Journal of Development Economics 164 (2023) 103145

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

Can technology improve the classroom experience in primary education? An African experiment on a worldwide program☆

Joana Cardim [a,b], Teresa Molina-Millán [c,b,d], Pedro C. Vicente [e,b,f,g,*]

[a] Education Policy Institute, United Kingdom
[b] NOVAFRICA, Portugal
[c] Universidad de Alicante, Spain
[d] IZA, Germany
[e] Nova School of Business and Economics, Portugal
[f] CEPR, United Kingdom
[g] BREAD, United States of America

ARTICLE INFO

Dataset link: http://dx.doi.org/10.17632/jdy8b542cs.1

JEL classification:
O12
I21

Keywords:
Primary education
Computer-assisted learning
CAL
Field experiment, RCT
Africa
Angola

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 but only in the most popular subject in the CAL platform.

1. Introduction

Human capital is widely considered to be vital for economic 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 et al. (2012), Molina-Millán et al. (2019). But the role of school inputs in the classroom should not be disregarded.

While some attention has been devoted to classical school inputs targeting equally all students in the classroom, these could be inadequate to help strengthening education quality.[1] Targeting more and better

☆ 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, Spain. This work was also funded by Fundação para a Ciência e a Tecnologia, Portugal (SFRH/BD/131707/2017, UID/ECO/00124/2013, UID/ECO/00124/2019), POR Lisboa, Portugal (LISBOA-01-0145-FEDER-007722, LISBOA-01-0145-FEDER-022209) and POR Norte, Portugal (LISBOA-01-0145-FEDER-022209). Teresa Molina-Millán acknowledges financial support from Generalitat Valenciana, Spain CDEIGENT/2020/016 and from the Spanish Ministry of the Economy and Competitiveness , Grant ECO2017-87069-P. Ethics clearance was given by Universidade Nova de Lisboa. All remaining errors are our responsibility.

* Correspondence to: Nova School of Business and Economics, Universidade Nova de Lisboa, Rua da Holanda 1, 2775-405 Carcavelos, Portugal.
E-mail addresses: joana.cardimdias@epi.org.uk (J. Cardim), teresa.molina@ua.es (T. Molina-Millán), pedro.vicente@novasbe.pt (P.C. Vicente).

[1] Glewwe et al. (2009) show that textbooks only help best-performing students in Kenya, where the educational system is geared towards elite students.

https://doi.org/10.1016/j.jdeveco.2023.103145
Received 17 February 2022; Received in revised form 22 June 2023; Accepted 30 June 2023
Available online 8 July 2023

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 et al. (2019). Important questions do however remain about how to link these CAL programs to regular teachers and classroom dynamics. In particular, involving teachers could help a variety of outcomes including decreasing their absenteeism.[2]

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, which are central to the analysis and contribution of this paper. First, ProFuturo's 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, which link all students and the teacher in a classroom. We ask in this paper whether the program is effective at familiarizing 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 also investigate the effects of the implementation of ProFuturo on children's cognitive skills.

The context of our study is the capital city of Luanda, Angola.[3] 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 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 systematic observation of classes, 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 0.59 standard deviations in days missed by teachers, which represents a 51 percent reduction. 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, in line with ProFuturo's emphasis on students' interaction and other-regarding behaviors. 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 significant treatment effects in students' test scores of other subjects despite the centrality of those contents in the design of ProFuturo.

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 et al. (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.[4]

One pedagogical approach that seems to have a role in student achievement is Computer Assisted Learning (CAL). This includes making hardware available to students along with a specific software designed to develop particular skills. 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 et al. (2009) find significant improvements in pre-algebra and algebra skills after a targeted CAL program was implemented. These authors hypothesize that the referred 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 employed technology that

[2] A recent example is Beg et al. (2022). However, educational policies involving teachers through peer guidance and through evaluation are generally effective (Jackson et al., 2014).

[3] 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.

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

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

Muralidharan et al. (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 Math 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 Math outcomes instead of Language. Some specific examples are Rouse and Krueger (2004) for an early American program focusing on Language, and Carrillo et al. (2011) for a more recent program in Ecuador with a large contrast between Math 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, not contributing directly to student–teacher interactions.[5] Escueta et al. (2017) claim that little is known about how CAL programs interact with teachers' efforts.[6] Beg et al. (2022) 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 Math and Science tests increased after four 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. And the mixing of teacher implementation with individual customization could be particularly promising: Berry et al. (2020) suggest that continuous assessment of students and the teaching practices should be jointly set.[7] 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 Math learning in low-performing primary schools in Chile. Students improved learning outcomes but the program increased Math 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 sufficient tablets for all students. This technology may be easily implemented in the context of a developing country, as batteries last for nine hours and its software runs only offline. Each tablet is equipped with a software, produced on purpose for ProFuturo.[8] 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 level of computer proficiency as perceived 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 1208 teachers and around 100,000 children in primary schools. These schools typically serve children from disadvantaged socioeconomic neighborhoods in the main cities of Angola.

[5] The recent study by Kerwin and Thornton (2021) in Uganda fleshes out the importance of input complementarity in pedagogical interventions.

[6] This is despite the fact that there is significant evidence of positive impacts of peer interaction and learning between teachers on student performance (Jackson and Bruegmann, 2009; Jackson et al., 2014; Papay et al., 2020).

