Maastricht University

UNITED NATIONS UNIVERSITY

UNU-MERIT

Working Paper Series

#2019-007

The impact of a mathematics computer-assisted learning platform on students' mathematics test scores

Marcelo Perera and Diego Aboal

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: info@merit.unu.edu | website: http://www.merit.unu.edu

Boschstraat 24, 6211 AX Maastricht, The Netherlands
Tel: (31) (43) 388 44 00

UNU-MERIT Working Papers
ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology
UNU-MERIT

The Impact of a Mathematics Computer-Assisted Learning Platform on Students' Mathematics Test Scores

Marcelo Perera[a,c] and Diego Aboal[a,b,c]

a/CINVE, b/Universidad ORT Uruguay, c/Facultad de Ciencias Económicas y Administración, Universidad de la República de Uruguay.

March 2019

Abstract

Since 2013, the Uruguayan educational system has been using an online adaptive learning tool for mathematics: The Mathematics Adaptive Platform (PAM for its Spanish acronym). PAM's content has been adapted to the national curriculum and it is a tool that – based on an analysis of students' experiences – offers personalised feedback according to each student's skill level. The use of PAM has been spreading throughout the education system. By 2016, approximately half of all students in 3rd through 6th grades of primary education had used the platform. The purpose of this study is to identify the effect of the use of PAM on the test score gain in mathematics based on longitudinal data from a sample of students in primary education. The results show a positive effect of 0.2 standard deviations on mathematics test scores. Results also show that the impact of PAM increases as the socioeconomic status of students decreases. There is no heterogeneous impact by gender. This is the first evidence at a country-wide level of the impact of a pedagogical tool of this type.

Keywords: Evaluation of Computer Assisted Learning systems, Mathematics, Uruguay.

JEL classification: I21, I28.

1. Introduction

In July 2013[1], Plan Ceibal, the public institution in charge of the One Laptop per Child program in Uruguay, provided PAM for students and teachers in primary education in Uruguay. This digital tool is designed for content development and for learning concepts and methods. The platform offers teachers different resources for creating their lessons, to establish learning objectives and to supply individual or group tasks for their students in order to enrich and to deepen their learning and practice. PAM offers more than 100,000 activities that give personalised assistance to each student that is adapted to his or her level of knowledge. The platform gives immediate feedback to the student after each answer, offering help and theoretical materials and showing alternative solutions to the problems[2].

Even though we continue to observe considerable expansion in the usage of PAM in primary education, there is still considerable variability in the use it and in the intensity of use of the platform by teachers and students.

The purpose of this study is to identify the effect that using PAM had on test score gains in mathematics based on longitudinal data from a sample of the cohort of students who were studying in the 3rd grade of primary education in 2013. The students' proficiency levels in mathematics and in reading were evaluated at the end of 2013 and again three years later in 2016. Combining this data with data about the use of PAM obtained from the platform's database, we analysed whether using the tool was a factor associated with the gain in students' tests scores.

We still know relatively little about what impact the use of information and communication technologies (ICT) in education has on learning in developing countries. This is particularly relevant for these countries where concerns about the quality of education have led to increasing adoption of ICT in class. Reviews of high quality studies show mixed evidence regarding the impact of using technology in education (see for example Bulman and Fairlie, 2016). This suggests technology's potential in education depends on the details of the intervention. These details determine to what extent a specific intervention is capable of overcoming the restrictions in a student's learning process.

Even though high quality research on the effects of technology-assisted learning are scarce, some experimental evaluations, such as those by Banerjee et al. (2007) and Muralidharan et al. (2016), have provided solid evidence of the high relative impact of such interventions on students in primary and secondary schools in developing countries. One of the findings of these studies is that the educational tools that accounted for the greatest gains appear to be those that use technology to personalise the learning process.

Muralidharan et al. (2016) evaluated the impact of an intervention that included using a learning software (for mathematics and language) called Mindspark. Like PAM, Mindspark is designed for

[1] The academic year in Uruguay starts in March and ends in December, therefore, the tool was available at mid academic year.

[2] The platform is developed by the company bettermarks (http://uy.bettermarks.com) based in Germany. It has been adapted to the national curriculum and currently covers the topics from 3rd grade of primary to 4th grade of secondary education. Further information about PAM and some experiences in schools in Uruguay can be found on the following site: https://www.youtube.com/user/canalceibal/search?query=PAMhttps://www.youtube.com/user/canalceibal/search?query=PAM

personalised student instruction[3]. This kind of software employs algorithms for error recognition and the possibility to give students immediate, personalised feedback (hints, alternative solutions, activities corresponding to their level, etc.). The authors found a positive impact on the learning of mathematics and language for a sample of students in middle school in India. They also state that the relative gain is greater for students with a lower academic level. They speculate that the main mechanism behind the efficacy of the intervention is the capability of the tool to direct the instruction in a personalised form, given the extremely heterogeneous nature of students' ability levels[4].

PAM is an example of a digital learning tool that adapts to a student's ability level. The tool was made available to all teachers and students in the educational system; however, the decision of whether and how to use it was left to them. Even though its use has been spreading, how teachers and students are using the tool is still relatively unknown.

The objective of this paper is to provide evidence of the impact on mathematics learning of a technology-assisted learning tool in the way it was applied in Uruguay. The objective is not only to estimate the average impact on test scores but also the impact by socio-economic status and gender.

The strategy to identify the impact of PAM is based on three pillars: i) the availability of longitudinal data on learning outcomes ii) the variability in the use of PAM between 2013 and 2016 and iii) the possibility of using the gain in reading test scores to control the potential selection bias due to the presence of unobservable factors which may affect the general performance of the student.

Although this is not an experimental study, and therefore the internal validity of the evaluation is based on some assumptions, it has an important advantage over experimental studies in terms of their external validity. We are analysing a country-wide intervention, and therefore, we are contemplating effects generally ignored in studies evaluating small scale interventions[5]. In addition, in this paper we are analysing heterogeneity in treatment effects by student characteristics (socioeconomic status and gender); which is rare in the literature (Bulman and Fairlie, 2016).

The remainder of this paper is organised as follows. Section 2 provides a very brief literature review. In section 3 we describe PAM. Section 4 discusses the data. In section 5 we present the methodology applied to estimate the effects of PAM, with a special emphasis on the challenges to identify a causal relation. Section 6 presents the results. In section 7 we summarise the main findings of the study.

2. Brief literature review

A comprehensive review of the literature is out of the scope of this paper. What follows is a summary of the findings of three recent reviews.

---

[3] The students used the software outside of class time for 90 minutes. Half of the time they used it in free activities and the other half under the guidance of an assistant in groups of 12-15 students.

[4] Adapting the instruction to the student's level of learning also explains the high effectiveness of other interventions documented in the literature. Such is the case of differentiated support for students with low performance (for example, see the evaluation of the "Balsakhi" Programme in India in Banerjee et al., 2007) or of grouping students according to their skill levels (see the assessment of the impact of "tracking" in Kenya in Duflo et al., 2011).

[5] A discussion on the challenges of scaling programs that have proven effective in experimental studies on small scale can be seen in Banerjee et al. (2016).

Bulman and Fairlie (2016) perform a narrative review of over 30 studies for developed and developing countries that analyse the impact of the use of technology with educational purposes in classroom and at home on educational outcomes. In particular, they focus on the impact of use of computers, internet and computer assisted instruction (CAI) software on grades, test scores, retention, graduation, and attendance. They conclude that the evidence show mixed results with a pattern of null effects. However, they highlight that in contrast with this general pattern, CAI targeting math in developing countries tend to have positive effects.

Sung et al. (2016) carry out a meta-analysis of 110 papers on the effects of integrating mobile devices with teaching and learning on students' learning performance. These papers included the analysis of the impact of different types of hardware, software, teaching method, domain subject, implementation setting and intervention duration on different types of learning achievement indicators and also on "affective variables" including motivation, interest, etc. They find that the effect size of interventions (of any type) on learning achievement in mathematics (12 studies) was 0.34.

Cheung and Slavin (2013) perform a meta-analysis of the effect of educational technology (supplemental CAI, computer-management learning, and comprehensive programs) on mathematics achievement of K-12 (primary and secondary education) students. The study includes 74 papers. They find an average effect of 0.15. However, when looking exclusively at CAI interventions (55 studies), the effect rises to 0.18. These authors also summarise the results of 21 previous major meta-analyses on the effects of educational technology on mathematics achievement. These studies find average effect sizes (of any type of educational technology intervention) ranging from 0.10 to 0.62.

