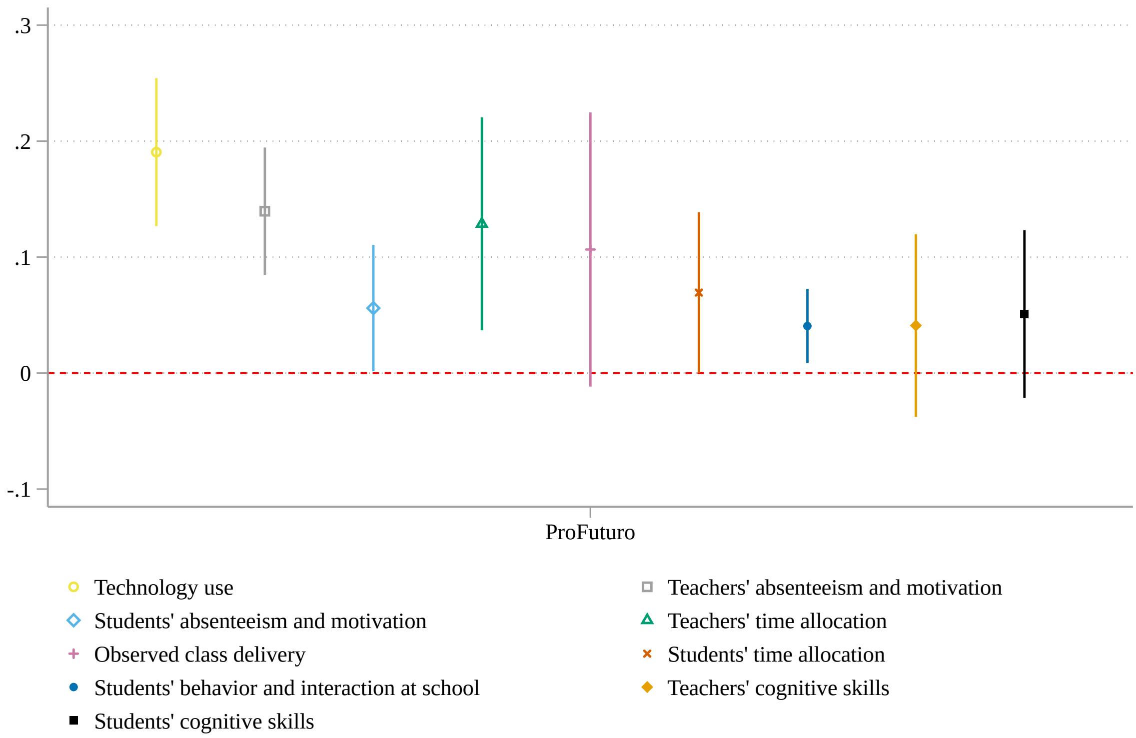
Rouse, Cecilia Elena and Alan B Krueger. 2004. "Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program." *Economics of Education Review* 23 (4):323–338.

Yang, Yihua, Linxiu Zhang, Junxia Zeng, Xiaopeng Pang, Fang Lai, and Scott Rozelle. 2013. "Computers and the academic performance of elementary school-aged girls in China's poor communities." *Computers & Education* 60 (1):335–346.

Young, Alwyn. 2019. "Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results." *The Quarterly Journal of Economics* 134 (2):557–598.

# Figures and tables

Figure 1: Main treatment effects - aggregated outcomes employing z-scores




Note: All estimates are based on OLS regressions using Equation 1, except estimates on Observed class delivery which are based on an OLS regression employing Equation 2. Outcomes are grouped in indices that are built using the procedure in Kling, Liebman, and Katz (2007). We calculate school-grade means for each individual outcome and we compute within-sample z-scores for each school-grade outcome, employing the mean and the standard deviation of the control group. We then obtain the unweighted average z-score for each category. The indices are defined by the following outcomes: (1) Technology use: built from outcomes of Table 1; (2) Teachers' absenteeism and motivation: built from outcomes of Table 2; (3) Students' absenteeism and motivation: built from outcomes of Table 3; (4) Teachers' time allocation: built from the outcomes in Table 4; (5) Observed class delivery: built from the outcomes in Table 5; (6) Students' time allocation: built from the outcomes in Table 6; (7) Students' behavior and interaction at school: built from the outcomes of Table 7; (8) Teachers' cognitive skills: built from the outcomes in Table 8; (9) Students' cognitive skills: built from the outcomes in Table 9. All specifications employed include grade and stratum fixed effects. Confidence intervals are built using statistical significance at the 10 percent level. Standard errors are clustered at the school level.

Table 1: Use of technology.

| | Teachers' survey | Students' survey | | | Guardians' survey |
|---|---|---|---|---|---|
| | Computer use | Index of technology usage – basic | Index of technology usage – advanced | Desire to use more technology at school | Time using technology at home |
| | (1) | (2) | (3) | (4) | (5) |
| ProFuturo | 0.090*** | 0.093* | -0.025 | 0.027*** | 0.232*** |
| | (0.020) | (0.052) | (0.056) | (0.007) | (0.078) |
| | [0.001] | [0.229] | [0.751] | [0.012] | [0.043] |
| Observations | 489 | 2290 | 2219 | 2316 | 631 |
| $R^2$ | 0.11 | 0.21 | 0.22 | 0.06 | 0.22 |
| Mean (control group) | 0.03 | -0.00 | 0.00 | 0.91 | -0.01 |

Note: Estimates based on OLS regressions. All columns present estimates using Equation 1. The estimation sample in column (1) consists of teachers interviewed in the corresponding endline survey; the estimation sample in columns (2)-(4) consists of students interviewed in the corresponding endline survey; the estimation sample in column (5) consists of guardians interviewed in the corresponding endline survey. Depending on the column the dependent variables are defined as follows. (1) Computer use: indicator variable equal to 1 if the teacher used a computer in the classroom during 2018. (2) Index of technology usage – basic: index variable averaging three indicator variables for knowing how to turn on/off a computer, to write in a computer, and to open/close programs and applications. This variable is normalized, i.e., as a z-score. (3) Index of technology usage – advanced: index variable averaging three indicator variables for knowing how to perform searches in the internet, to save a file, and to print documents. This variable is normalized, i.e., as a z-score. (4) Desire to use more technology at school: indicator variable equal to 1 if the student agreed with the statement 'I wish I could use more computers or technology at school.'. (5) Time using technology at home: time allocated to play and study with technology at home. This variable is normalized, i.e., as a z-score. All specifications include strata fixed effects. Column (1) includes teacher-level controls; Columns (2)-(4) include student-level controls; Column (5) includes guardian-level controls. The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at school level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. P-values from randomization-inference tests are reported in brackets.

Table 2: Motivation. Teachers.

| | School administrative data | | Teachers' survey | | | Guardians' survey | Class observation data | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Round 2 | Round 3 |
| | Number of days missing school | | Number of days | | | | | |
| | (previous month) | (academic year) | missing school | arriving late to school | Teachers' motivation | Teachers' motivation | Teachers' motivation | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| ProFuturo | -0.685*** | -0.396*** | -0.102** | 0.165 | 0.056 | 0.075 | 0.052 | 0.230* |
| | (0.179) | (0.078) | (0.039) | (0.101) | (0.061) | (0.080) | (0.138) | (0.132) |
| | [0.037] | [0.004] | [0.103] | [0.234] | [0.494] | [0.509] | [0.661] | [0.160] |
| Observations | 268 | 468 | 504 | 479 | 514 | 635 | 108 | 120 |
| $R^2$ | 0.62 | 0.61 | 0.04 | 0.13 | 0.06 | 0.11 | 0.20 | 0.18 |
| Mean (control group) | -0.01 | 0.03 | 0.03 | -0.00 | -0.03 | 0.05 | 0.00 | 0.01 |

Note: Estimates based on OLS regressions. Columns (1)-(6) present estimates using Equation 1 and columns (7)-(8) present estimates using Equation 2. The estimation sample in columns (1)-(2) consists of school records of teachers on abseenteism; the estimation sample in columns (3)-(5) consists of teachers interviewed in the corresponding endline survey; the estimation sample in column (6) consists of guardians interviewed in the corresponding endline survey; the estimation sample in columns (7)-(8) consists of classrooms observed in November 2018 (column 7) and in March 2019 (column 8). Depending on the column the dependent variables are defined as follows. (1) Number of days missing school (previous month): number of days missing school, in the month previous to the survey. This variable is normalized, i.e., as a z-score. (2) Number of days missing school (academic year): number of days missing school, in the academic year of 2018. This variable is normalized, i.e., as a z-score. (3) Number of days missing school: number of days missing school, namely in the month previous to the survey. This variable is normalized, i.e., as a z-score. (4) Number of days arriving late to school: number of days arriving late to school, namely in the month previous to the survey. This variable is normalized, i.e., as a z-score. (5) Teachers' motivation: index averaging three variables measuring teachers' motivation towards the school. The statements employed a 5-point Likert-type scale from 1 (strongly disagree) to 5 (strongly agree). The statements are the following: 'I like to create my own content instead of following classes that are already prepared;' 'I am usually up to date with the latest pedagogical content,' and 'I usually study at home the contents I will teach in class.' This variable is normalized, i.e., as a z-score. (6) Teachers' motivation: measures the degree of teachers' motivation as reported by parents. It employs a 4-point scale from 1 (not motivated at all) to 4 (very motivated). This variable is normalized, i.e., as a z-score. (7) and (8) Teachers' motivation: motivation of teachers in the classroom reported by the enumerators. These measures employ a 5-point scale from 1 (not motivated at all) to 5 (very motivated), in round 2 and round 3, respectively. These variables are normalized, i.e., as z-scores. Columns (3)-(5) include teacher-level controls; column (6) includes guardian-level controls; columns (7)-(8) include classroom-level controls. The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. P-values from randomization-inference tests are reported in brackets.