[7] Another study (Naslund-Hadley et al., 2014) focusing on an interactive audio aid in Math for pre-school teachers in Paraguay has found positive and significant improvements in standardized test scores. Jackson and Makarin (2018) show that middle-school teachers in the US integrate off-the-shelf Math online content in their classes with realized improvements in student performance (particularly for the weaker teachers).

[8] The ProFuturo protocol implied that no other software was installed in their tablets.

2.1. Program adoption in Luanda

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.

The school principals that had ProFuturo introduced in their schools decided on which classes received the treatment. The majority of classes selected belonged to grades 4, 5, and 6, i.e., the highest grades of primary school. Seventy-six percent of all classes in these grades received the intervention.[9]

On average, the duration of a ProFuturo class, as reported by school principals, was 106 min. 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 min per week according to principals. ProFuturo coordinators reported similar levels of exposure: 136 min per week, on average. Note however that these numbers apply only to students attending classes assigned to the program.

It is important to note that Science contents were the ones most frequently selected by teachers using the ProFuturo platform.[10] 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). There was a clear emphasis on group work and active participation of students in classroom activities.

3. Experimental design

In the end of 2017, ProFuturo selected 42 Catholic schools in Luanda to be included in this study.[11] In Angola, we have no evidence that students in private or religious schools are systematically different from those attending public schools.[12] 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,[13] 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).

3.1. 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 the ProFuturo's software package. Figure B1 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 interviews, 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 November 2017 to March 2018, and for the endline surveys, 11 months later, from November 2018 to May 2019.[14]

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 and attitudes 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 on the guardian's perception of teacher motivation, his/her satisfaction with the school, as well as the student's 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.[15] 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

[9] In the Online Appendix, Table A2, we show the differences in the characteristics of treated vs. untreated teachers, students, and guardians in the ProFuturo schools. We do not find clear signs of selection of treatment within schools with the exception of baseline student performance, which appears to be higher in treated classes.

[10] In surveys of teachers implementing ProFuturo, the content in Science was also found to be most aligned with the Angolan curriculum, when compared to other subjects.

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

[12] We employ a nationally representative household survey, collected by the National Statistics Office in Angola (IBEP,2008–2009), to assess differences between households with children attending public schools and children attending private or religious schools. In this survey, households corresponding to 1051 students attending public schools and to 1050 attending private or religious schools in Luanda were interviewed. We find no systematic differences between the two sub-samples when considering education of household head, ownership of a house, of a cell phone, and of a computer, as well as access to piped water and the internet.

[13] 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.

[14] At baseline, all teachers were surveyed before the academic year began, while data on students, caregivers, and class observations were collected from February to March 2018. At endline, close to half of the teachers and all students from sixth grade and their caregivers were interviewed before the end of the academic year (December 2018), while the remaining teachers, students, and caregivers were interviewed in 2019. Grade 6 is the last grade in Angolan primary schools: students who pass this grade and continue studying go on to other (secondary) schools.

[15] 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.

our measurement design, one shortly after the beginning of treatment 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 Math 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 targets children from 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.[16] 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. Still, we employ these data to formulate a better understanding of the main activities and subjects studied through the platform, as well as to construct a treatment variable at the level of the classroom.[17]

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.[18] 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 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 an improved learning environment for 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 ProFuturo_s is a treatment indicator taking value 1 for schools which were assigned to receive ProFuturo and 0 otherwise. y_{is0} is the baseline value of the dependent variable. 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.[19] ϵ_{is} 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 ANCOVA specification maximizes statistical power in field experiments (McKenzie, 2012).

Given that baseline values of the outcome variable are not available for classroom observation data, we employ the following specification when using 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.

[16] 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.

[17] 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 with 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.

[18] For details on the proposed research design prior to data access, see Research design, June 2017.

[19] Control variables are as follows. When analyzing outcome variables at the level of the teacher: gender, age, age squared, indicator variables for whether the respondent has at least 12 years of education, a professional or technical degree, or university studies, and respondent's month of interview. 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, indicators variables for whether the respondent has at least 9 years of education, or university studies, and respondent's month of interview, as well as number of members of the household.

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 36 years old, 45 percent 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, house ownership, access to piped water and ownership of IT goods. The exception is that teachers in the treatment group are 11 percentage points more likely to have completed university studies.

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 10 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. There are no significant differences with respect to students' baseline test scores.[20]

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 significantly younger, by 1.2 years. Guardians from the treatment group are also 5 percentage points less likely to have completed university studies, from a baseline of 17 percent for the control group. In term of households' wealth, on average 60 percent of the caregivers own a house and 38 percent of them have piped water. Ninety four percent of the sample owns a mobile phone, 26 percent has access to internet at home, and 48 percent owns a computer.[21] There are no significant differences with respect to ownership of assets and IT goods.[22]

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 control 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).[23]

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 frequency of computer use by teachers during classes. Columns (2) to (4) refer 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 employ computers more frequently during their classes when compared to teachers in the control group. The magnitude of the effect is 0.79 standard deviations, significant at the 1 percent level. The exact p-value from randomization inference is also 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 0.11 standard deviations, significant at the 5 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 2.5 percentage points and it is significant at the 1 percent level. The exact p-value is 0.025. Guardians also report an increase in the use of technology at home by their children. The size of this effect is 0.24 standard deviations, statistically significant at the 5 percent level. The randomization inference p-value is 0.08.

