Are Public Schools in Developing Countries Ready to Integrate EdTech into Regular Instruction?

Ferman, Bruno and Finamor, Lucas and Lima, Lycia

Sao Paulo School of Economics - FGV, Yale, Sao Paulo School of Business Administration

27 June 2019

Are Public Schools in Developing Countries Ready to Integrate EdTech into Regular Instruction?*

Bruno Ferman[†] Lucas Finamor [‡] Lycia Lima[§]

First Draft: June 2019

This Draft: August 2021

*We would like to thank the staff from the Lemann Foundation and from Khan Academy for their support to the project. Lemann Foundation funded the collection of the survey data. An initial draft of this paper was part of Lycia Lima's PhD dissertation at the Sao Paulo School of Business Administration - FGV and it circulated under the title "*Are Public Schools Ready to Integrate Math Classes with Khan Academy?*". This project was pre-registered at AEA RCT (ID 0002456).

[†]Sao Paulo School of Economics - FGV, bruno.ferman@fgv.br, *corresponding author*. Rua Itapeva, 474. Sao Paulo, Brazil. +(55) (11) 3799 33 50

[‡]Yale University - Deparment of Economics, lucas.finamor@yale.edu

[§]São Paulo School of Business Administration - FGV, lycia.lima@fgv.br

Abstract

We study the impacts of a program that introduced a computer-assisted learning platform into regular math classes using a randomized control trial in Brazilian primary public schools. Once a week, teachers would take their students to the school's computer lab and teach using a dynamically adaptive platform, instead of their standard math classes. We find no average treatment effect on students' math proficiency. However, we find positive effects of the program on measures of attitudes towards math. Moreover, we find suggestive evidence that the program may have positive effects on proficiency when infrastructure is better, so that it can be implemented using one computer per student. These results highlight the implementation challenges associated with education technology interventions in developing countries.

JEL Codes: C93, I21, O15

Keywords: Computer-aided learning, Education Technology, Program Implementation

1 Introduction

Primary school enrollment in the different regions of the developing world has substantially increased over the past decades, but evidence shows that converting higher enrollment into improved human capital is a challenge (Glewwe and Muralidharan, 2016; WorldBank, 2018). Among the many different approaches for addressing educational deficiency, the use of technology-enhanced instruction has been growing in popularity as an approach for improving the quality of teaching and learning.

Reviews by Glewwe and Muralidharan (2016), Bulman and Fairlie (2016), Rodriguez-Segura (2020), Escueta et al. (2020) and Ganimian et al. (2020) show that results are largely varied among different categories of education technology interventions. While the evidence suggests that simply granting hardware to students in developing countries does not lead to gains in proficiency, interventions that provide students with a given software/platform as a learning aid (the so-called computer-assisted learning, *henceforth* CAL) generally show more promising results. Such programs seem to be particularly promising if the platform has the ability to dynamically adapt content to each student's needs, addressing heterogeneous learning levels within a classroom (Ganimian et al., 2020; Muralidharan et al., 2019; de Barros and Ganimian, 2021). However, most of the available evidence on CAL interventions is related to programs that increase the number of hours students are exposed to academic instruction, complementing traditional teaching. Less is known about the effects of CAL interventions during school time, as an integrated resource into regular teaching.

We present the findings of a large-scale randomized evaluation in Brazil of a program that integrates a CAL platform into regular math classes. Once a week, teachers would take their students to the school's computer lab and the students would use a dynamically adaptive platform, under their supervision, instead of conducting a regular lecture. The use of a dynamically adaptive platform is important in the context of developing countries to address the common issue of large heterogeneity within class, in which there are often several students lagging behind the curriculum.[1] Moreover, considering an implementation integrated into regular teaching is particularly important for developing countries, where capacity constraints may preclude the use of such programs as a complement to regular activities.[2] The program was also designed to address infrastructure constraints in case schools do not have one computer available per student. In such cases, students could rotate between the platform and alternative activities not carried out in the computer.[3]

Using a national standardized exam, we find no evidence that the program enhanced math proficiency, on average. However, we find evidence that the program improved students' attitudes towards math, using a validated instrument to capture these perceptions. In an attempt to understand these results, we perform an exercise to explore the role of the quality of implementation. This exercise suggests that such null effect on students' test scores may hide a positive effect in schools with better infrastructure to receive the program, but counterbalanced by negative effects in schools with worse infrastructure, where the computer lab did not have one computer per student, so the program had to be implemented in a rotation mode. While we do not have direct experimental variation to estimate such heterogeneous effects, we are able to carry out this comparison by leveraging the design of the experiment, which delivered one treated grade at every participant school. We explain in detail the limitations of such exercise in Section 6.3, and why we consider such heterogeneity results as only suggestive. Taken together, these results suggest that the program is very efficient in engaging students and changing attitudes towards math even when we consider a more limited use of the platform, but that gains in proficiency may require a more consistent use, with better implementation.

[1] Andrabi et al. (2007); Das and Zajonc (2010); Duflo et al. (2011); Pritchett and Beatty (2015); Muralidharan et al. (2019); de Barros and Ganimian (2021)

[2] In Brazil, only 4.8% of the public schools' students are enrolled in schools that have at least 80% of its students enrolled just in one period of the day (morning or afternoon) making it possible to use the school in the other period of the day for the additional classes (using data from the Brazilian School Census of 2017). Moreover, approximately 46.8% students do not have personal computers at home (using data from national standardize exam, SAEB, in 2017 for students in 5th and 9th grades in public schools).

[3] 90% of elementary public students in Brazil are in classes with more students than the number of personal computers in the school's computer lab (using the Brazilian School Census of 2017, for all public schools and grades between 1st-9th.)

While there is mounting evidence of positive effects when CAL programs are implemented as extra hours of instruction, less is known about their relative efficacy when compared to traditional instruction.[4] Our results provide one of the first and few pieces of evidence on the effects of CAL programs with dynamic adjustments directly integrated into regular classroom. Ma et al. (2020), Büchel et al. (2020), and Bettinger et al. (2020) isolate the effect of technology by including a treatment arm with additional instruction exposure, but without technology. However, in their setting, the treatment is before or after school hours, rather than during school day. Linden (2008) study the implementation of a CAL program during regular school hours, but the platform used in this study was not dynamically adaptive. Mo et al. (2020) find null effects of a CAL program when it was implemented by a government agency, in contrast with positive results when implemented by a NGO. Their interpretation is that government officials were more likely to substitute CAL for regular instruction while the NGO always used as an additional resource. However, in their setting, the program was not designed to be integrated into regular instruction, and such implementation was contrary to the protocol. Most related to our study, Muralidharan and Singh (2021) present the pre-analysis plan of a randomized evaluation, conducted in parallel to ours, in which a CAL program with dynamic adjustments was integrated into regular classroom in India.

Our results point out to the fact that implementation challenges are particularly relevant when the program is integrated into regular school hours because, in this case, the opportunity cost of implementing this program is a regular class. We provide evidence that the program improves relative to a regular class when the implementation is better, but may be worse than a regular class when the implementation is worse. In contrast, when a CAL program is implemented as extra hours of instruction, such negative effects should be less likely to appear. We also find evidence that other infra-structure challenges, such as internet/hardware problems, that led to similar effects on total time exposure to the plat-

[4] Banerjee et al. (2007); Linden (2008); Lai et al. (2012, 2013, 2015); Yang et al. (2013); Mo et al. (2014); Muralidharan et al. (2019); Mo et al. (2020); Ganimian et al. (2020)

form, did not lead to such stark heterogeneous effects. This suggests that the effects of the program are not simply a linear function of exposure time. This would be consistent with the fact that opportunity costs of using the program are different depending on the kind of infra-structure challenges, and/or that there may be increasing returns to scale when we move from using the platform for half of a class to a full class, at least at this range of exposure time.

We also provide one of the first experimental studies of an in-school implementation of Khan Academy, which is one of the most popular online platforms focused on delivering educational content tailored at each students' level. Other studies have previously tried to investigate the effects of the Khan Academy platform use on math achievement. However the majority of the existing evidence relies on quasi-experimental approaches and/or small samples.[5] A notable exception is an experimental study by Büchel et al. (2020), who studied a randomized control trial in El Salvador that was implemented slightly after ours, in 2018. In their setting, Khan Academy entered as an additional resource that increased the duration of math exposure, while in our setting it followed the guidelines from the *Khan Academy in Schools* program, which integrated the platform into regular math classes, so it did not increase the total number of hours students were exposed to math content. They report an increase in math proficiency of 0.21σ when comparing with control students and 0.09σ when comparing with students that were exposed to the same additional hours of math classes without the technology.

2 Background and Context

Elementary education in Brazil is mandatory and goes from 1st to 9th grades, with entry ages from 6 to 14. Elementary education is in its majority publicly provided. In 2017, among

[5]For example, Chu et al. (2018) use an encouragement design to show Khan Academy led to significant improvement in students' test performance, based on a sample of 103 middle school students in the US. Using non-experimental methods, Adams (2016) and Kelly and Rutherford (2017) find no association between Khan Academy use and math test scores, while Manaus (2016), Phillips and Cohen (2015) and Weeraratne and Chin (2018) find positive results. Adams (2016) reviews other studies with qualitative evaluations.

the 183,743 schools offering elementary education, 78.8% of them were public, accounting for 83.2% of the 27 million enrolled students.[6] Public education in Brazil is completely tuition free but, similarly to other developing countries, Brazil struggles to offer good quality of education. In the 2018 PISA exam, Brazilian students had an average score of 384 in math, compared to an average of 489 for the OECD countries, placing the country in the 72th position among the 80 participant countries.[7]

Our implementing partner, the Lemann Foundation, is a non-profit organization that runs several programs with the purpose of enhancing the quality of public education in Brazil. One of their initiatives is to promote the use of CAL in public schools through the program *Khan Academy in Schools*.[8] The program engages Government's Secretaries of Education which, after signing a participation agreement, receive the support from the Lemann Foundation to implement Khan Academy in schools. The program had three main pillars: i) delivering a one day training for Math teachers to present the platform and their functionalities; ii) advising teachers to carry out one of their weekly math classes (50 minutes per week) at the school's computer lab using the platform,[9] and iii) close monitoring of intervention's implementation by Lemann Foundation staff, which acted as promoters of the program, providing assistance for solving any potential difficulties schools/teachers were facing.

Khan Academy is an online interactive platform offering free instruction and practice in mathematics as well as other subjects, such as science, computer programming, history, economics, among others. The platform offers contents in a personalized environment, adapting the user's experience to identify strengths and tackle learning gaps. The level of math con-

[6]According to the 2017 School Census.

[7]PISA stands for *Programme for International Student Assessment* and its an exam organized by the OECD to measure the abilities of language, math and science of 15 years old students in several countries. It is organized every 3 years since 2000. The scores have a mean of 500 points and a 100 points standard deviation.

[8]*"Khan Academy nas Escolas"*, later renamed to *"Innovation in Schools"* or *"Inovação nas escolas"*

[9]The 50 minutes correspond to 1 out 5 or 6 weekly math classes. Teachers select modules in the platform related to the content covered in the math curriculum. The platform then is dynamically adaptive, depending on how the students perform in the platform.

tents available ranges from basic addition and subtraction to more advanced topics, such as differential equations and multivariable calculus.

The program may enhance students' math performance through four main channels. First, it may increase the quality of math content accessed by students by offering quality material developed by specialists. The second potential channel is by increasing students' learning through offering content and exercises tailored to each students' level, addressing students' heterogeneity within class. A third channel through which the platform may have an impact on a students' performance is by shifting the students' perceptions regarding math, turning the studying experience more attractive. By presenting the math content in an interactive and friendly way, designed to promote a fun and exciting learning experience, the platform may change the students' attitudes towards math, which may be ultimately translated into an increased math performance. Finally, since teachers can access reports on students' usage in the platform, it may enhance their teaching impact as they are able to familiarize themselves with each student's level and learning gaps. Our experiment is not designed to tease apart the effects from each of these channels and it should then be seen as the composite effect of the platform. Teasing them apart would require a much larger sample, and a design in which access to each component of the treatment is independently randomized, which would not be feasible given our implementation constraints.

The implementation of the program requires a good technology infrastructure, including a sufficiently high-speed internet connection. To guarantee an adequate implementation of the program, schools that had less than 0.5 computer per student were granted additional computers from the Lemann Foundation. There was also information technology support for schools in the city of Manaus, which had weaker baseline infrastructure, to guarantee that the computers and internet were functioning. Importantly, since we are not interested in the effects of such improvements in the computer lab *per se*, all schools, irrespective of treatment status, received these benefits. Therefore, differences between treated and control grades should reflect solely the use of the platform. For the evaluation sample, we can observe two different modalities of program implementation: i) individual use of the computer and ii) rotational usage of the computer between two students. In the rotational mode, each student used the computer during half of the class, and was assigned by the teacher other math activities during the remainder of the class. The possibility of implementing the program in a rotational mode is relevant because many schools in Brazil do not have infra-structure to implement the program with one computer per student.

If this program is scaled, we should expect variation in the implementation across schools. Schools with a higher rate of computers per students should be more likely to implement the program with one students per computer, while other schools may be more likely to implement the program based on a rotation mode. Given that some schools in the experiment received additional computers, we should expect an implementation with fewer computers per student in case the program were scaled without this aid. Likewise, given that some schools received support for internet connectivity, we should expect that the schools in our experiment experienced less connectivity problems than if the program were scaled without this support. Finally, we may expect the effects of the program to be lower if there is no monitoring from an implementer like Lemann Foundation. Therefore, we see our average treatment effects as an upper bound on the effects we should expect if the program were scaled up, given that we should expect a better implementation in our experiment. We emphasize again that schools received the hardware and IT support regardless of the treatment status. Therefore, our estimates capture the effects of the program, and are *not* confounded with the effects of this additional support.

3 Study Design

3.1 Sample Selection

This experiment was conducted in primary public schools of five cities in three different regions of Brazil in the 2017 school calendar year. The cities of Barueri, Mogi das Cruzes and Sao Bernardo do Campo were selected from the Southeast region; Pelotas from the South;

and Manaus from the North region. Cities were selected based on previous relationship between the city government and the implementing partner (Lemann Foundation), and conditional on the existence of a satisfactory level of municipal school infrastructure (existence of a computer lab and internet connection).

In the five cities selected, all primary education schools were invited to voluntarily apply to the program. Among all applicants, the Lemann Foundation determined a final list composed of 166 schools that were initially eligible to participate in the treatment randomization. Out of these, before the treatment was assigned, nine schools left the evaluation sample due to lack of the necessary infrastructure or because they did not have a matching pair to form a stratum. This resulted in 157 schools in the final evaluation sample.[10] In Figure A.1 we compare the schools in the experiment with all Brazilian urban public schools in terms of the math proficiency and number of computers per student in school before the experiment. While the distribution of schools in our experiment has better schools in terms of test scores and number of computers relative to the unconditional distribution of schools in Brazil, it spams roughly the whole range of schools in these two dimensions. Still, it is not possible to rule out that the schools in our sample differ from those that did not participate in the study in terms of unobservables. For example, it may be that schools that voluntarily apply to participate in this program are more open to ed tech interventions. In this case, we interpret our results as an upper bound on the effects of this program in case it is scaled up. Note that schools were not aware about the hardware and IT support described in Section 2 at the time of enrollment. Therefore, we can rule out the possibility of negative selection in our sample due to schools with worse infrastructure enrolling to receive this support.

3.2 Experimental design

Schools may be of three different types, based on the grades they offer: (a) Cycle I schools, which offer grades 1-5 (students between 6-10 years old); (b) Cycle II schools, which

[10] There were 29 schools in Pelotas, 63 schools in Manaus, 21 schools in Barueri, 27 schools in Mogi das Cruzes and 17 in Sao Bernardo do Campo.

correspond to 6th-9th grades (students between 11-14 years old); and (c) Both cycles schools, which have students from 1st to 9th grades (students aged 6-14 years old).

In addition to the municipality and the grades offered (cycle I, II or both), schools were stratified based on two additional criteria: whether they had ever received the Khan Academy program in the years preceding the experiment;[11] and whether Math proficiency data for the 2015 national standardized exam was available. For the cases in which the resulting strata were composed of more than 5 schools, further stratification was carried out based on the math scores for the standardized national exam.

With the purpose of increasing engagement and reducing attrition, every school in our sample received the program in some grade, which was assigned randomly. Only 5th and 9th grade students participate in our study, since for these grades we have math proficiency data from a national standardized exam. Therefore, we consider as treated schools those that received the program in the 5th or 9th grade and as control schools those that received it in a different grade. Figure 1 illustrates the randomization for the three groups of schools: those that only have (i) 1st-5th, (ii) 1st-9th, and (iii) 6th-9th grade students. Treatment was randomized between 3rd and 5th grades for schools in the first group, 5th and 9th grades for the second, and 6th and 9th grades for the third group of schools. It is worth noticing that all Cycle I and II schools serve as treatment for one grade and control for the other.[12] The 157 schools in our study were divided into 35 strata (which had from 2 to 11 schools each). Since schools with both cycles had 5th and 9th grades participating in the study, our sample is composed of a total of 217 school $\times$ grades in 47 strata-grade pairs.