Table 3: Motivation. Students and guardians.

| | Students' survey | | | Guardians' survey | Students' survey | Class observation data | |
|---|---|---|---|---|---|---|---|
| | | | | | | Round 2 | Round 3 |
| | Positive attitude towards school | Likes to study Maths | Likes to read | Overall school satisfaction | Number of days missing school | Students' motivation | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| ProFuturo | 0.005 | 0.015** | 0.004 | 0.133* | 0.012 | -0.114 | 0.220 |
| | (0.004) | (0.006) | (0.006) | (0.077) | (0.015) | (0.151) | (0.194) |
| | [0.358] | [0.091] | [0.610] | [0.207] | [0.618] | [0.581] | [0.364] |
| Observations | 2323 | 2323 | 2310 | 665 | 2293 | 108 | 83 |
| $R^2$ | 0.06 | 0.03 | 0.02 | 0.10 | 0.05 | 0.30 | 0.20 |
| Mean (control group) | 0.89 | 0.95 | 0.97 | 0.11 | 0.36 | 0.00 | 0.00 |

Note: Estimates based on OLS regressions. Columns (1)-(5) present estimates using Equation 1 and columns (6)-(7) present estimates using Equation 2. The estimation sample in columns (1), (2), (3), and (5) consist of students interviewed in the corresponding endline survey; the estimation sample in column (4) consists of guardians interviewed in the corresponding endline survey; the estimation sample in columns (6)-(7) consists of classrooms observed in November 2018 (column 6) and in March 2019 (column 7). Depending on the column the dependent variables are defined as follows. (1) Positive attitude towards school: index variable averaging nine indicator variables that measure if students agreed with statements regarding their school satisfaction. The statements are the following: 'I feel safe when I am at school;' 'I feel I belong to this school;' 'It is easy for me to pay attention in class;' 'I usually ask questions out loud in class;' 'If I do not understand something in class, I ask the teacher;' 'What I am learning in class will help me in the future;' 'I care about my grades;' 'It is not hard for me to interact with my peers;' 'I like school.'. (2) Likes to study Maths: indicator variable equal to 1 if the student agreed with the statement 'I like Maths.' (3) Likes to read: indicator variable equal to 1 if the student agreed with the statement 'I like reading.' This variable is normalized, i.e., as a z-score. (4) Overall school satisfaction: index averaging four variables, namely indicator variables equal to 1 if the guardian got information from the school about: the student's performance, the teacher's motivation, the teacher's evaluation, and the school principal's evaluation. (5) Number of days missing school: number of days the student reported to have missed school in the two previous weeks before the survey. (6) and (7) Students' motivation: motivation of students in the classroom reported by the enumerators. It employs a 5-point Likert-type scale from 1 (not motivated at all) to 5 (very motivated), in round 2 and round 3 respectively. These variables are normalized, i.e., as z-scores. Columns (1),(2),(3) and (5) include student-level controls; column (4) includes guardian-level controls; columns (6)-(7) include classroom-level controls. The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. P-values from randomization-inference tests are reported in brackets.

Table 4: Time allocation. Teachers.

| | Teachers' survey | | | | |
|---|---|---|---|---|---|
| | | Time allocation by activity in a regular week | | | |
| | Quality of class preparation index | Teaching | Planning at school | Planning at home | Administrative activities |
| | (1) | (2) | (3) | (4) | (5) |
| ProFuturo | 0.121** | 0.042* | -0.057* | -0.041 | -0.013 |
| | (0.057) | (0.023) | (0.032) | (0.040) | (0.042) |
| | [0.114] | [0.225] | [0.220] | [0.463] | [0.852] |
| Observations | 511 | 513 | 479 | 500 | 370 |
| $R^2$ | 0.07 | 0.04 | 0.08 | 0.06 | 0.05 |
| Mean (control group) | 0.02 | 0.85 | 0.40 | 0.44 | 0.33 |

Note: Estimates based on OLS regressions. All columns present estimates using Equation 1. The estimation sample in columns (1)-(5) consists of teachers interviewed in the corresponding endline survey. Depending on the column the dependent variables are defined as follows. (1) Quality of class preparation index: index variable averaging three indicator variables for the teacher having: a plan of the subjects to teach during the academic year, a book of class registries and summaries, as well as a notebook in which he/she prepares classes. This variable is normalized, i.e., as a z-score. (2) Teaching: indicator variables equals to 1 if the teacher allocated more than 5 hours per week to teach. (3) Planning at school: indicator variable equals to 1 if the teacher allocated more than 5 hours per week to planning classes at school. (4) Planning at home: indicator variable equals to 1 if the teacher allocated more than 5 hours per week to planning classes at home. (5) Administrative activities: indicator variable equals to 1 if the teacher allocated more than 5 hours per week to administrative work. Columns (1)-(5) include teacher-level controls. The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at the school level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. P-values from randomization-inference tests are reported in brackets.

Table 5: Observed class delivery.

| | Reading, instruction and discussion | Practice & drill | Monitoring | Knowledge |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Round 2** | | | | |
| ProFuturo | 0.009 | 0.001 | 0.065** | -0.047 |
| | (0.023) | (0.005) | (0.032) | (0.094) |
| | [0.767] | [0.832] | [0.108] | [0.689] |
| Observations | 123 | 123 | 123 | 123 |
| $R^2$ | 0.20 | 0.47 | 0.33 | 0.29 |
| Mean (control group) | 0.14 | 0.02 | 0.16 | 0.00 |
| **Round 3** | | | | |
| ProFuturo | -0.001 | 0.012** | -0.009 | 0.346** |
| | (0.033) | (0.006) | (0.031) | (0.136) |
| | [0.984] | [0.117] | [0.829] | [0.058] |
| Observations | 129 | 129 | 129 | 128 |
| $R^2$ | 0.23 | 0.22 | 0.15 | 0.32 |
| Mean (control group) | 0.20 | 0.00 | 0.23 | 0.00 |

Note: Estimates based on OLS regressions. All columns present estimates using Equation 2. Estimation sample in columns (1)-(4) consists of classroom-level observations in round 2 and round 3. Dependent variables in columns (1) to (3) represent the number of times the two enumerators coded a certain activity during the whole class over the number of times the two enumerators coincided in their coding. Depending on the column the dependent variables are defined as follows. (1) Reading, instruction and discussion: proportion of times the teacher was involved in an activity in which he/she or the students were reading out loud, were engaged in an activity consisting of instruction of academic content, or were involved in an academic discussion or debate. (2) Practice & drill: proportion of times the teacher was dedicated to activities that were undertaken with the objective of memorizing and practicing material such as multiplication tables, vocabulary or spelling words. (3) Monitoring: proportion of times the teacher was involved in actively monitoring seatwork done by students and in monitoring students performing a copying activity. (4) Knowledge: Degree to which the teacher in the observed class is ranked by enumerators as having a deep knowledge of the subject being taught. These variables employ a 5-point Likert-type scale from 1 (very poor knowledge) to 5 (very deep knowledge). This variable is normalized, i.e., as a z-score. Columns (1)-(4) include classroom-level controls. The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. P-values from randomization-inference tests are reported in brackets.

Table 6: Time allocation. Students.

| | Guardians' survey | | | | |
|---|---|---|---|---|---|
| | Time allocation by activity in a regular week | | | Shared time between guardians and their children | |
| | Reading | Studying | Playing | Using technology | Studying |
| | (1) | (2) | (3) | (4) | (5) |
| ProFuturo | 0.228*** | 0.069 | 0.151** | 0.109* | 0.061 |
| | (0.069) | (0.074) | (0.057) | (0.064) | (0.040) |
| | [0.032] | [0.512] | [0.053] | [0.278] | [0.300] |
| Observations | 597 | 656 | 659 | 665 | 665 |
| $R^2$ | 0.14 | 0.14 | 0.23 | 0.23 | 0.12 |
| Mean (control group) | -0.03 | -0.03 | 0.00 | 0.04 | 0.09 |

Note: Estimates based on OLS regressions. All columns present estimates using Equation 1. The estimation sample in columns (1)-(5) consists of guardians interviewed in the corresponding endline survey. Depending on the column the dependent variables are defined as follows. Columns (1)-(3) regard variables describing the time allocated by the student to different activities in a regular week. Each of those variables employs a 4-point Likert-type scale defined from 1 (30 minutes or less) to 4 (2 or more hours). The activities are the following. (1) Reading: time spent reading. This variable is normalized, i.e., as a z-score. (2) Studying: time spent studying. This variable is normalized, i.e., as a z-score. (3) Playing: time spent playing. This variable is normalized, i.e., as a z-score. Columns (4)-(5) concern variables depicting the activities performed during the time spent together by guardians and their children in the previous month before the survey. Each of those variables employs a 4-point Likert-type scale defined from 1 (never or almost never) to 4 (everyday). The activities are the following. (4) Using technology: time devoted to using technology. This variable is normalized, i.e., as a z-score. (5) Studying: time devoted to doing homework. This variable is normalized, i.e., as a z-score. Columns (1)-(5) include guardian-level controls. The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. P-values from randomization-inference tests are reported in brackets.

Table 7: Students' behavior and interactions at school.

| | Students' survey | | |
|---|---|---|---|
| | Students' altruism | Students get help from peers | Students have friends at school |
| | (1) | (2) | (3) |
| ProFuturo | 0.025*** | 0.019* | 0.002 |
| | (0.009) | (0.010) | (0.010) |
| | [0.067] | [0.200] | [0.866] |
| Observations | 2316 | 2323 | 2323 |
| $R^2$ | 0.04 | 0.05 | 0.07 |
| Mean (control group) | 0.89 | 0.85 | 0.76 |

Note: Estimates based on OLS regressions. All columns present estimates using Equation 1. The estimation sample consists of students interviewed in the corresponding endline survey. Depending on the column the dependent variables are defined as follows. (1) Students' altruism: Indicator variable equal to 1 if the student identifies with the statement 'I am similar to students that like to share with others.'. (2) Students get help from peers: indicator variable equal to 1 if the student agreed with the statement 'My peers help me in class if I need.'. (3) Students have friends at school: indicator variable equal to 1 if the student agreed with the statement 'I have many friends at school.'. Columns (2)-(3) include student-level controls. The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. P-values from randomization-inference tests are reported in brackets.

Table 8: Cognitive skills. Teachers.

| | Teachers' survey | | | | | |
|---|---|---|---|---|---|---|
| | Cognitive tests | | Self-assessment | | | |
| | Portuguese | Maths | Overall | Portuguese | Maths | Science |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| ProFuturo | -0.151** | -0.093 | 0.054 | 0.090 | 0.090 | 0.129** |
| | (0.059) | (0.078) | (0.051) | (0.067) | (0.066) | (0.056) |
| | [0.099] | [0.363] | [0.451] | [0.325] | [0.315] | [0.124] |
| Observations | 491 | 491 | 508 | 463 | 441 | 412 |
| $R^2$ | 0.09 | 0.15 | 0.16 | 0.11 | 0.17 | 0.13 |
| Mean (control group) | 0.03 | 0.02 | -0.00 | -0.03 | 0.01 | 0.00 |

Note: Estimates based on OLS regressions. Columns (1)-(6) present estimates using Equation 1. The estimation sample in columns (1)-(6) consists of teachers interviewed in the corresponding endline survey. Depending on the column the dependent variables are defined as follows. (1) Portuguese: Score of teachers' Language test. This variable is normalized, i.e., as z-score. (2) Maths: Score of teachers' Mathematics test. This variable is normalized, i.e., as a z-score. (3) Overall: Variable depicting teachers' self-reported assessment of their overall performance. It employs a 5-point Likert-type scale from 1 (Very bad) to 5 (very good). This variable is normalized, i.e., as a z-score. (4) Portuguese: Variable depicting teachers' self-reported assessment of their Language performance. It employs a 5-point Likert-type scale from 1 (Very bad) to 5 (very good). This variable is normalized, i.e., as a z-score. (5) Maths: Variable depicting teachers' self-reported assessment of their Mathematics performance. It employs a 5-point Likert-type scale from 1 (Very bad) to 5 (very good). This variable is normalized, i.e., as a z-score. (6) Science: Variable depicting teachers' self-reported assessment of their Science performance. It employs a 5-point Likert-type scale from 1 (Very bad) to 5 (very good). This variable is normalized, i.e., as a z-score.
The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. P-values from randomization-inference tests are reported in brackets.