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 during the classes they teach, but also at home, as is the case for students. Although schools had a reasonable level of discretion on how to implement ProFuturo, this is reassuring evidence that the program led to a wide range of effects on behaviors and perceptions related to the use of technology.

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. The significance of these effects is confirmed by self-reported survey data on missing days of classes. The effect

[20] Figure D1 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.

[21] We employ the Demographic and Health Survey (DHS) data for Angola (2015–2016) to assess how our sample of households with children enrolled in fourth to sixth grade in our sample of Catholic schools in Luanda compares to children enrolled in the same grades regardless of the type of school. Numbers in the DHS are comparable. About 44 percent of households surveyed in the DHS have access to piped water. Around 96 percent of households own a mobile phone, 36 percent have internet, and 30 percent a computer.

[22] Note that sample averages for the guardians interviewed in the baseline survey should be taken with caution. We managed to interview 1070 caregivers from a targeted sample of 2520. 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.

[23] 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 38 percent, while in the students' survey it was 7 percent. In the guardians' survey the attrition rate was larger: in the endline we interviewed 44 percent of the baseline sample. Overall, attrition rates are not significantly different between the treatment and the control groups. Appendix Table E2 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.

Table 1
Use of technology.

	Teachers' survey	Students' survey			Guardians' survey
	Frequency of 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	0.786***	0.112**	−0.029	0.025***	0.239**
	(0.087)	(0.055)	(0.061)	(0.007)	(0.090)
	[0.000]	[0.178]	[0.762]	[0.025]	[0.077]
Observations	489	2314	2314	2307	631
R^2	0.27	0.23	0.21	0.06	0.18
Mean (control group)	0.00	0.00	0.00	0.91	0.00

Note: Estimates based on OLS regressions. All columns present estimates using Eq. (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) Frequency of computer use: variable averaging a set of categorical variables measuring the frequency at which the teacher used a computer in each of the classes he/she taught during 2018. The question employed a scale from 1 (Never) to 5 (Almost always or always). This variable is normalized, i.e., as a z-score. (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.

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 0.59 standard deviations, which represents a 51 percent reduction. This result is statistically significant at the 1 percent level. The randomization inference *p*-value is 0.05. We also observe that ProFuturo reduces the number of days that teachers were absent during the complete school year by 0.43 standard deviations (representing a 22 percent reduction), 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 0.09 standard deviations, which is significant at the 5 percent level. The randomization inference *p*-value is 0.11.[24] Note that teachers in treated schools are more likely to self-report to be late when arriving at school. This effect is significant at the 10 percent level, and 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 0.28 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.[25]

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 on students' motivation.

Students in treatment schools report they like Mathematics more than in control schools. ProFuturo increased the likelihood students report liking Mathematics by 1.6 percentage points, which is statistically significant at the 1 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 17 percentage points, 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.

24 The estimation sample in columns 1 and 2 consists of teachers interviewed in the baseline and for whom we have administrative records. The probability of having administrative records is not significantly different between the treatment and control groups. Overall, teachers tend to under report the number of days missing school. Administrative records show than on average teachers in treated and control schools at baseline missed 5 days of school in the month previous to the survey. Using self-reported data this average goes down to 1 day. Self-report biases could explain the difference in treatment effects on days missed in the administrative relative to the survey data. For clarity, we include in the Online Appendix, the same regressions as in columns (1)-(4) of Table 2 but with the outcome variables expressed in days (Table F1). Beg et al. (2022) find positive but smaller effects on teacher attendance in Pakistan, of 1 percentage point. Note, however, that the level of teacher attendance in the control group of that study is 94 percent, much larger than in our setting, which is approximately 66 percent in our control group. In addition, the intervention in Beg et al. (2022), being based on videos, had a clear potential for substitutability with the teacher presence as mentioned in their paper. This potential for substitutability was arguably higher than in ProFuturo, which was designed for submission by the teachers.

25 The increasing motivation pattern for teachers could be related to increased productivity with the program. Kerwin and Thornton (2021) find a similar pattern.

Table 2
Motivation. Teachers.

	School administrative data		Teachers' survey			Guardians' survey	Class observation data	
	Number of days missing school		Number of days				Round 2	Round 3
	(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	−0.592***	−0.430***	−0.094**	0.179*	0.026	0.088	0.065	0.281*
	(0.166)	(0.086)	(0.045)	(0.104)	(0.059)	(0.089)	(0.107)	(0.141)
	[0.052]	[0.004]	[0.114]	[0.217]	[0.744]	[0.478]	[0.672]	[0.141]
Observations	268	468	504	479	514	634	123	128
R^2	0.66	0.62	0.06	0.16	0.10	0.10	0.12	0.15
Mean (control group)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: Estimates based on OLS regressions. Columns (1)–(6) present estimates using Eq. (1) and columns (7)–(8) present estimates using Eq. (2). The estimation sample in columns (1)–(2) consists of school records of teachers on absenteeism; 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. All specifications include strata fixed effects. Columns (1)–(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 Math	Likes to read	Overall school satisfaction	Number of days missing school	Students' motivation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ProFuturo	0.006	0.016***	0.004	0.169*	0.011	−0.077	0.218
	(0.004)	(0.006)	(0.006)	(0.100)	(0.015)	(0.188)	(0.197)
	[0.317]	[0.052]	[0.623]	[0.182]	[0.664]	[0.757]	[0.382]
Observations	2314	2314	2301	664	2284	123	128
R^2	0.06	0.03	0.02	0.09	0.05	0.15	0.11
Mean (control group)	0.89	0.95	0.97	0.00	0.36	0.00	0.00

Note: Estimates based on OLS regressions. Columns (1)–(5) present estimates using Eq. (1) and columns (6)–(7) present estimates using Eq. (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 Math: indicator variable equal to 1 if the student agreed with the statement 'I like Math.' (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. All specifications include strata fixed effects. 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.