Even though spillovers within schools could raise a concern, schools were instructed to use Khan Academy only in treated grades and to explicitly prevent the usage in control

[11]In our evaluation sample, only 14 schools in the city of Pelotas had Khan Academy implementation in the previous years. Students in our experiment sample, however, were never exposed to the Khan Academy platform in school. In Section 5 we check whether control students were ever exposed to the platform.

[12]Some Cycle I control schools received the program in the 4th instead of the 3rd grade. This is not a problem for our experimental design, because the only relevant point here is that 5th grade students in these schools did not participate in the program, so they serve as a control group.

grades. Furthermore, in control schools, we avoided treating grades that were close to the treated ones, in terms of students' age, in order to minimize the risk of contamination. In section 5, we show that we do not detect any sign of spillovers for control students. Moreover, using the administrative data from the Ministry of Education, we verified that there are 240 teachers lecturing math in control grades, among which only 12 (5%) also teach in a grade receiving Khan Academy. Therefore, we do not see spillovers as a major concern. Finally, in Appendix A we contrast, in a difference-in-differences framework, control school-grades that participated in the RCT with other school-grades from the same municipality, but that did not participate in the experiment. We do not find any systematic differences between these two groups — if there were spillovers then we should see positive effects for control-grades. This exercise also minimizes other threats to our identification strategy, such as the possibility that school principals could treat control grades differently than if there were no RCT, to compensate these students from not participating in the program.

4 Data and Empirical Strategy

4.1 Data

Data for this study stems from two main sources. First, we use administrative data from the 2017 Ministry of Education's Basic Education's Evaluation System (*Sistema de Avaliacao da Educacao Basica - SAEB*). Every two years, at the end of the school calendar year, the government implements standardized exams to measure students' academic proficiency in the 5th and 9th grades, compulsory for all Brazilian public schools with 10 or more students. The SAEB exam also collects data on students' characteristics, including demographics, household characteristics, leisure and studying habits, parents' education, employment status and school retention record. Although this exam is implemented in all public schools in Brazil with more than 10 enrolled students, the Ministry of Education only releases proficiency data for those school grades that had at least 80 percent of enrolled students taking the test. We have administrative data for all schools in our sample (including those that left the study after treatment assignment), with the exception of those school grades that did not meet the minimum attendance requirement. The exam is high stakes for the schools' principals and local politicians, since it corresponds to the major part of a school-quality index (*IDEB*) released bi-annually by the Ministry of Education.

We also collected survey data over two rounds: a baseline carried out in March 2017, before the beginning of the program, and a follow-up in November 2017, right before the end of the school calendar year. Baseline data was not collected for one municipality (Sao Bernardo do Campo), because this municipality joined the evaluation late.[13] We collected data for an instrument that measured students' attitudes towards mathematics Brito (1998), who translated to Portuguese and validated in Brazil the instrument originally developed by Aiken Jr and Dreger (1961). This instrument was composed of a questionnaire with 20 questions that presented different statements about an individuals' feelings regarding Math, with Agree/Disagree four point Likert Scale answer options. The different statements express either a positive or a negative connection with Math (such as *"Mathematics is enjoyable and stimulating to me"* or *"Mathematics makes me feel uneasy and confused"*).[14] An index for attitudes towards math was created by summing up all scores for positive statements, and adding the reverse score for negative statements, and then standardized to have zero mean and standard deviation one within the control group, by grade level.[15]

We also collected data on students' demographic characteristics, students' self reported access and usage of computer and internet both at home and at school as well as their preference in relation to school subjects. On the follow-up survey, information on the knowledge and usage of Khan Academy was also collected to assess program compliance and contamination in the control group. Unfortunately, we are not able to link individual level administrative data with survey data because the SAEB dataset is de-identified.

[13] No data was collected for the 7 schools that dropped the program right after treatment assignment.

[14] See the original papers for the full list of questions. Aiken Jr and Dreger (1961) have the original questions in English and Brito (1998) the translated sentences to Portuguese.

[15] An answer of 4 in a negative statement was recoded into 1 to reflect the reaction to an opposite positive statement, and so on. For details on the construction of the index consult the original papers.

We complement the two main data sets with reports from the Lemann foundation on the status of implementation in each school. Since every school had one treated grade, we have this information for all schools in our sample. We also use information extracted from the Khan Academy platform on the usage by treated students. This information is useful for a descriptive view of the implementation of the program, and it is not available for students in the control group.

4.2 Balance and Attrition

4.2.1 SAEB data

Table 1 presents the baseline balance for the 14 covariates reported in the SAEB data set. The first column shows the mean for the control control group for each variable and the standard deviation in square brackets. The second column shows that regression adjusted differences between control and treatment groups, displaying the estimates from a regression for each covariate on an indicator variable for the treatment and strata-grade fixed effects, with standard errors (in squared brackets) clustered at the strata level. In the third column we display the number of valid observations. We present these results for the pooled sample and separately for the 5th and 9th grade. We also present the p-value for joint significance of all variables and we do not have evidence of significant differences between treatment arms in any of the samples considered.

There are two potential sources of attrition in the SAEB dataset: i) school-grade-level attrition, since proficiency data is only released by the Ministry of Education for those school-grades that had at least 80% of student attendance in the exam and ii) student-level attrition for those students that did not take the SAEB exam. In Panel A of Table 2, we show school-grade level attrition results for the SAEB exam. For this dimension, we define attrition as the absence of math proficiency data in the SAEB exam, at the school-grade level. We report the control group mean, regression adjusted differences between treatment and control groups, the number of observations and number of clusters, for the pooled sample, and for the 5th and 9th grades subsample respectively.[16] There are no significant differences in attrition rates between treatment and control groups, showing that the intervention is not correlated with the likelihood of the schools having SAEB data reported. In Panel B, we use student-level data in the SAEB exam to show that there are no differences between treatment and control groups on the proportion of students not taking the SAEB test (for those grades that had the results reported). In both cases, the attrition rate is low and not correlated with the treatment status.

4.2.2 Survey

Table 3 presents survey student-level baseline characteristics and the balance tests, following the same structure as Table 1. The results demonstrate randomization was successful as characteristics are balanced across treatment arms (the p-value of a joint test that there is no difference between treatment and control for all baseline covariates is equal to 0.696, 0.275 and 0.790 respectively for the three samples considered).

There are two potential sources of attrition in the survey, school-level and student-level attrition. Our first source of attrition is associated with schools that left the program after treatment assignment. Seven schools out of our sample of 157 schools - both in treatment and control groups - left the study after randomization took place for various reasons, mostly unrelated with treatment assignment. The small number of school dropouts and the different reasons associated with the withdraw minimize our concerns with differential selective attrition.[17] The second source is student-level attrition which is related to students either not being present in class during the survey application or failing to complete the answers for the attitudes towards math instrument. In Panel C of Table 2, we show that survey attrition rate is around 29% when we consider as attrited students those that did not answer all of

[16]The dependent variable is an indicator whether there is no outcome data available.

[17]Two out of seven schools left the program after randomization but before communication of treatment assignment. Out of the other 5 schools that dropped out, only 2 dropped out due to problems with the treatment assignment (one school assigned treatment in the 5th grade and one school assigned control in the 5th grade), and one school due to lack of teachers' engagement. The remaining 2 schools left the program due to unavailability of the computer lab and absence of computer lab instructor.

the 20 questions on attitudes towards math. This is mostly driven by students not being in school on the day the survey was administered. In Panel D, we consider a stricter measure of attrition when students failed to complete any of the 20 questions and therefore we cannot construct the index measure of attitudes towards math as defined by Brito (1998).

In both cases, attrition in treatment group is 2.5-2.8 percentage points lower than that in the control group (p-values 0.058 and 0.083). Even though the differential attrition is small and just marginally significant, we conduct several robustness and validation checks to assess if there is any evidence that it might threaten the results using the survey data (*attitudes towards math*). In Appendix Table A.1, we show covariates remain balanced between treatment and control groups even after conditioning on the sample of non attritors in the follow-up survey round. We also show that treatment effects estimates in subsamples defined by grade-municipality is not correlated with neither the attrition level nor the differential attrition. In appendix Table A.5 we also show that, as expected, attritors have worse baseline outcomes than non-attritors which will be useful when interpreting the results using the bounds procedure proposed by Lee (2009). Finally, while we focus on an attitudes measure that requires a non-missing answer for all 20 questions, we also consider a different form of aggregation that generates non-missing values if the student answered at least one of the questions.

4.3 Empirical Strategy

The experimental design generated random variation on which school $\times$ grades had their teachers assigned to receive a Khan Academy training from the Lemann Foundation, and to use the platform integrated to one math class every week. We define the "treatment" as the teacher being assigned to receive this training and follow up from the Lemann Foundation, and the class being assigned to use the Khan Academy platform as recommended in the intervention, which was expected to last for approximately 24 weeks.[18]

It is not possible to guarantee, however, that all teachers followed the exact plan of the intervention (that is, integrating the platform in one math class per week for the treated grades). Moreover, while every school in the sample had at least one treatment and one control grades, and every school declared they were committed to avoiding control grades' usage of the platform, the Khan Academy platform is free and openly available. It is, therefore, possible, although improbable, that control students and teachers were using it. For these reasons, our estimates should be considered as an intention to treat effect (ITT) of the intervention. In Section 5, we show that contamination to the control students was minimal, and that the intervention significantly increased the exposure of treated school students to the platform.

Our ITT estimates are based on the following regression:

$$y_{igs} = \alpha + \beta_{\text{ITT}} Z_{gs} + \Gamma \mathbf{X}_{igs} + \epsilon_{igs}, \tag{1}$$

where y_{igs} is an outcome of interest for individual i, who belongs to grade g in a school s, Z_{gs} is an indicator variable that takes value 1 if students in grade g and school s were treated, $\mathbf{X}_{igs}$ is a set of baseline controls, which includes strata fixed effects, and ϵ_{igs} is an error term.[19] β_{ITT} is the average treatment effect of the program. We report both results pooling 5th and 9th grades (in which case we interact the strata fixed effects with grade), and separately for each grade. Standard errors are clustered at the strata level, following a recent recommendation by de Chaisemartin and Ramirez-Cuellar (2019). Note that, this way, we allow for the error of different students within the same school to be correlated. We assess the reliability of such standard errors using the assessment proposed by Ferman (2021).

[18]There was some variation on the start date of the intervention in the different cities. Pelotas, Barueri and Mogi had 24 weeks of exposure, while Manaus had 20 weeks and Sao Bernardo had 16 weeks. Results are similar if we drop observations from Sao Bernardo.

[19]All variables present in the balance tables are included in the regressions with controls.

We consider two main outcomes: math proficiency and attitudes towards math.[20] Our math proficiency results are based on the SAEB data, which covers all schools of our sample, including the 7 schools that left the study after treatment assignment (although excluding the school-grades for which data was not released). For attitudes towards math, we rely on survey data, for which we only have information for the subsample of compliers (150 schools). All scores were standardized to have zero mean and standard deviation one within the control group, by grade level.

5 Program Implementation and Compliance

5.1 Evidence from students' survey

Before presenting the treatment effects on the main outcomes of interest, we present in this section evidence that the students allocated into treatment group were exposed to Khan Academy, and that we find no evidence of contamination in the control group. Table 4 shows results for the follow-up survey which, in addition to collecting data on attitudes towards math, gathered information on other variables, such as student's familiarity with Khan Academy, reported use during school, use of computer and preferences regarding subjects. The table displays, for the pooled sample and 5th and 9th grades separately, the control group mean, the regression adjusted differences between treatment arms and the number of observations for different variables collected on the follow up survey round.

Our results show that around 97% of the students in treated grades report using Khan Academy (around 82% report using it in school). In the control group, only 6.3% of the students report using the platform (4.4% report using in school), so contamination does not raise major concerns. The intervention increased the probability that students report using the computer lab at schools, both during and outside class. The coefficient for using the computer lab during math classes is very large and significant, as expected. Students in

[20] Math proficiency and attitudes towards math were the main outcomes registered in the paper's pre-analysis plan. AEA RCT Resgistry: AEARCTR-0002456.

treated grades were 44.5pp more likely to report that they use computer lab during math classes. There is evidence that the intervention has not substantially crowded out other school activities happening in the computer lab, as the results suggest the probability of using the computer lab in other classes decreased by a very small magnitude (-5.5pp) relative to the increased use during math class. The intervention also increased the probability that students report using the school computer lab not during classes, which is consistent with treated students using the platform even after school hours. While we do not find an increase in the proportion of students who use computer at home, this does not imply that treated students are not using Khan Academy at home, as the program may have increased the probability of using Khan Academy at home for those who report frequently using computer at home regardless of the treatment status.

5.2 Evidence from implementation and usage monitoring

Lemann Foundation's staff visited all schools five times throughout the school year, and during these visits they collected information on the usage of the Khan Academy platform. We use this information to assess the quality of implementation and how it affects students usage. While virtually all treated students were exposed to platform, many schools experienced some implementation problems during the program. In about 31% of those visits, they reported that the implementation was inadequate. In 71% of those cases, inadequate implementation was due to infrastructure problems. Of those cases with infrastructure problem, around 78% was due to internet connectivity problems, while around 15% was due to problems with the computers. Overall, 51% of the schools reported inadequate implementation due to infrastructure problems in at least one month.[21]

Another important information collected by Lemann Foundation's staff was about the modality of implementation in terms of number of students per computer. In around 37% of the schools, there was one computer for each student, so that students could spend the whole

[21] Around 7% of the cases with inadequate usage were because there were no math teachers during that period, and around 5% of the cases were because teachers were not motivated with the project.

math class in the platform. For the other schools, there was a rotation system, in which students would use Khan Academy for half of the class, and work on other math-related activities not in the computer for the remainder of the class.[22]

Such implementation challenges had important consequences for the total time of exposure to the platform. Based on the recommended implementation of one class per week, we would expect to see in the rotational modality approximately 600 minutes of use for the duration of the study, roughly 25 minutes per week, while in the modality of one computer per student the expectation was for students to have twice this exposure.[23]

In columns 1 to 3 of Table 5, we show how the total number of minutes logged in the platform correlates with infrastructure problems and with the type of implementation. In schools that implemented the program with rotation and had infrastructure problems, 5th graders spent 540 minutes logged in the platform from April to October.[24] When a school did not present internet problems, 5th graders spent approximately 30% more minutes in the platform, while in schools with one computer per student 5th graders spent 42% more minutes. 9th graders spent substantially fewer minutes in the platform relative to 5th graders, spending a total of 386 minutes in schools with infrastructure problems and with rotation. This number was 48% higher in schools with one computer per student, but no higher in schools with no infrastructure problems.

We also present in columns 4 to 6 of Table 5 the number of weeks students logged in the platform. 5th grade students logged in more weeks than 9th graders, and 5th graders in schools with no infrastructure problems logged in more times. However, there is no significant difference in the number of weeks logged in for schools with one computer per student, suggesting that the larger number of minutes in such schools come mainly from the intensive

[22]There is no information on the type of implementation for 9 out of 150 schools. For these schools, the staff from the Lemann Foundation did not collect this information during the visits.

[23]The communication with schools principal and teachers emphasized the usage for one weekly math class. Expectations for total usage in the academy year were not communicated.

[24]We consider usage from the beginning of the implementation until the SAEB exam. If we considered until the end of the school year, then these students would have a total of 687 minutes in the platform.

margin of usage. In Appendix Figure A.2 we show that the infrastructure problems (related to internet connection and other hardware issues) were concentrated in the beginning of the intervention up until June. As mentioned in Section 2, the implementing partner provided support for schools with internet connection problems so, in the last months of intervention, most of the schools no longer experienced connectivity challenges. Therefore, it is important to highlight again that we should expect a worse implementation of the program in case it is taken to scale without such support. It is important to recall that, given our research design in which all schools had at least one treated grade, such internet support from the implementing partner does not affect the internal validity of our findings.

Finally, we also have information on usage of teachers throughout the year. We find that around 90% of the teachers logged in the platform at least once. Around 40% to 60% of the teachers logged in the platform in each month of the study, and, conditional on logging in, they used the platform for 9.7 minutes in each session.

6 Results

6.1 Treatment Effects on Math Proficiency

Columns 1 and 2 of Table 6 show intent to treat effect estimates of the program on math proficiency for the pooled sample (Panel A), and for the 5th and 9th grades separately (Panels B and C), using the administrative data from the national exam. The first column shows the results for the regression on the treatement indicator and strata fixed effects while the second column includes, additionally, the covariates specified in equation 1.