Table 9: Cognitive skills. Students.

| | Students' survey | | | |
|---|---|---|---|---|
| | | Test scores | | |
| | Digit span | Portuguese | Maths | Science |
| | (1) | (2) | (3) | (4) |
| ProFuturo | 0.037 | -0.007 | -0.015 | 0.066** |
| | (0.048) | (0.061) | (0.061) | (0.028) |
| | [0.602] | [0.944] | [0.841] | [0.100] |
| Observations | 2323 | 1011 | 1011 | 1011 |
| $R^2$ | 0.15 | 0.25 | 0.39 | 0.39 |
| Mean (control group) | -0.00 | 0.00 | 0.00 | -0.01 |

Note: Estimates based on OLS regressions. All columns present estimates using Equation 1. Estimation sample in columns (1)-(4) consists of students interviewed in the corresponding endline survey. Depending on the column the dependent variables are defined as follows. (1) Digit span: Score of the memory for digit span test. This variable is normalized, i.e., as a z-score. (2) Portuguese: Score of students' Language test. This variable is normalized, i.e., as a z-score. (3) Maths: Score of students' Mathematics test. This variable is normalized, i.e., as a z-score. (4) Science: Score of students' Science test. This variable is normalized, i.e., as a z-score. Columns (1)-(4) include student-level controls. The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. P-values from randomization-inference tests are reported in brackets.

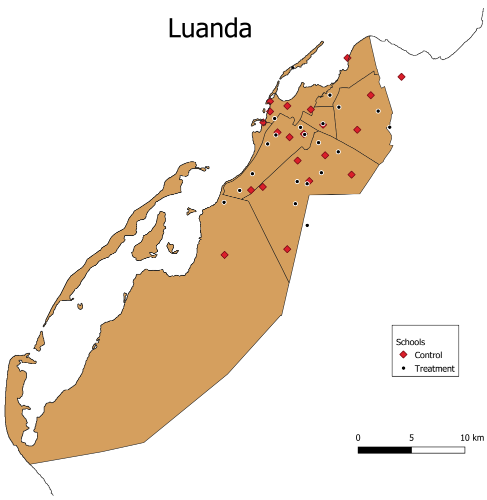
# A Program details

This Appendix provides a more detailed description of how ProFuturo was implemented in the schools included in our study, of program adoption in treated schools, of the content repository, i.e., the specifics of the Computer-assisted Learning (CAL) software of the ProFuturo program, of content adoption, as well as of classroom observation of ProFuturo classes.

## A.1 Implementation of the program in Luanda

School principals and teachers from the 21 treatment schools in Luanda were trained during a week at the end of 2017. School principals selected an average of eight teachers from their school to receive training. These teachers were chosen based on their motivation, technological skills, and availability. During the training, ProFuturo coordinators recommended that each treatment class within a treated school used the program for around 90 minutes per week. Teachers were instructed to prepare ProFuturo classes inside the school, and ask for ProFuturo's support when needed. Once the program was introduced, ProFuturo coordinators were expected to visit each program school every week. See Figure A1 for the geographical distribution of the schools in our study.

Figure A1: Geographic distribution of the schools in our study




Suitcases were distributed to schools (one per school) during the first month of the following school year (February 2018), each containing 48 tablets, one computer, and one projector. Figure A2 shows one of these suitcase including all equipment inside.

Figure A2: ProFuturo suitcase


## A.2 Program adoption

The limited equipment per school that ProFuturo had available in the setting we study conditioned the intensity of the program. In face of these restrictions, school principals decided to target the provision of the program to a sub-sample of classes. The majority of classes selected belonged to grades 4, 5, and 6 (the highest grades of primary school). This way, it was only a matter of time that most children in treated schools would be beneficiaries of the program.

We have two sources of data regarding program coverage and usage per class: reports by school principals and reports by ProFuturo coordinators.

On average, the duration of a ProFuturo class, as reported by school principals, was 106 minutes. The number of times per week treated students had a ProFuturo class was 1.2. Combining these two figures, we can conclude that on average students from classes benefiting from ProFuturo were exposed to the program 132 minutes per week according to principals. ProFuturo coordinators reported similar levels of exposure: 136 minutes per week. Note however that these numbers apply only to students attending classes assigned to the program.

Indeed, restricting the distribution of the suitcases to grades 4, 5, and 6 was not enough to ensure that all students of those grades would have access to the program. School principals typically selected out some classes from those grades not to be targeted by ProFuturo. According to reports of these principals, the reasons for these exclusions included: the high number of children in those classes; none of the teachers from those classes having had ProFuturo training; the weekly schedule of their classroom not matching the schedule of trained teachers; among others.

Table A1 shows program take-up. On average, 76 percent of classes from grades 4, 5, and 6 were using ProFuturo in October 2018. Note that, although the program was launched in March 2018, it was not fully functioning in many schools until May/June, which means that our October measures are for just a few months after the program started in some cases. Furthermore, there is significant heterogeneity per school, as the percentage of classes from the relevant grades using ProFuturo measured in October goes from 11 to 100 percent.

Table A1: Intensity of treatment

| School | Treated Classes - May | Treated classes - October |
|---|---|---|
| 1 | 83% | 100% |
| 2 | 63% | 100% |
| 3 | 45% | 45% |
| 4 | 67% | 100% |
| 5 | 100% | 100% |
| 6 | 100% | 100% |
| 7 | 43% | 43% |
| 8 | 56% | 11% |
| 9 | 100% | 100% |
| 10 | 33% | 11% |
| 11 | 57% | 86% |
| 12 | 44% | 100% |
| 13 | 50% | 88% |
| 14 | 56% | 56% |
| 15 | 40% | 33% |
| 16 | 17% | 22% |
| 17 | 100% | 100% |
| 18 | 100% | 100% |
| 19 | 57% | 100% |
| 20 | 50% | 100% |
| 21 | 100% | 100% |
| Average intensity | 65% | 76% |

## A.3 Repository

ProFuturo contains a repository with contents and digital activities with the purpose of developing students' core competencies. There are three main content blocks: Language, STEM (Mathematics, Science, and Technology), and Ethics and Citizenship.

The repository aims to work alongside the learning curricula of each country, while helping teachers to plan their classes. It is the teacher that selects the content to be accessed by the students, based on the country curricula and on the specific needs of their students.

The repository includes content and activities with different levels of difficulty. Each level of difficulty contains several didactic units. Each of these units includes contents and digital activities on a specific subject with corresponding learning objectives. Didactic units have a common structure which is composed of the following: introduction; learning content (a theoretical and interactive part); practical activities (exercises that provide feedback to students); assessment (includes a specific threshold that students must achieve to complete the unit); summary (synthesis of what was learned in the unit); didactic support (only for teachers, gives suggestions on how to best incorporate the unit in the students' learning experience and how to evaluate students).

On the latter, ProFuturo proposes a number of strategies to maximize students' learning using the platform. These suggestions often have the objective of increasing students' interactions - for example by rearranging the table disposition inside the classroom or organizing students in small groups.

Figure A3 shows a Science activity, figure A4 shows a Mathematics activity, and figure A5 shows a Portuguese activity.

Figure A3: Science activity


Figure A4: Mathematics activity

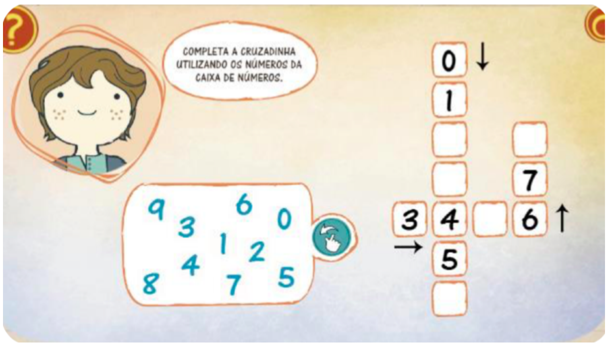
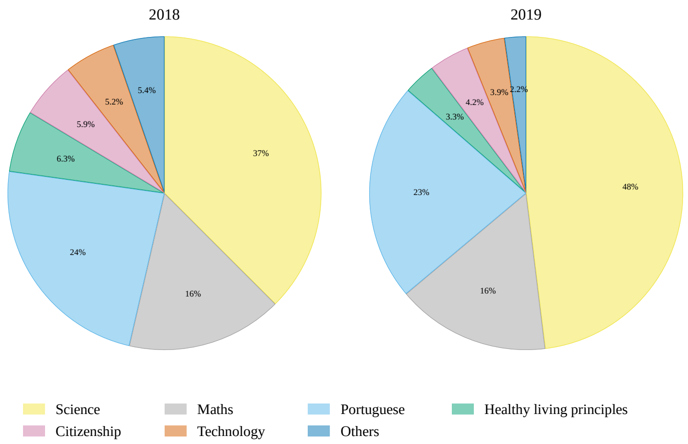



Figure A5: Portuguese activity

## A.4 Content adoption

Administrative data at the school level provided by the ProFuturo platform give us information about the content of the activities performed by students in 19 treated schools.[18] The data refer to the period between February 2018 and October 2019. On average, around 40 percent of students in these treated schools used ProFuturo at least once in 2018 and 2019. We do not have information on the time each student spent in the platform. We do know however that on average students started 62 activities in 2018 and 22 activities in 2019. Overall, schools employed ProFuturo primarily for Science, Portuguese, and Maths. Figure A6 shows the distribution of activities by subject. Around 37 percent of all activities in 2018 and 48 percent in 2019 were in Science. Apart from Science, in those years, 23-24 percent of the activities were dedicated to Portuguese, and 16 percent of the activities were devoted to Maths. These statistics are in line with the findings of the surveys we conducted to evaluate the implementation of the ProFuturo program. Based on the corresponding interviews with the teachers and school principals in August and November 2018, the content in Science was found to be most aligned with the Angolan curriculum, when compared to other subjects.

Figure A6: Content adoption per subject




Note: Data from ProFuturo platform on students' activities by subject. The database does not include information on usage by students in two schools.

## A.5 Classroom observations

In August 2018, while the program was functioning in treatment schools, we selected a class using the ProFuturo program from each school to be observed using the 'Classroom observation questionnaire,' which was marginally adapted from the Stallings classroom instrument to fit classes employing ProFuturo. From table A2 we can see that in these

[18] The ProFuturo database does not include data on usage by students in two treated schools.