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' index of quality in preparing classes as well as teachers' self-reported allocation of time in a regular working week by activity. Specifically, the class 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 0.11 standard deviation units, significant at the 10 percent level, although not significant using randomization inference. For a regular week, teachers report to spend more time teaching. The effect magnitude is 5 percentage points, significant at the 10 percent level, although not significant with randomization inference. There are no statistical significant differences in time planning and undertaking administrative activities.

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	0.111*	0.046*	−0.059	−0.041	0.014
	(0.056)	(0.024)	(0.037)	(0.039)	(0.042)
	[0.148]	[0.227]	[0.255]	[0.466]	[0.831]
Observations	511	513	479	500	370
R^2	0.08	0.05	0.10	0.09	0.08
Mean (control group)	0.00	0.85	0.40	0.44	0.33

Note: Estimates based on OLS regressions. All columns present estimates using Eq. (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 h per week to teach. (3) Planning at school: indicator variable equals to 1 if the teacher allocated more than 5 h per week to planning classes at school. (4) Planning at home: indicator variable equals to 1 if the teacher allocated more than 5 h per week to planning classes at home. (5) Administrative activities: indicator variable equals to 1 if the teacher allocated more than 5 h per week to administrative work. All specifications include strata fixed effects, and 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)
	[0.758]	[0.830]	[0.110]	[0.708]
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)
	[0.981]	[0.096]	[0.810]	[0.072]
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 Eq. (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. All specifications include strata fixed effects, and 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 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 0.11). 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

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	0.233***	0.099	0.161**	0.134*	0.045
	(0.075)	(0.071)	(0.062)	(0.069)	(0.045)
	[0.048]	[0.372]	[0.056]	[0.173]	[0.476]
Observations	596	655	658	664	664
R^2	0.09	0.12	0.21	0.18	0.13
Mean (control group)	0.00	0.00	0.00	0.00	0.00

Note: Estimates based on OLS regressions. All columns present estimates using Eq. (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 min 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. All specifications include strata fixed effects, and 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.

during the classroom observation. The magnitudes 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 0.1 and 0.07.

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.[26] 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.[27]

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 0.16 standard deviations more time playing. Results are significant at the 1 percent and 5 percent levels, respectively. The corresponding randomization inference p-values are 0.05 and 0.06. It is possible that these changes compensated for decreased time working to help the family. Regarding the time treated guardians and their children spend together, it is more likely to be spent using technology, by additional 0.13 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 2.4 percentage points, significant at the 5 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 2 percentage points more likely to have received help from other students, which is statistically significant at the 5 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.[28]

Overall, we find some patterns consistent 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.

[26] We could observe in ProFuturo classes that there was no time devoted to monitoring copying activities (see Table A3 in Appendix A). A possibility is that, 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.

[27] Table F2 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 3.7 percentage points, significant at the 5 percent level. The exact p-value is 0.07.

[28] We find a positive correlation between students' behavior and interactions with their teachers and peers at school and students' performance within treated schools at the endline, namely for the best students. This is shown in Table F3 in Appendix F. It is suggestive that teamwork could work with the rest of ProFuturo in a complementary way.

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	0.024**	0.021**	0.003
	(0.009)	(0.010)	(0.010)
	[0.086]	[0.139]	[0.852]
Observations	2307	2314	2314
R^2	0.04	0.05	0.08
Mean (control group)	0.89	0.85	0.76

Note: Estimates based on OLS regressions. All columns present estimates using Eq. (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.'. All specifications include strata fixed effects, and 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	Math	Overall	Portuguese	Math	Science
	(1)	(2)	(3)	(4)	(5)	(6)
ProFuturo	−0.213***	−0.139	0.060	0.078	0.045	0.128**
	(0.065)	(0.083)	(0.048)	(0.068)	(0.065)	(0.063)
	[0.026]	[0.199]	[0.410]	[0.439]	[0.596]	[0.178]
Observations	491	491	508	463	441	412
R^2	0.13	0.19	0.17	0.12	0.19	0.13
Mean (control group)	0.00	0.00	0.00	0.00	0.00	0.00

Note: Estimates based on OLS regressions. Columns (1)–(6) present estimates using Eq. (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) Math: 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) Math: 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. All specifications include strata fixed effects, and 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.

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 −0.21 standard deviations, statistically significant at the 1 percent level (also significant but at the 5 percent level when using 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 specific self-assessments in Portuguese or Mathematics, but there is a positive treatment effect of 0.13 standard deviations in the self-assessment of knowledge in Science, significant at the 5 percent level, but 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 a positive treatment effect 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 statistical significance when employing randomization inference (the exact p-value is 0.07).[29] We

[29] Consistently, we show in Table F4 of the Online Appendix, that there is a significantly positive correlation between the number of months of exposure to Science contents under ProFuturo and the test scores in Science.