3. Plan Ceibal and PAM

In 2010, Uruguay became the first country in the world to provide one laptop to each public primary-school child, free of charge. In a country of 3.4 million people, almost 800,000 laptops were handed over to children and teachers so far under the One Laptop Per Child (OLPC) program[6]. The laptop is owned by the child, who can take it home, and use it for learning inside and outside of the classroom.

Plan CEIBAL –the name the OLPC programme receives in Uruguay- is one of the world's most ambitious rollouts of educational technologies. Since the introduction of Plan Ceibal in Uruguay (starting in 2007), the proportion of internet users in Uruguay has shot up (from 18% in 2007 to 66% in 2016), overtaking countries like Italy (61%). The proportion of internet users in Uruguay is now closer to the average for Central Europe and Baltic countries (71%) than to the average for Latin America and the Caribbean countries (56%) (International Telecommunication Union database). Since the introduction of Plan Ceibal, the gap in the access to a computer and to an internet connection by individuals of different income levels (the 'digital divide') has decreased abruptly. The access to computers in the poorest 10% of the households went from 6% in 2007 to 75% in 2014 and in the richest 10% from 73% to 90%. In the same period, the access to internet in the 10% poorest households went from 1% to 27% and in the richest 10% from 63% to 90% (Plan Ceibal, 2015).

[6] Worldwide in 2017 over 3 million children and teachers have received laptops as part of OLPC's programme (http://one.laptop.org/map accessed 10/27/2017). OLPC is a USA-based non-profit organization set up to oversee the creation of an affordable educational device for use in the developing world.

However, Plan Ceibal is more than just one laptop per child. On top of providing one laptop per child, it includes free connectivity in schools and some other public places; it trains teachers to use the laptop in the classroom, and provides software and support to facilitate learning. PAM is one of the learning tools that Plan Ceibal offers to children and teachers.

PAM is an online mathematics learning tool that is available for free for children and teachers. It includes 100,000 exercises, 10,000 exercise types, 1,600 learning objectives (750 introductory and elementary, 780 core knowledge, 440 advanced and 140 text problem based; some overlap occurs among these categories) and 140,000 learning paths. Learning from mistakes is the core concept behind PAM. The idea is that children can learn by working on the exercises, making mistakes, receiving automatic feedback, trying again and applying what they have learned.

PAM has some characteristics. First, it covers the national curriculum. Second, it has interaction tools (such as tips, formula input, drag and drop, sorting, colouring shapes, constructing charts and diagrams, plotting functions, drawing points, lines, etc.). Third, it gives feedback to students through 25,700 exercise step based feedbacks, 2,800 error pattern based feedbacks and solution explained for every exercise. Finally, it is an adaptive tool that suggests learning paths for each student and learning objective following the sequence: do exercises, identify failed exercise series, follow up, do something easier, revise previous knowledge, close knowledge gap.[7]

Although almost 50% of students of 3rd to 6th grade of primary education use the platform, the use of it by teachers and students is not compulsory and there is no canonical use suggested by the authorities or Bettermarks (the firm providing the software). Therefore, each teacher and student can use it in the best way it fits her/his needs. Plan Ceibal provides training to teachers in a voluntary base. The use of the tool by children is highly correlated to the use of the tool by teachers. Those teachers that use PAM usually assign problem sets to students through the platform. Not much is known about the exact way in which teachers use the tool. Plan Ceibal is currently studying the ways in which teachers use the platform as a learning tool in and out of the classroom.

4. Data

This paper uses two databases. The first database contains information about tests scores in mathematics and reading. This data is provided by the Department of Research and Educational Statistics of the Central Directive Council of the National Administration for Public Education (DIEE/CODICEN/ANEP, for its Spanish acronyms).

The data comes from a representative sample of the cohort of students who were in their third year of primary school in 2013. The first assessment was performed at the end of 2013 and the second at the end of 2016, when the majority of those children were in their 6th year of primary school.

The first evaluation corresponds to the Third Regional Comparative and Explanatory Study (TERCE), a large-scale study of learning achievements coordinated by the Latin American Laboratory for Assessment of the Quality of Education (LLECE/UNESCO), in which Uruguay and 14 other countries in Latin America took part (see UNESCO-OREALC, 2016, for details on the design of the instruments of evaluation)[8].

The second learning assessment of was applied by DIEE/CODICEN/ANEP in 2016 on the same cohort of students who participated in TERCE in 2013 and was based on the same methodology applied by

[7] http://bettermarks.com/wp-content/uploads/sites/10/whitepaper_learning.pdf Accessed 10/27/2017.

[8] The TERCE evaluation focuses on the curriculum of each country. All documentation on this evaluation can be found at: http://www.unesco.org/new/en/santiago/education/education-assessment-llece/third-regional-comparative-and-explanatory-study-terce/ (accessed 11/09/2017).

LLECE/UNESCO in the regional evaluations. The DIEE created a procedure for comparing the results from the 3rd year test to the metric of the 6th year test, thereby allowing the analysis of the students' learning progress in mathematics and reading between 2013 and 2016[9].

The second database is provided by Plan Ceibal and contains information on PAM use by students. We have information for the years 2013-2016. In particular, we have information about the number of exercises performed each year by the students who participated in the learning assessments[10]. There is also another indicator for the year 2016 regarding the participation of the student in the Competition in Mathematics (a special event organised by Plan Ceibal) and the number of exercises performed by the student during the competition. In this way it is possible to reconstruct a measure of the number of exercises performed outside of the competition, which is the variable used to approximate PAM usage in 2016.

The two databases were merged using a unique student identifier[11].

All the results of the impact evaluation are based on the information from the 2143 students attending 237 public or private schools all over the country that were evaluated in mathematics and reading in 2013 and 2016[12] (see Table 1). This group of students, for which the test score gain between both evaluations could be measured, make up the sample used in this paper.

Table 1: Number of students evaluated in mathematics and reading and number of schools that participated in the evaluation in each year

	Students	**Schools**
Evaluation 2013	3761	242
Evaluation 2016	2349	237
Both evaluations, 2013 and 2016 (*)	**2143**	**237**

Notes: the number of students is a little bit larger if it includes the students who participated in only one of the evaluations, either in mathematics or reading. The data given in this table corresponds to the students who took both of the tests (reading and mathematics). (*) The actual number of students evaluated is slightly larger (2174), however 31 students could not be found in the PAM database.

A PAM user was defined alternatively as a student who performed at least one, at least 10 or at least 20 exercises in a given year (or as the annual average over a given period). The percentages of students falling into these categories are shown in Table 2. In 2016, we find that 44% of the students performed at least one exercise on the platform outside of the competition, 39% performed 10 or more and one third of the students performed 20 or more exercises. No matter which definition of user we consider, there is strong growth in the use of the platform over the period of interest. For example, the ratio of the students who performed 20 or more exercises grew by a factor of 3, increasing from 11% to 34% in the period 2013-2016.

---

[9] See the Annex A.2.1 of the following document: http://www.ceip.edu.uy/documentos/2017/varios/1971/Libro.pdf

[10] For the year 2016, there is an indicator that identifies a student's participation or non-participation in the Competition in Mathematics as well as the number of exercises performed during the competition. This made it possible to reconstruct the number of exercises performed outside the competition. The latter was the variable finally used to approximate the educational use of PAM during the year 2016.

[11] The databases were previously anonymised in order to protect the identity of the students in the sample.

[12] The test of differences between the original TERCE sample and the sample finally analysed were not significant, so bias cannot be expected due to loss of sample.

Table 2: Percentage of students who performed at least 1, 10 or 20 exercises per year (from a sample of 2143 students from 237 schools)

	% of users		
Year	1 or more exercises	10 or more exercises	20 or more exercises
2016 (outside the competition)	44.4	39.0	33.8
2016 (total number of exercises)	48.2	42.7	39.0
2015 (total number of exercises)	42.9	37.9	28.8
2014 (total number of exercises)	36.4	20.7	14.1
2013 (total number of exercises)	20.7	13.4	11.3

Note: the percentage of users in 2014-2016 refers to the percentage of students with an annual average of at least "N" exercises

Table 3 shows some descriptive statistics of the number of exercises performed by students on PAM. If we consider the year 2016 and the activity that took place outside of the competition, we find that a quarter of the users performed up to 22 exercises, half of them performed up to 84 exercises and three quarters of them performed up to 274 exercises. The median has increased over the years and especially in 2016. The use of PAM increased both, as a consequence of the increase in the number of users and, the increase in the number of exercises per student.