On average, we find no differences in math proficiency between students attending grades assigned to treatment and control groups. In this dimension, there is no effect of the program on average for the pooled sample or for the 5th and 9th grades individually. The estimates are precise enough to rule out large positive treatment effects on math proficiency. The pooled sample standard error implies that the study was well powered to detect effects of 0.09 standard deviations. The 95% confidence interval is given by [-0.063,0.031]. The inference assessment based on Ferman (2021) does not detect large problems with the inference procedure when we consider the full sample or when we restrict to 5th grade students, but suggests that we should consider inference for the subsample of 9th graders with more caution.[25]

6.2 Treatment Effects on Attitudes Towards Math

In columns 3 and 4 of Table 6 we present the results for the attitudes towards math index. Our results indicate that students attending treatment grades had slightly higher, and significant, scores in the attitudes towards math index (0.061σ for the pooled sample, 0.062σ for the 5th grade and 0.057σ for the 9th grade, for the specification including covariates).

While differential attrition is marginally significant for this outcome, we show evidence that such differential attrition does not explain these results. In Appendix Figure A.3, we contrast the point estimates of the effects for each each region $\times$ grade with the differential attrition in this cell (plot on the left). If our results were driven by differential attrition, then we should expect larger effects in cells such that the differential attrition is higher. We do not find such evidence. In the panel on the right, we also show that point estimates are not systematically related with attrition rates in the control group. In particular, the effect is highest exactly for the cell with lowest attrition and with close to zero differential attrition. Combined with the information from Appendix Table A.1 that treated and control students are well balanced even when we condition on being a non-attritor, we believe our positive effects on attitudes are not driven by attrition. Additionally, we present in Appendix Table A.3 the bounds proposed by Lee (2009), which yields a lower bound of 0.03 and an upper bound of 0.13. While we cannot reject that the lower bound is different from zero, given the

[25] The assessment proposed by Ferman (2021) calculates the size of the inference method if we consider that the null is true and errors are iid normal, ranges from 6% to 7% when we consider the full sample or the sample of 5th graders. This suggests that the number of strata is reasonably large enough to justify inference based on standard errors clustered at the strata level. The assessment, however, is higher for regressions using the sample of 9th graders, reaching up to 8.9% in the specification including covariates using math proficiency as outcome variable. This suggests that inference based on this subsample should be considered with caution.

evidence above, we believe the true effect is far from such lower bound.

As we discussed in Section 4.2.2, the attrition is relatively larger when we consider only students that answered all the 20 questions. Approximately 10% of the students responded some questions, but not all of them. We consider an alternative measure of the index that only takes into account valid questions for each student and re-weight them to have the same support as the original index. We present this measure in the appendix table A.4, showing that we obtain similar results.

In the Appendix Table A.6, we repeat the same exercises from sections 6.1 and 6.2 using only a common sample of school-grades that are observed in the data. The results are very similar, given the high overlap between the two data sources: from the original 217 school-grade pairs, 86.1% are observed in the national exam, 96.3% in the survey and 82.5% in both of them.

6.3 Potential explanations for the results

The above results point in the direction of modest effects on attitudes towards math that were not translated in average proficiency gains. As we discussed in section 5, there were implementation challenges in several schools. In particular, infrastructure problems such as non-reliable internet connection, and schools that implemented the program on the modality based on rotation of students, prevented a more consistent use of the platform in some schools. We consider whether those schools with worse implementation are driving the null result on math proficiency.

While we do not have experimental variation on the modality and quality of implementation, we take advantage of the fact that all schools implemented the program in at least one grade and use school-level implementation information that covers our entire sample to perform a heterogeneity exercise. Following our instructions, Lemann Foundation staff visited all schools in our sample, collecting data on implementation in all schools in exactly the same way, irrespective of the grade that received the program. Given that, within each school, we extrapolate the information on infrastructure problems and type of implementation from the treated to the control grade, so that we can use these variables to estimate whether the treatment effect was different depending on these implementation variables. Such empirical strategy relies on the assumption that, conditional on type (or quality) of implementation, potential outcomes of students are independent of the grade that received the program in their schools. This assumption could be invalid if, for example, school principals put more effort in guaranteeing that the infrastructure is working well when the program is assigned to one of the grades that will be evaluated in the SAEB exam. Alternatively, the type of implementation may depend on the grade if grades have substantially different number of students.

If those aspects were relevant, then we should expect better implementation in schools that received the program in a grade that was evaluated in the SAEB exam. In Table 7, we provide evidence that this is not the case. In Panel A, we show the results of a school-grade-level regression of a dummy variable that takes value one if the there are no internet connection/hardware problems on the treatment indicator and strata fixed effects. In columns 1-2, we display the results for 5th and 9th grades for all schools. For example, the results presented in column 1 compare the proportion of schools with no infrastructure problem in the 5th grade control schools (so this information comes from implementation in the other grades in these schools) to this information for 5th grade treated schools. Columns 3-4 and 5-6 show estimates for 5th and 9th grades in two cycle and one cycle schools respectively. In Panel B, we perform the same exercise using an indicator of one computer per student as a dependent variable. None of the estimated coefficients are significant, providing support to the validity of the assumption our extrapolation exercise relies on.[26] In Appendix Table A.2, we also show that, controlling for school fixed effects, the number of students per classroom does not significantly vary by grade. This provides further evidence that we should expect that the computer lab of a given school would comport the same modality

[26] Standard errors are not reported for the 9th grade in the subsample of one cycle schools, as the dependent variable reflecting good infrastructure was equal to zero for all 14 schools in this group.

of treatment (rotation versus one computer per student) regardless of the treated grade. Finally, in Appendix Tables A.7, A.8, A.9 and A.10, we compare our baseline variables for treated and control schools conditioning on the quality of the implementation. While the p-values for the joint tests are always large, there are some significant differences in baseline test scores. Therefore, we always control for these variables when we consider this exercise.

Table 8 presents the results for this heterogeneity exercise. We present the results for math proficiency in Panel A, and for attitudes towards math in Panel B. In column 1, we interact the treatment with information on whether the implementation had one computer per student or computers were shared by two students. We find positive effects for students in schools with one computer per student, while negative effects for students in schools with a shared implementation. We reject the null that the effects were the same for students in these different types of school. As we show in columns 3 and 5, the positive effect in schools with better infrastructure was mainly driven by 5th grade students. In Appendix Table A.11 we show that such heterogeneous effects remain even when we consider only within-municipality variation in the type of implementation. For 9th graders, we find negative effects for both types of implementation, and we cannot reject the null that the effects were the same for these two groups.

These results are consistent with the fact that, as discussed in section 5.2, 9th graders had a substantially lower use of the platform. Another possibility is that 5th grade students were exposed to this material at an earlier stage of their school life. It may also be that, for 5th graders, there is a greater correspondence between the platform and the curriculum in that grade. However, the idea of the program is that teachers choose the modules students would access based on the grade curriculum, and the platform covers the material from both grades. Finally, it is important to note that we analyze the results for 9th graders with caution, because the sample is much smaller. Indeed, the inference assessment proposed by Ferman (2021) suggests that the standard errors for the regressions restricted to 9th graders may be substantially underestimated, and this is aggravated when we consider heterogeneous effects for this subsample.

In columns 2, 4, and 6 of Table 8, we also include in the regression an interaction between the treatment dummy and a dummy indicating whether there were internet problems in the school. There is no evidence that treatment effects differed when there were internet/hardware problems, once we condition on the modality of implementation.[27] While internet problems and shared implementation had a similar negative effect on the number of minutes exposed to the platform (Table 5), it is interesting that those infrastructure problems had different consequences on the treatment effects. Such difference can be rationalized once we consider that a reduction in the amount of exposure to the platform because there were internet problems is very different from a reduced exposure because students had to share the computers. In the first case, teachers could still fall back to a standard class without technology. In this case, the infrastructure problems would reduce the exposure to the platform, but would not reduce the total exposure to math content. In contrast, in the second case students would experience only half of the time with the platform in each session, and the other half with other math activities. Our results suggest that, at least around this range of exposure, there are relevant gains of scale of studying in the platform in a given session, and spending half of a class on the platform is not sufficient to attain such gains of scale. In this case, spending half of a class in the platform and the other half in other activities could be worse than spending the full class in the platform, and also worse than spending the full time in a regular class with no technology.[28]

Note also that the implementing partner provided IT assistance to deal with internet problems, so that after the beginning of the implementation the proportion of schools with internet problems sharply decreased. Without such aid, it is conceivable that the positive

[27]In a previous version of the paper, we considered a regression including only the interaction between internet problems and treatment assignment. There, we found that schools with no internet problems had larger treatment effects. The results in Table 8 show that those earlier results came from the fact that internet problems are positively correlated with shared implementation.

[28]It is not possible to reliably estimate treatment effects for the full interaction of internet/hardware problems and type of implementation, because we have only 8 schools with internet/hardware problems that implemented the program with one computer per student.

effects we found for the schools with one computer per students would be smaller. Therefore, our results should be seen as an upper bound on the effects of this program in schools with one computer per student, when we have in place an additional effort to deal with internet connection and hardware problems.

In Panel B, we consider the heterogeneous effects on attitudes towards math. In the pooled data (columns 1 and 2) and when we restrict to 5th graders (columns 3 and 4), we have similar point estimates and we cannot reject the null hypothesis that the effects are the same for schools with better and worse implementation. For 9th graders, we find a small negative effect for schools with one computer per student, and positive effects for schools with shared computers. While the p-value of an asymptotic test that these effects are equal is 0.085, we recall that our sample is smaller for 9th graders, so such asymptotic tests should be considered with caution. Therefore, we do not take that as a strong evidence that there are significant differences in treatment effects across these groups. Indeed, we present in Appendix Table A.12 p-values based on permutation tests for these heterogeneous exercises, confirming that we do not have strong evidence of heterogeneous treatment effects on attitudes for 9th graders.

It is possible to rationalize these results if we take into account that virtually all treated students were exposed to the platform, regardless of the type of implementation. Students in the rotation implementation, despite having to rotate between studying in the platform and doing other math activities, were also significantly exposed to the platform. Therefore, even students in schools with worse implementation used the platform and were exposed to math in a potentially more exciting and interactive manner. This may explain why we find similar positive effects on both groups (with rotational or single use implementation) on attitudes towards math.[29] The program can therefore be seen as an effective way to change attitudes towards math, even in the short-run and with a shared implementation.

[29] However, given our experiment design, we cannot guarantee that this was the main mechanism driving these results.

Overall, these heterogeneous patterns can be rationalized in a model in which perceptions about math can be affected by the program, even in a more limited implementation in which students would have to rotate between the platform and other activities. However, to achieve proficiency gains, the platform may require more consistent and longer usages.

Finally, it is important to note that we see such heterogeneous results as only suggestive evidence that the program, if well implemented, can have positive effects on students' test scores. First, as explained above, the heterogeneous effects are not estimated based on experimental variation, and such analysis was not pre-registered at the AEA registry. We present these results even though they were not pre-registered because they are important to provide a better understanding of the results on main outcomes presented in Sections 6.1 and 6.2 (see Duflo et al. (2020) for a discussion on the potential benefits of presenting analyses that were not pre-registered). Second, even if the assumptions for extrapolation of the information on infrastructure problems and implementation modality are valid, the heterogeneous effects would only identify the treatment effects for different types of schools. Therefore, it is not possible to guarantee that a school that implemented the program with shared computers would have had the same expected effect of a school with one computer per student if it had more computers. For example, it may be that there are other variables, such as motivation of the school principal, that explains both the the implementation modality and the lower treatment effects. In this case, even if we improve the infrastructure of these schools, we should not necessarily expect better results. Finally, estimating effects for sub-samples essentially means a lower effective number of observations, so inference based on asymptotic approximations become less reliable (see, for example, Young (2018)). Consistent with that, the inference assessment proposed by Ferman (2021) detects that the inference methods considered in the estimation of the heterogeneous effects (Table 8) are less reliable than the ones considered in the estimation of the main effects (Table 6). Inference is particularly unreliable when we consider the heterogeneous effects for the sample of 9th graders.

6.4 Student-Level Heterogeneous Effects

We take advantage of the fact that in both data sets (the national standardized exam and the survey), we have individual information on gender, race, parental education, and whether the students have personal computers at home. We use this to compute heterogenous effects along these dimensions. Table 9 presents the results for Math Proficiency. The first column presents the heterogeneity analysis being performed and the second column identifies each sub-group. The first 2 columns of results show the overall effects, and the next 4 columns the effects of the treatment when interacted with the modality of implementation (shared or one computer for each student). The first line coincides with the results from Tables 6 and 8.

In the overall sample, we can see a small contrast of effects depending on parental education — students whose parents have less than high school degree present *relatively* larger effects. When we interact with the implementation modality, we find clearer patterns. Students with lower socioeconomic status, proxied by parental education, race, or availability of computers at home, show larger point estimates. Additionally, the point estimates for girls is larger than for boys.

Table 10 shows the heterogeneity results for the outcome of attitudes towards math, with the same structure as Table 9. We can see the exact same patterns. The effects are larger for girls and for those with lower socioeconomic status proxied by race or computer ownership. We do not have questions about parental education in the survey. We can also test how the effects depend on the intensity of computer usage *before* the intervention. The effects are larger for those with no or moderated use of computers at home.[30] The estimates are qualitatively the same when we interact treatment with the implementation modality. Taken together, these heterogeneities are consistent with the results by de Barros and Ganimian (2021), given the correlation between socioeconomic status and treatment effects.

[30]Despite having this question in the national exam as well, we cannot use it because it was measured after our intervention. In the survey this question was asked in the baseline survey, before the intervention.

7 Conclusion

In this paper, we present novel experimental evidence on the impacts of integrating a dynamically adaptive computer-assisted technology into regular math instruction in public schools across five cities in three different regions of Brazil. We find that, on average, the program does not have an impact on students' math scores, although we find significant effects on attitudes towards math. We also explore differences by quality of implementation, providing suggestive evidence that the program may have positive effects when the schools' infrastructure allows for a modality based on one computer per student. We also find evidence that the program seems to be more effective for girls and for students with lower socioeconomic status, when implemented with one computer per student.

Combining our results with the available evidence on CAL programs suggest that the effectiveness of such programs depend crucially on a series of implementation details, as also argued by Muralidharan et al. (2019) and Mo et al. (2020). A first important implementation issue regards whether the CAL program increases or maintains constant the total number of hours students are exposed to math content. In the second case, the effect of a CAL program depends crucially on the net effectiveness of the CAL program relative to a standard math class. This helps explain why the literature converged in pointing out the benefits of CAL programs in supplementing traditional teaching, while there is mixed evidence on the potential for CAL as effective substitutes (Glewwe and Muralidharan (2016) or Bulman and Fairlie (2016)).

Overall, these results point out that the external validity of experimental results on CAL programs should be considered with caution, particularly for policies aiming at scaling these interventions in developing countries. Our evidence shows that implementation challenges may lead to null effects even in a context in which there were substantial efforts aiming at implementation support (see discussion in Section 2). In this sense, we see our results as an important contribution to the literature in that it provides evidence on some key determinants that are relevant in the extrapolation of experimental results on CAL programs and provides guidance for future research.

References

Adams, D. L. (2016). *Action research on the effectiveness of Khan Academy as a tier I intervention*. PhD thesis, Miami University.

Aiken Jr, L. R. and Dreger, R. M. (1961). The effect of attitudes on performance in mathematics. *Journal of Educational Psychology*, 52(1):19.

Andrabi, T., Das, J., Khwaja, A. I., Vishwanath, T., and Zajonc, T. (2007). Learning and educational achievements in punjab schools (leaps): Insights to inform the education policy debate. *World Bank, Washington, DC*.

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

Bettinger, E., Fairlie, R. W., Kapuza, A., Kardanova, E., Loyalka, P., and Zakharov, A. (2020). Does edtech substitute for traditional learning? experimental estimates of the educational production function. Technical report, National Bureau of Economic Research.

Brito, M. R. F. (1998). Adaptação e validação de uma escala de atitudes em relação à matemática). *Zetetiké: Revista de Educação Matemática*, 6(9).

Büchel, K., Jakob, M., Christoph, K., Steffen, D., and Brunetti, A. (2020). The relative effectiveness of teachers and learning software: Evidence from a field experiment in el salvador. Technical report, University of Bern, Department of Social Sciences.

Bulman, G. and Fairlie, R. W. (2016). Technology and education: Computers, software, and the internet. In *Handbook of the Economics of Education*, volume 5, pages 239–280. Elsevier.

Callaway, B. and Sant'Anna, P. H. (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.

Chu, L., Nautiyal, A., Rais, S., and Yamtich, H. (2018). Does khan academy work?

Das, J. and Zajonc, T. (2010). India shining and bharat drowning: Comparing two indian states to the worldwide distribution in mathematics achievement. *Journal of Development Economics*, 92(2):175–187.

de Barros, A. and Ganimian, A. J. (2021). Which students benefit from personalized learning? experimental evidence from a math software in public schools in india.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

de Chaisemartin, C. and Ramirez-Cuellar, J. (2019). At What Level Should One Cluster Standard Errors in Paired Experiments, and in Stratified Experiments with Small Strata? *arXiv e-prints*, page arXiv:1906.00288.