Table 9
Cognitive skills. Students.

	Students' survey			
		Test scores		
	Digit span	Portuguese	Math	Science
	(1)	(2)	(3)	(4)
ProFuturo	0.038	−0.003	−0.011	0.073**
	(0.048)	(0.061)	(0.062)	(0.027)
	[0.606]	[0.975]	[0.895]	[0.067]
Observations	2314	1008	1008	1008
R^2	0.15	0.25	0.39	0.39
Mean (control group)	0.00	0.00	0.00	0.00

Note: Estimates based on OLS regressions. All columns present estimates using Eq. (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) Math: 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. All specifications include strata fixed effects, and 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.

do not find any significant effects on other subjects.[30] We note that our study relates to a relatively low intensity treatment over a relatively short time window. Still other comparable studies in terms of intensity and duration find significant effects on students' test scores.[31]

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 Online 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 not systematic.[32]

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 et al. (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 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: Students' technology use, built from outcomes in columns (2)-(5) of Table 1; Teachers' absenteeism and motivation, built from outcomes in columns (2)-(8) 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.

Fig. 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 5 and 10 percent levels.[33] 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 teachers' absenteeism and motivation, and teachers' time allocation. The magnitudes of these effects are 0.11 and 0.15 standard deviations, significant at the 10 or 5 percent levels when employing exact p-values. We find effects on students' technology use, their absenteeism and motivation, their time allocation, and their behavior and interaction at school. However, these are not robust to randomization inference. We do not find significant effects on the remaining aggregates.

5.4. Local average treatment effects

Because not all students and teachers within ProFuturo schools were actually treated, it is important to ask whether our ITT effects differ from actual treatment effects. We focus on Local Average Treatment Effects (LATE) provided the plausible identification strategy offered by employing the binary variable defined by the initial assignment of ProFuturo to schools as an instrument variable. We employ two versions of actual treatment at the class level: the first employs reports from principals at the school level; the second is built from the ProFuturo

[30] Six randomly selected classes in each school performed the student tests. We consider the endline survey sample of students for which we have test scores data. We do not have evidence of significant treatment effects on whether students took the test. Table F5 in Appendix F shows treatment effects on students' self-assessment in Portuguese and Mathematics, and on students' ability to estimate their own performance. In Portuguese, students are 4.1 percentage points more likely to correctly estimate their performance, and 5.6 percentage points less likely to overestimate it, significant at the 10 percent level and at the 1 percent level, respectively.

[31] This is the case of Beg et al. (2022) and Banerjee et al. (2007). Other relevant studies that have higher intensity (and sometimes higher duration) are Muralidharan et al. (2019), Naslund-Hadley et al. (2014), and Araya et al. (2019).

[32] If one does not value the finding in Science stemming from the low intensity ProFuturo treatment, it is possible that the differences in context and involvement of teachers of ProFuturo in Angola could explain the less systematic results when comparing to the more established literature on extracurricular CAL, which reports clear test score effects (Banerjee et al., 2007; Muralidharan et al., 2019).