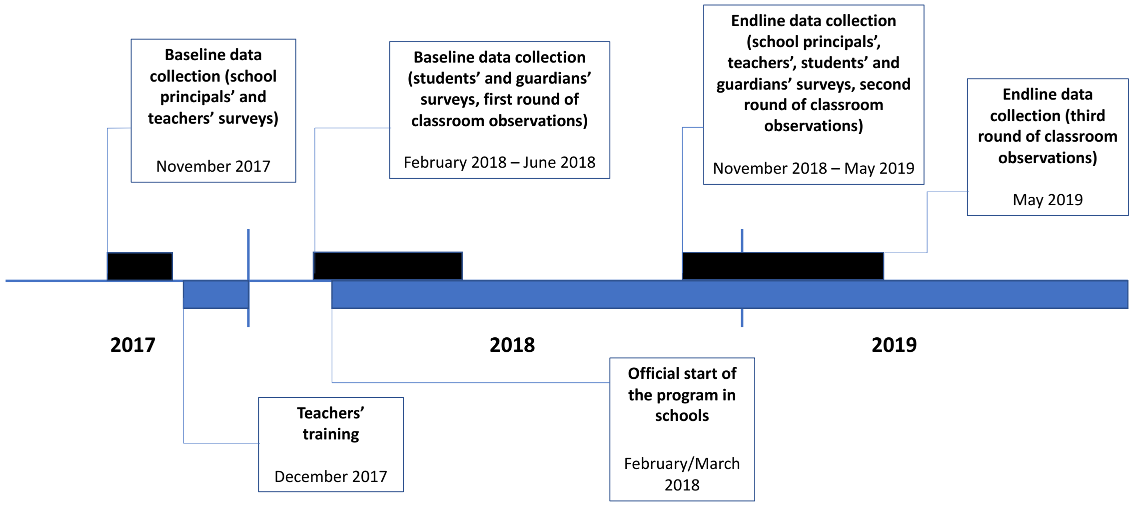
classes, teachers spend the majority of classroom time monitoring seatwork (32 percent) followed by managing the classroom with students (16 percent) and discussion (14 percent). They do not devote any time to monitoring copying or to activities involving practice and drill.

Table A2: Class observation. ProFuturo classes.

| | Obs. | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Reading | 21 | 0.01 | 0.04 | 0.00 | 0.11 |
| Instruction | 21 | 0.07 | 0.12 | 0.00 | 0.43 |
| Discussion | 21 | 0.14 | 0.16 | 0.00 | 0.60 |
| Practice and drill | 21 | 0.00 | 0.00 | 0.00 | 0.00 |
| Monitoring seatwork | 21 | 0.32 | 0.24 | 0.00 | 0.80 |
| Monitoring copying | 21 | 0.00 | 0.00 | 0.00 | 0.00 |
| Disciplining | 21 | 0.02 | 0.05 | 0.00 | 0.20 |
| Managing with students | 21 | 0.16 | 0.19 | 0.00 | 0.80 |
| Managing alone | 21 | 0.10 | 0.16 | 0.00 | 0.70 |
| Socializing | 21 | 0.02 | 0.04 | 0.00 | 0.11 |
| Visitor | 21 | 0.10 | 0.15 | 0.00 | 0.40 |
| Not present | 21 | 0.05 | 0.13 | 0.00 | 0.50 |

# B Timeline

Figure A7: Timeline



# C List of outcome variables

## C.1 Use of technology

- **Teachers. Computer use:** indicator variable equal to 1 if the teacher used a computer in the classroom during 2018. Source: teachers' survey.

- **Students. Index of technology usage – basic:** index variable averaging three indicator variables for knowing how to turn on/off a computer, to write in a computer, and to open/close programs and applications. This variable is normalized, i.e., as a z-score. Source: students' survey.

- **Students. Index of technology usage – advanced:** index variable averaging three indicator variables for knowing how to perform searches in the internet, to save a file, and to print documents. This variable is normalized, i.e., as a z-score. Source: students' survey.

- **Students. Desire to use more technology at school:** indicator variable equal to 1 if the student agreed with the statement 'I wish I could use more computers or technology at school.'. Source: students' survey.

- **Students. Time using technology at home:** time allocated to play and study with technology at home. This variable is normalized, i.e., as a z-score. Source: guardians' survey.

## C.2 Motivation - teachers

- **Number of days missing school (previous month) - administrative:** number of days missing school, in the month previous to the survey. This variable is normalized, i.e., as a z-score. Source: schools' administrative data.

- **Number of days missing school (academic year) - administrative:** number of days missing school, in the academic year of 2018. This variable is normalized, i.e., as a z-score. Source: schools' administrative data.

- **Number of days missing school - self-reported:** number of days missing school, namely in the month previous to the survey. This variable is normalized, i.e., as a z-score. Source: teachers' survey.

- **Number of days arriving late to school:** number of days arriving late to school, namely in the month previous to the survey. This variable is normalized, i.e., as a z-score. Source: teachers' survey.

- **Teachers' motivation - self-reported:** index averaging three variables measuring teachers' motivation towards the school. The statements employed a 5-point Likert-type scale from 1 (strongly disagree) to 5 (strongly agree). The statements are the following: 'I like to create my own content instead of following classes that are already prepared;' 'I am usually up to date with the latest pedagogical content,' and 'I usually study at home the contents I will teach in class.' This variable is normalized, i.e., as a z-score. Source: teachers' survey.

- **Teachers' motivation - perceived by guardians:** measures the degree of teachers' motivation as reported by parents. It employs a 4-point scale from 1 (not motivated at all) to 4 (very motivated). This variable is normalized, i.e., as a z-score. Source: guardians' survey.

- **Teachers' motivation - observed:** motivation of teachers in the classroom reported by the enumerators. These measures employ a 5-point scale from 1 (not motivated at all) to 5 (very motivated), in round 2 and round 3. These variables are normalized, i.e., as z-scores. Source: classroom observation.

## C.3 Motivation - students and guardians

- **Students. Positive attitude towards school:** index variable averaging nine indicator variables that measure if students agreed with statements regarding their school satisfaction. The statements are the following: 'I feel safe when I am at school;' 'I feel I belong to this school;' 'It is easy for me to pay attention in class;' 'I usually ask questions out loud in class;' 'If I do not understand something in class, I ask the teacher;' 'What I am learning in class will help me in the future;' 'I care about my grades;' 'It is not hard for me to interact with my peers;' 'I like school.'. Source: students' survey.

- **Students. Likes to study Maths:** indicator variable equal to 1 if the student agreed with the statement 'I like Maths.' Source: students' survey.

- **Students. Likes to read:** indicator variable equal to 1 if the student agreed with the statement 'I like reading.' This variable is normalized, i.e., as a z-score. Source: students' survey.

- **Guardians. Overall school satisfaction:** index averaging four variables, namely indicator variables equal to 1 if the guardian got information from the school about: the student's performance, the teacher's motivation, the teacher's evaluation, and the school principal's evaluation. Source: guardians' survey.

- **Students. Number of days missing school:** number of days the student reported to have missed school in the two previous weeks before the survey. Source: students' survey.

- **Students' motivation:** motivation of students in the classroom reported by the enumerators. It employs a 5-point Likert-type scale from 1 (not motivated at all) to 5 (very motivated), in round 2 and round 3. These variables are normalized, i.e., as z-scores. Source: classroom observation.

## C.4 Time allocation - teachers

- **Quality of class preparation index:** index variable averaging three indicator variables for the teacher having: a plan of the subjects to teach during the academic year, a book of class registries and summaries, as well as a notebook in which he/she prepares classes. This variable is normalized, i.e., as a z-score. Source: teachers' survey.

- **Teaching:** indicator variable equals to 1 if the teacher allocated more than 5 hours per week to teach. Source: teachers' survey.

- **Planning at school:** indicator variable equals to 1 if the teacher allocated more than 5 hours per week to planning classes at school. Source: teachers' survey.

- **Planning at home:** indicator variable equals to 1 if the teacher allocated more than 5 hours per week to planning classes at home. Source: teachers' survey.

- **Administrative activities:** indicator variable equals to 1 if the teacher allocated more than 5 hours per week to administrative work. Source: teachers' survey.

## C.5 Observed class delivery

- **Reading, instruction and discussion:** proportion of times the teacher was involved in an activity in which he/she or the students were reading out loud, were engaged in an activity consisting of instruction of academic content, or were involved in an academic discussion or debate. Source: classroom observation.

- **Practice & drill:** proportion of times the teacher was dedicated to activities that were undertaken with the objective of memorizing and practicing material such as multiplication tables, vocabulary or spelling words. Source: classroom observation.

- **Monitoring:** proportion of times the teacher was involved in actively monitoring seatwork done by students and in monitoring students performing a copying activity. Source: classroom observation.

- **Knowledge:** Degree to which the teacher in the observed class is ranked by enumerators as having a deep knowledge of the subject being taught. These variables employ a 5-point Likert-type scale from 1 (very poor knowledge) to 5 (very deep knowledge), in round 2 and round 3. These variables are normalized, i.e., as z-scores. Source: classroom observation.
- **Disciplining:** proportion of times the teacher was reprimanding students for poor behavior. Source: classroom observation.
- **Managing:** proportion of times the teacher was involved in organizational or managerial tasks. Source: classroom observation.
- **Off-task:** proportion of times the teacher was absent from the classroom or involved with social interaction. Source: classroom observation.

## C.6 Time allocation - students

- **Students. Reading:** time spent reading. This variable is normalized, i.e., as a z-score. Source: guardians' survey.
- **Students. Studying:** time spent studying. This variable is normalized, i.e., as a z-score. Source: guardians' survey.
- **Students. Playing:** time spent playing. This variable is normalized, i.e., as a z-score. Source: guardians' survey.
- **Guardians and their children - using technology:** time spent together by guardians and their children in the previous month before the survey using technology. This variable employs a 4-point Likert-type scale defined from 1 (never or almost never) to 4 (everyday).This variable is normalized, i.e., as a z-score. Source: guardians' survey.
- **Guardians and their children - studying:** time spent together by guardians and their children in the previous month before the survey doing homework. This variable employs a 4-point Likert-type scale defined from 1 (never or almost never) to 4 (everyday).This variable is normalized, i.e., as a z-score. Source: guardians' survey.

## C.7 Students' behavior and interactions at school

- **Students' altruism:** Indicator variable equal to 1 if the student identifies with the statement 'I am similar to students that like to share with others.'. Source: students' survey.

- **Students get help from peers:** indicator variable equal to 1 if the student agreed with the statementMy peers help me in class if I need.'. Source: students' survey.

- **Students have friends at school:** indicator variable equal to 1 if the student agreed with the statement 'I have many friends at school.'. Source: students' survey.