Duflo, E., Banerjee, A., Finkelstein, A., Katz, L. F., Olken, B. A., and Sautmann, A. (2020). In praise of moderation: Suggestions for the scope and use of pre-analysis plans for rcts in economics. Working Paper 26993, National Bureau of Economic Research.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101(5):1739–74.

Escueta, M., Nickow, A. J., Oreopoulos, P., and Quan, V. (2020). Upgrading education with technology: Insights from experimental research. *Journal of Economic Literature*, 58(4):897–996.

Ferman, B. (2021). Assessing inference methods. *arXiv e-prints*, page arXiv:1912.08772.

Ganimian, A., Hess, F., and Vegas, E. (2020). Realizing the promise: How can education technology improve learning for all. Technical report, Technical Report. Brookings Institution. Washington, DC URL: https://www . . . .

Glewwe, P. and Muralidharan, K. (2016). *Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications*, volume 5. Elsevier B.V., 1 edition.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Kelly, D. P. and Rutherford, T. (2017). Khan academy as supplemental instruction: A controlled study of a computer-based mathematics intervention. *The International Review of Research in Open and Distributed Learning*, 18(4).

Lai, F., Luo, R., Zhang, L., Huang, X., and Rozelle, S. (2015). Does computer-assisted learning improve learning outcomes? evidence from a randomized experiment in migrant schools in beijing. *Economics of Education Review*, 47:34–48.

Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., Boswell, M., and Rozelle, S. (2013). Computer assisted learning as extracurricular tutor? evidence from a randomised experiment in rural boarding schools in shaanxi. *Journal of Development Effectiveness*, 5(2):208–231.

Lai, F., Zhang, L., Qu, Q., Hu, X., Shi, Y., Boswell, M., and Rozelle, S. (2012). Does computer-assisted learning improve learning outcomes? evidence from a randomized experiment in public schools in rural minority areas in qinghai, china. *Rural Education Action Project Working paper*, 237.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Linden, L. L. (2008). *Complement or substitute?: The effect of technology on student achievement in India*. InfoDev Working Paper, Columbia University.

Ma, Y., Fairlie, R. W., Loyalka, P., and Rozelle, S. (2020). Isolating the “tech” from edtech: Experimental evidence on computer assisted learning in china. Technical report, National Bureau of Economic Research.

Manaus (2016). Assessing the use of technology and khan academy to improve educational outcomes in sacatepequez, guatemala. Technical report.

Mo, D., Bai, Y., Shi, Y., Abbey, C., Zhang, L., Rozelle, S., and Loyalka, P. (2020). Institutions, implementation, and program effectiveness: Evidence from a randomized evaluation of computer-assisted learning in rural china. *Journal of Development Economics*, page 102487.

Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., Qiao, Y., Boswell, M., and Rozelle, S. (2014). Integrating computer-assisted learning into a regular curriculum: Evidence from a randomised experiment in rural schools in shaanxi. *Journal of development effectiveness*, 6(3):300–323.

Muralidharan, K. and Singh, A. (2021). Evaluation of a technology-aided instruction platform (mindspark) in government schools in rajasthan. *AEA RCT Registry 0002546*.

Muralidharan, K., Singh, A., and Ganimian, A. J. (2019). Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review*, 109(4):1426–60.

Phillips, D. and Cohen, J. (2015). Learning gets personal. how idaho students and teachers are embracing personalized learning through khan academy. Technical report.

Pritchett, L. and Beatty, A. (2015). Slow down, you're going too fast: Matching curricula to student skill levels. *International Journal of Educational Development*, 40:276–288.

Rodriguez-Segura, D. (2020). Educational technology in developing countries: A systematic review.

Weeraratne, B. and Chin, B. (2018). Can khan academy e-learning video tutorials improve mathematics achievement in sri lanka?. *International Journal of Education and Development using Information and Communication Technology*, 14(3):93–112.

WorldBank (2018). World development report 2018 : Learning to realize education's promise.

Yang, Y., Zhang, L., Zeng, J., Pang, X., Lai, F., and Rozelle, S. (2013). Computers and the academic performance of elementary school-aged girls in china's poor communities. *Computers & Education*, 60(1):335–346.

Young, A. (2018). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results*. *The Quarterly Journal of Economics*, 134(2):557–598.

Tables

Table 1: Baseline Covariates Balance - SAEB

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
Male	0.504	-0.008	14411	0.512	-0.010	10072	0.485	-0.001	4339
	[0.500]	[0.010]		[0.500]	[0.012]		[0.500]	[0.016]	
White	0.283	-0.009	14423	0.293	-0.013	10047	0.255	0.002	4376
	[0.450]	[0.012]		[0.455]	[0.016]		[0.436]	[0.013]	
Black	0.073	-0.005	14423	0.070	-0.007	10047	0.082	0.000	4376
	[0.261]	[0.005]		[0.255]	[0.008]		[0.274]	[0.008]	
Mixed	0.527	0.007	14423	0.517	0.015	10047	0.551	-0.014	4376
	[0.499]	[0.011]		[0.500]	[0.014]		[0.497]	[0.025]	
Asian	0.028	0.004	14423	0.023	0.002	10047	0.041	0.007	4376
	[0.166]	[0.002]		[0.151]	[0.003]		[0.198]	[0.005]	
Native	0.025	-0.001	14423	0.025	0.000	10047	0.026	-0.001	4376
	[0.157]	[0.003]		[0.157]	[0.004]		[0.158]	[0.005]	
Race not declared	0.064	0.004	14423	0.071	0.004	10047	0.045	0.006	4376
	[0.244]	[0.005]		[0.257]	[0.007]		[0.207]	[0.009]	
Age	12.007	-0.005	14625	10.821	0.018	10220	15.099	-0.063	4405
	[2.087]	[0.020]		[0.795]	[0.025]		[0.916]	[0.034]	
Mother has completed at least	0.625	0.025	9606	0.636	0.019	6034	0.606	0.037	3572
high school	[0.484]	[0.013]		[0.481]	[0.022]		[0.489]	[0.022]	
Mother literate	0.985	-0.002	14564	0.989	-0.005	10173	0.976	0.006	4391
	[0.120]	[0.002]		[0.106]	[0.003]		[0.152]	[0.005]	
Father has completed at least	0.571	0.017	8006	0.565	0.007	4990	0.582	0.034	3016
high school	[0.495]	[0.015]		[0.496]	[0.021]		[0.493]	[0.024]	
Father literate	0.958	0.001	14373	0.962	0.001	10007	0.948	0.001	4366
	[0.201]	[0.004]		[0.192]	[0.004]		[0.222]	[0.007]	
Teacher younger than 50 years old	0.760	0.008	12805	0.761	0.012	10530	0.752	-0.017	2275
	[0.427]	[0.049]		[0.426]	[0.057]		[0.432]	[0.171]	
2015 Prova Brasil math grade	0.095	0.029	16820	0.090	-0.066	11654	0.107	0.266	5166
	[1.023]	[0.089]		[0.934]	[0.084]		[1.216]	[0.132]	
P value joint	0.799			0.420			0.892		

Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the control group mean, ii) the results of student-level regressions of covariates available in the SAEB dataset on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects and iii) Number of observations. Standard errors clustered at the strata level are in brackets. P-values for a test that all covariates are balanced are reported at the bottom of the table for each of the three samples considered.

Table 2: Attrition

Pooled sample					5th grade					9th grade				
Mean (control)	Diff	N Obs.	N Schools	N Strata	Mean (control)	Diff	N Obs.	N Schools	N Strata	Mean (control)	Diff	N Obs.	N Schools	N Strata
Panel A: School-grade-level Attrition in the SAEB exam														
0.142	-0.008	217	157	35	0.099	-0.002	143	143	32	0.229	-0.020	74	74	15
	[0.038]					[0.050]					[0.085]			
	(0.829)					(0.968)					(0.813)			
Panel B: Student-level Attrition in the SAEB exam														
0.132	0.005	17151	143	34	0.123	0.006	11906	129	31	0.156	0.002	5245	58	14
	[0.008]					[0.009]					[0.011]			
	(0.558)					(0.532)					(0.852)			
Panel C: Student-level Attrition in the in the Survey (Answered no question)														
0.298	-0.028	18065	150	35	0.275	-0.031	12220	136	32	0.356	-0.020	5845	136	15
	[0.015]					[0.020]					[0.030]			
	(0.058)					(0.112)					(0.503)			
Panel D: Student-level Attrition in the Survey (Did not answered the 20 questions)														
0.393	-0.025	18065	150	35	0.377	-0.030	12220	136	32	0.433	-0.015	5845	136	15
	[0.015]					[0.020]					[0.028]			
	(0.083)					(0.133)					(0.589)			

Notes: This table reports differences in attrition between treatment and control groups in the SAEB exam (school-grade-level in Panel B and student-level in Panel C) and in the survey (Panels C and D). We report for the pooled sample and for the 5h grade and 9th grades samples separately: i) the control group mean, ii) the results of regressions of our indicator of attrition (which takes value one if there is no follow-up data available) on a dummy variable indicating treatment assignment and strata fixed effects, iii) Number of observations and iv) Number of clusters. Standard errors, in brackets, are clustered at the strata level. P-values are in parenthesis. In Panel C attrition is defined as the students that did not answered any of the 20 questions while in Panel D if they missed any question.

Table 3: Baseline Covariates Balance - Survey

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
Attitudes towards math	0.000	0.004	11422	0.000	-0.007	7203	0.000	0.024	4219
	[1.000]	[0.029]		[1.000]	[0.030]		[1.000]	[0.063]	
Male	0.505	-0.005	12369	0.513	-0.015	7871	0.488	0.012	4498
	[0.500]	[0.009]		[0.500]	[0.010]		[0.500]	[0.015]	
Year of Birth	2,004.6	-0.010	12381	2,005.9	-0.053	7872	2,001.8	0.066	4509
	[2.298]	[0.035]		[1.396]	[0.053]		[1.013]	[0.038]	
White	0.327	-0.014	10703	0.364	-0.028	6540	0.256	0.008	4163
	[0.469]	[0.009]		[0.481]	[0.014]		[0.437]	[0.011]	
Black	0.107	-0.013	10703	0.111	-0.010	6540	0.100	-0.017	4163
	[0.309]	[0.006]		[0.314]	[0.010]		[0.300]	[0.012]	
Native	0.038	0.002	10703	0.041	0.004	6540	0.033	0.000	4163
	[0.192]	[0.004]		[0.198]	[0.006]		[0.180]	[0.005]	
Mixed	0.488	0.026	10703	0.450	0.034	6540	0.563	0.012	4163
	[0.500]	[0.011]		[0.498]	[0.018]		[0.496]	[0.013]	
Asian	0.039	-0.001	10703	0.034	0.001	6540	0.048	-0.004	4163
	[0.194]	[0.006]		[0.182]	[0.006]		[0.214]	[0.008]	
Has computer at home	0.580	-0.007	12396	0.572	-0.014	7892	0.596	0.005	4504
	[0.494]	[0.012]		[0.495]	[0.016]		[0.491]	[0.026]	
Frequently uses computer at home	0.455	-0.003	12380	0.454	-0.007	7884	0.457	0.006	4496
	[0.498]	[0.010]		[0.498]	[0.013]		[0.498]	[0.019]	
Has internet at home	0.736	-0.008	12360	0.741	-0.022	7867	0.726	0.017	4493
	[0.441]	[0.014]		[0.438]	[0.020]		[0.446]	[0.019]	
Uses computer at home for school activities	0.520	-0.006	12365	0.518	-0.018	7872	0.526	0.016	4493
	[0.500]	[0.012]		[0.500]	[0.015]		[0.499]	[0.024]	
Uses computer lab at school	0.367	-0.011	12374	0.419	-0.013	7879	0.255	-0.008	4495
	[0.482]	[0.044]		[0.493]	[0.056]		[0.436]	[0.048]	
Uses computer lab at school during portuguese classes	0.237	0.023	12403	0.290	0.019	7896	0.123	0.031	4507
	[0.426]	[0.039]		[0.454]	[0.052]		[0.329]	[0.040]	
Uses computer lab at school during math classes	0.255	0.048	12368	0.318	0.035	7873	0.119	0.071	4495
	[0.436]	[0.055]		[0.466]	[0.054]		[0.323]	[0.084]	
Uses computer lab at school during other classes	0.332	-0.052	12334	0.335	-0.018	7852	0.327	-0.112	4482
	[0.471]	[0.031]		[0.472]	[0.038]		[0.469]	[0.056]	
Uses computer lab at school not during class	0.144	-0.013	12377	0.148	-0.018	7878	0.135	-0.005	4499
	[0.351]	[0.010]		[0.355]	[0.012]		[0.342]	[0.025]	

(cont)

Table 3 Cont. - Baseline Covariates Balance - Survey

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
(cont)									
Has mobile phone	0.715	-0.001	12265	0.683	0.000	7808	0.783	-0.001	4457
	[0.452]	[0.012]		[0.466]	[0.018]		[0.412]	[0.013]	
Has internet on mobile phone	0.706	-0.003	11286	0.680	-0.004	6925	0.759	-0.003	4361
	[0.455]	[0.010]		[0.467]	[0.014]		[0.428]	[0.014]	
Lives with mother	0.893	0.005	12362	0.902	0.007	7864	0.874	0.001	4498
	[0.309]	[0.007]		[0.298]	[0.008]		[0.332]	[0.014]	
Lives with father	0.617	0.003	12360	0.640	-0.002	7861	0.569	0.013	4499
	[0.486]	[0.010]		[0.480]	[0.014]		[0.495]	[0.017]	
Has books at home	0.767	-0.009	12394	0.740	-0.021	7890	0.826	0.013	4504
	[0.422]	[0.011]		[0.439]	[0.015]		[0.379]	[0.014]	
Parents talk about school	0.844	-0.001	12394	0.867	-0.012	7891	0.795	0.019	4503
	[0.363]	[0.006]		[0.339]	[0.008]		[0.404]	[0.007]	
Works outside home	0.082	0.000	12388	0.080	-0.004	7882	0.084	0.008	4506
	[0.274]	[0.007]		[0.272]	[0.008]		[0.278]	[0.012]	
Has ever repeated a grade	0.238	-0.006	12304	0.186	0.011	7830	0.349	-0.036	4474
	[0.426]	[0.013]		[0.389]	[0.017]		[0.477]	[0.011]	
Math is the preferred subject	0.428	0.008	12389	0.506	0.007	7894	0.260	0.009	4495
	[0.495]	[0.015]		[0.500]	[0.017]		[0.439]	[0.027]	
Portuguese is the preferred subject	0.249	0.008	12389	0.267	0.007	7894	0.208	0.010	4495
	[0.432]	[0.012]		[0.443]	[0.013]		[0.406]	[0.021]	
Other subject is preferred	0.323	-0.016	12389	0.226	-0.014	7894	0.532	-0.018	4495
	[0.468]	[0.013]		[0.418]	[0.012]		[0.499]	[0.030]	
Participated in Math Olympics	0.192	0.000	11340	0.074	0.005	7192	0.444	-0.009	4148
	[0.394]	[0.010]		[0.262]	[0.012]		[0.497]	[0.022]	
P value joint	0.696			0.275			0.790		

Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately, three columns respectively with the control group mean, the regression adjusted differences between treatment and control groups, and number of observations for 27 covariates. We report estimates from a regression for each covariate on an indicator variable for the treatment and strata-grade fixed effects. Standard errors clustered at the strata level are in brackets. P-values for a test that all covariates are balanced are reported at the bottom of the table for each of the three samples considered.

Table 4: Follow-up Survey

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
Has computer at home	0.622 [0.485]	-0.012 [0.014] (0.398)	12816	0.631 [0.483]	-0.013 [0.015] (0.381)	9004	0.595 [0.491]	-0.008 [0.034] (0.809)	3812
Frequently uses computer at home	0.472 [0.499]	0.015 [0.013] (0.237)	12808	0.484 [0.500]	0.020 [0.016] (0.209)	9004	0.438 [0.496]	0.004 [0.025] (0.884)	3804
Has internet at home	0.795 [0.404]	-0.002 [0.011] (0.875)	12745	0.804 [0.397]	-0.002 [0.014] (0.910)	8953	0.770 [0.421]	-0.002 [0.025] (0.923)	3792
Uses computer at home for school activities	0.519 [0.500]	0.004 [0.014] (0.775)	12764	0.526 [0.499]	0.001 [0.018] (0.953)	8962	0.502 [0.500]	0.011 [0.030] (0.699)	3802
Uses computer lab at school	0.488 [0.500]	0.285 [0.057] (0.000)	12820	0.555 [0.497]	0.192 [0.059] (0.001)	9010	0.300 [0.458]	0.513 [0.062] (0.000)	3810
Uses computer lab at school during portuguese classes	0.317 [0.465]	-0.039 [0.046] (0.388)	12801	0.370 [0.483]	-0.057 [0.057] (0.325)	8994	0.167 [0.373]	0.003 [0.038] (0.939)	3807
Uses computer lab at school during math classes	0.340 [0.474]	0.445 [0.057] (0.000)	12743	0.398 [0.490]	0.330 [0.057] (0.000)	8951	0.175 [0.380]	0.728 [0.055] (0.000)	3792
Uses computer lab at school during other classes	0.368 [0.482]	-0.055 [0.038] (0.145)	12703	0.386 [0.487]	-0.066 [0.047] (0.158)	8923	0.316 [0.465]	-0.027 [0.057] (0.632)	3780
Uses computer lab at school not during class	0.151 [0.358]	0.051 [0.017] (0.004)	12791	0.140 [0.347]	0.037 [0.016] (0.024)	8985	0.181 [0.385]	0.084 [0.047] (0.069)	3806
Uses Khan Academy	0.063 [0.244]	0.903 [0.021] (0.000)	12673	0.078 [0.268]	0.882 [0.030] (0.000)	8924	0.022 [0.145]	0.956 [0.006] (0.000)	3749
Uses Khan Academy during school	0.044 [0.204]	0.782 [0.031] (0.000)	12549	0.055 [0.228]	0.707 [0.036] (0.000)	8833	0.010 [0.100]	0.967 [0.004] (0.000)	3716

Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the control group mean, ii) the results of a student-level regression of different measures collected in the follow-up survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects and iii) Number of observations. Standard errors are in brackets and p-values in parenthesis. Standard errors are clustered at the strata level.