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

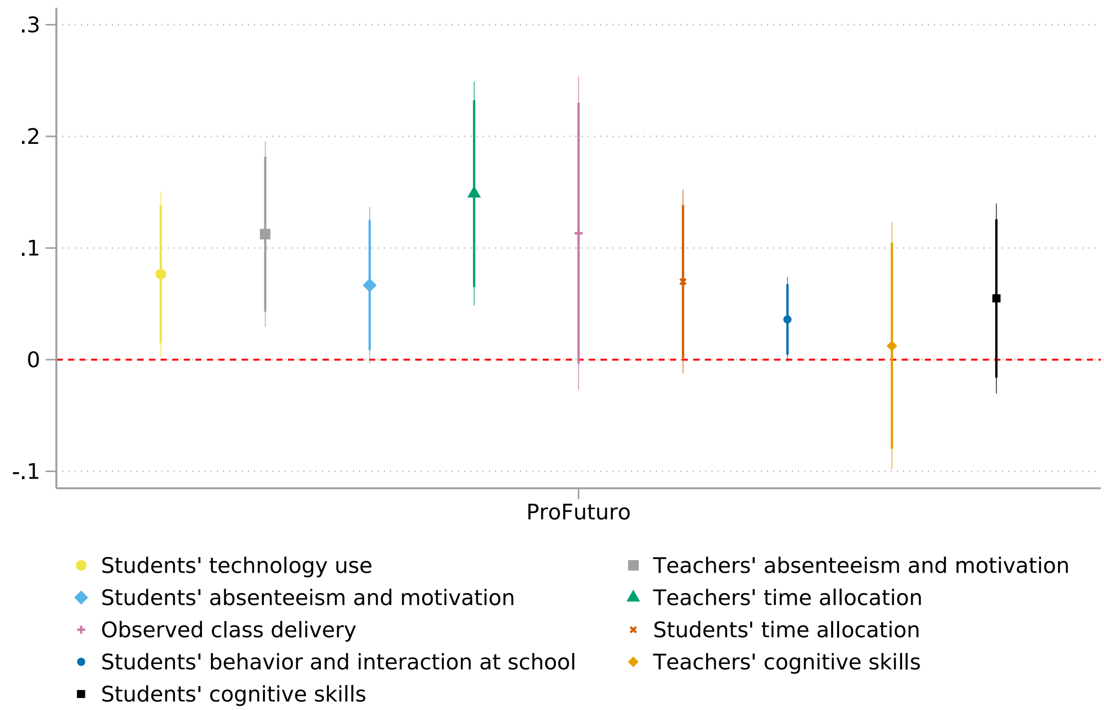

Fig. 1. Main treatment effects - aggregated outcomes employing z-scores.
Note: All estimates are based on OLS regressions using Eq. (1), except estimates on Observed class delivery which are based on an OLS regression employing Eq. (2). Outcomes are grouped in indices that are built using the procedure in Kling et al. (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) Students' technology use: built from outcomes in columns (2)-(5) of Table 1; (2) Teachers' absenteeism and motivation: built from outcomes in columns (2)-(8) 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 5 and 10 percent level. Standard errors are clustered at the school level.

platform data.[34] We show the LATE estimation for our main aggregate outcomes in Fig. 2.[35]

We find qualitatively similar effects when comparing the LATE (built from schools' reports or the ProFuturo platform) with the ITT effect of ProFuturo. Statistical significance is achieved for a wider set of outcome variables and effect magnitudes are comparable, although expectedly larger for the LATE. In the Online Appendix, Tables F7 and F8 we show the LATE for each individual outcome variable in the main tables of the paper and find the same pattern.[36] We conclude that selection of treatment within treated schools did not imply dramatic consequences over estimates of the impact of ProFuturo.[37]

5.5. Additional results and robustness

We now mention a few additional results and robustness that complement our main analysis in this paper. We first devote some attention to the distinction between motivation and monitoring in driving the impact on teachers, namely in terms of their absenteeism. To draw some light into these mechanisms, beyond our direct evidence on motivation of Table 2, we estimate heterogeneous effects of ProFuturo employing the share of ProFuturo teachers in each school. The hypothesis would be that faced with a lower share of treated teachers in the school, treated teachers would be more monitored. These results are shown in Table F12 of the Online Appendix. We find no clear evidence of significant interaction effects, i.e., of potential monitoring pressures, except when considering the self-reported measure of absenteeism, which could be more prone to bias. The same pattern arises when estimating the conditional correlation between visits by ProFuturo coordinators and absenteeism. This is shown in Table F13 in the Online Appendix.[38]

Second, we conduct a few robustness exercises, namely relating to the choice of control variables for teachers, students, and guardians, to the comparability of the treatment and control schools (by applying a weighting exercise), and to additional multiple hypothesis testing. Specifically, in Section G of the Appendix we show the replication of

[34] Both these data sources have imperfections. The first is gathered from reports in two points in time. The second does not include two treatment schools, which do not have platform data available, and has limitations in the ability to match classes.

[35] Since the results employing the data from the ProFuturo platform are very similar to ones employing the schools' reports, we leave the version employing the schools' reports to the Online Appendix, Figure F1.

[36] In Tables F9 and F10 of the Online Appendix, we also show impacts of treatment intensity on the individual outcome variables. Treatment intensity is measured as number of months treated, employing both the reports from principals and the ProFuturo platform. We find the sign and the statistical significance of intensity to be comparable to the ones found in our main estimates.

[37] In Appendix F, Table F11, we also show estimates of the treatment on the treated effect for teacher-level outcomes employing platform data for both treated and control schools, as for control teachers we use treatment data from 2019, after our experiment finished. Results are again comparable when considering the ITT.

[38] We also estimate heterogeneous effects of ProFuturo on teachers' outcome variables, employing the level of teachers' cognitive skills at the baseline. These results are shown in Table F14 of the Online Appendix. We find no clear evidence of significant and negative interaction effects, which implies we do not have supportive evidence in favor of substitutability between teachers' quality and ProFuturo.