## C.8 Cognitive skills - teachers

- **Cognitive tests - Portuguese:** Score of teachers' Language test. This variable is normalized, i.e., as z-score. Source: cognitive tests submitted as part of the teachers' survey.

- **Cognitive tests - Maths:** Score of teachers' Mathematics test. This variable is normalized, i.e., as a z-score. Source: cognitive tests submitted as part of the teachers' survey.

- **Self-assessment - Overall:** Variable depicting teachers' self-reported assessment of their overall performance. It employs a 5-point Likert-type scale from 1 (Very bad) to 5 (very good). This variable is normalized, i.e., as a z-score. Source: teachers' survey.

- **Self-assessment - Portuguese:** Variable depicting teachers' self-reported assessment of their Language performance. It employs a 5-point Likert-type scale from 1 (Very bad) to 5 (very good). This variable is normalized, i.e., as a z-score. Source: teachers' survey.

- **Self-assessment - Maths:** Variable depicting teachers' self-reported assessment of their Mathematics performance. It employs a 5-point Likert-type scale from 1 (Very bad) to 5 (very good). This variable is normalized, i.e., as a z-score. Source: teachers' survey.

- **Self-assessment - Science:** Variable depicting teachers' self-reported assessment of their Science performance. It employs a 5-point Likert-type scale from 1 (Very bad) to 5 (very good). Source: teachers' survey.

## C.9 Cognitive skills - students

- **Digit span:** Score of the memory for digit span test. This variable is normalized, i.e., as a z-score. Source: cognitive test submitted as part of the students' survey.

- **Test scores - Portuguese:** Score of students' Language test. This variable is normalized, i.e., as a z-score. Source: students' cognitive tests.

- **Test scores - Maths:** Score of students' Mathematics test. This variable is normalized, i.e., as a z-score. Source: students' cognitive tests.
- **Test scores - Science:** Score of students' Science test. This variable is normalized, i.e., as a z-score. Source: students' cognitive tests.
- **Self-assessment - Portuguese:** Measures how the student self-evaluates in Language. The question employed a Likert-type scale from 1 (I am a very bad student in Language) to 5 (I am a very good student in Language). This variable is normalized, i.e., as a z-score. Source: students' survey.
- **Correct - Portuguese:** Indicator variable equal to 1 if the student's evaluation of his/her performance in Portuguese using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) coincides with his/her position in the distribution of the test scores in Portuguese inside the corresponding school. Source: students' survey and cognitive tests.
- **Over estimates performance - Portuguese:** Indicator variable equal to 1 if the student's evaluation of his/her performance in Portuguese using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) is above his/her position in the distribution of the test scores in Portuguese inside the corresponding school. Source: students' survey and cognitive tests.
- **Under estimates performance - Portuguese:** Indicator variable equal to 1 if the student's evaluation of his/her performance in Portuguese using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) is below his/her position in the distribution of the test scores in Portuguese inside the corresponding school. Source: students' survey and cognitive tests.
- **Self-assessment - Maths:** Measures how the student self-evaluates in Mathematics. The question employed a Likert-type scale from 1 (I am a very bad student in Mathematics) to 5 (I am a very good student in Mathematics). This variable is normalized, i.e., as a z-score. Source: students' survey.
- **Correct - Maths:** Indicator variable equal to 1 if the student's evaluation of his/her performance in Mathematics using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) coincides with his/her position in the distribution of the test scores in Mathematics inside the corresponding school. Source: students' survey and cognitive tests.
- **Over estimates performance - Maths:** Indicator variable equal to 1 if the student's evaluation of his/her performance in Mathematics using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) is above his/her position in the distribution

of the test scores in Mathematics inside the corresponding school. Source: students' survey and cognitive tests.

- **Under estimates performance - Maths:** Indicator variable equal to 1 if the student's evaluation of his/her performance in Mathematics using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) is below his/her position in the distribution of the test scores in Mathematics inside the corresponding school. Source: students' survey and cognitive tests.

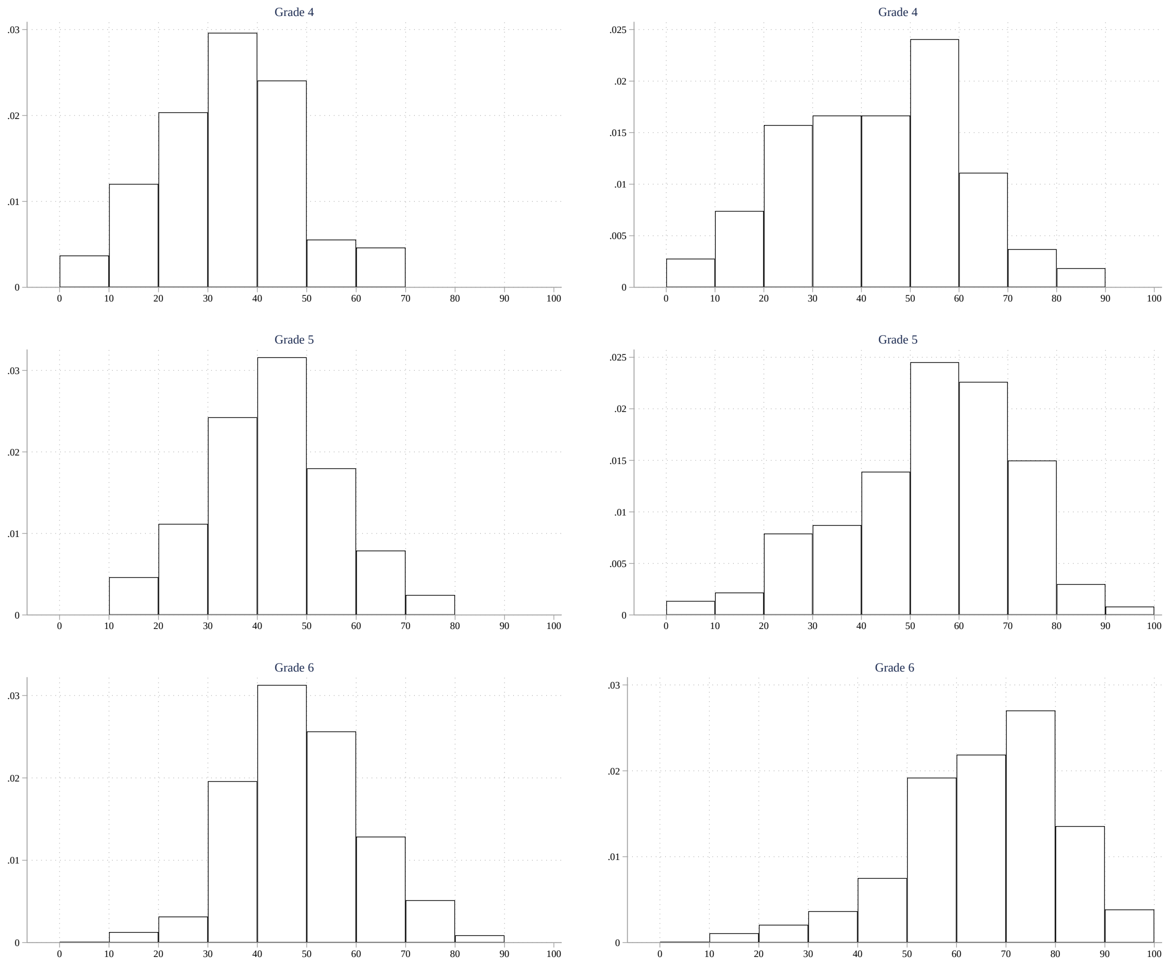
# D Descriptive statistics and balance

Table D1: Balance

| | Observations | Control mean | Difference between treatment and control (s.e.) |
|---|---|---|---|
| Teachers | | | |
| Female (=1) | 629 | 0.40 | -0.04<br>(0.04) |
| Age | 629 | 36.49 | 0.39<br>(1.10) |
| Married (=1) | 623 | 0.45 | 0.03<br>(0.05) |
| Has children (=1) | 628 | 0.74 | 0.06<br>(0.05) |
| Completed 12 years of education (=1) | 629 | 0.79 | -0.01<br>(0.03) |
| Professional/technical degree (=1) | 629 | 0.21 | 0.01<br>(0.03) |
| University studies (=1) | 629 | 0.36 | 0.11***<br>(0.03) |
| Teaching experience (years) | 626 | 3.02 | 0.12<br>(0.09) |
| Owns house (=1) | 517 | 0.52 | 0.06<br>(0.04) |
| Has piped water (=1) | 517 | 0.40 | 0.02<br>(0.04) |
| Index IT goods (0-1) | 517 | 0.46 | 0.01<br>(0.02) |
| Students | | | |
| Female (=1) | 2519 | 0.53 | -0.02<br>(0.02) |
| Age | 2323 | 11.09 | 0.10<br>(0.08) |
| Attended kindergarden (=1) | 2512 | 0.79 | -0.03*<br>(0.02) |
| Failed at least one course (=1) | 2519 | 0.30 | 0.04*<br>(0.02) |
| Test score - overall (standardized) | 1476 | -0.00 | 0.05<br>(0.10) |
| Guardians | | | |
| Female (=1) | 1066 | 0.56 | -0.01<br>(0.03) |
| Age | 1070 | 39.00 | -1.19**<br>(0.54) |
| Number of children | 1049 | 1.04 | 0.04**<br>(0.02) |
| Completed 12 years of education (=1) | 1070 | 0.77 | -0.02<br>(0.03) |
| University studies (=1) | 1070 | 0.17 | -0.05**<br>(0.02) |
| Owns house (=1) | 1070 | 0.60 | -0.02<br>(0.03) |
| Has piped water (=1) | 1070 | 0.38 | 0.06<br>(0.06) |
| Index IT goods (0-1) | 1070 | 0.38 | -0.01<br>(0.03) |

Note: Column (1) reports the number of observations. Column (2) reports the mean of the control group. Column (3) reports estimates for the coefficient of the treatment indicator variable in Equation 2, controlling only for strata fixed effects. Standard errors clustered at the school level are reported in parentheses.*** p<0.01, ** p<0.05, * p<0.1.

Figure A1: Distributions of test scores at baseline




(a) Portuguese (b) Maths

Note: Figures show the test score distributions by grade, corresponding to the questions that were administered at baseline in both Portuguese and Maths. Test scores rank from 0 to 100.