Table 3: Number of exercises performed by students who used PAM in a given period (from a sample of 2143 students from 237 schools)

	Annual Number of Exercises of Users (students who did 1 or more exercises)			
Year	Percentile 25th	Median	Percentile 75th	Percentile 90th
2016 (outside the competition)	22	84	274	402
2016 (total number of exercises)	25	102	461	931
2015 (total number of exercises)	17	39	140	546
2014 (total number of exercises)	5	12	39	67
2013 (total number of exercises)	7	24	61	89
2014-2016 (total number of exercises)	11	43	167	329

The Household Socioeconomic Status (SES by its Spanish acronym) used in the paper was constructed by LLECE/UNESCO based on the information from a questionnaire given to the families of the students of TERCE 2013. This variable indicates the student's SES and comes from a combination (a factor) of variables related to the educational and labour background of the mother, household income, the goods and services available in the house, and the number of books available. This variable has a mean near 0 and a standard deviation close to 1 in the analysed sample. Higher values indicate higher socioeconomic status.

5. Empirical strategy

The identification of the causal effect of an intervention, and in particular in the educational area, poses several challenges. First, the need to have good measurements of the variables of interest, in particular of the progress the students have made in their learning process. The estimation of the effect of PAM in this study is based on reliable data. On one hand, the measurement of the student's learning progress originates from evaluations based on international standards and is a national representative sample of the cohort of all students who attended the 3rd year of primary school in

2013. On the other hand, the indicators of use of PAM rely on objective information provided by the platform itself.

The second challenge is the strategy for identifying the causal effect itself, that is, the way in which the data is analysed in order to find out whether there is a causal relationship between the use of PAM and the learning outcomes. This is particularly challenging in observational studies, since it is not possible to assign the treatment experimentally, which would allow for the inference of a causal effect through the simple comparison of outcomes of the treatment and the control group.

PAM is a digital tool which was made available to all teachers and students in the educational system. However, the decision to whether and how the tool would be used was left to them and possibly to the adults responsible for the child. Considering the students who attended the 6^{th} year in 2016, there are as many children who performed at least one exercise on the platform, as there are those who did not. Suppose that PAM were used by the students who are most motivated or skilled, or whose usage is encouraged by the most effective teachers (or parents). By simply comparing users (treatment group) and non-users (control group), we would risk assigning a false causal effect to PAM. In statistical terms, we would obtain a biased estimation of the impact of PAM.

Our starting point is working with the gains in mathematics achieved by the students between 2013 and the end of 2016. Once we had defined the indicator of the use of PAM in the years 2014 to 2016, we were interested in whether the test score gain for students using PAM was different from the gain for students who did not use PAM. Nevertheless, this comparison can be biased if there are other differences between users and non-users which affect their performance in mathematics and have nothing to do with PAM. Our strategy is to compare the gains between users and non-users by "controlling" or "conditioning" for this type of variables.

The most complex problem however is the potential difference between users and non-users in unobservables. For example, the educational variables preceding the evaluation period which partially account for the student's current performance, or dimensions which are difficult to measure such as the student's motivation to learn, natural ability or the effectiveness of his or her teachers.

We can use two characteristics of our data to control for unobservables. First, we have two consecutive measurements of learning progress (test scores) in mathematics for the same students. This allows us to use the initial scores as an indirect method for controlling the educational factors which had contributed to the student's performance in the base year of the evaluation. In the second place, under the plausible assumption that PAM only affects performance in mathematics and not in reading, we suggest using the gain in reading achieved by a student as a measure to summarise the unobservable variables for the student and the teacher which have influenced his or her general performance in both mathematics and reading.

In order to clarify this last idea, imagine the following situation. Let us suppose that the use of PAM does not have any effect on the mathematics test score gains, but the most motivated students (who are motivated for other reasons than using PAM) were more inclined to use it. In turn suppose motivation is an explanatory factor of the student's gain in both mathematics and reading. If we proceed and compare the test score gain in mathematics of the students who use PAM with the gain of those who do not, we will find that the first group performed better. In this case we would erroneously assign PAM as the causal factor of that difference, when in fact this result is due to bias by omission of the underlying causal factor which is motivation.

Now, consider a slightly different gain measurement: let us compare the difference between the gain in mathematics and the gain in reading for PAM users and non-users. In this case we will not find an bias in favour of PAM users, since the motivated students outperform non-users in both mathematics and reading and therefore both effects are compensated for in this new measurement.

In this case, we will conclude correctly (according to our assumption) that the causal effect is zero. On the other hand, suppose that PAM has a positive impact on performance in mathematics and not in reading, it can be reasoned that the first measure (gains in mathematics) will lead to a positive result larger than the real effect (i.e. biased upwards) because it includes the motivation effect, whereas the second measure (gain in mathematics minus gains in reading) will allow us to infer the true effect of PAM on learning in mathematics.

Suppose now that the differences between students who used PAM and non-users consisted in attributes or abilities that are only related to mathematics. If this is the case, these differences cannot be controlled by the gain in reading and the estimations could still be positively biased. An alternative methodology is the application of an instrumental variable (IV). In other words, a variable which indirectly measures (is highly correlated to) the use of PAM and which at the same time is not related to this type of unobservable variable.

We suggest measuring the use of PAM by one student by means of the use of PAM by his or her classmates. Although this can be seen as the application of an instrumental variable, in the current study we present it as a proxy for the teacher's use of the platform[13].

Nevertheless, this instrumental variable will not correct the possible bias due to the presence of a specific type of unobservable characteristics of the teacher. For example, the teachers who encourage the use of PAM may be the teachers who are on average most effective in teaching mathematics, but not necessarily the best in teaching reading[14]. In other words, even though our aim is to perform the most rigorous analysis given the available data, we cannot completely rule out a subtle argumentation such as the one stated above, that links special characteristics of teachers with the usage of PAM, which could bias our estimates.

In short, our identification strategy consists in comparing the gain in mathematics between users and non-users of PAM, controlling by a set of information summarised in the following variables: the test results obtained in 2013, the gain in reading, the socioeconomic status, the gender, the region (Montevideo/Other regions), the school type, the participation in the competition and the usage of PAM in 2013[15].

A separate discussion is needed for the definition of PAM users. As mentioned above, we present results under the assumption that a PAM user is a student who has performed a given number of exercises per year (at least one, at least 10 or at least 20 exercises per year). Another aspect is whether we have to consider the use by the student or the use by the teacher. The available data

---

[13] The estimations that include a measure of group use correspond to the reduced form that directly links the gain in mathematics with the instrument. The estimation by instrumental variables is not reported here, but its results are very similar in terms of significance and sign of the impact. They are available upon request.

[14] We would need a valid instrument for teacher's use of PAM.

[15] PAM was introduced in 2013, i.e. the base year of our evaluation. This is why we include this variable as an additional control in the estimations. An alternative is to exclude the students who used PAM in 2013 and perform the analysis with the rest of the students. This was done, and the results are similar to the ones reported in this study.

corresponds to use by the student and shows an important intra-group correlation, which suggests that the decision of the teacher to use PAM is a determining factor in the student's use.

The discussion whether to include a measure of student use or a measure of use by a group of students (as a proxy for use by the teacher) depends on the mechanism by which PAM can affect learning. This mechanism on the other hand depends on the way teachers incorporate the tool into teaching. In the current study, we confine ourselves to presenting results based on measures of student use and of mean use by class.

Formally, the effect of the use of PAM is estimated with the following equation:

$$T^m_{ij(t)} - T^m_{ij(t-1)} = \delta\left(T^l_{ij(t)} - T^l_{ij(t-1)}\right) + (\rho^m - 1)T^m_{ij(t-1)} + (\rho^l - 1)T^l_{ij(t-1)} + \beta PAM_{ij} + \gamma X_{ij} + \varepsilon_{ij} \quad (1)$$

$T^a_{ij(t)}$ is the score in topic a (m = mathematics, l = language) obtained by child i from school j in year t (2013, 2016), therefore the term on the left-hand side is the gain in mathematics. PAM is a measure for the use of the platform for some year or some time period between $t-1$ and t, X is a vector of characteristics of the children and schools (gender, SES, participation in the competition in 2016, all-day school, Montevideo/Other regions) and ε_{ij} is a random error that can be correlated between observations of the same school but is independent between observations of different schools (i.e. clustered within schools).

A variation of the above equation is one where the dependent (the left-hand side) variable is the difference between the gain in mathematics and the gain in reading:

$$\left(T^m_{ij(t)} - T^m_{ij(t-1)}\right) - \left(T^l_{ij(t)} - T^l_{ij(t-1)}\right) = (\rho^m - 1)T^m_{ij(t-1)} + (\rho^l - 1)T^l_{ij(t-1)} + \beta PAM_{ij} + \gamma X_{ij} + \varepsilon_{ij} \quad (2)$$

Columns 1 and 2 of tables A1, A2 and A3 in the Appendix show the results of the estimation of this specification.