Table 5: Descriptive Statistics - Usage of Khan Academy

	Total number of minutes			Total number of weeks logged in		
	Pooled	Grade 5	Grade 9	Pooled	Grade 5	Grade 9
	(1)	(2)	(3)	(4)	(5)	(6)
No infrastructure problem	147.3	169.3	-18.3	2.888	3.979	-2.357
s.e.	[60.2]	[75.9]	[65.3]	[1.726]	[1.775]	[1.723]
p-value	(0.014)	(0.026)	(0.779)	(0.094)	(0.025)	(0.171)
One computer per student	195.0	224.2	183.9	1.669	2.082	1.741
s.e.	[77.6]	[100.5]	[45.7]	[1.560]	[1.586]	[1.676]
p-value	(0.012)	(0.026)	(0.000)	(0.284)	(0.189)	(0.299)
9th grade	-178.3	-	-	-3.206	-	-
	[46.9]			[0.947]		
	(0.000)			(0.001)		
Municipality fixed effects	Y	Y	Y	Y	Y	Y
Mean (with infrastructure problem and rotation)	471.5	540.0	386.3	12.495	13.407	11.359
	[42.31]	[64.8]	[34.4]	[0.813]	[1.221]	[0.771]
Number of Students	8302	5325	2977	8302	5325	2977
Number of Schools	103	65	38	103	65	38
Number of Strata	33	30	15	33	33	33

Notes: This table reports, in columns 1-3, results from a student-level regression of the total number of minutes spent in the platform on an indicator of no infrastructure problems, an indicator of modality of implementation based on one computer per student, and municipality fixed effects, for the pooled sample, and 5th and 9th grades subsamples respectively. In column 1 we also include an indicator of the 9th grade. Standard errors are clustered at the strata level. In columns 4-6, we report results for the same specifications using the total number of weeks logged in as the dependent variable.

Table 6: Results on Math Proficiency and Attitudes towards math

	Math test scores		Attitudes towards math	
	(1)	(2)	(3)	(4)
	Panel A: Full sample			
Treatment	-0.023	-0.016	0.056	0.061
s.e.	[0.035]	[0.024]	[0.033]	[0.022]
p-value	(0.513)	(0.515)	(0.090)	(0.006)
N obs	14846	14846	11157	11157
N schools	143	143	151	151
N strata	34	34	35	35
	Panel B: 5th grade			
Treatment	-0.036	-0.002	0.044	0.062
s.e.	[0.046]	[0.033]	[0.033]	[0.027]
p-value	(0.427)	(0.948)	(0.176)	(0.021)
N obs	10388	10388	7806	7806
N schools	129	129	137	137
N strata	31	31	32	32
	Panel C: 9th grade			
Treatment	0.011	-0.051	0.086	0.057
s.e.	[0.060]	[0.044]	[0.058]	[0.030]
p-value	(0.853)	(0.248)	(0.137)	(0.057)
N obs	4458	4458	3351	3351
N schools	58	58	72	72
N strata	14	14	15	15
Includes covariates	No	Yes	No	Yes

Notes: This table reports the results of a student-level regression of math proficiency (columns 1-2) and attitudes towards math (columns 3-4) on an dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects. Panels A, B and C refer to the pooled sample, and 5th and 9th grades subsamples separately. For the pooled regressions, we interact the strata fixed effects with grade. The specifications reported in column 2 include the covariates presented in Table 1, while the specifications reported in column 2 include the covariates presented in Table 3. Standard errors are clustered at the strata level. The inference assessment proposed by Ferman (2021) ranges from 0.061 to 0.075 in the regressions presented in Panels A and B, while it ranges from 0.071 to 0.092 in Panel C. We run 1000 draws of iid normal random variables to construct these assessments.

Table 7: Validity of Measures for Heterogeneity Exercises

	All schools		Two cycle schools		One cycle schools	
	5th grade	9th grade	5th grade	9th grade	5th grade	9th grade
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: No Infrastructure Problem			
T	-0.024	0.019	-0.023	0.023	-0.025	0.000
s.e	[0.065]	[0.082]	[0.101]	[0.101]	[0.085]	-
p-value	(0.705)	(0.815)	(0.816)	(0.816)	(0.765)	-
Mean (omitted group)	0.551	0.471	0.567	0.571	0.538	0.000
	[0.060]	[0.087]	[0.092]	[0.095]	[0.081]	-
Number of schools	136	72	58	58	78	14
			Panel B: One Computer per Student			
T	0.034	-0.022	0.027	-0.027	0.040	0.000
s.e	[0.057]	[0.071]	[0.087]	[0.087]	[0.076]	-
p-value	(0.555)	(0.755)	(0.757)	(0.757)	(0.595)	-
Mean (omitted group)	0.403	0.529	0.567	0.643	0.250	0.000
	[0.063]	[0.087]	[0.092]	[0.092]	[0.078]	-
Number of schools	127	72	58	58	69	14

Notes: This table reports, in Panel A, results of a school-grade-level regression of a dummy variable that takes value one if the there are no infrastructure problems on the treatment indicator and strata fixed effects. In columns 1-2, we display the results for 5th and 9th grades for all schools, while columns 3-4 and 5-6 show estimates for 5th and 9th grades in two cycle and one cycle schools respectively. Panel B shows results for the indicator of one computer per student as the dependent variable. The means for the omitted groups in columns 1 and 2 of Panel B (40% for 5th grade and 53% for 9th grade) are not inconsistent with the number reported in the text, that 37% of schools are based on one computer per student modality. In the table, two cycle schools are accounted twice, since our estimates are at the school-grade level.

Table 8: **ITT Heterogeneity**

	Full sample		5th grade		9th grade	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Math proficiency						
$T \times$ "one comp per stud" (β_1)	0.081	0.082	0.127	0.128	-0.102	-0.109
	[0.047]	[0.049]	[0.053]	[0.055]	[0.052]	[0.057]
	(0.088)	(0.094)	(0.016)	(0.020)	(0.048)	(0.057)
$T \times$ "shared" (β_2)	-0.076	-0.087	-0.082	-0.081	-0.075	-0.241
	[0.034]	[0.070]	[0.041]	[0.077]	[0.068]	[0.106]
	(0.027)	(0.215)	(0.044)	(0.290)	(0.271)	(0.023)
$T \times$ "internet/hardware prob"		0.010		-0.006		0.169
		[0.085]		[0.099]		[0.116]
		(0.904)		(0.953)		(0.146)
p-value ($\beta_1 = \beta_2$)	0.014	0.038	0.005	0.026	0.781	0.141
N	13231	13231	9088	9088	4143	4143
Panel B: Attitudes towards math						
$T \times$ "one comp per stud" (β_1)	0.036	0.034	0.070	0.069	-0.031	-0.002
	[0.046]	[0.046]	[0.052]	[0.052]	[0.072]	[0.101]
	(0.426)	(0.463)	(0.179)	(0.182)	(0.661)	(0.986)
$T \times$ "shared" (β_2)	0.053	0.057	0.035	0.051	0.108	0.191
	[0.021]	[0.087]	[0.028]	[0.092]	[0.018]	[0.200]
	(0.010)	(0.508)	(0.207)	(0.584)	(0.000)	(0.338)
$T \times$ "internet/hardware prob"		-0.003		-0.019		-0.079
		[0.101]		[0.117]		[0.202]
		(0.980)		(0.871)		(0.694)
p-value ($\beta_1 = \beta_2$)	0.729	0.791	0.531	0.858	0.085	0.283
N	10710	10710	7359	7359	3351	3351

Notes: This table reports results for student-level regressions of math proficiency (Panel A) and attitudes towards math (Panel B) on interaction terms between the treatment dummy and dummy variables for type of implementation. Columns (2), (4), and (6) also include interactions between the treatment dummy and a dummy indicating if there was internet/hardware problems. All specifications also include the type of implementation dummy, strata fixed effects, and the covariates reported in Table 1. We also include the internet/hardware problems dummy in the specifications from columns (2), (4), and (6). Standard errors are clustered at the strata level are presented in brackets. We present p-values in parenthesis. The inference assessment proposed by Ferman (2021) for a 5% nominal test ranges from 6% to 11% for the specifications using the full sample and for the specifications restricting to 5th graders. For the specifications restricting to 9th graders, the inference assessment ranges from 9% to 20%, so we consider the results from columns (5) and (6) with caution.

Table 9: Results: Heterogeneity — Math proficiency

		All		**Shared**		**1 comp/student**	
Analysis	Effect for	Effect (1)	SE (2)	Effect (3)	SE (4)	Effect (5)	SE (6)
Baseline	All	-0.009	(0.026)	-0.076	(0.032)	0.081	(0.052)
Gender	Boy	-0.015	(0.033)	-0.085	(0.038)	0.066	(0.069)
Gender	Girl	-0.013	(0.025)	-0.078	(0.031)	0.107	(0.051)
Race	White	-0.019	(0.037)	-0.074	(0.039)	0.017	(0.059)
Race	Black	-0.012	(0.027)	-0.082	(0.032)	0.141	(0.052)
Computer	No	-0.003	(0.036)	-0.052	(0.040)	0.103	(0.068)
Computer	Yes,1	-0.030	(0.033)	-0.105	(0.044)	0.069	(0.075)
Computer	Yes,2+	0.011	(0.041)	-0.075	(0.045)	0.075	(0.074)
Mother Educ	$<$HS	0.006	(0.041)	-0.060	(0.045)	0.123	(0.074)
Mother Educ	$\geq$ HS	-0.059	(0.032)	-0.122	(0.036)	0.030	(0.045)
Father Educ	$<$ HS	-0.025	(0.037)	-0.074	(0.038)	0.061	(0.060)
Father Educ	$\geq$ HS	-0.056	(0.035)	-0.117	(0.039)	0.033	(0.062)

Notes: This table shows the results for treatment effect heterogeneity on math proficiency. The first line shows the baseline results for the overall effect (column 1) and for the regression interacting the implementation modality: shared in column 3 and one student per computer in column 5. Columns 2, 4 and 6 shows the respective standard errors. Each block of rows presents the results of one heterogeneity. In order, heterogeneities by gender (boys x girls), race (white/asian x black/brown/native), number of computers at home (none,one, two or more) mother's education (less than high school x high school degree or more) and father's education (less than high school x high school degree or more). For each heterogeneity block one regression was estimated for the overall effect interacted with the covariate of interest (columns 1-2) and another further interacting implementation modality of the covariate of interest (columns 3-6). All regression include the full set of covariates and the standard errors are estimated using cluster at the strata level.

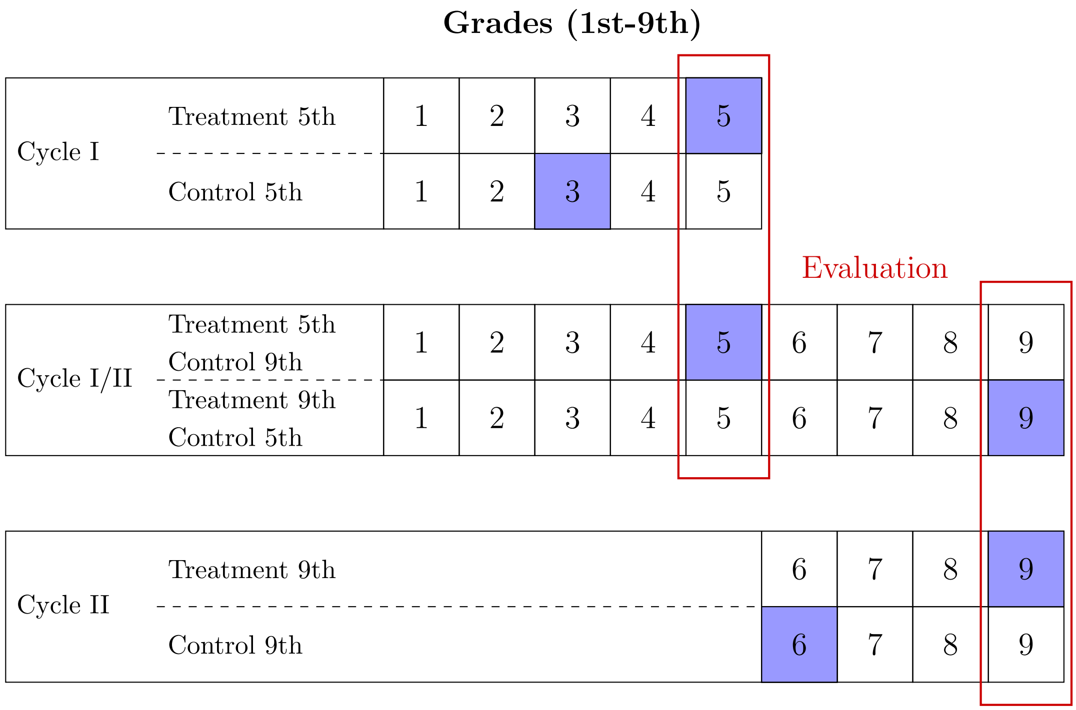
Table 10: Results: Heterogeneity — Attitudes Towards Math

		All		**Shared**		**1 comp/student**	
Analysis	Effect for	Effect	SE	Effect	SE	Effect	SE
Baseline	All	0.061	(0.022)	0.054	(0.020)	0.038	(0.047)
Gender	Boy	0.046	(0.027)	0.036	(0.024)	-0.008	(0.035)
Gender	Girl	0.119	(0.037)	0.100	(0.032)	0.143	(0.085)
Race	White	0.072	(0.025)	0.065	(0.031)	0.059	(0.038)
Race	Black	0.084	(0.032)	0.069	(0.027)	0.063	(0.078)
Computer	No	0.131	(0.030)	0.108	(0.025)	0.104	(0.081)
Computer	Yes,1	0.069	(0.034)	0.050	(0.034)	0.080	(0.053)
Computer	Yes,2+	0.002	(0.049)	-0.004	(0.051)	-0.009	(0.072)
Use Computer	No	0.125	(0.030)	0.113	(0.024)	0.084	(0.069)
Use Computer	Yes,moderately	0.043	(0.034)	0.013	(0.037)	0.074	(0.051)
Use Computer	Yes,always	0.014	(0.109)	0.086	(0.105)	-0.118	(0.148)

Notes: This table shows the results for treatment effect heterogeneity on math proficiency. The first line shows the baseline results for the overall effect (column 1) and for the regression interacting the implementation modality: shared in column 3 and one student per computer in column 5. Columns 2, 4 and 6 shows the respective standard errors. Each block of rows presents the results of one heterogeneity. In order, heterogeneities by gender (boys x girls), race (white/asian x black/brown/native), number of computers at home (none,one, two or more) and computer use at home in the baseline survey (never, moderately and always). For each heterogeneity block one regression was estimated for the overall effect interacted with the covariate of interest (columns 1-2) and another further interacting implementation modality of the covariate of interest (columns 3-6). All regression include the full set of covariates and the standard errors are estimated using cluster at the strata level.

Figures

Figure 1: Randomization Procedure by type of school



Notes: Blue squares represent grades that receive the program. The red rectangles indicate the grades that are in the evaluation (5th and 9th grades).

Appendix A Effects on control school $\times$ grades

We consider whether students in schools that participated in the RCT, but in grades that did not receive the program, were affected by the fact that there were other grades in their schools that received the program. This could be the case, for example, if there are relevant spillovers, due to control students using the platform after seeing students in other grades using it. In Section 5.2, we provide evidence that this is not a major concern. It may be that school principals compensate students in control grades due to the fact that they did not participate in the program. If this is a change that only occurred with control students because their schools participated in the experiment, then these control students would not provide a proper counterfactual for the case in which there were no program whatsoever.