# E Attrition

Table E1: Attrition

| | Teachers | | Students | | Guardians | |
|---|---|---|---|---|---|---|
| | Panel | Total | Panel | Total | Panel | Total |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| ProFuturo | -0.001 | -0.008 | -0.016 | -0.016 | 0.009 | -0.014 |
| | (0.027) | (0.021) | (0.011) | (0.011) | (0.031) | (0.021) |
| Observations | 822 | 822 | 2520 | 2520 | 1460 | 1460 |
| $R^2$ | 0.03 | 0.04 | 0.02 | 0.02 | 0.06 | 0.04 |
| Mean (control group) | 0.62 | 0.86 | 0.93 | 0.93 | 0.44 | 0.74 |

Note: Estimates based on OLS regressions. All columns present estimates using Equation 2. Estimation sample in columns (1) and (2) consists of teachers, estimation sample in columns (3) and (4) consists of students in fourth, fifth and sixth grade, estimation sample in columns (5) and (6) consists of the guardians of the sample of students. Depending on the column the dependent variables are defined by the following. Columns (1), (3) and (5): indicator variable that takes value 1 for individuals interviewed at baseline and at endline, and 0 for individuals not found at endline. Columns (2), (4) and (6): indicator variable that takes value 1 for individuals interviewed at endline, regardless of whether they were interviewed at baseline or not. All specifications include strata fixed effects. Standard errors, reported in parentheses, are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. P-value RI reports randomization inference p-values, clustered by school and stratified by strata pair.

Table A1: Balance after attrition

| | Observations | Control mean | Difference between treatment and control (s.e.) |
|---|---|---|---|
| Teachers | | | |
| Female (=1) | 517 | 0.41 | -0.04<br>(0.04) |
| Age | 517 | 36.51 | 0.18<br>(1.22) |
| Married (=1) | 513 | 0.45 | 0.03<br>(0.05) |
| Has children (=1) | 516 | 0.75 | 0.05<br>(0.05) |
| Completed 12 years of education (=1) | 517 | 0.78 | -0.02<br>(0.03) |
| Professional/technical degree (=1) | 517 | 0.22 | 0.03<br>(0.03) |
| University studies (=1) | 517 | 0.34 | 0.12***<br>(0.03) |
| Teaching experience (years) | 514 | 2.99 | 0.19*<br>(0.10) |
| Owns house (=1) | 517 | 0.52 | 0.06<br>(0.04) |
| Has piped water (=1) | 517 | 0.40 | 0.02<br>(0.04) |
| Index IT goods (0-1) | 517 | 0.46 | 0.01<br>(0.02) |
| Students | | | |
| Female (=1) | 2323 | 0.54 | -0.03**<br>(0.01) |
| Age | 2323 | 11.09 | 0.10<br>(0.08) |
| Attended kindergarden (=1) | 2315 | 0.80 | -0.04**<br>(0.02) |
| Failed at least one course (=1) | 2322 | 0.30 | 0.03<br>(0.02) |
| Test score - overall (standardized) | 1379 | 0.02 | 0.05<br>(0.10) |
| Guardians | | | |
| Female (=1) | 671 | 0.57 | 0.01<br>(0.04) |
| Age | 672 | 39.00 | -0.35<br>(0.69) |
| Number of children | 660 | 1.04 | 0.02<br>(0.02) |
| Completed 12 years of education (=1) | 672 | 0.77 | -0.02<br>(0.03) |
| University studies (=1) | 672 | 0.17 | -0.05**<br>(0.02) |
| Owns house (=1) | 672 | 0.61 | -0.02<br>(0.04) |
| Has piped water (=1) | 672 | 0.37 | 0.06<br>(0.06) |
| Index IT goods (0-1) | 672 | 0.36 | -0.00<br>(0.02) |

Note: Estimation sample consists of teachers, students, and guardians interviewed at endline. Column (1) reports the number of observations. Column (2) reports the mean of the control group. Column (3) reports estimates for the coefficient of the treatment indicator variable in Equation 1, controlling for strata fixed effects. Standard errors clustered at the school level are reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

# F Additional Results

Table F1: Class observation.

| | Reading | Instruction | Discussion | Practice & drill | Monitoring seatwork | Monitoring copying | Disciplining | Managing | Off-task |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9)) |
| **Round 2** | | | | | | | | | |
| ProFuturo | 0.003 | -0.006 | 0.011 | 0.001 | 0.016 | 0.049*** | -0.001 | 0.015 | -0.038 |
| | (0.009) | (0.011) | (0.020) | (0.005) | (0.023) | (0.014) | (0.004) | (0.025) | (0.041) |
| | [0.795] | [0.634] | [0.641] | [0.832] | [0.585] | [0.013] | [0.888] | [0.688] | [0.486] |
| Observations | 123 | 123 | 123 | 123 | 123 | 123 | 123 | 123 | 123 |
| $R^2$ | 0.21 | 0.25 | 0.16 | 0.47 | 0.36 | 0.39 | 0.16 | 0.21 | 0.33 |
| Mean (control group) | 0.02 | 0.04 | 0.08 | 0.02 | 0.13 | 0.04 | 0.01 | 0.11 | 0.18 |
| **Round 3** | | | | | | | | | |
| ProFuturo | 0.001 | -0.013 | 0.011 | 0.012** | -0.023 | 0.014 | -0.001 | 0.037** | 0.013 |
| | (0.017) | (0.019) | (0.017) | (0.006) | (0.025) | (0.015) | (0.006) | (0.016) | (0.040) |
| | [0.979] | [0.612] | [0.610] | [0.117] | [0.463] | [0.484] | [0.888] | [0.065] | [0.791] |
| Observations | 129 | 129 | 129 | 129 | 129 | 129 | 129 | 129 | 129 |
| $R^2$ | 0.27 | 0.23 | 0.16 | 0.22 | 0.14 | 0.19 | 0.14 | 0.27 | 0.27 |
| Mean (control group) | 0.04 | 0.09 | 0.06 | 0.00 | 0.17 | 0.06 | 0.01 | 0.07 | 0.19 |

Note: Estimates based on OLS regressions. All columns present estimates using Equation 2. Estimation sample in columns (1)-(9) consists of classroom-level observations in round 2 and round 3. All dependent variables represent the number of times the two enumerators coded a certain activity during the whole class over the number of times the two enumerators coincided in their coding. Depending on the column the dependent variables are defined as follows. (1) Reading: proportion of times the teacher was involved in an activity in which he/she or the students were reading out loud. (2) Instruction: proportion of times the teacher was involved in an activity consisting of instruction of academic content. (3) Discussion: proportion of times teacher and students were involved in an academic discussion or debate. (4) Practice & drill: proportion of times the teacher was dedicated to activities that were undertaken with the objective of memorizing and practicing material such as multiplication tables, vocabulary or spelling words. (5) Monitoring seatwork: proportion of times the teacher was involved in actively monitoring seatwork done by students. (6) Monitoring copying: proportion of times the teacher was involved in monitoring students performing a copying activity. (7) Disciplining: proportion of times the teacher was reprimanding students for poor behavior. (8) Managing: proportion of times the teacher was involved in organizational or managerial tasks. (9) Off-task: proportion of times the teacher was absent from the classroom or involved with social interaction. Columns (1)-(9) include classroom-level controls. The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at the school level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. P-values from randomization-inference tests are reported in brackets.

Table F2: Self-evaluation of cognitive skills. Students.

| | Students' survey | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Portuguese | | | | Maths | | | |
| | Self-assessment | Correct | Over estimates | Under estimates | Self-assessment | Correct | Over estimates | Under estimates |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| ProFuturo | -0.037 | 0.039 | -0.052** | 0.010 | -0.117*** | -0.012 | -0.009 | 0.024* |
| | (0.034) | (0.024) | (0.020) | (0.011) | (0.034) | (0.028) | (0.023) | (0.014) |
| | [0.451] | [0.249] | [0.071] | [0.493] | [0.030] | [0.769] | [0.781] | [0.235] |
| Observations | 2319 | 1344 | 1344 | 1344 | 2317 | 1344 | 1344 | 1344 |
| $R^2$ | 0.09 | 0.06 | 0.05 | 0.03 | 0.13 | 0.04 | 0.07 | 0.04 |
| Mean (control group) | -0.00 | 0.46 | 0.48 | 0.06 | -0.00 | 0.47 | 0.38 | 0.14 |

Note: Estimates based on OLS regressions. All columns present estimates using Equation 1. The estimation sample consists of students interviewed in the corresponding endline survey. Depending on the column the dependent variables are defined as follows. (1) Self-assessment - Portuguese: Measures how the student self-evaluates in Language. The question employed a Likert-type scale from 1 (I am a very bad student in Language) to 5 (I am a very good student in Language). This variable is normalized, i.e., as a z-score. (2) Correct: Indicator variable equal to 1 if the student's evaluation of his/her performance in Portuguese using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) coincides with his/her position in the distribution of the test scores in Portuguese inside the corresponding school. (3) Over estimates performance: Indicator variable equal to 1 if the student's evaluation of his/her performance in Portuguese using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) is above his/her position in the distribution of the test scores in Portuguese inside the corresponding school. (4) Under estimates performance: Indicator variable equal to 1 if the student's evaluation of his/her performance in Portuguese using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) is below his/her position in the distribution of the test scores in Portuguese inside the corresponding school. (5) Self-assessment - Maths: Measures how the student self-evaluates in Mathematics. The question employed a Likert-type scale from 1 (I am a very bad student in Mathematics) to 5 (I am a very good student in Mathematics). This variable is normalized, as a z-score. (6) Correct: Indicator variable equal to 1 if the student's evaluation of his/her performance in Mathematics using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) coincides with his/her position in the distribution of the test scores in Mathematics inside the corresponding school. (7) Over estimates performance: Indicator variable equal to 1 if the student's evaluation of his/her performance in Mathematics using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) is above his/her position in the distribution of the test score in Mathematics inside the corresponding school. (8) Under estimates performance: Indicator variable equal to 1 if the student's evaluation of his/her performance in Mathematics using a 3-point Likert-type scale variable from 1 (bad) to 3 (good) is below his/her position in the distribution of the test score in Mathematics inside the corresponding school. Columns (1)-(8) include student-level controls. The full list of controls is presented in Section 4. Standard errors, reported in parentheses, are clustered at the school level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. P-values from randomization-inference tests are reported in brackets.