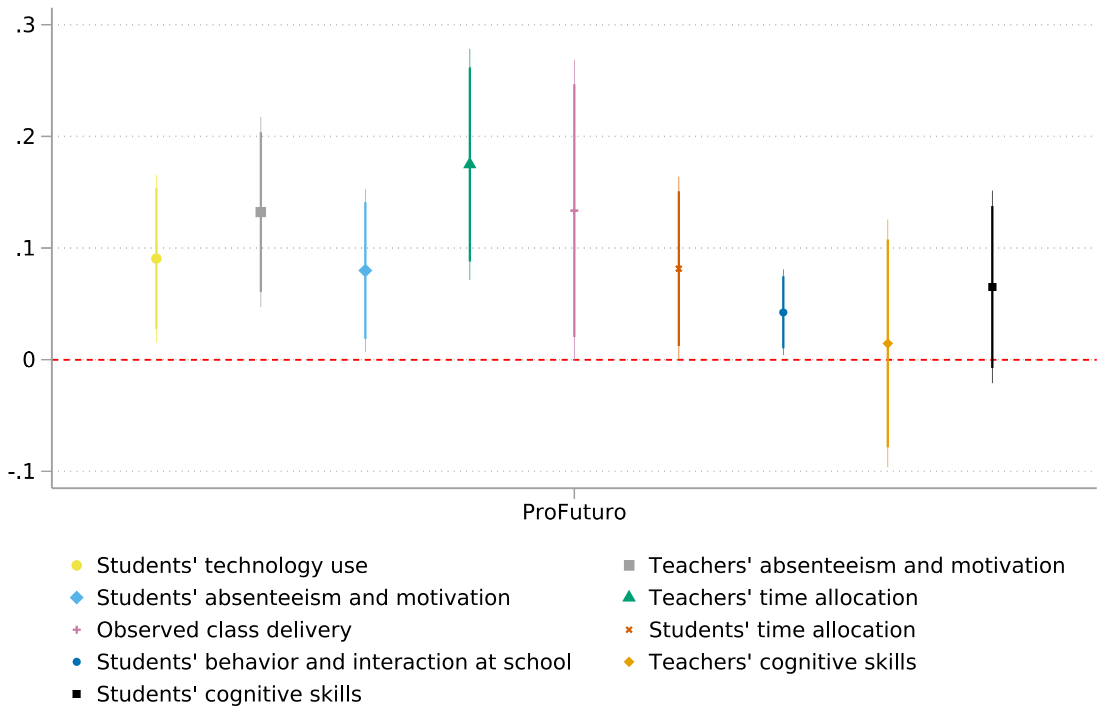

Fig. 2. Local average treatment effects - aggregated outcomes employing z-scores. ProFuturo platform.
Note: All estimates are based on Two-stage Least Squares where treatment assignment at the school level is used as an instrument for the rate of treatment of classes at the grade level. Data from the ProFuturo platform is used to assess which classes where treated. Outcomes are grouped in indices that are built using the procedure in Kling et al. (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) Students' technology use: built from outcomes in columns (2)-(5) of Table 1; (2) Teachers' absenteeism and motivation: built from outcomes in columns (2)-(8) 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 5 and 10 percent level. Standard errors are clustered at the school level.

the main results of the paper while employing the Post-double Selection LASSO procedure for selecting the referred control variables.[39] 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.[40] Results applying Post-double Selection LASSO and IPTW are comparable to our benchmark findings. In Section I of the Appendix we report p-values of the procedure described in Romano and Wolf (2016), which we employ to account for multiple hypothesis at the table level for Tables 1–9. This is an alternative to the outcome aggregation procedure we applied above. Overall, as expected, the p-values estimated through Romano-Wolf for our individual treatment effects increase. Still, statistical significance is maintained primarily for use of technology, teachers' absenteeism, students' time allocation, students' interaction at school, and marginally for cognitive skills.

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. Although we do not find significant impacts on test scores in Portuguese or Mathematics, we 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.

Note that the estimated cost of the ProFuturo program in Angolan schools is approximately USD 2230 per school per year, which corresponds to USD 3 per student in our treatment schools, and to USD 7 per student in the targeted grades of our treatment schools. This cost can be restated as cost-effectiveness: USD 10 per increase in 0.1 standard

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

[40] 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 class level.

deviations in a student's test score (Science).[41] The referred cost covers basic support activities, training of coaches (coordinators) and lead teachers, monitoring and follow-up. It also includes the equipment costs, assuming a 6-year lifetime. This cost estimate per student places the ProFuturo program in the lower half (McEwan, 2015) and close to the median in terms of cost-effectiveness (McEwan, 2015; Mbiti et al., 2019) of comparable interventions in schools of developing countries.

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.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Replication files are available at: http://dx.doi.org/10.17632/jdy8b542cs.1.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2023.103145.