In order to test whether these are relevant concerns in our setting, we contrast the test scores from students in school $\times$ grades that participated in the RCT, with students in schools that did not participate in the RCT. Since schools that participated in the RCT were not randomly assigned, we consider a difference-in-differences (DID) methodology to estimate these effects. In this DID framework, control school $\times$ grades that participated in the RCT comprise the "treatment group", while school $\times$ grades in schools that did not participate in the RCT comprise the "control group". We restrict to municipality schools in the same municipalities that participated in the RCT. The post-treatment period is 2017, while we have information from every other year from 2005 to 2015 as pre-treatment periods.[31]

Figure A.4 presents the evolution of these two groups, separately for 5th and 9th grades. For 5th grade, we see that the two groups follow a very similar path over the years, and that there is no evidence of a significant change in 2017. For 9th grade, there are some oscillations in the series for schools that participated in the experiment, which is likely due to the fact that we have relatively fewer schools with 9th grade in the RCT. While there is a difference between treated and control schools in 2017, we also see other years with similar differences in the pre-treatment periods.

In Table A.13, we present the DID regression results, for both the pooled sample, and for 5th and 9th grades separately. We also include a more flexible specification in which we allow for municipality $\times$ time specific trends.[32] Overall, we do not find any evidence that students in schools that participated in the RCT, but in grades that did not receive the program, were significantly affected. This provides further evidence that the concerns above regarding potential contamination due to the fact that we considered a within-school design are not first order. Note that in our DID specification there is no variation in the timing of treatment, so recent concerns about DID regression when there is variation in timing do not apply to our setting (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2020).[33]

[31] Before 2005, the SAEB exam only had information from a sample of schools, so we are not be able to use the data for these years for this exercise.

[32] For the pooled regression in column 2, we allow for municipality $\times$ time $\times$ grade specific trends.

[33] Our results also remain similar in case we weight schools using the average number of students across years (instead of the current year number of students), or if we do not use weights. Therefore, our results are also robust to the concerns raised by de Chaisemartin and D'Haultfœuille (2020).

Appendix B Appendix Tables

Table A.1: Balance conditional on non-attritors

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
Attitudes towards math	0.030 [1.004]	0.010 [0.037]	7243	0.049 [1.006]	-0.023 [0.039]	4688	-0.012 [0.998]	0.071 [0.070]	2555
Male	0.502 [0.500]	0.004 [0.012]	7761	0.501 [0.500]	-0.001 [0.013]	5056	0.504 [0.500]	0.014 [0.023]	2705
Year of Birth	2,004.7 [2.232]	-0.022 [0.040]	7764	2,006.0 [1.266]	-0.083 [0.056]	5054	2,001.9 [0.949]	0.093 [0.050]	2710
White	0.336 [0.472]	-0.015 [0.010]	6692	0.369 [0.483]	-0.017 [0.013]	4194	0.269 [0.444]	-0.011 [0.014]	2498
Black	0.099 [0.299]	-0.005 [0.007]	6692	0.104 [0.305]	-0.009 [0.011]	4194	0.090 [0.286]	0.001 [0.012]	2498
Native	0.040 [0.196]	0.003 [0.006]	6692	0.043 [0.203]	0.004 [0.008]	4194	0.034 [0.181]	0.002 [0.009]	2498
Mixed	0.486 [0.500]	0.019 [0.014]	6692	0.447 [0.497]	0.021 [0.020]	4194	0.563 [0.496]	0.014 [0.014]	2498
Asian	0.039 [0.194]	-0.002 [0.007]	6692	0.037 [0.188]	0.001 [0.007]	4194	0.044 [0.205]	-0.007 [0.009]	2498
Has computer at home	0.602 [0.490]	-0.016 [0.016]	7772	0.597 [0.491]	-0.020 [0.016]	5065	0.613 [0.487]	-0.009 [0.034]	2707
Frequently uses computer at home	0.468 [0.499]	-0.004 [0.013]	7765	0.465 [0.499]	-0.003 [0.015]	5062	0.476 [0.500]	-0.008 [0.026]	2703
Has internet at home	0.740 [0.439]	-0.005 [0.015]	7749	0.751 [0.433]	-0.024 [0.019]	5049	0.716 [0.451]	0.030 [0.025]	2700
Uses computer at home for school activities	0.531 [0.499]	-0.010 [0.014]	7750	0.528 [0.499]	-0.018 [0.017]	5050	0.539 [0.499]	0.005 [0.035]	2700
Uses computer lab at school	0.372 [0.483]	-0.016 [0.041]	7751	0.419 [0.494]	-0.005 [0.057]	5051	0.266 [0.442]	-0.035 [0.045]	2700
Uses computer lab at school during portuguese classes	0.245 [0.430]	0.010 [0.040]	7773	0.301 [0.459]	0.002 [0.054]	5065	0.122 [0.328]	0.024 [0.038]	2708
Uses computer lab at school during math classes	0.263 [0.440]	0.047 [0.055]	7758	0.333 [0.471]	0.035 [0.056]	5055	0.108 [0.311]	0.070 [0.081]	2703
Uses computer lab at school during other classes	0.337 [0.473]	-0.055 [0.029]	7732	0.337 [0.473]	-0.020 [0.037]	5039	0.337 [0.473]	-0.123 [0.059]	2693
Uses computer lab at school not during class	0.138 [0.345]	-0.014 [0.010]	7760	0.142 [0.349]	-0.021 [0.011]	5057	0.130 [0.337]	-0.002 [0.027]	2703

(cont)

Table A.1 Cont : Balance conditional on non-attritors

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
				(cont)					
Has mobile phone	0.711	0.010	7699	0.680	0.008	5018	0.779	0.013	2681
	[0.454]	[0.014]		[0.467]	[0.021]		[0.415]	[0.015]	
Has internet on mobile phone	0.710	0.007	7026	0.689	0.004	4401	0.752	0.013	2625
	[0.454]	[0.013]		[0.463]	[0.018]		[0.432]	[0.014]	
Lives with mother	0.902	0.001	7752	0.908	0.007	5048	0.888	-0.010	2704
	[0.297]	[0.007]		[0.289]	[0.009]		[0.315]	[0.013]	
Lives with father	0.639	0.001	7748	0.658	-0.010	5047	0.595	0.021	2701
	[0.480]	[0.015]		[0.474]	[0.019]		[0.491]	[0.028]	
Has books at home	0.777	-0.009	7771	0.748	-0.013	5064	0.841	0.000	2707
	[0.416]	[0.012]		[0.434]	[0.015]		[0.366]	[0.017]	
Parents talk about school	0.837	0.009	7772	0.859	-0.002	5066	0.787	0.030	2706
	[0.370]	[0.009]		[0.348]	[0.010]		[0.410]	[0.014]	
Works outside home	0.067	0.004	7772	0.064	0.006	5063	0.075	0.000	2709
	[0.251]	[0.006]		[0.245]	[0.008]		[0.263]	[0.012]	
Has ever repeated a grade	0.211	-0.001	7724	0.163	0.011	5033	0.319	-0.025	2691
	[0.408]	[0.014]		[0.369]	[0.020]		[0.466]	[0.019]	
Math is the preferred subject	0.440	0.007	7769	0.521	0.003	5064	0.260	0.015	2705
	[0.496]	[0.020]		[0.500]	[0.022]		[0.439]	[0.031]	
Portuguese is the preferred subject	0.238	-0.001	7769	0.250	0.002	5064	0.212	-0.007	2705
	[0.426]	[0.014]		[0.433]	[0.017]		[0.409]	[0.022]	
Other subject is preferred	0.321	-0.006	7769	0.229	-0.005	5064	0.528	-0.008	2705
	[0.467]	[0.015]		[0.420]	[0.017]		[0.499]	[0.030]	
Participated in Math Olympics	0.182	-0.001	7086	0.063	0.007	4606	0.446	-0.018	2480
	[0.386]	[0.012]		[0.243]	[0.013]		[0.497]	[0.027]	
P value joint	0.820			0.854			0.327		

Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the control group mean, ii) the results of student-level regressions of covariates collected in the baseline survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects and iii) Number of observations. The sample is composed of non-attritors, individuals for which there is follow-up data available. Standard errors clustered at the strata level are presented in brackets. P-values are presented in parenthesis. P-values for a test that all variables are balanced are reported at the bottom of the table for each of the three samples considered.

Table A.2: Number of Students Enrolled per Classroom

	Cycle I schools	Cycle II schools	Two cycle schools
	(1)	(2)	(3)
3rd grade	0.526 [0.429] (0.220)		
4th grade	-0.588 [0.450] (0.192)		
6th grade		2.357 [1.474] (0.110)	
9th grade			0.190 [0.743] (0.799)
Mean (omitted group)	28.936 [0.649]	28.936 [0.649]	27.328 [0.949]
Omitted group	5th grade	9th grade	5th grade
Number of schools	78	14	58

Notes: This table reports results of a regression of maximum number of students enrolled per class in each grade on i) indicator variables of 3rd and 4th grades (in column 1 - Cycle I schools); ii) 6th grade (in column 2 - Cycle II schools) and iii) 9th grade (in column 3 - Two cycle schools) and school fixed effects.

Table A.3: Lee bounds - Attitudes towards math

	Full Sample	5th grade	9th grade
Lower bound	0.029	-0.010	0.045
SE	(0.040)	(0.050)	(0.071)
Upper bound	0.128	0.153	0.115
SE	(0.063)	(0.080)	(0.083)
Nobs	18109	12262	5847

Notes: This tables shows the results for the Lee bounds procedure (Lee, 2009) for the outcome attitudes towards math. The standard errors were estimated using cluster-bootstrap with 1,000 replications at the strata level. In the first column we show the result for the full sample and in the second and third columns, respectively, for the 5th and 9th grade students separately.

Table A.4: Results on Math Proficiency and Attitudes towards math (re-normalized index)

	(1)	(2)
	Panel A: Full sample	
Treatment	0.046	0.053
s.e.	[0.033]	[0.022]
p-value	(0.161)	(0.016)
N obs	12849	12849
N schools	150	150
N strata	35	35
	Panel B: 5th grade	
Treatment	0.032	0.052
s.e.	[0.033]	[0.026]
p-value	(0.332)	(0.045)
N obs	9031	9031
N schools	136	136
N strata	32	32
	Panel C: 9th grade	
Treatment	0.081	0.058
s.e.	[0.055]	[0.029]
p-value	(0.138)	(0.042)
N obs	3818	3818
N schools	72	72
N strata	15	15
Includes covariates	No	Yes

Notes: This table reports the results of a student-level regression on the re-normalized attitudes towards math index on an dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects. Panels A, B and C refer to the pooled sample, and 5th and 9th grades subsamples separately. For the pooled regressions, we interact the strata fixed effects with grade. The specifications reported in column 2 include the covariates presented in Table 1. Standard errors are clustered at the strata level. The renormalized index consider all questions with non-missing responses and renormalizes to have the same support as the original index.

Table A.5: Characteristics of Survey Attritors x Non-Attritors

Variable	Non-Attritors	Diff	N	Variable	Non-Attritors	Diff	N
Attitudes towards math	0.035 [1.001]	-0.089 [0.025]	11422	Uses computer lab at school during other classes	0.312 [0.464]	0.003 [0.008]	12334
Male	0.504 [0.500]	-0.003 [0.009]	12369	Uses computer lab at school not during class	0.132 [0.338]	0.012 [0.005]	12377
Year of Birth	2,004.5 [2.292]	-0.184 [0.026]	12381	Has mobile phone	0.713 [0.452]	0.004 [0.009]	12265
White	0.314 [0.464]	-0.009 [0.008]	10703	Has internet on mobile phone	0.710 [0.454]	-0.010 [0.009]	11286
Black	0.095 [0.294]	0.013 [0.006]	10703	Lives with mother	0.901 [0.299]	-0.016 [0.007]	12362
Native	0.040 [0.196]	-0.002 [0.003]	10703	Lives with father	0.636 [0.481]	-0.050 [0.009]	12360
Mixed	0.513 [0.500]	-0.005 [0.009]	10703	Has books at home	0.774 [0.419]	-0.029 [0.010]	12394
Asian	0.037 [0.190]	0.003 [0.004]	10703	Parents talk about school	0.840 [0.366]	0.007 [0.008]	12394
Has computer at home	0.586 [0.493]	-0.042 [0.010]	12396	Works outside home	0.071 [0.257]	0.031 [0.005]	12388
Frequently uses computer at home	0.459 [0.498]	-0.030 [0.009]	12380	Has ever repeated a grade	0.216 [0.411]	0.060 [0.009]	12304
Has internet at home	0.731 [0.443]	-0.013 [0.011]	12360	Math is the preferred subject	0.433 [0.495]	-0.023 [0.007]	12389
Uses computer at home for school activities	0.520 [0.500]	-0.028 [0.010]	12365	Portuguese is the preferred subject	0.243 [0.429]	0.034 [0.011]	12389
Uses computer lab at school	0.365 [0.481]	-0.003 [0.009]	12374	Other subject is preferred	0.325 [0.468]	-0.011 [0.008]	12389
Uses computer lab at school during portuguese classes	0.251 [0.434]	0.004 [0.012]	12403	Participated in Math Olympics	0.199 [0.399]	0.009 [0.008]	11340
Uses computer lab at school during math classes	0.285 [0.451]	-0.009 [0.010]	12368				

Notes: This table reports, for the pooled sample the mean and standard deviation (in brackets) for individuals that appeared in the baseline survey and in the survey follow-up (non-attritors). The second and seventh columns present the estimated differences for attritors, coming from the regression of the outcome on a dummy for attrition and strata fixed effects. The standard error, in brackets, are clustered at the strata level. The third and eigth columns show the number of observations with valid responses for each variable.

Table A.6: Results on Math Proficiency and Attitudes towards math — Common Sample

	Math test scores		Attitudes towards math	
	Study sample	Common Sample	Study sample	Common Sample
	(1)	(2)	(3)	(4)
	Panel A: Baseline Effects			
Treatment	-0.016	-0.009	0.061	0.067
s.e	[0.024]	[0.026]	[0.022]	[0.023]
p-value	(0.519)	(0.719)	(0.009)	(0.006)
N obs	14,846	13,855	11,157	10,361
	Panel B: Implementation			
T × shared	-0.076	-0.076	0.054	0.058
s.e	[0.032]	[0.032]	[0.020]	[0.022]
p-value	(0.023)	(0.023)	(0.010)	(0.011)
T × one comp per stud	0.081	0.081	0.038	0.058
s.e	[0.052]	[0.052]	[0.047]	[0.047]
p-value	(0.131)	(0.131)	(0.423)	(0.232)
N obs	13,231	13,231	10,710	9,927

Notes: Columns (1) and (3) replicates the baseline results and the interaction with implementation modality for, respectively math proficiency and attitudes towards math. Columns (2) and (4) replicates columns (1) and (3) for the subsample of school-grades that we observe both the math proficiency and the attitudes towards math. All regressions include covariates and standard errors are computed at the strata level. The number of observations in columns 1 and 2 of Panel B are the same because we have information from both the national standardized exam and from our survey for all schools in which implementation data was collected.