Table F3: Main treatment effects - aggregated outcomes employing z-scores

| | Technology use | Teachers' absenteeism & motivation | Students' absenteeism & motivation | Teachers' time allocation | Observed class delivery | Students' time allocation | Students' behavior and interaction at school | Teachers' cognitive skills | Students' cognitive skills |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| ProFuturo | 0.191*** | 0.140*** | 0.056* | 0.129** | 0.107 | 0.069* | 0.041** | 0.041 | 0.051 |
| | (0.038) | (0.033) | (0.032) | (0.055) | (0.070) | (0.041) | (0.019) | (0.047) | (0.043) |
| | [0.002] | [0.003] | [0.193] | [0.070] | [0.231] | [0.204] | [0.118] | [0.481] | [0.352] |
| Observations | 124 | 124 | 124 | 116 | 93 | 117 | 124 | 116 | 124 |
| $R^2$ | 0.54 | 0.45 | 0.32 | 0.20 | 0.34 | 0.30 | 0.36 | 0.29 | 0.77 |
| Mean (control group) | 0.00 | 0.03 | 0.03 | 0.00 | -0.02 | 0.06 | 0.00 | -0.03 | -0.01 |

Note: Estimates based on OLS regressions. Columns (1) to (4) and columns (6) to (9) present estimates using Equation 1. Column (5) presents estimates using Equation 2. Outcomes are grouped in indices that are built using the procedure in Kling, Liebman, and Katz (2007). We calculate school-grade means for each individual outcome and we compute within-sample z-scores for each school-grade outcome, employing the mean and the standard deviation of the control group. We then obtain the unweighted average z-score for each category. The indices are defined by the following outcomes: (1) Technology use: built from outcomes of Table 1; (2) Teachers' absenteeism and motivation: built from outcomes of Table 2; (3) Students' absenteeism and motivation: built from outcomes of Table 3; (4) Teachers' time allocation: built from the outcomes in Table 4; (5) Observed class delivery: built from the outcomes in Table 5; (6) Students' time allocation: built from the outcomes in Table 6; (7) Students' behavior and interaction at school: built from the outcomes of Table 7; (8) Teachers' cognitive skills: built from the outcomes in Table 8; (9) Students' cognitive skills: built from the outcomes in Table 9. All specifications employed include grade and stratum fixed effects. Standard errors, reported in parentheses, are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. P-values from randomization-inference tests are reported in brackets.

# G Robustness of estimates to control variables: Post-Double Selection LASSO procedure

Table G1: Comparison with Post-Double Selection LASSO.

| Outcome | Data source | Coeff. | p-value | Exact p-value | Post-Double Selection LASSO Coef. | Post-Double Selection LASSO p-value | Post-Double Selection LASSO Exact p-value |
|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| Table 1: Use of technology. | | | | | | | |
| Computer use | Teachers' survey | 0.090 | 0.000 | 0.001 | 0.083 | 0.000 | 0.004 |
| Index of technology usage - basic | Students' survey | 0.093 | 0.079 | 0.229 | 0.097 | 0.079 | 0.254 |
| Index of technology usage - advanced | Students' survey | -0.025 | 0.664 | 0.751 | 0.000 | 0.998 | 0.998 |
| Desire to use more technology at school | Students' survey | 0.027 | 0.001 | 0.012 | 0.029 | 0.000 | 0.015 |
| Time using technology at home | Guardians' survey | 0.232 | 0.005 | 0.043 | 0.171 | 0.025 | 0.135 |
| Table 2: Motivation. Teachers. | | | | | | | |
| Number of days missing school (previous month) | School admin. data | -0.685 | 0.001 | 0.037 | -0.709 | 0.000 | 0.025 |
| Number of days missing school (academic year) | School admin. data | -0.396 | 0.000 | 0.004 | -0.332 | 0.000 | 0.012 |
| Number of days missing school | Teachers' survey | -0.102 | 0.013 | 0.103 | -0.091 | 0.025 | 0.157 |
| Number of days arriving late to school | Teachers' survey | 0.165 | 0.110 | 0.234 | 0.156 | 0.128 | 0.232 |
| Teachers' motivation | Teachers' survey | 0.056 | 0.361 | 0.494 | 0.053 | 0.345 | 0.474 |
| Teachers' motivation | Guardians' survey | 0.075 | 0.351 | 0.509 | 0.091 | 0.195 | 0.341 |
| Table 3: Motivation. Students and guardians. | | | | | | | |
| Positive attitude towards school | Students' survey | 0.005 | 0.169 | 0.358 | 0.006 | 0.134 | 0.286 |
| Likes to study Maths | Students' survey | 0.015 | 0.011 | 0.091 | 0.016 | 0.008 | 0.068 |
| Likes to read | Students' survey | 0.004 | 0.469 | 0.610 | 0.008 | 0.183 | 0.374 |
| Overall school satisfaction | Guardians' survey | 0.133 | 0.093 | 0.207 | 0.089 | 0.188 | 0.356 |
| Number of days missing school | Students' survey | 0.012 | 0.434 | 0.618 | 0.008 | 0.630 | 0.724 |
| Table 4: Time allocation. Teachers. | | | | | | | |
| Quality of class preparation index | Teachers' survey | 0.121 | 0.040 | 0.114 | 0.115 | 0.048 | 0.155 |
| Time allocated to teaching | Teachers' survey | 0.042 | 0.084 | 0.225 | 0.029 | 0.255 | 0.433 |
| Time allocated to planning at school | Teachers' survey | -0.057 | 0.078 | 0.220 | -0.047 | 0.207 | 0.387 |
| Time allocated to planning at home | Teachers' survey | -0.041 | 0.319 | 0.463 | -0.046 | 0.236 | 0.428 |
| Time allocated to administrative activities | Teachers' survey | -0.013 | 0.764 | 0.852 | -0.009 | 0.833 | 0.905 |

Note: Estimates based on OLS regression (see equation 1). Exact p-values from randomization-inference tests are reported in columns (3) and (6). In columns (1)-(3), the specifications are constant across outcome variables by source of data (see Section 4. In columns (4)-(6), the specifications are outcome-specific and include controls which are selected using the Post-Double Selection LASSO procedure.

Table G2: Comparison with Post-Double Selection LASSO.

| Outcome | Data source | Coeff. | p-value | Exact p-value | Post-Double Selection LASSO Coef. | p-value | Exact p-value |
|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| Table 6: Time allocation. Students. | | | | | | | |
| Time allocated to reading | Guardians' survey | 0.228 | 0.002 | 0.032 | 0.270 | 0.000 | 0.011 |
| Time allocated to studying | Guardians' survey | 0.069 | 0.358 | 0.512 | 0.077 | 0.196 | 0.360 |
| Time allocated to playing | Guardians' survey | 0.151 | 0.011 | 0.053 | 0.131 | 0.036 | 0.167 |
| Shared time using technology | Guardians' survey | 0.109 | 0.095 | 0.278 | 0.029 | 0.680 | 0.784 |
| Shared time studying | Guardians' survey | 0.061 | 0.137 | 0.300 | -0.009 | 0.869 | 0.908 |
| Table 7: Students' behavior and interactions at school. | | | | | | | |
| Students' altruism | Students' survey | 0.025 | 0.009 | 0.067 | 0.028 | 0.003 | 0.039 |
| Students get help from peers | Students' survey | 0.019 | 0.058 | 0.200 | 0.022 | 0.042 | 0.163 |
| Students have friends at school | Students' survey | 0.002 | 0.809 | 0.866 | 0.003 | 0.768 | 0.835 |
| Table 8: Cognitive skills. Teachers. | | | | | | | |
| Cognitive test Portuguese | Teachers' survey | -0.151 | 0.015 | 0.099 | -0.155 | 0.018 | 0.112 |
| Cognitive test Maths | Teachers' survey | -0.093 | 0.241 | 0.363 | -0.098 | 0.224 | 0.363 |
| Self-assessment overall | Teachers' survey | 0.054 | 0.287 | 0.451 | 0.031 | 0.539 | 0.647 |
| Self-assessment Portuguese | Teachers' survey | 0.090 | 0.188 | 0.325 | 0.125 | 0.066 | 0.191 |
| Self-assessment Maths | Teachers' survey | 0.090 | 0.176 | 0.315 | 0.082 | 0.159 | 0.344 |
| Self-assessment Science | Teachers' survey | 0.129 | 0.026 | 0.124 | 0.082 | 0.058 | 0.203 |
| Table 9: Cognitive skills. Students. | | | | | | | |
| Digit span | Students' survey | 0.037 | 0.439 | 0.602 | 0.006 | 0.904 | 0.927 |
| Test score Portuguese | Students' survey | -0.007 | 0.914 | 0.944 | 0.043 | 0.432 | 0.579 |
| Test score Maths | Students' survey | -0.015 | 0.803 | 0.841 | -0.022 | 0.730 | 0.804 |
| Test score Science | Students' survey | 0.066 | 0.022 | 0.100 | 0.091 | 0.000 | 0.013 |

Note: Estimates based on OLS regression (see equation 1). Exact p-values from randomization-inference tests are reported in columns (3) and (6). In columns (1)-(3), the specifications are constant across outcome variables by source of data (see Section 4. In columns (4)-(6), the specifications are outcome-specific and include controls which are selected using the Post-Double Selection LASSO procedure.

# H Robustness of estimates to Inverse Probability of Treatment Weighting

In order to improve the comparability of the treatment and control schools at the individual level, we estimate each person's likelihood of belonging to the treatment group, using individuals' characteristics at baseline. To estimate this propensity score we estimate the following probit model:

$$P_{is}(\text{ProFuturo}) = \Phi(\alpha + X'_{is}\gamma + \upsilon_{is}) \tag{3}$$

where $P_{is}(\text{ProFuturo})$ is the probability of being in a treated school at baseline for individual $i$ in school $s$. Note that individual $i$ can be a teacher, a student, or a student's guardian. $\Phi()$ represents the cumulative distribution function of the standard normal, $X_{is}$ is a set of individual characteristics including strata fixed effects, for either teachers, students, and/or students' guardians depending on the sample at stake. As the potential set of individual characteristics is large, we use the LASSO procedure to obtain the propensity score. Penalty levels for the LASSO were determined using information criteria, in particular the bias-corrected Akaike Information Criteria (AIC), developed for small samples, and shown to have good selection performance (Ahrens, Hansen, and Schaffer, 2018).

Since the distribution of baseline characteristics is similar between treated and untreated individuals conditional on the propensity score (see Tables H1 - H2), we can obtain unbiased treatment effects by using the inverse of the probability of receiving the treatment as weights to create an artificial population in which demographic characteristics are not difference between comparison groups (Joffe et al., 2004). To do so, we calculate the Inverse Probability of Treatment Weighting (IPTW) for each individual as:

$$IPTW_{is} = \frac{A_s}{\widehat{P_{is}}(\text{ProFuturo})} + \frac{1 - A_s}{1 - \widehat{P_{is}}(\text{ProFuturo})} \tag{4}$$

where $A_s$ is equal to 1 if school s was assigned to the treatment group, and zero otherwise.

Tables H3 - H4 show Weighted Least Squares estimates accounting for IPTW for the outcomes variables in Tables 1-9.