Although equation (2) may be more intuitive in aiming at explaining the differential gain between mathematics and reading, we believe it is more appropriate to work with equation (1) which involves controlling by the gain in reading but in a more flexible way.

In both equations, $\boldsymbol{\beta}$ is the effect of PAM on the gain in mathematics.

To identify a heterogeneous effect of the use of PAM by SES, a regression term is added to account for the interaction between the use of PAM and SES. Similarly, to identify a heterogeneous effect by gender an interaction term of use of PAM and a dummy variable representing gender was used. Formally, we estimated equations of the following type:

$$T^m_{ij(t)} - T^m_{ij(t-1)} = \delta\left(T^l_{ij(t)} - T^l_{ij(t-1)}\right) + (\rho^m - 1)T^m_{ij(t-1)} + (\rho^l - 1)T^l_{ij(t-1)} + \beta_0 PAM_{ij} + \beta_1 PAM_{ij} * G_{ij} + \beta_2 PAM_{ij} * SES_{ij} + \gamma X_{ij} + \varepsilon_{ij} \quad (3)$$

where G is a dummy variable which indicates the gender (1 = female), SES is a variable which indicates the socioeconomic status, $\boldsymbol{\beta}_0$ is the mean effect of PAM, $\boldsymbol{\beta}_1$ the differential effect by gender and $\boldsymbol{\beta}_2$ captures the heterogeneity of the effect depending on the socioeconomic status of the children.

6. Results

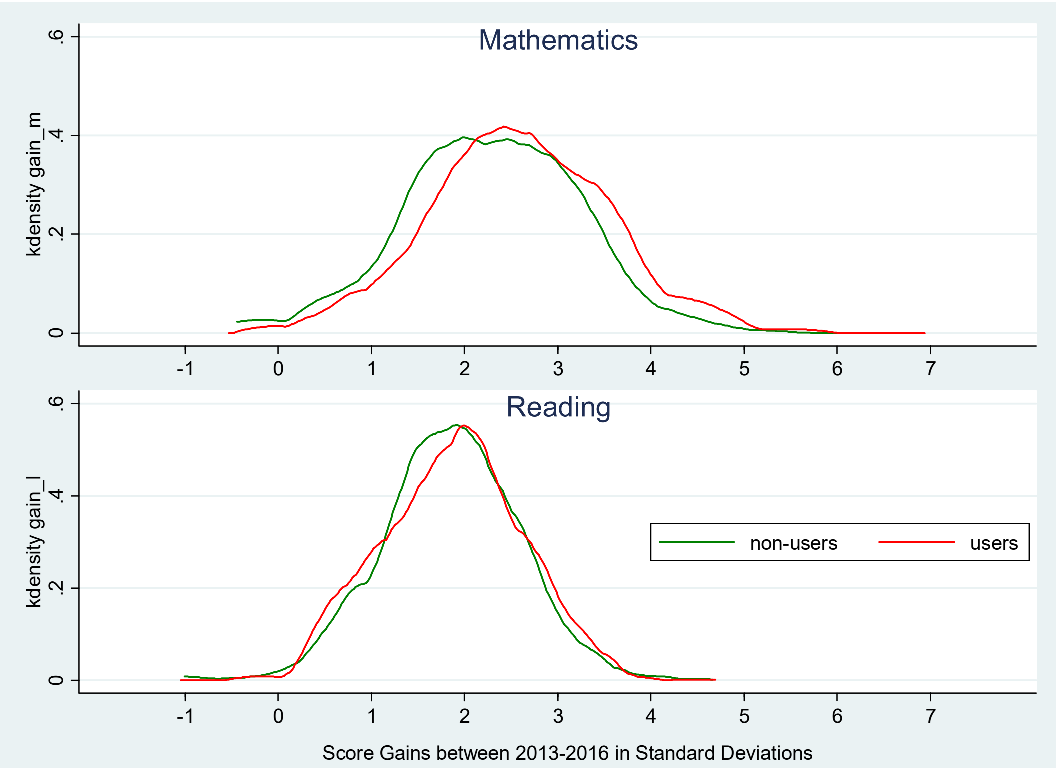
6.1 Average impact

All estimates result from standardised values of the test scores. Therefore, the gains in math and reading are expressed in standard deviations (SD). To have a reference value, the average gain in mathematics by students in the sample over their three years of schooling from 3rd to 6th grade was approximately 2.6 SD, while the average gain in reading was 1.9 SD.

Figure 1 shows the density function of the gain in mathematics (upper plot) and of the gain in reading (lower plot). Users are defined as students who performed at least one exercise between years 2014 and 2016.

There is a clear difference in the distribution of the gain in mathematics between users and non-users. The density function for students who used the platform is shifted to the right, that is, towards higher scores. This is not observed in reading.

Figure 1: Distribution of the score gain in mathematics (upper plot) and the score gain in reading (lower plot) comparing PAM-users (red line) and non-users (green line)



Note: Users are defined as students who performed exercises between 2014 and 2016

Instead of seeing the distribution of the gains, let us compare the median of each of those distributions and compare users and non-users. This comparison is a measure of the shift in the previously shown density functions. A statistical contrast is also performed to determine whether such a shift is statistically significant. All of this is presented in for different definitions of PAM user.

Table 4 shows, first, that PAM users show an average gain in mathematics that is higher than the one obtained by non-users. Second, there are no significant differences between users and non-users with respect to the gain in reading (except for one of the PAM user definitions where the difference is less than half the observed difference in mathematics).

Table 4: Average Score Gain in Mathematics and Reading 2013-2016, PAM-users and non-users, according to different user definitions

	Score Gain in Mathematics (standard deviations)			Score Gain in Reading (standard deviations)		
	Users	Non-users	Difference	Users	Non-users	Difference
Exercises in 2016						
1 or more	2.63	2.33	0.30***	1.91	1.86	0.05
10 or more	2.59	2.37	0.22***	1.94	1.85	0.09***
20 or more	2.57	2.40	0.17***	1.90	1.87	0.03
Exercises between 2014 and 2016 (*)						
1 or more	2.56	2.24	0.32***	1.88	1.88	0.00
10 or more	2.60	2.31	0.30***	1.90	1.86	0.05
20 or more	2.56	2.38	0.18***	1.90	1.87	0.03

(*) The number of exercises between 2014 and 2016 indicates the annual average of exercises over those three years. Significance level: * p <0.10, ** p <0.05, *** p <0.01.

Below, we show the results of the econometric estimations comparing the results in mathematics of users and non-users controlling for the possible differences between the two groups by a set of observable variables (initial score, gender, socioeconomic status, region, type of school) as well as by the gain in reading.

The first three columns of Table 5 report the gain in mathematics obtained by students using PAM according to the three user definitions. The left chart of Figure 2 illustrates these three estimates. The differences are significant and positive. The estimated coefficients indicate that PAM has an impact of approximately 0.2 SD on the gain in mathematics.

Column 4 of Table 5 changes the user condition from a student's attribute to an attribute of his or her group. Methodologically, this involves changing the variable of interest in the regression model. The indicator variable of user is replaced by a variable indicating the proportion of the students in the group who are users. This definition of use of the platform can be interpreted as a measure of use by the part of the classroom teacher.

The coefficient of this variable is illustrated in the chart on the right-hand side of Figure 2. It turns out to be significant and positive while showing that the higher the use at a group level, the greater the gain in mathematics obtained by the students in the group[16]. This result shows the possible existence of some type of external effect at the class level and therefore the importance of using the tool at a group level[17].

A final analysis consists of contrasting the relationship between the intensity of platform use – measured as the number of exercises performed – and the gain in mathematics. The analysis is similar to the previous ones with the difference that users are grouped according to the number of exercises performed during the year (column 5 of Table 5 and Figure 3). Similarly, we can define the intensity of group use by defining four groups according to the number of exercises performed by the students in the group (column 6 of Table 5 and Figure 4). It is important to state that the relationship between the number of exercises and the gain in mathematics is not linear.

[16] The magnitude of the coefficient indicates that the gain from belonging to a group where 100% of the students use PAM is 0.27 SD with respect to a group where no one uses it.

[17] Working with group data reduces the sample size because for some private schools in the sample this data is not available.

Let us take a look at the results regarding the intensity of group use (Figure 4). The coefficients of the 4 groups are positive and significant. A non-linear relationship between the number of exercises and the gain in mathematics is shown. The maximum impact is achieved by groups with a number of exercises between 40 and 100. Table A4 in the Appendix shows an alternative estimation of the relationship between the number of exercises and the gain. The results are qualitatively similar.