Table A.7: Balance Heterogeneity: Survey - Infrastructure

	Pooled Sample		5th grade		9th grade	
	X	$(1-X)$	X	$(1-X)$	X	$(1-X)$
	(β_1)	(β_2)	(β_1)	(β_2)	(β_1)	(β_2)
Attitudes towards math	-0.069	0.061	-0.085	0.069	-0.038	0.048
	[0.038]	[0.039]	[0.048]	[0.036]	[0.052]	[0.081]
Male	-0.028	0.006	-0.028	-0.005	-0.023	0.025
	[0.013]	[0.013]	[0.015]	[0.015]	[0.017]	[0.020]
Year of Birth	0.0	0.030	0.0	-0.063	0.0	0.163
	[0.045]	[0.050]	[0.073]	[0.061]	[0.077]	[0.054]
White	-0.002	-0.011	-0.009	-0.023	0.023	0.006
	[0.018]	[0.012]	[0.025]	[0.016]	[0.019]	[0.012]
Black	0.003	-0.006	0.013	-0.009	-0.026	0.000
	[0.010]	[0.007]	[0.015]	[0.011]	[0.016]	[0.009]
Native	0.011	-0.006	0.014	-0.008	-0.001	-0.004
	[0.006]	[0.005]	[0.008]	[0.007]	[0.011]	[0.007]
Mixed	-0.015	0.023	-0.021	0.036	0.004	0.001
	[0.019]	[0.015]	[0.024]	[0.023]	[0.022]	[0.018]
Asian	0.002	0.001	0.003	0.003	0.001	-0.004
	[0.008]	[0.005]	[0.008]	[0.007]	[0.016]	[0.008]
Has computer at home	-0.007	-0.006	-0.007	-0.016	-0.005	0.009
	[0.020]	[0.019]	[0.031]	[0.025]	[0.024]	[0.036]
Frequently uses computer at home	0.001	-0.003	0.001	-0.007	0.008	0.005
	[0.018]	[0.014]	[0.022]	[0.019]	[0.027]	[0.025]
Has internet at home	0.004	-0.014	-0.006	-0.031	0.033	0.011
	[0.018]	[0.022]	[0.029]	[0.032]	[0.028]	[0.026]
Uses computer at home for school activities	-0.004	-0.006	-0.015	-0.016	0.026	0.011
	[0.021]	[0.015]	[0.030]	[0.018]	[0.032]	[0.032]
Uses computer lab at school	-0.024	-0.004	0.011	-0.031	-0.095	0.037
	[0.100]	[0.041]	[0.097]	[0.070]	[0.139]	[0.042]
Uses computer lab at school during portuguese classes	-0.012	0.046	0.011	0.031	-0.053	0.074
	[0.092]	[0.030]	[0.104]	[0.046]	[0.112]	[0.047]
Uses computer lab at school during math classes	0.069	0.034	0.075	0.009	0.074	0.076
	[0.121]	[0.037]	[0.092]	[0.053]	[0.245]	[0.046]
Uses computer lab at school during other classes	-0.113	-0.012	0.004	-0.032	-0.386	0.018
	[0.063]	[0.031]	[0.067]	[0.050]	[0.108]	[0.066]
Uses computer lab at school not during class	-0.031	-0.001	-0.038	-0.001	-0.013	-0.001
	[0.013]	[0.015]	[0.022]	[0.013]	[0.020]	[0.037]

(cont)

Table A.5 Cont : Balance Heterogeneity: Survey — Infrastructure

	Pooled Sample		5th grade		9th grade	
	X	$(1-X)$	X	$(1-X)$	X	$(1-X)$
	(β_1)	(β_2)	(β_1)	(β_2)	(β_1)	(β_2)
Has mobile phone	0.006	-0.003	0.002	0.002	0.016	-0.010
	[0.015]	[0.016]	[0.025]	[0.026]	[0.016]	[0.018]
Has internet on mobile phone	-0.002	-0.004	0.003	-0.007	-0.012	0.000
	[0.013]	[0.014]	[0.021]	[0.022]	[0.029]	[0.015]
Lives with mother	0.002	0.008	0.017	0.002	-0.032	0.017
	[0.012]	[0.009]	[0.018]	[0.009]	[0.015]	[0.013]
Lives with father	0.003	0.004	0.015	-0.015	-0.025	0.029
	[0.018]	[0.014]	[0.026]	[0.019]	[0.012]	[0.023]
Has books at home	0.001	-0.015	0.004	-0.037	0.003	0.020
	[0.014]	[0.017]	[0.019]	[0.023]	[0.032]	[0.015]
Parents talk about school	0.004	-0.003	-0.011	-0.009	0.043	0.008
	[0.007]	[0.008]	[0.010]	[0.011]	[0.019]	[0.009]
Works outside home	-0.023	0.015	-0.022	0.010	-0.025	0.023
	[0.008]	[0.009]	[0.012]	[0.012]	[0.009]	[0.014]
Has ever repeated a grade	-0.013	-0.002	0.001	0.018	-0.048	-0.029
	[0.020]	[0.018]	[0.026]	[0.027]	[0.017]	[0.016]
Math is the preferred subject	-0.015	0.024	-0.016	0.029	-0.015	0.018
	[0.023]	[0.022]	[0.030]	[0.022]	[0.035]	[0.037]
Portuguese is the preferred subject	0.004	0.011	-0.009	0.019	0.034	-0.002
	[0.019]	[0.017]	[0.026]	[0.018]	[0.018]	[0.031]
Other subject is preferred	0.011	-0.035	0.026	-0.048	-0.019	-0.016
	[0.020]	[0.019]	[0.019]	[0.015]	[0.042]	[0.041]
Participated in Math Olympics	0.016	-0.010	0.014	-0.001	0.017	-0.023
	[0.017]	[0.014]	[0.018]	[0.018]	[0.038]	[0.022]
joint p-value ($\beta_1 = 0$)	0.736		0.738		0.534	
joint p-value ($\beta_2 = 0$)	0.697		0.729		0.108	
joint p-value ($\beta_1 = \beta_2 = 0$)	0.692		0.672		0.551	

Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the results of student-level regressions of covariates collected in the baseline survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment interacted with a heterogeneity variable. X is an indicator variable which takes value one if there were no infrastructure problems. Standard errors are clustered at the strata level.

Table A.8: Balance Heterogeneity: Survey - One computer per student

	Pooled Sample		5th grade		9th grade	
	X	$(1-X)$	X	$(1-X)$	X	$(1-X)$
	(β_1)	(β_2)	(β_1)	(β_2)	(β_1)	(β_2)
Attitudes towards math	-0.015	0.028	-0.025	0.030	-0.001	0.026
	[0.074]	[0.034]	[0.072]	[0.033]	[0.094]	[0.079]
Male	-0.017	-0.008	-0.020	-0.023	-0.012	0.016
	[0.015]	[0.011]	[0.014]	[0.013]	[0.031]	[0.018]
Year of Birth	0.1	0.014	0.1	-0.087	0.0	0.180
	[0.046]	[0.045]	[0.105]	[0.047]	[0.080]	[0.054]
White	0.006	-0.009	0.019	-0.027	-0.017	0.026
	[0.023]	[0.011]	[0.038]	[0.016]	[0.025]	[0.012]
Black	0.000	-0.004	0.007	-0.004	-0.016	-0.005
	[0.011]	[0.007]	[0.028]	[0.010]	[0.028]	[0.012]
Native	0.004	-0.001	0.005	0.002	0.002	-0.006
	[0.005]	[0.005]	[0.008]	[0.006]	[0.010]	[0.007]
Mixed	-0.018	0.014	-0.042	0.029	0.033	-0.011
	[0.023]	[0.015]	[0.030]	[0.022]	[0.020]	[0.016]
Asian	0.007	0.000	0.011	0.001	-0.003	-0.004
	[0.010]	[0.005]	[0.011]	[0.006]	[0.012]	[0.009]
Has computer at home	-0.019	-0.006	-0.030	-0.019	-0.008	0.012
	[0.019]	[0.016]	[0.033]	[0.020]	[0.024]	[0.037]
Frequently uses computer at home	-0.005	-0.001	-0.014	-0.005	0.005	0.005
	[0.017]	[0.012]	[0.021]	[0.016]	[0.032]	[0.025]
Has internet at home	-0.008	-0.008	-0.023	-0.025	0.011	0.022
	[0.009]	[0.020]	[0.027]	[0.026]	[0.022]	[0.027]
Uses computer at home for school activities	-0.009	-0.003	-0.035	-0.015	0.021	0.014
	[0.014]	[0.014]	[0.024]	[0.017]	[0.026]	[0.034]
Uses computer lab at school	-0.068	0.010	-0.130	0.034	0.054	0.000
	[0.099]	[0.043]	[0.178]	[0.065]	[0.081]	[0.047]
Uses computer lab at school during portuguese classes	-0.001	0.018	-0.044	0.031	0.101	0.030
	[0.086]	[0.031]	[0.135]	[0.041]	[0.073]	[0.036]
Uses computer lab at school during math classes	0.097	0.026	0.088	0.023	0.144	0.059
	[0.136]	[0.037]	[0.107]	[0.045]	[0.180]	[0.057]
Uses computer lab at school during other classes	-0.187	-0.008	-0.176	0.036	-0.175	-0.061
	[0.051]	[0.029]	[0.125]	[0.036]	[0.175]	[0.043]
Uses computer lab at school not during class	-0.039	-0.002	-0.063	0.000	-0.004	-0.004
	[0.015]	[0.014]	[0.032]	[0.014]	[0.023]	[0.037]

(cont)

Table A.6 Cont: Balance Heterogeneity: Survey - One computer per student

	Pooled Sample		5th grade		9th grade	
	X	$(1-X)$	X	$(1-X)$	X	$(1-X)$
	(β_1)	(β_2)	(β_1)	(β_2)	(β_1)	(β_2)
Has mobile phone	0.000	0.000	-0.013	0.006	0.020	-0.008
	[0.018]	[0.016]	[0.036]	[0.024]	[0.023]	[0.018]
Has internet on mobile phone	-0.002	-0.003	-0.005	0.000	0.004	-0.007
	[0.013]	[0.013]	[0.011]	[0.020]	[0.032]	[0.014]
Lives with mother	-0.008	0.009	0.015	0.003	-0.039	0.020
	[0.013]	[0.008]	[0.019]	[0.008]	[0.019]	[0.012]
Lives with father	0.012	-0.004	0.039	-0.029	-0.037	0.029
	[0.025]	[0.013]	[0.036]	[0.016]	[0.030]	[0.023]
Has books at home	0.011	-0.014	0.003	-0.028	0.021	0.011
	[0.015]	[0.016]	[0.021]	[0.022]	[0.028]	[0.018]
Parents talk about school	0.003	0.001	-0.019	-0.009	0.031	0.016
	[0.011]	[0.007]	[0.014]	[0.010]	[0.012]	[0.007]
Works outside home	-0.018	0.009	-0.013	0.000	-0.028	0.023
	[0.007]	[0.008]	[0.010]	[0.010]	[0.009]	[0.014]
Has ever repeated a grade	-0.052	0.006	-0.059	0.030	-0.039	-0.032
	[0.024]	[0.016]	[0.036]	[0.022]	[0.018]	[0.016]
Math is the preferred subject	-0.023	0.019	-0.029	0.022	-0.013	0.016
	[0.027]	[0.021]	[0.033]	[0.020]	[0.037]	[0.038]
Portuguese is the preferred subject	0.000	0.013	0.001	0.010	-0.003	0.016
	[0.025]	[0.015]	[0.044]	[0.016]	[0.018]	[0.032]
Other subject is preferred	0.023	-0.032	0.029	-0.032	0.016	-0.031
	[0.030]	[0.018]	[0.035]	[0.011]	[0.052]	[0.042]
Participated in Math Olympics	0.018	-0.008	0.019	-0.002	0.016	-0.020
	[0.022]	[0.013]	[0.028]	[0.015]	[0.029]	[0.022]
joint p-value ($\beta_1 = 0$)	0.135		0.377		0.502	
joint p-value ($\beta_2 = 0$)	0.963		0.234		0.138	
joint p-value ($\beta_1 = \beta_2 = 0$)	0.647		0.312		0.495	

Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the results of student-level regressions of covariates collected in the baseline survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment interacted with a heterogeneity variable. X is an indicator variable which takes value one if the school had one computer per student. Standard errors are clustered at the strata level.

Table A.9: Balance Heterogeneity: Prova Brasil — Infrastructure

	Pooled Sample		5th grade		9th grade	
	X	$(1-X)$	X	$(1-X)$	X	$(1-X)$
	(β_1)	(β_2)	(β_1)	(β_2)	(β_1)	(β_2)
Male	-0.025	-0.004	-0.028	-0.008	-0.009	0.003
	[0.018]	[0.014]	[0.021]	[0.016]	[0.016]	[0.024]
White	-0.006	-0.011	-0.008	-0.023	0.005	0.011
	[0.024]	[0.015]	[0.029]	[0.018]	[0.012]	[0.020]
Black	0.0	-0.014	0.0	-0.016	0.0	-0.011
	[0.009]	[0.007]	[0.012]	[0.010]	[0.015]	[0.008]
Mixed	-0.017	0.026	-0.019	0.050	-0.011	-0.017
	[0.017]	[0.019]	[0.021]	[0.026]	[0.011]	[0.039]
Asian	0.006	0.005	0.005	0.001	0.011	0.012
	[0.004]	[0.004]	[0.005]	[0.005]	[0.005]	[0.007]
Native	0.003	-0.002	0.005	-0.002	-0.004	0.000
	[0.003]	[0.004]	[0.004]	[0.005]	[0.011]	[0.006]
Race not declared	0.010	-0.004	0.012	-0.010	-0.003	0.006
	[0.009]	[0.007]	[0.013]	[0.009]	[0.008]	[0.013]
Age	0.036	-0.034	0.033	0.011	0.046	-0.115
	[0.023]	[0.033]	[0.033]	[0.044]	[0.053]	[0.041]
Mother has completed at least high school	0.010	0.049	0.002	0.044	0.030	0.057
	[0.022]	[0.020]	[0.037]	[0.031]	[0.068]	[0.024]
Mother literate	-0.003	0.000	-0.007	-0.002	0.011	0.005
	[0.004]	[0.003]	[0.005]	[0.004]	[0.014]	[0.006]
Father has completed at least high school	-0.023	0.059	-0.043	0.062	0.026	0.056
	[0.021]	[0.020]	[0.033]	[0.026]	[0.053]	[0.030]
Father literate	-0.003	0.004	-0.003	0.005	0.001	0.002
	[0.006]	[0.006]	[0.007]	[0.006]	[0.010]	[0.010]
Teacher younger than 50 years old	0.072	-0.047	0.099	-0.065	-0.192	0.010
	[0.075]	[0.077]	[0.079]	[0.102]	[0.185]	[0.239]
2015 Prova Brasil math grade	-0.224	0.146	-0.237	0.015	-0.155	0.366
	[0.087]	[0.086]	[0.099]	[0.081]	[0.200]	[0.136]
joint p-value ($\beta_1 = 0$)	0.501		0.278		0.943	
joint p-value ($\beta_2 = 0$	0.206		0.597		0.776	
joint p-value ($\beta_1 = \beta_2 = 0$)	0.293		0.305		0.937	

Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the results of student-level regressions of covariates collected in the baseline survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment interacted with a heterogeneity variable. X is an indicator variable which takes value one if there were no infrastructure problems. Standard errors are clustered at the strata level.

Table A.10: Balance Heterogeneity: Prova Brasil — One computer per student

	Pooled Sample		5th grade		9th grade	
	X	$(1-X)$	X	$(1-X)$	X	$(1-X)$
	(β_1)	(β_2)	(β_1)	(β_2)	(β_1)	(β_2)
Male	-0.014	-0.012	-0.021	-0.019	-0.011	-0.005
	[0.021]	[0.013]	[0.025]	[0.016]	[0.035]	[0.022]
White	-0.010	-0.010	-0.008	-0.025	-0.040	0.012
	[0.034]	[0.014]	[0.045]	[0.018]	[0.022]	[0.021]
Black	0.0	-0.012	0.0	-0.015	0.0	-0.006
	[0.014]	[0.006]	[0.021]	[0.009]	[0.018]	[0.008]
Mixed	-0.005	0.024	-0.008	0.052	0.031	-0.025
	[0.017]	[0.017]	[0.026]	[0.022]	[0.017]	[0.041]
Asian	0.003	0.006	0.001	0.003	0.006	0.011
	[0.004]	[0.004]	[0.005]	[0.005]	[0.007]	[0.007]
Native	0.009	-0.003	0.013	-0.005	-0.004	0.001
	[0.004]	[0.004]	[0.005]	[0.005]	[0.010]	[0.006]
Race not declared	0.006	-0.005	0.005	-0.011	0.006	0.007
	[0.010]	[0.007]	[0.014]	[0.008]	[0.012]	[0.013]
Age	-0.013	-0.005	-0.034	0.043	0.055	-0.106
	[0.030]	[0.030]	[0.048]	[0.035]	[0.065]	[0.046]
Mother has completed at least	-0.002	0.041	-0.010	0.031	0.030	0.067
high school	[0.029]	[0.019]	[0.051]	[0.028]	[0.054]	[0.025]
Mother literate	-0.002	-0.001	-0.006	-0.004	0.008	0.004
	[0.004]	[0.003]	[0.006]	[0.003]	[0.014]	[0.006]
Father has completed at least	-0.026	0.045	-0.043	0.038	0.032	0.067
high school	[0.025]	[0.020]	[0.041]	[0.027]	[0.053]	[0.031]
Father literate	0.005	0.000	0.005	-0.001	-0.004	-0.001
	[0.008]	[0.005]	[0.010]	[0.006]	[0.017]	[0.010]
Teacher younger than 50 years old	0.030	-0.066	0.046	-0.072	0.019	-0.015
	[0.080]	[0.070]	[0.080]	[0.090]	[0.268]	[0.239]
2015 Prova Brasil math grade	-0.211	0.071	-0.231	-0.065	-0.039	0.383
	[0.151]	[0.085]	[0.138]	[0.077]	[0.261]	[0.137]
joint p-value ($\beta_1 = 0$)	0.872		0.564		0.720	
joint p-value ($\beta_2 = 0$)	0.090		0.213		0.399	
joint p-value ($\beta_1 = \beta_2 = 0$)	0.342		0.349		0.622	

Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the results of student-level regressions of covariates collected in the baseline survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment interacted with a heterogeneity variable. X is an indicator variable which takes value one if there were no infrastructure problems. Standard errors are clustered at the strata level.