Table H1: Balance. IPTW.

| | Observations | Control mean | Difference between treatment and control (s.e.) |
|---|---|---|---|
| Teachers | | | |
| Female (=1) | 629 | 0.40 | -0.03<br>(0.04) |
| Age | 629 | 36.49 | 0.35<br>(1.12) |
| Married (=1) | 623 | 0.45 | 0.04<br>(0.05) |
| Has children (=1) | 628 | 0.74 | 0.05<br>(0.04) |
| Completed 12 years of education (=1) | 629 | 0.79 | -0.02<br>(0.03) |
| Professional/technical degree (=1) | 629 | 0.21 | 0.02<br>(0.03) |
| University studies (=1) | 629 | 0.36 | 0.08***<br>(0.03) |
| Teaching experience (years) | 626 | 3.02 | 0.11<br>(0.09) |
| Owns house (=1) | 517 | 0.52 | 0.06<br>(0.05) |
| Has piped water (=1) | 517 | 0.40 | 0.02<br>(0.04) |
| Index IT goods (0-1) | 517 | 0.46 | 0.01<br>(0.02) |
| Students | | | |
| Female (=1) | 2519 | 0.53 | -0.02<br>(0.02) |
| Age | 2323 | 11.09 | 0.03<br>(0.08) |
| Attended kindergarden (=1) | 2512 | 0.79 | -0.02<br>(0.02) |
| Failed at least one course (=1) | 2519 | 0.30 | 0.01<br>(0.02) |
| Test score - overall (standardized) | 1476 | -0.00 | 0.08<br>(0.10) |
| Guardians | | | |
| Female (=1) | 1066 | 0.56 | -0.02<br>(0.04) |
| Age | 1066 | 39.00 | -0.98*<br>(0.54) |
| Number of children | 1045 | 1.04 | 0.03*<br>(0.02) |
| Completed 12 years of education (=1) | 1066 | 0.77 | -0.03<br>(0.03) |
| University studies (=1) | 1066 | 0.17 | -0.04**<br>(0.02) |
| Owns house (=1) | 1066 | 0.60 | -0.01<br>(0.03) |
| Has piped water (=1) | 1066 | 0.38 | 0.03<br>(0.05) |
| Index IT goods (0-1) | 1066 | 0.38 | -0.02<br>(0.02) |

Note: This table replicates Table D1. The difference is that the specifications apply IPTW to account for sample imbalance at the individual level. Standard errors, reported in parentheses, are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1.

Table H2: Balance after attrition. IPTW.

| Teachers | Observations | Control mean | Difference between treatment and control (s.e.) |
|---|---|---|---|
| Female (=1) | 517 | 0.41 | -0.03<br>(0.04) |
| Age | 517 | 36.51 | 0.13<br>(1.30) |
| Married (=1) | 513 | 0.45 | 0.04<br>(0.05) |
| Has children (=1) | 516 | 0.75 | 0.04<br>(0.05) |
| Completed 12 years of education (=1) | 517 | 0.78 | -0.04<br>(0.03) |
| Professional/technical degree (=1) | 517 | 0.22 | 0.04<br>(0.03) |
| University studies (=1) | 517 | 0.34 | 0.08**<br>(0.03) |
| Teaching experience (years) | 514 | 2.99 | 0.17<br>(0.11) |
| Owns house (=1) | 517 | 0.52 | 0.06<br>(0.05) |
| Has piped water (=1) | 517 | 0.40 | 0.02<br>(0.04) |
| Index IT goods (0-1) | 517 | 0.46 | 0.01<br>(0.02) |
| Students | | | |
| Female (=1) | 2323 | 0.54 | -0.04**<br>(0.01) |
| Age | 2323 | 11.09 | 0.03<br>(0.08) |
| Attended kindergarden (=1) | 2315 | 0.80 | -0.03**<br>(0.02) |
| Failed at least one course (=1) | 2322 | 0.30 | 0.01<br>(0.02) |
| Test score - overall (standardized) | 1379 | 0.02 | 0.08<br>(0.10) |
| Guardians | | | |
| Female (=1) | 671 | 0.57 | -0.02<br>(0.05) |
| Age | 671 | 39.00 | -0.08<br>(0.64) |
| Number of children | 659 | 1.04 | 0.01<br>(0.02) |
| Completed 12 years of education (=1) | 671 | 0.77 | -0.03<br>(0.03) |
| University studies (=1) | 671 | 0.17 | -0.05**<br>(0.02) |
| Owns house (=1) | 671 | 0.61 | -0.00<br>(0.04) |
| Has piped water (=1) | 671 | 0.37 | 0.01<br>(0.06) |
| Index IT goods (0-1) | 671 | 0.36 | -0.02<br>(0.02) |

Note: This table replicates Table A1. The difference is that the specifications apply IPTW to account for sample imbalance at the individual level. Standard errors, reported in parentheses, are clustered at school level. *** p<0.01, ** p<0.05, * p<0.1.

Table H3: Comparison with IPTW.

| Outcome | Data source | Coeff. | p-value | Exact p-value | Inverse Probability of Treatment Weighting Coef. | p-value | Exact p-value |
|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| Table 1: Use of technology. | | | | | | | |
| Computer use | Teachers' survey | 0.090 | 0.000 | 0.001 | 0.091 | 0.000 | 0.002 |
| Index of technology usage - basic | Students' survey | 0.093 | 0.079 | 0.229 | 0.106 | 0.059 | 0.205 |
| Index of technology usage - advanced | Students' survey | -0.025 | 0.664 | 0.751 | -0.012 | 0.848 | 0.891 |
| Desire to use more technology at school | Students' survey | 0.027 | 0.001 | 0.012 | 0.028 | 0.001 | 0.024 |
| Time using technology at home | Guardians' survey | 0.232 | 0.005 | 0.043 | 0.213 | 0.019 | 0.101 |
| Table 2: Motivation. Teachers. | | | | | | | |
| Number of days missing school (previous month) | School admin. data | -0.685 | 0.001 | 0.037 | -0.695 | 0.001 | 0.036 |
| Number of days missing school (academic year) | School admin. data | -0.396 | 0.000 | 0.004 | -0.388 | 0.000 | 0.009 |
| Number of days missing school | Teachers' survey | -0.102 | 0.013 | 0.103 | -0.104 | 0.014 | 0.095 |
| Number of days arriving late to school | Teachers' survey | 0.165 | 0.110 | 0.234 | 0.158 | 0.176 | 0.319 |
| Teachers' motivation | Teachers' survey | 0.056 | 0.361 | 0.494 | 0.065 | 0.328 | 0.447 |
| Teachers' motivation | Guardians' survey | 0.075 | 0.351 | 0.509 | 0.073 | 0.329 | 0.514 |
| Table 3: Motivation. Students and guardians. | | | | | | | |
| Positive attitude towards school | Students' survey | 0.005 | 0.169 | 0.358 | 0.008 | 0.067 | 0.209 |
| Likes to study Maths | Students' survey | 0.015 | 0.011 | 0.091 | 0.020 | 0.002 | 0.030 |
| Likes to read | Students' survey | 0.004 | 0.469 | 0.610 | 0.008 | 0.203 | 0.370 |
| Overall school satisfaction | Guardians' survey | 0.133 | 0.093 | 0.207 | 0.134 | 0.119 | 0.246 |
| Number of days missing school | Students' survey | 0.012 | 0.434 | 0.618 | 0.010 | 0.527 | 0.658 |
| Table 4: Time allocation. Teachers. | | | | | | | |
| Quality of class preparation index | Teachers' survey | 0.121 | 0.040 | 0.114 | 0.128 | 0.071 | 0.185 |
| Time allocated to teaching | Teachers' survey | 0.042 | 0.084 | 0.225 | 0.050 | 0.036 | 0.146 |
| Time allocated to planning at school | Teachers' survey | -0.057 | 0.078 | 0.220 | -0.037 | 0.220 | 0.353 |
| Time allocated to planning at home | Teachers' survey | -0.041 | 0.319 | 0.463 | -0.046 | 0.287 | 0.457 |
| Time allocated to administrative activities | Teachers' survey | -0.013 | 0.764 | 0.852 | 0.011 | 0.797 | 0.858 |

Estimates based on OLS regression in columns (1)-(3) (see equation 1). Estimates based on IPTW regression in columns (4)-(6). Exact p-values from randomization-inference tests are reported in columns (3) and (6). The full list of controls is presented in Section 4.

Table H4: Comparison with IPTW.

| Outcome | Data source | Coeff. | p-value | Exact p-value | Inverse Probability of Treatment Weighting Coef. | p-value | Exact p-value |
|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| Table 6: Time allocation. Students. | | | | | | | |
| Time allocated to reading | Guardians' survey | 0.228 | 0.002 | 0.032 | 0.258 | 0.003 | 0.038 |
| Time allocated to studying | Guardians' survey | 0.069 | 0.358 | 0.512 | 0.044 | 0.568 | 0.722 |
| Time allocated to playing | Guardians' survey | 0.151 | 0.011 | 0.053 | 0.138 | 0.028 | 0.083 |
| Shared time using technology | Guardians' survey | 0.109 | 0.095 | 0.278 | 0.077 | 0.267 | 0.498 |
| Shared time studying | Guardians' survey | 0.061 | 0.137 | 0.300 | 0.132 | 0.016 | 0.096 |
| Table 7: Students' behavior and interactions at school. | | | | | | | |
| Students' altruism | Students' survey | 0.025 | 0.009 | 0.067 | 0.022 | 0.035 | 0.133 |
| Students get help from peers | Students' survey | 0.019 | 0.058 | 0.200 | 0.021 | 0.050 | 0.180 |
| Students have friends at school | Students' survey | 0.002 | 0.809 | 0.866 | 0.008 | 0.500 | 0.650 |
| Table 8: Cognitive skills. Teachers. | | | | | | | |
| Cognitive test Portuguese | Teachers' survey | -0.151 | 0.015 | 0.099 | -0.185 | 0.003 | 0.055 |
| Cognitive test Maths | Teachers' survey | -0.093 | 0.241 | 0.363 | -0.116 | 0.203 | 0.347 |
| Self-assessment overall | Teachers' survey | 0.054 | 0.287 | 0.451 | 0.069 | 0.219 | 0.378 |
| Self-assessment Portuguese | Teachers' survey | 0.090 | 0.188 | 0.325 | 0.114 | 0.113 | 0.245 |
| Self-assessment Maths | Teachers' survey | 0.090 | 0.176 | 0.315 | 0.099 | 0.148 | 0.284 |
| Self-assessment Science | Teachers' survey | 0.129 | 0.026 | 0.124 | 0.132 | 0.045 | 0.151 |
| Table 9: Cognitive skills. Students. | | | | | | | |
| Digit span | Students' survey | 0.037 | 0.439 | 0.602 | 0.028 | 0.565 | 0.680 |
| Test score Portuguese | Students' survey | -0.007 | 0.914 | 0.944 | 0.011 | 0.858 | 0.894 |
| Test score Maths | Students' survey | -0.015 | 0.803 | 0.841 | 0.002 | 0.971 | 0.977 |
| Test score Science | Students' survey | 0.066 | 0.022 | 0.100 | 0.071 | 0.015 | 0.086 |

Note: Estimates based on OLS regression in columns (1)-(3) (see equation 1). Estimates based on IPTW regression in columns (4)-(6). Exact p-values from randomization-inference tests are reported in columns (3) and (6). The full list of controls is presented in Section 4.