Identifying the relation between the number of exercises and the test score gain is a complex task. It certainly depends on many factors such as the way in which teachers include the tool in his or her pedagogical approach or whether or not the teacher guides the student in the use of the platform[18]. Therefore, the above commented results should not be interpreted as an estimation of the optimal number of exercises.

Tables in the Appendix present additional estimations. The positive relationship between the condition of being a platform user and the gain in mathematics is a robust result with respect to different specifications.

[18] If this distinction is relevant, the existence of competitions generates a distortion in the measure of the intensity of usage. Concerning the data from 2016, although working with the number of exercises outside of the competition, it is possible that this measure may not completely correct the number of exercises induced by the competition. In fact, even with this corrected measure, we observe that the participants of the competition have on average a larger number of exercises than the rest.

Table 5: PAM effect by user definition and number of exercises

	(1) Gains Math	(2) Gains Math	(3) Gains Math	(4) Gains Math	(5) Gains Math	(6) Gains Math
[>1]	0.209**					
	(0.08)					
[>10]		0.201**				
		(0.08)				
[>20]			0.209**			
			(0.08)			
%[>1]group				0.267**		
				(0.11)		
[1,20]					0.189**	
					(0.09)	
[20,40]					0.21	
					(0.13)	
[>40]					0.270***	
					(0.09)	
g[1,40]						0.184**
						(0.08)
g[40,100]						0.365**
						(0.18)
g[100,300]						0.312*
						(0.17)
g[>100]						0.236*
						(0.14)
Stand. Error	0.68	0.68	0.68	0.68	0.68	0.68
Schools	237	237	237	210	237	210
Students	2143	2143	2143	1842	2143	1842

Notes: Columns (1), (2) and (3) consider as PAM-users students who performed at least 1, 10 and 20 exercises in 2016 (outside the competition), respectively. Column (4) considers the proportion of students in the group who performed exercises in 2016 (outside the competition). Column (5) includes a set of dummy variables that identify three intervals of the number of exercises performed by the student in 2016 (outside the competition). Column (6) includes a set of dummy variables that identify four intervals of the median number of exercises of the group in 2016 (outside the competition). Dependent Variable: Score Gains in Math. Control Variables: Score in Math and Reading 2013, Gain in Reading, Use in 2013, Female, SES, Region, Type of School, Participation in Competition. Standard Errors clustered by School in parentheses. Significance level: *p<0.10, **p<0.05,***p<0.01.

Figure 2: Test Scores Gains in Math by user definitions

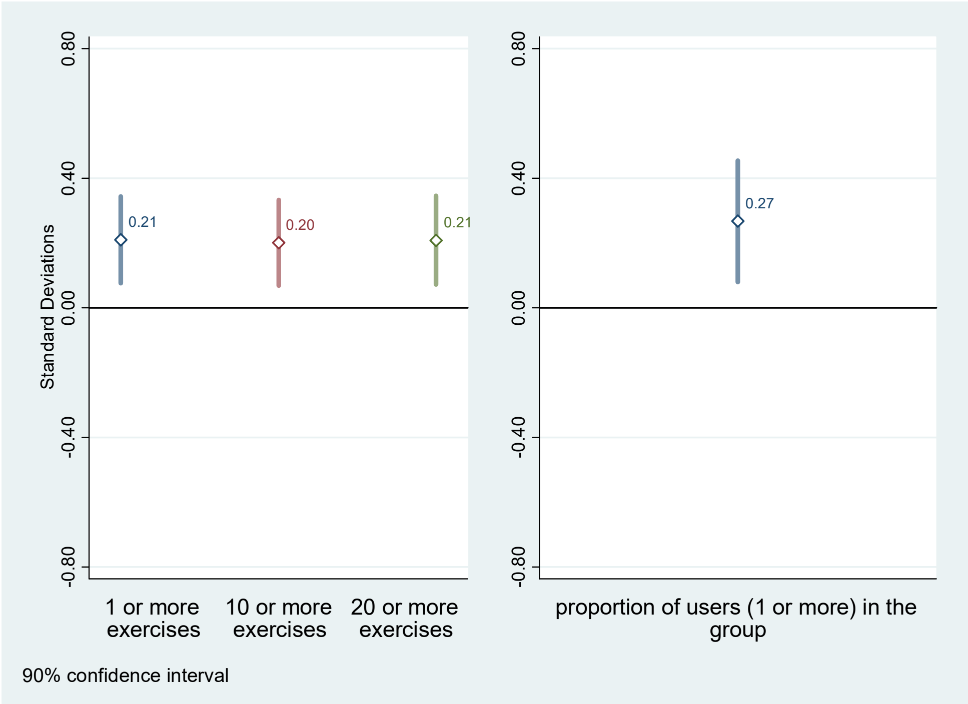


Figure 3: Test Scores Gains in Math by number of exercises performed

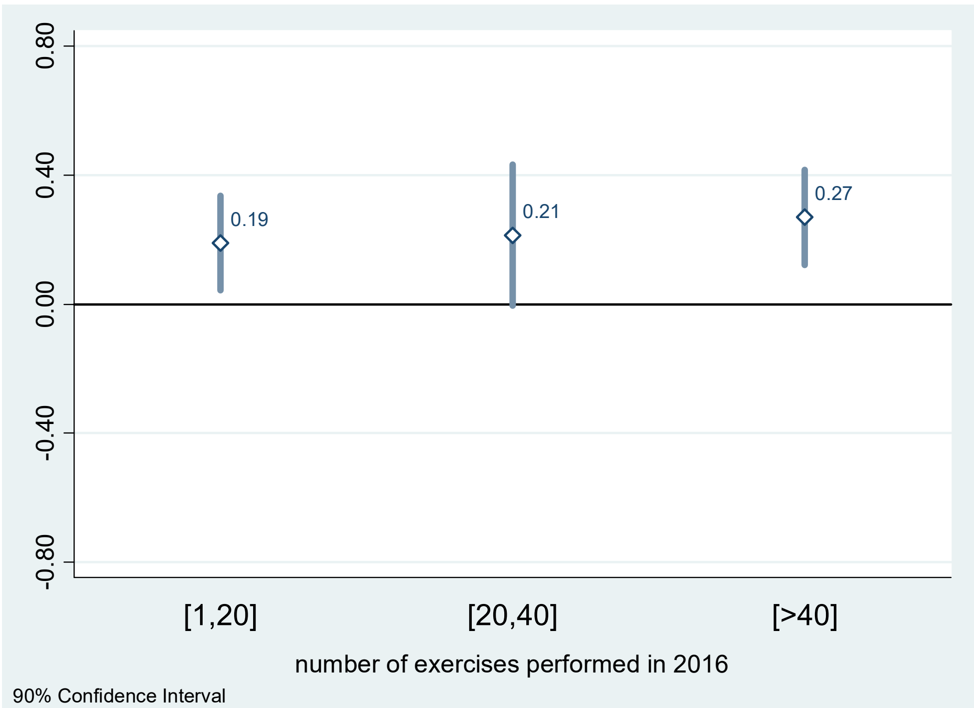
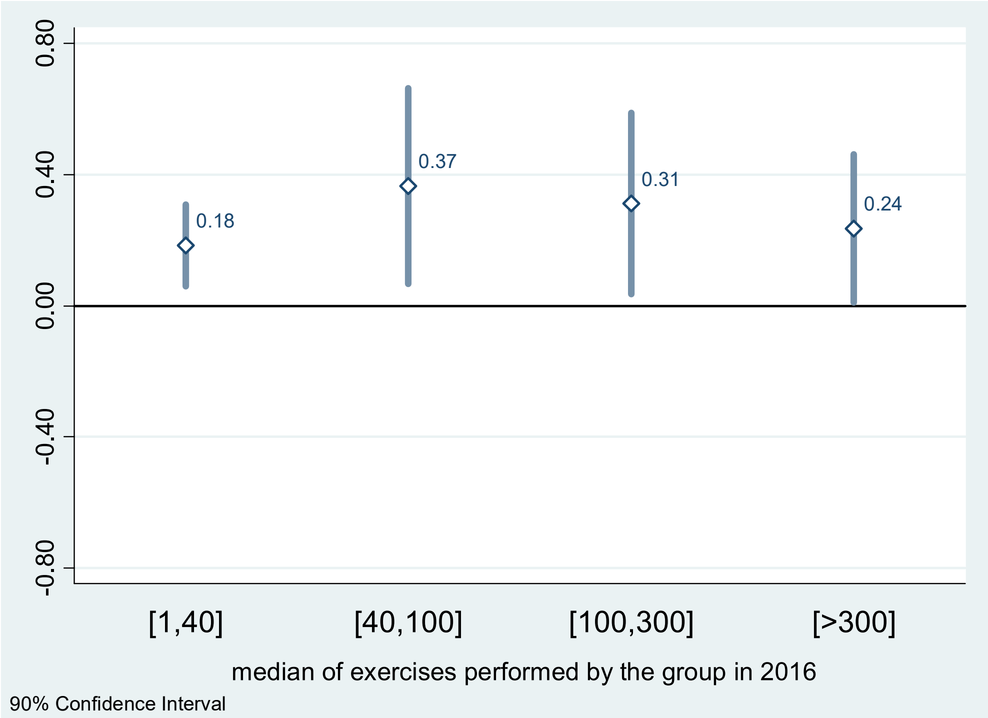