Table A.11: **ITT Heterogeneity - Within Municipality Results**

	Pooled		5th grade	
	(1)	(2)	(3)	(4)
Panel A: Math proficiency				
$T \times$ "one comp per stud"	0.255	0.245	0.325	0.306
	[0.099]	[0.105]	[0.100]	[0.110]
	(0.010)	(0.020)	(0.001)	(0.005)
$T \times$ "internet/hardware prob"		-0.049		-0.145
		[0.116]		[0.120]
		(0.672)		(0.228)
N	13231	13231	9088	9088
Panel B: Attitudes towards math				
$T \times$ "one comp per stud"	-0.053	-0.040	-0.009	0.003
	[0.075]	[0.087]	[0.078]	[0.088]
	(0.485)	(0.649)	(0.906)	(0.971)
$T \times$ "internet/hardware prob"		0.058		0.075
		[0.180]		[0.226]
		(0.746)		(0.741)
N	10710	10710	7359	7359

Notes: Notes: This table reports results for student-level regressions of math proficiency (Panel A) and attitudes towards math (Panel B) on interaction terms between the treatment dummy and dummy variables for type of implementation. We also include interactions between the treatment dummy and dummy variables for each municipality. Therefore, the coefficients on $T \times$ "one comp per stud" should be interpreted as the differential treatment effect when we consider implementation with one computer per student relative to when computers are shared, when we consider only within municipality variation in the type of implementation. Columns (2), and (4) also include interactions between the treatment dummy and a dummy indicating if there was internet/hardware problems. All specifications also include the type of implementation dummy, strata fixed effects, and the covariates reported in Table 1. We also include the internet/hardware problems dummy in the specifications from columns (2), and (4). Standard errors are clustered at the strata level are presented in brackets. We present p-values in parenthesis. We omit the estimates for 9th graders because there is not much within-municipality variation for type of implementation, implying unreliable results. Consistent with that, the assessment proposed by Ferman (2021) for a 5% nominal-level test suggests rejection rates of around 20% under the null.

Table A.12: **ITT Heterogeneity - Permutation p-values**

	Full sample	5th grade	9th grade
	(1)	(2)	(3)
Panel A: Math proficiency			
$T \times$ "one comp per stud" (β_1)	0.081	0.127	-0.102
	(0.183)	(0.070)	(0.133)
$T \times$ "internet/hardware prob"	-0.076	-0.082	-0.075
	(0.033)	(0.068)	(0.368)
p-value ($\beta_1 = \beta_2$)	(0.030)	(0.010)	(0.806)
Panel B: Attitudes towards math			
$T \times$ "one comp per stud" (β_1)	0.036	0.070	-0.031
	(0.504)	(0.244)	(0.740)
$T \times$ "internet/hardware prob"	0.053	0.035	0.108
	(0.017)	(0.251)	(0.003)
p-value ($\beta_1 = \beta_2$)	(0.755)	(0.562)	(0.170)

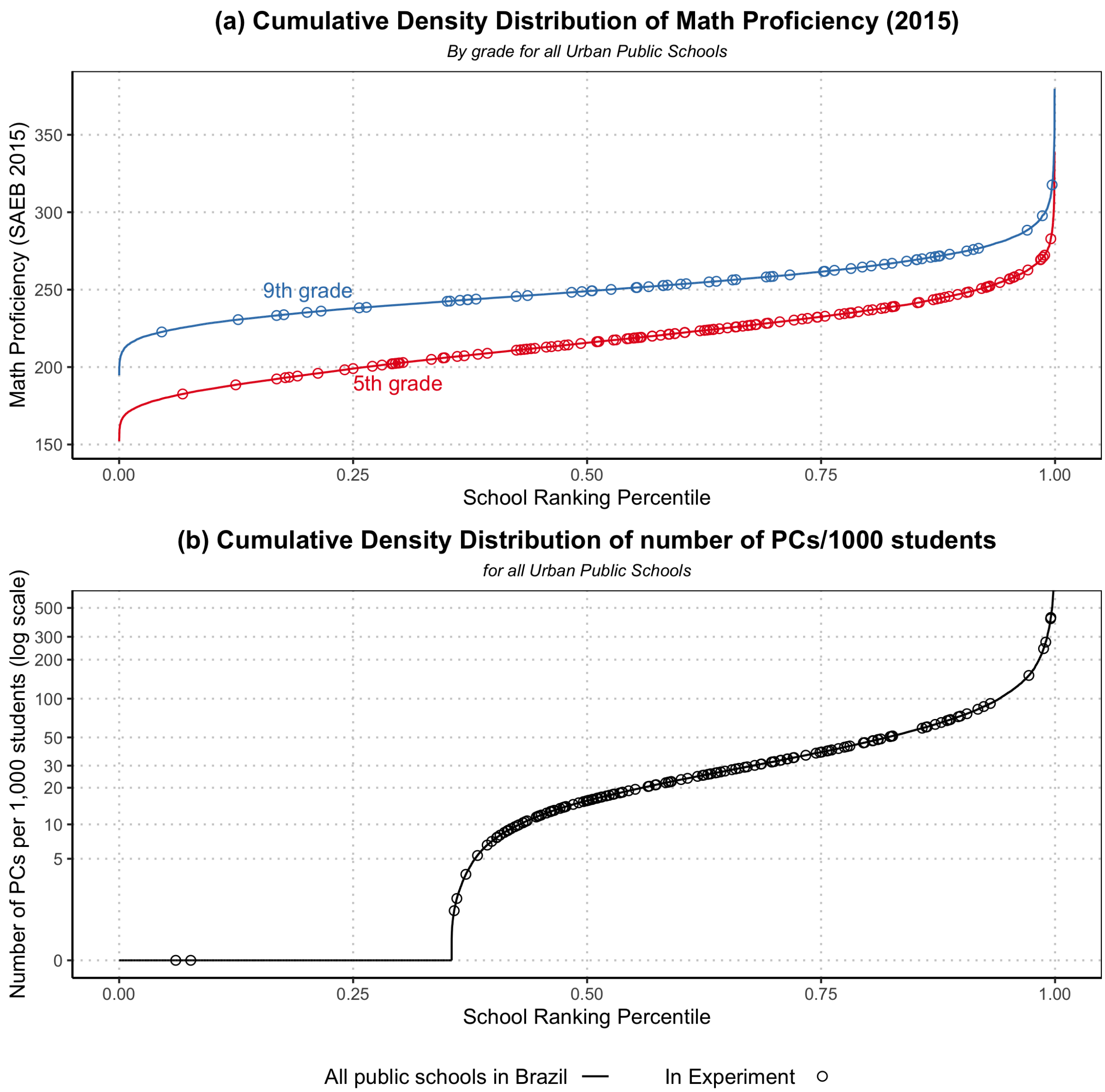
Notes: this table replicates the results from columns 1, 3, and 5 of Table 8, but calculating the p-values using a permutation test. The p-values are presented in parenthesis

Table A.13: DID: Control grades x Schools not in the experiment

	Pooled		5th grade		9th grade	
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment × 2017	-0.042	0.026	-0.025	0.028	-0.002	0.021
	[0.040]	[0.033]	[0.043]	[0.040]	[0.071]	[0.058]
	(0.293)	(0.426)	(0.563)	(0.480)	(0.979)	(0.716)
Municipality × time FE	No	Yes	No	Yes	No	Yes
Number of schools	568	568	506	506	206	206
Number of observations	3348	3348	2484	2484	864	864

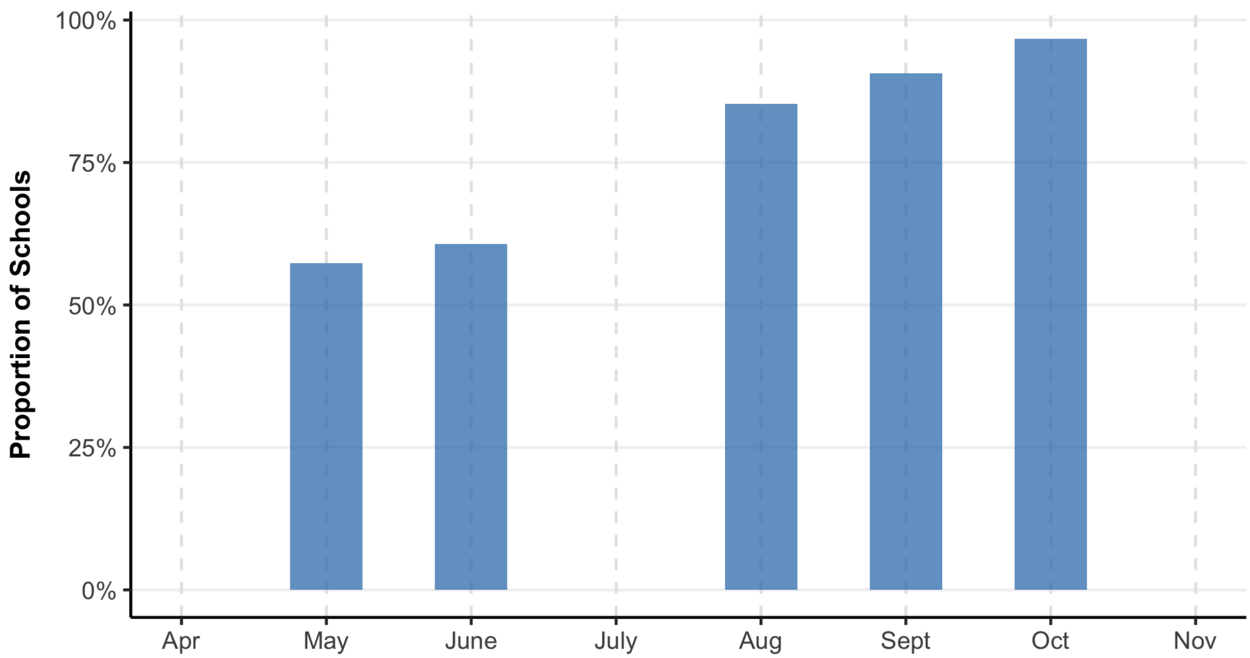
Notes: This table reports the results of difference-in-differences estimates comparing the control school-grades in our experiment with municipality school-grades that did not participate in the RCT, but are from the municipalities that participated in the study. The outcome variable is the school × grade math proficiency in the national standardized exam. The pre-treatment data is from every other year from 2005 to 2015, while the post-treatment data is from 2017. We standardized this variable using the same mean and standard deviation as in our main specification in the paper. All regressions include school × grade fixed effects, year fixed effects, and the interaction between an indicator variable for control grades in the experiment with a dummy for the 2017. Columns 1 and 2 report results pooling 5th and 9th grades. In these regressions, we also include an interaction between year and grade. We report results for 5th grade in columns 3 and 4, and for 9th grade in columns 5 and 6. In columns 1, 3, and 5, we also include municipality × time fixed effects (those fixed effects are also interacted with grade in column 2). Standard errors clustered at the school level are reported in brackets. P-values are reported in parenthesis.

Figure A.1: Comparison of schools in the experiment with all urban public schools in Brazil



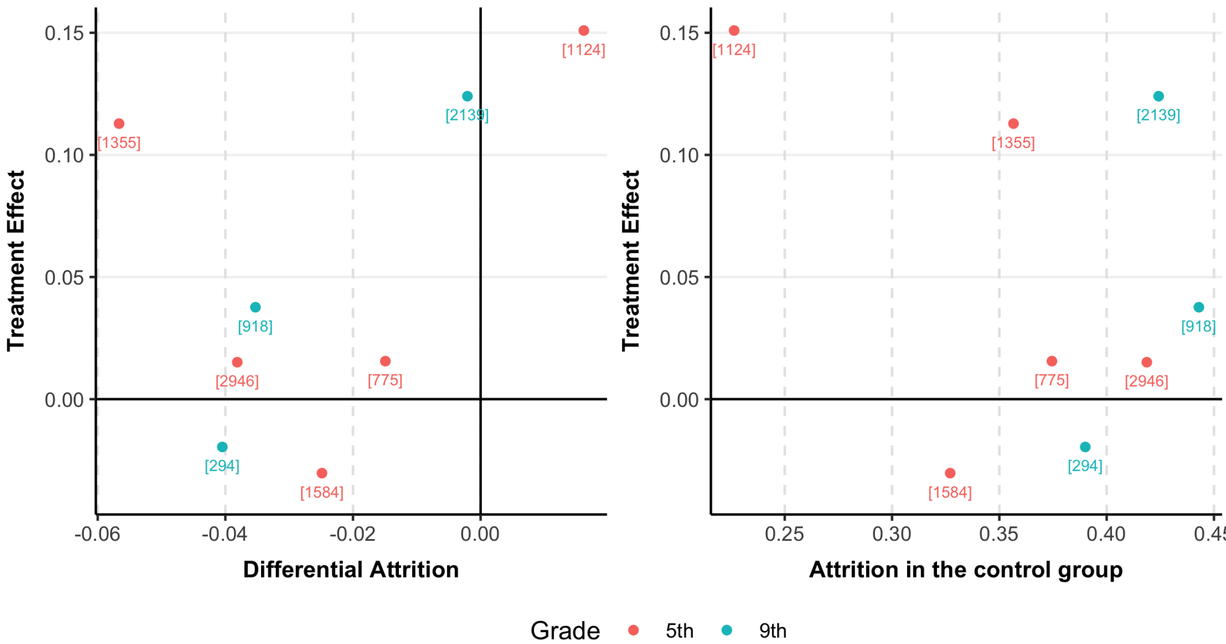
Notes: The solid lines represent the cumulative density function (CDF) for the Math Proficiency in the SAEB 2015 exam (panel a) and for the number of computers per 1,000 students using the School Census in 2016 (panel b). The number of computers consider only the number of computers dedicated to student use. In both panels all urban public schools in Brazil are considered. The circles represent the schools in the experiment. In panel (a) the blue color is for 9th graders and the red color for 5th graders. In panel (b), for better visualization we transform the y-axis using $log(y+1)$ instead of y.

Figure A.2: Proportion of Schools with Adequate Infrastructure



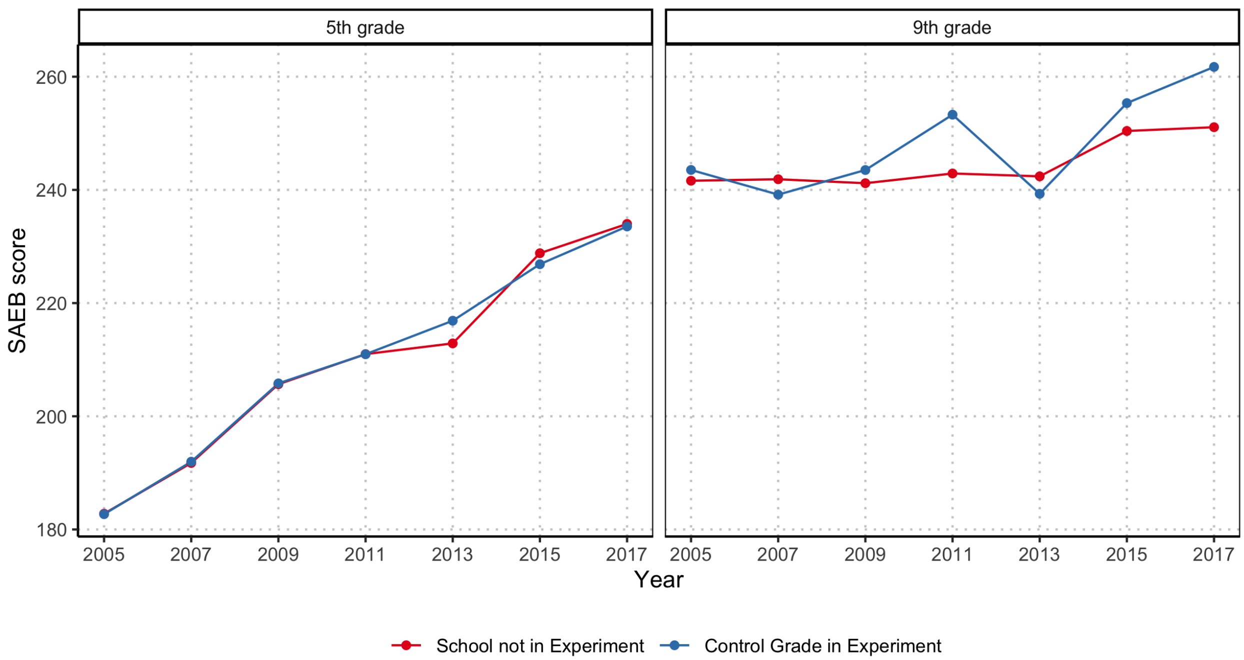
Notes: Each bar represents the proportion of school with adequate infrastructure for each month. There is no data for July because it is the month of winter recess.

Figure A.3: Differential Attrition and Treatment Effects on Attitudes towards math



Notes: Each dot represents a municipality-grade sub-sample. On the left panel, we computed for each sub-sample the differential attrition and treatment effect of attitudes towards math, controlling for strata fixed effects. The numbers in brackets are the number of non-missing observations for each sub-sample. The panel on the right is a similar exercise but using the attrition level in the control group instead of the differential attrition.

Figure A.4: Average SAEB scores



Notes: The blue curve represents the average math proficiency in the SAEB exam between 2005 and 2017 for schools in the experiment, but only for grades that did not receive the treatment. The red curve represents the same average for all the other public schools in the same municipalities of the schools in the experiment. The left plot graphs the scores for 5th graders and the right plot for 9th graders. We restrict the sample to schools for which we have math proficiency for all years.