References

Angrist, J., Lavy, V., 2002. New evidence on classroom computers and pupil learning. Econ. J. 112 (482), 735–765.
Araya, R., Arias Ortiz, E., Bottan, N.L., Cristia, J., 2019. Does Gamification in Education Work? Experimental Evidence from Chile. IDB Working Paper Series.
Athey, S., Imbens, G.W., 2017. The econometrics of randomized experiments. In: Handbook of Economic Field Experiments. Vol. 1. Elsevier, pp. 73–140.
Banerjee, A.V., Cole, S., Duflo, E., Linden, L., 2007. Remedying education: Evidence from two randomized experiments in India. Q. J. Econ. 122 (3), 1235–1264.
Barrera-Osorio, F., Linden, L.L., 2009. The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia. Vol. 4836. World Bank Policy Research Working Paper.
Barrow, L., Markman, L., Rouse, C.E., 2009. Technology's edge: The educational benefits of computer-aided instruction. Am. Econ. J. Econ. Policy 1 (1), 52–74.
Beg, S., Halim, W., Lucas, A.M., Saif, U., 2022. Engaging teachers with technology increased achievement, bypassing teachers did not. Am. Econ. J. Econ. Policy 14 (2), 61–90.
Berry, J., Kannan, H., Mukherji, S., Shotland, M., 2020. Failure of frequent assessment: An evaluation of India's continuous and comprehensive evaluation program. J. Dev. Econ. 143 (102406).
Beuermann, D.W., Cristia, J., Cueto, S., Malamud, O., Cruz-Aguayo, Y., 2015. One laptop per child at home: Short-term impacts from a randomized experiment in Peru. Am. Econ. J. Appl. Econ. 7 (2), 53–80.
Carrillo, P.E., Onofa, M., Ponce, J., 2011. Information Technology and Student Achievement: Evidence from a Randomized Experiment in Ecuador. IDB Working Paper.
Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., Severín, E., 2017. Technology and child development: Evidence from the one laptop per child program. Am. Econ. J. Appl. Econ. 9 (3), 295–320.
Duflo, E., Hanna, R., Ryan, S.P., 2012. Incentives work: Getting teachers to come to school. Amer. Econ. Rev. 102 (4), 1241–1278.
Engle, R.W., 2002. Working memory capacity as executive attention. Curr. Direct. Psychol. Sci. 11 (1), 19–23.
Escueta, M., Quan, V., Nickow, A.J., Oreopoulos, P., 2017. Education Technology: An Evidence-Based Review. Tech. rep., National Bureau of Economic Research.
Fairlie, R.W., Robinson, J., 2013. Experimental evidence on the effects of home computers on academic achievement among schoolchildren. Am. Econ. J. Appl. Econ. 5 (3), 211–240.
Glewwe, P., Kremer, M., Moulin, S., 2009. Many children left behind? Textbooks and test scores in Kenya. Am. Econ. J. Appl. Econ. 1 (1), 112–135.
Hanushek, E.A., Woessmann, L., 2008. The role of cognitive skills in economic development. J. Econ. Lit. 46 (3), 607–668.
Heß, S., 2017. Randomization inference with Stata: A guide and software. Stata J. 17 (3), 630–651.
Jackson, C.K., Bruegmann, E., 2009. Teaching students and teaching each other: The importance of peer learning for teachers. Am. Econ. J. Appl. Econ. 1 (4), 85–108.
Jackson, K., Makarin, A., 2018. Can online off-the-shelf lessons improve student outcomes? Evidence from a field experiment. Am. Econ. J. Econ. Policy 10 (3), 226–254.
Jackson, C.K., Rockoff, J.E., Staiger, D.O., 2014. Teacher effects and teacher-related policies. Annu. Rev. Econ. 6 (1), 801–825.
Kerwin, J.T., Thornton, R.L., 2021. Making the grade: The sensitivity of education program effectiveness to input choices and outcome measures. Rev. Econ. Stat. 103 (2), 251–264.
Kling, J.R., Liebman, J.B., Katz, L.F., 2007. Experimental analysis of neighborhood effects. Econometrica 75 (1), 83–119.
Lai, F., Luo, R., Zhang, L., Huang, X., Rozelle, S., 2015. Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing. Econ. Educ. Rev. 47, 34–48.
Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., Boswell, M., Rozelle, S., 2013. Computer assisted learning as extracurricular tutor? Evidence from a randomised experiment in rural boarding schools in Shaanxi. J. Dev. Effect. 5 (2), 208–231.
Machin, S., McNally, S., Silva, O., 2007. New technology in schools: Is there a payoff? Econ. J. 117 (522), 1145–1167.
Malamud, O., Pop-Eleches, C., 2011. Home computer use and the development of human capital. Q. J. Econ. 126 (2), 987–1027.
Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., Rajani, R., 2019. Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania. Q. J. Econ. 134 (3), 1627–1673.
McEwan, P.J., 2015. Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. Rev. Educ. Res. 85 (3), 353–394.
McKenzie, D., 2012. Beyond baseline and follow-up: The case for more T in experiments. J. Dev. Econ. 99 (2), 210–221.
Mo, D., Swinnen, J., Zhang, L., Yi, H., Qu, Q., Boswell, M., Rozelle, S., 2013. Can one-to-one computing narrow the digital divide and the educational gap in China? The case of Beijing migrant schools. World Dev. 46, 14–29.
Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., Qiao, Y., Boswell, M., Rozelle, S., 2014. Integrating computer-assisted learning into a regular curriculum: Evidence from a randomised experiment in rural schools in Shaanxi. J. Dev. Effect. 6 (3), 300–323.
Molina-Millán, T., Barham, T., Macours, K., Maluccio, J.A., Stampini, M., 2019. Long-term impacts of conditional cash transfers: review of the evidence. World Bank Res. Obs. 34 (1), 119–159.
Muralidharan, K., Singh, A., Ganimian, A.J., 2019. Disrupting education? Experimental evidence on technology-aided instruction in India. Amer. Econ. Rev. 109 (4), 1426–1460.
Naslund-Hadley, E., Parker, S.W., Hernandez-Agramonte, J.M., 2014. Fostering early math comprehension: Experimental evidence from Paraguay. Glob. Educ. Rev. 1 (4), 135–154.
Papay, J.P., Taylor, E.S., Tyler, J.H., Laski, M.E., 2020. Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data. Am. Econ. J. Econ. Policy 12 (1), 359–388.
Romano, J.P., Wolf, M., 2016. Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. Statist. Probab. Lett. 113, 38–40.
Rouse, C.E., Krueger, A.B., 2004. Putting computerized instruction to the test: A randomized evaluation of a "scientifically based" reading program. Econ. Educ. Rev. 23 (4), 323–338.
Yang, Y., Zhang, L., Zeng, J., Pang, X., Lai, F., Rozelle, S., 2013. Computers and the academic performance of elementary school-aged girls in China's poor communities. Comput. Educ. 60 (1), 335–346.
Young, A., 2019. Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. Q. J. Econ. 134 (2), 557–598.

[41] Using data from individual treatment as provided by schools' report and the ProFuturo platform, we have that on average in each school 304 or 248 students used the program (respectively for the schools' report and the ProFuturo platform). The cost per treated student was on average USD 7 or 9. However, there is significant heterogeneity, as this number ranges from USD 4 or 5 in a school where 540 or 440 students were treated to USD 16 or 22 in a school where only 135 or 99 students used the platform.