Figure 4: Test Scores Gain in Mathematics by number of exercises performed in average in the group in 2016



6.2 The impact of PAM by socioeconomic status of students

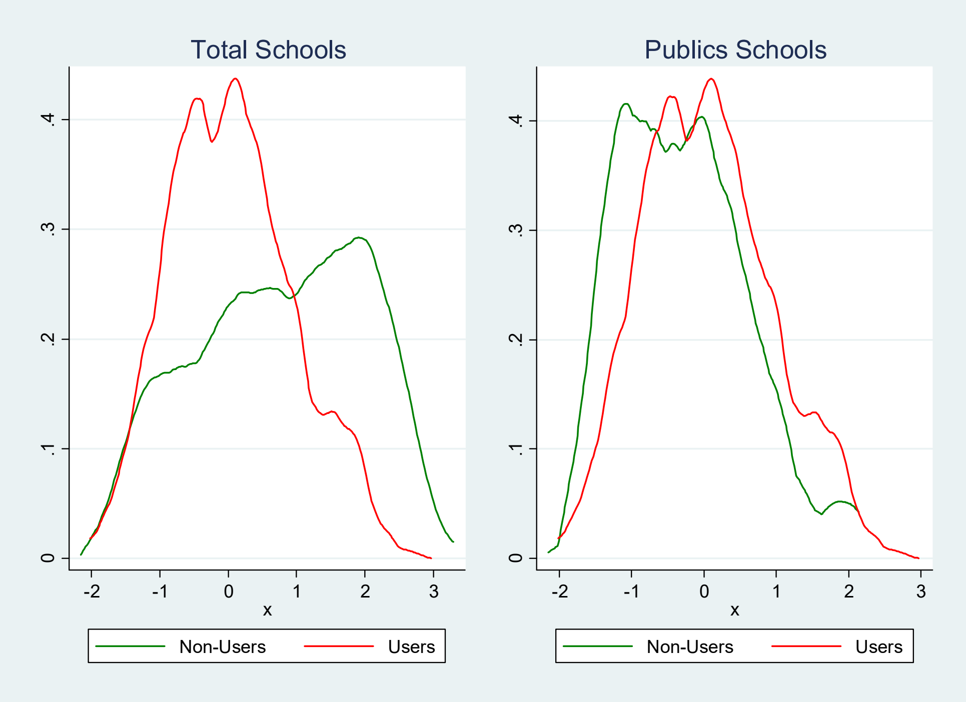
Figure 5 shows the distribution of the SES of users and non-users. The left chart shows that in the whole sample (public and private schools) there is a higher concentration of non-users in the higher SES values. This result, however, is explained by the lower propensity to use PAM among the private schools in the sample which have predominantly high SES students. Excluding private schools (right chart) shows a slightly biased user concentration towards higher SES values.

Figure 6 shows the distribution of the gain in mathematics of users and non-users, grouping the students by tertile of SES. We observe that the shift to the right of the distribution of the test scores gain of users (relative to non-users), is more pronounced in the 1st tertile (the third of students with the lower SES values) than in the 2nd and 3rd tertile.

The first row of Table 6 presents the coefficient of the interaction term of the dummy identifying a user of PAM and SES in the explanatory equation of the test score gain in mathematics. For all three user definitions, this term is statistically significant and negative. This indicates that the impact of the use of PAM decreases with the SES.

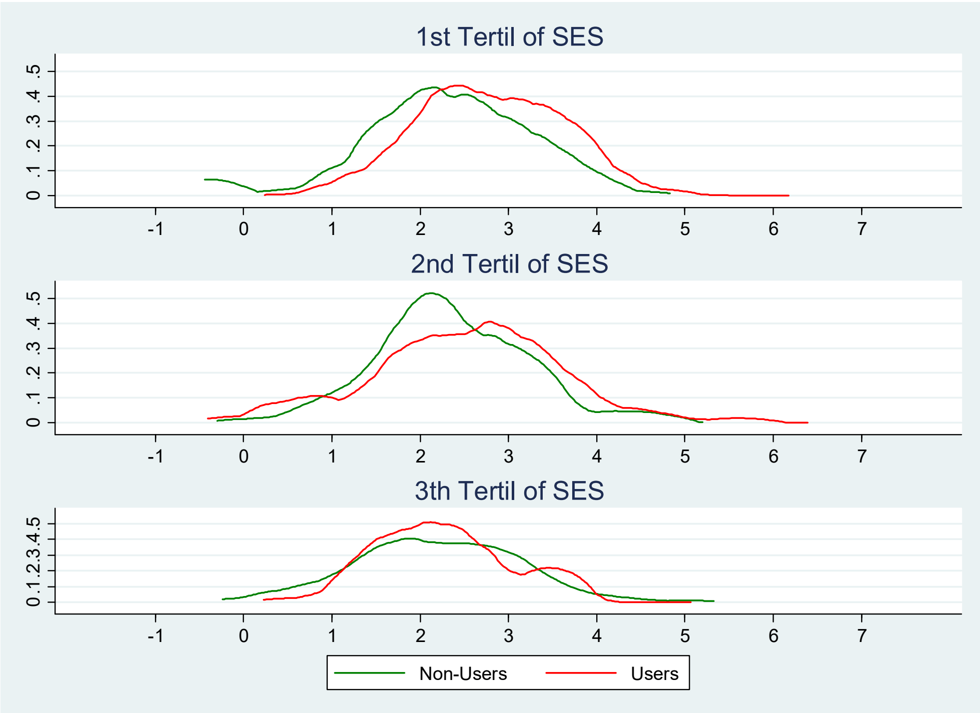
The remaining rows in Table 6 show the marginal effect of the use of PAM for an average student according to the decile of the SES to which it belongs. Figure 7 illustrates this result showing the monotone decreasing relationship between the PAM effect and SES.

Figure 5: Distribution of SES, PAM-users and non-users: total sample (on the left) and public schools only (on the right)



Note: users defined as students who performed exercises between 2014 and 2016.

Figure 6: Distribution of test scores gain in mathematics by tertile of SES



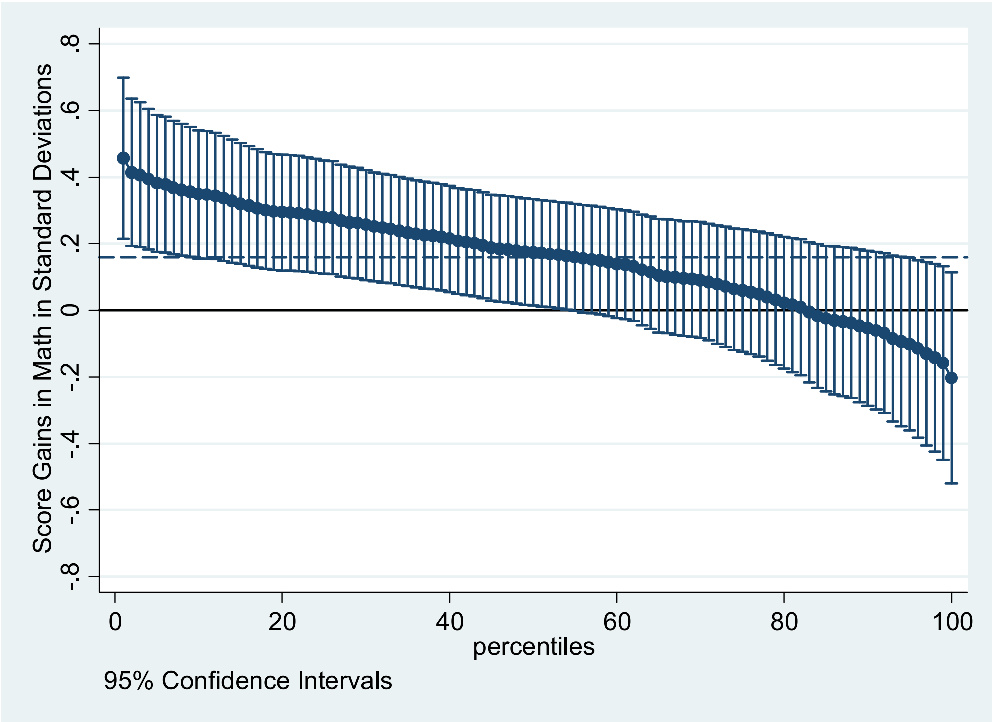
Note: users defined as students who performed exercises between 2014 and 2016.

Table 6: Estimation of the effect of PAM use on math test scores gain by level of SES

	1 or more exercises	10 or more exercises	20 or more exercises
Interacction of Use of PAM and S	**-0.128** **	**-0.141** ***	**-0.132** ***
	(0.0503)	**(0.0496)**	**(0.0422)**
Marginal Effect of Use of PAM by Decile of SES			
1st	0.408 ***	0.394 ***	0.374 ***
	(0.1229)	(0.1073)	(0.1076)
2nd	0.356 ***	0.337 ***	0.321 ***
	(0.1088)	(0.0953)	(0.1)
3rd	0.317 ***	0.293 ***	0.280 ***
	(0.0995)	(0.0881)	(0.0959)
4th	0.282 ***	0.255 ***	0.244 **
	(0.0927)	(0.0836)	(0.0937)
5th	0.247 ***	0.216 ***	0.208 **
	(0.0875)	(0.0812)	(0.0928)
6th	0.211 **	0.177 **	0.172 *
	(0.0842)	(0.081)	(0.0934)
7th	0.180 **	0.143 *	0.140
	(0.0832)	(0.0827)	(0.095)
8th	0.129	0.087	0.087
	(0.0854)	(0.089)	(0.1)
9th	0.055	0.005	0.011
	(0.0962)	(0.1043)	(0.1114)
10th	-0.044	-0.104	-0.090
	(0.1204)	(0.1319)	(0.1322)
Average	0.209 **	0.201 **	0.209 **
	(0.0809)	(0.0803)	(0.0825)

Note: user definitions based on the exercises performed in 2016 outside the competition. Clustered standard errors at school level in parentheses. Significance level: * p <0.10, ** p <0.05, *** p <0.01.

Figure 7: Marginal effect of use of PAM on math test scores by percentile of SES



Note: users defined as students who performed at least 10 exercises between 2014 and 2016. The dashed blue line indicates the average marginal effect.

6.3 The impact of PAM by gender

In Table 7, we observe that there are no significant differences in the percentage of PAM users according to gender. In the case of the participation in the Competition in Mathematics in 2016, we observe that girls have a slightly higher participation rate.

Table 7: Percentages of users of PAM by gender and by participation in the competition in mathematics in 2016

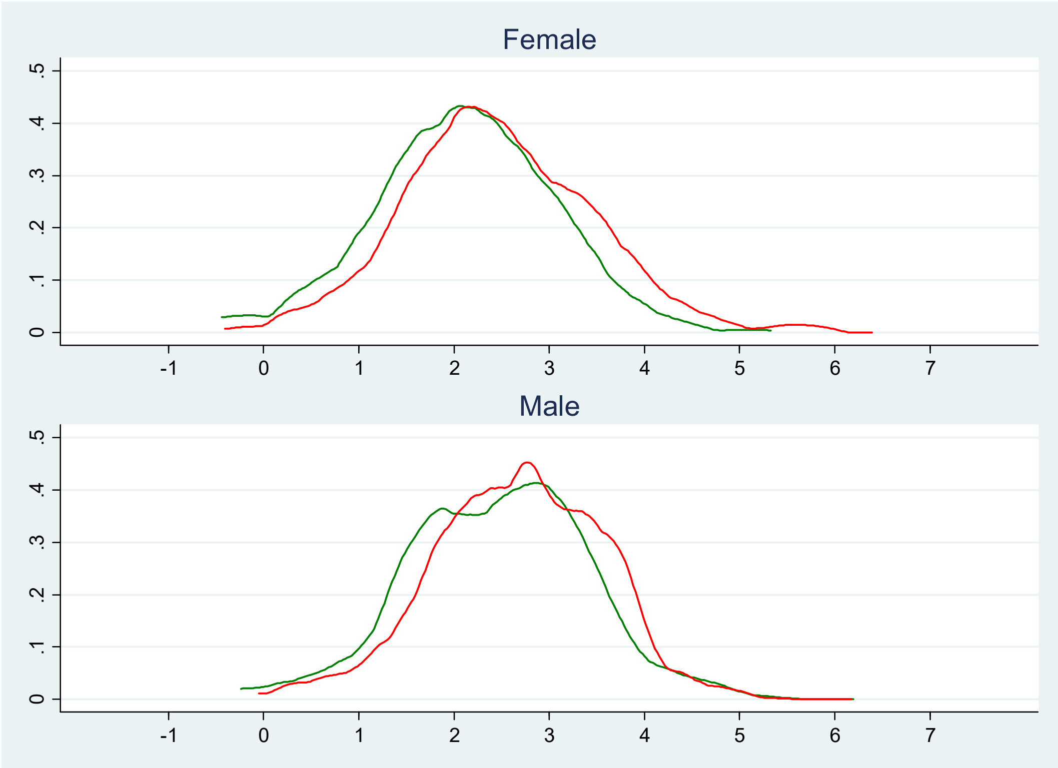
	Male	Female	Total
User definition:			
Exercises in 2016 outside the competition			
1 or more	44.4	44.4	44.4
10 or more	39.1	39.0	39.1
20 or more	35.8	31.8	33.8
Exercises between 2014 and 2016 (*)			
1 or more	67.2	67.7	67.5
10 or more	51.2	52.4	51.8
20 or more	42.3	43.1	42.7
Participants in the Competitic	31.9	33.2	32.5

Note: The number of exercises between 2014 and 2016 refers to an annual average for those three years.

Figure 8 shows the distribution of the test score gain in mathematics for users and non-users and male and female. We do not observe a different effect of PAM in girls and boys (i.e. the shift of the red curve is similar for boys and girls).

From the econometric estimations (Table 8) we conclude that there is not statistically significant differential effect of the use of PAM by gender. The term that shows the effect of the interaction between the dummy variable PAM use and gender is not statistically significant in any case. In other words, the hypothesis that PAM use affects equally the gain in math test scores of boys and girls cannot be rejected.

Figure 8: Distribution of the gain in mathematics of PAM users and non-users by gender



Note: users defined as students who performed exercises between 2014 and 2016.

Table 8: Estimation of the differential effect of the use of PAM by gender

	1 or more exercises	10 or more exercises	20 or more exercises
Interacction of Use of PAM and Gender	**-0.146**	**-0.094**	**-0.029**
	(0.153)	**(0.1505)**	**(0.183)**
Marginal Effect of Use of PAM by Gender			
Male	0.299 ***	0.267 **	0.223 *
	(0.1092)	(0.1264)	(0.1185)
Female	0.153	0.173 *	0.193
	(0.1117)	(0.0952)	(0.1276)
Average	0.209 **	0.201 **	0.209 **
	(0.0809)	(0.0803)	(0.0825)

Note: user definitions based on the exercises performed in 2016 outside the competition. Clustered standard errors at school level in parentheses. * p <0.10, ** p <0.05, *** p <0.01.

7. Conclusions

The objective of this paper was estimating the effect of using a mathematics computer-assisted learning platform on the gain in mathematics test scores based on longitudinal data from a sample of primary school students in Uruguay.

The available data allowed an empirical strategy to identify the causal effect that used relatively weak assumptions. This was possible thanks to the availability of relatively reach longitudinal data. This has allowed both to observe an important variability in the use of PAM by the students, and to have the possibility of using the gain in reading to control for the possible selection bias (that is usually assumed in this type of analysis). The results are robust to different specifications of the model and definitions of PAM user.

We found that PAM has a positive effect on math test scores. The estimated effect on math test scores gain is of the order of 0.2 standard deviations. A reference value for this result is the average progress achieved by the same students after three years of schooling which is estimated at 2.6 standard deviations.

When analysing the impact of the platform with a definition of use at the group level, we observe that the higher the use in the class the greater the gain students obtain. This result suggests the importance of the role of the teacher and the possibility of the existence of some type of positive externality at the class level.

The results of the econometric exercises also indicate that the impact of PAM use increases as the socioeconomic level decreases and that there is no statistically significant differential effect of the use of PAM with respect to gender.

Before concluding, some remarks are needed. It is important to highlight the nature of the intervention being evaluated. PAM was made available to teachers and students with few guidelines as to how it should be used[19]. The ways in which the tool is used for pedagogical purposes vary and are relatively unknown. Teachers who have adopted the tool have made their own assessment of how and how much to use it and, consequently, could have altered other pedagogical actions. Therefore, the average impact of PAM use can reflect differences in content, pedagogy, learning time, etc., linked to the decision to use PAM. A deeper study of these practices may shed light on the best ways of using the tool.

Finally, it is important to point out, as Muralidharan et al. (2016) do, that the effectiveness of the incorporation of technology into the education should not be interpreted as a loss of importance of the teacher in education. On the contrary, when the technology is incorporated to perform routine tasks (such as skill classification) and intensive tasks in data analysis (such as identifying patterns in students´ responses and providing differentiated feedback and orientation), it can complement and enhance the teacher's efforts. For example, teachers can dedicate more time to those aspects of education where they have a comparative advantage over technology, such as support for group learning strategies that can help develop social and other non-cognitive skills.

[19] There have been incentives to use it by Plan Ceibal such as promotional and informative offerings, workshops, courses for teachers and competitions for students but not specific guidelines.

References

Duflo, E., Dupas, P., and M. Kremer. 2011. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101: 1739-1744

Banerjee, A., Cole, S., Duflo, E. and L. Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122 (3): 1235-64.

Banerjee, A. , Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M. and M. Walton. 2016. "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of 'Teaching at the Right Level' in India". CEPR Discussion Paper 11530, Center for Economic Policy Research, London.

Bulman, G. and R.W. Fairlie. 2016. "Technology and Education: Computers, Software and the Internet," in Eric Hanushek, Stephen Machin, and Ludger Woessmann, eds., Handbook of the Economics of Education, Elsevier, pp. 239-280.

Cheung, A. and R. Slavin. 2013. The Effectiveness of Educational Technology Applications for Enhancing Mathematics Achievement in K-12 Classrooms: A Meta-Analysis. *Educational Research Review*, 9: 88-113.

European Commission. 2013. “Survey of Schools: ICT in Education”. Digital Agenda for Europe. Final Report.

Linden, L. 2008. "Complement or Substitute? The Effect of Technology on Student Achievement in India." Edited by Michael Trucano. InfoDev Working Paper 17, World Bank, Washington, DC.

Muralidharan, K., Singh, A., and A. Ganimian. 2016. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." NBER Working Paper 22923, National Bureau of Economic Research, Cambridge, MA.

Sung, Y., Chang, K. and T. Liu. 2016. “The Effects of Integrating Mobile Devices with Teaching and Learning on Students’ Learning Performance: A Meta-Analysis and Research Synthesis”. Computers & Education, 94: 252-275.

UNESCO-OREALC. 2016. Technical Report. Third Regional Comparative Study and Explanatory, TERCE. Santiago, Chile.

Appendix

Table A1: Effect of using PAM at least once between 2014 and 2016

	(1) Gains Math - Gains Read.	(2) Gains Math - Gains Read.	(3) Gains Math	(4) Gains Math
User in 2014-16	0.3242**	0.2164*	0.3194***	0.2496**
	(0.13)	(0.11)	(0.10)	(0.10)
Stand. Error	1.19	0.82	0.94	0.68
Schools	237	237	237	237
Students	2143	2143	2143	2143

Notes: User defined as a student who performs one or more exercises during 2014-2016. Dependent Variable: Columns (1) and (2): Score Gains in Math - Score Gains in Reading; Columns (3) and (4): Score Gains in Math. Control Variables in columns: Column (2): score in Math and Read 2013, Use 2013, Female, SES, Region, Type of School, Participation in Competition; Column (4): score in Math and Read 2013, Gain in Reading, Use 2013, Female, SES, Region, Type of School, Participation in Competition. Standard Errors clustered by School in parentheses. Significance level: *p<0.10,**p<0.05,***p<0.01.

Table A2: Effect of using PAM in the years 2016, 2015 and/or 2014

	(1) Gains Math - Gains Read.	(2) Gains Math - Gains Read.	(3) Gains Math	(4) Gains Math
User 2016	0.1744	0.1814**	0.2602*	0.2092**
	(0.13)	(0.07)	(0.14)	(0.08)
User 2015	0.1384	-0.0239	0.0756	0.0017
	(0.17)	(0.09)	(0.14)	(0.08)
User 2014	0.1064	0.1351**	0.0582	0.1035
	(0.17)	(0.07)	(0.13)	(0.08)
Stand. Error	1.19	0.82	0.94	0.68
Schools	237	237	237	237
Students	2143	2143	2143	2143

Notes: User defined as a student who performs one or more exercises during the year (for 2016 it refers to the exercises outside the competition). Dependent Variable: Columns (1) and (2): Score Gains in Math – Score Gains in Reading; Columns (3) and (4): Score Gains in Math. Control Variables in columns: Column (2): score in Math and Read 2013, Use 2013, Female, SES, Region, Type of School, Competition; Column (4): score in Math and Read 2013, Gain in Reading, Use 2013, Female, SES, Region, Type of School, Participation in Competition. Standard Errors clustered by School in parentheses. Significance level: *p<0.10, **p<0.05,***p<0.01.

Table A3: Effect of the proportion of students in the group of 2016 who used PAM in different years

	(1) Gains Math - Gains Read.	(2) Gains Math - Gains Read.	(3) Gains Math	(4) Gains Math
Users in the group 2016	0.0578	0.2089*	0.2383	0.2667**
	(0.20)	(0.11)	(0.19)	(0.11)
User 2015	0.0367	0.0031	0.0187	0.0263
	(0.19)	(0.10)	(0.15)	(0.10)
User 2014	0.0628	0.1538**	0.0138	0.1115
	(0.19)	(0.08)	(0.15)	(0.08)
Stand. Error	1.19	0.81	0.95	0.68
Schools	210	210	210	210
Students	1842	1842	1842	1842

Notes: Users in the group = proportion of the students in the group of 2016 who performed exercises in this year (refers to the exercises outside the competition). Dependent Variable: Columns (1) and (2): Score Gains in Math - Score Gains in Reading; Columns (3) and (4): Score Gains in Math. Control Variables in columns: Column (2): score in Math and Read 2013, Use 2013, Female, SES, Region, Type of School, Competition; Column (4): score in Math and Read 2013, Gain in Reading, Use 2013, Female, SES, Region, Type of School, Competition. Standard Errors clustered by School in parentheses. Significance level: *p<0.10, **p<0.05,***p<0.01.

Table A4: Effect of the intensity of PAM use (= log of 1 + number of performed exercises)

	(1) Gains Math	(2) Gains Math	(3) Gains Math
lej1	0.0411*		
	(0.02)		
lej1_2016f		0.0518**	
		(0.02)	
lej1_grupo			0.0437**
			(0.02)
Stand. Error	0.68	0.68	0.68
Schools	237	237	210
Students	2143	2143	1842

Notes: lej1 = logarithm of 1 + the annual average number of exercises between 2014 and 2016; lej1_2016f = logarithm of 1 + the number of exercises in 2016 (outside the competition); lej1_grupo = log of 1 + the median number of exercises of the group 2016 (outside the competition). Dependent Variable: Score Gains in Math. Control Variables in columns: Score in Math and Read 2013, Gain in Reading, Use 2013, Female, SES, Region, Type of School, Participation in Competition. Standard Errors clustered by School in parentheses. Significance level: *p<0.10, **p<0.05, ***p<0.01.

The UNU-MERIT WORKING Paper Series

2019-01 *From "destructive creation" to "creative destruction": Rethinking Science, Technology and innovation in a global context* by Luc Soete

2019-02 *Do young innovative companies create more jobs? Evidence from Pakistani textile firms* by Waqar Wadho, Micheline Goedhuys and Azam Chaudhry

2019-03 *What gains and distributional implications result from trade liberalization?* by Maria Bas and Caroline Paunov

2019-04 *FDI, multinationals and structural change in developing countries* by André Pineli, Rajneesh Narula and Rene Belderbos

2019-05 *The race against the robots and the fallacy of the giant cheesecake: Immediate and imagined impacts of artificial intelligence* Wim Naudé

2019-06 *The middle-technology trap: The case of the automotive industry in Turkey* by Ibrahim Semih Akçomak and Serkan Bürken

2019-07 *The impact of a mathematics computer-assisted learning platform on students' mathematics test scores* by Marcelo Perera and Diego Aboal