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: December 2020

*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 the online Khan Academy 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 there are no infrastructure problems and when the implementation modality is based on one computer per student. These results highlight the implementation challenges associated with educational tech-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. Overall, learning levels in developing countries remain critically low, with too many children and adolescents leaving school with insufficient literacy and numeracy skills (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. Different interventions rely on a range of approaches, such as introducing computers and internet connection in public schools, distributing laptops to students, and promoting the adoption of educational software that are able to deal with within-class heterogeneity in students' learning levels by delivering content adapted to each students' needs (Bulman and Fairlie, 2016).

In this paper, we present the findings of a large-scale randomized evaluation designed to evaluate a program that integrates a computer-assisted learning platform into regular math classes of Brazilian public primary schools. Once a week, teachers would take their students to the school's computer lab and the students would use the online Khan Academy platform for instructional content and exercises for 50 minutes, under their supervision. The main advantage of this program is that the platform is adaptive, tailoring the exercises for each particular student based on their performance.

Khan Academy is one of the most popular online platforms focused on delivering educational content tailored at each students' level, offering free instructional videos and personalized exercises both in math as well as in other subject areas, ranging from kindergarten to college levels. The platform stands out for its worldwide popularity, having reached 71 million individuals in 190 countries since its foundation in 2008. Through partnerships with several organizations in different countries, Khan Academy has increasingly expanded its reach to different audiences in various languages. The evaluated program was an initiative implemented in Brazil since 2012 as a partnership between Khan Academy and the nonprofit Lemann Foundation.

We present results from a field experiment based on 5th and 9th grade students from 157 schools (approximately 15000 students) located across three different regions of Brazil.[1] We estimate the impacts of the intervention, carried out in 2017, on math proficiency using a standardized national exam, and also on a measure of attitudes towards math. We first show that students in treated grades report to use the platform in math classes, and that this increase did not crowd out the use of computer lab by other subjects. In terms of outcomes, we find no evidence that *Khan Academy in Schools* enhanced math proficiency, on average. However, using a survey designed to measure student's attitudes towards math, we find evidence that the interactive and playful environment of the platform translates into more positive attitudes towards math.

In an attempt to understand these results, we perform an exercise to explore the role of the quality of implementation, which 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 students spent significantly less time in the platform when compared with the first group of schools. 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 in Section 6.3 the limitations of such exercise, 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 in a short time spam, but gains in proficiency may require a more consistent use.

Other studies have previously tried to investigate the effects of the Khan Academy

[1] These are the grades that participate in the standardized national evaluation in Brazil.

platform use on math achievement. However the majority of the existing evidence relies on quasi-experimental approaches and/or small samples.[2] 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.

While this paper is one of the first large-scale randomized evaluations, with more than 150 schools and almost 15,000 students, of an implementation of the Khan Academy platform, there has been a series of studies investigating the effects of technology-enhanced instruction interventions in developing countries on learning outcomes. Reviews by Glewwe and Muralidharan (2016) and Bulman and Fairlie (2016) show the results are largely varied, with estimates ranging from significantly negative to significantly positive magnitudes. Overall, the evidence suggests that simply granting hardware to students in developing countries do not lead to gains in proficiency.[3] On the other hand, interventions that provide students with a given software/platform as a learning aid generally show positive effects on learning, particularly if it has the ability to tailor content to the student's needs.[4]

Most of the available evidence is associated with computer-aided learning (*henceforth* CAL) interventions that increase the number of hours students are exposed to academic

[2]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.

[3]Barrera-Osorio and Linden (2009); Beuermann et al. (2015); Cristia et al. (2017); Malamud and Pop-Eleches (2011); Mo et al. (2013)

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

instruction, complementing traditional teaching. Less is known on the effects of CAL interventions during school time, as an integrated resource into regular teaching, and the few existing studies show mixed results. Linden (2008), for instance, finds negative effects of a CAL program implemented as a substitute for regular teaching in India, while Carrillo et al. (2011) find promising results in Ecuador, where a government-implemented large-scale CAL program in primary schools had a positive impact on mathematics test scores. While Büchel et al. (2020) implement an intervention that increased the number of hours, they also have a control arm that received the additional hours without the platform, allowing for a comparison between the technology and regular teaching, yielding substantially smaller effects. Bettinger et al. (2020) also examine the effects of different dosages of a CAL platform as a direct substitute for traditional teaching in Russia, finding positive effects on test scores. Their treatment was administrated as a substitute to homework, which differs from the treatment we analyze, where it was implemented during class hours. Finally, Mo et al. (2020) find that when schools (disrespecting the intervention protocol) used CAL as substitute to traditional learning, there were no effects, while on those schools that used as complements there were positive effects.

The modality of CAL implementation is rarely a choice for developing countries. Capacity restrictions may limit the ability of implementing CAL as a complement to traditional teaching. Indeed, in 2017, 90% of Brazilian public schools had classes in more than one period (morning, afternoon or evening).[5] Therefore, we contribute to the literature by implementing a large-scale randomized control trial to investigate the effects of a CAL adaptive intervention integrated into regular teaching in Brazil. In addition, our results may explain the mixed results found to date in this strain of the literature. When such programs are implemented during class hours, their effects will depend on their efficacy *relative* to a standard math class. We provide evidence that, in such cases, the net effect might range from

[5]We use the Brazilian educational census from 2017 and analyzed all schools with students from the 1st-9th grades. If at least 15% of the students had classes in more than one period we classified the school as having classes in more than one period.

negative to positive, depending on whether there are implementation challenges. Therefore, assessing the adequacy of the implementation conditions and the technology infrastructure is crucial before scaling up such programs in a developing country context.

This paper is organized as follows. Section 2 describes the background and the program. Section 3 presents the experimental design. Section 4 describes our data and empirical strategy. Section 5 presents details on the implementation of the program and compliance with the experimental design. Section 6 discusses the results and section 7 concludes.

2 Background and Context: Khan Academy in Schools Program

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

The initiative has greatly expanded over the years and currently reaches millions of students in over 190 countries with resources available in 36 languages. The Brazilian version of the platform was a joint effort between Khan Academy and Lemann Foundation, a Brazilian nonprofit focused at enhancing the quality of public schools in Brazil, which are mostly attended by children coming from lower income families. Focused on math education, the partnership translated the contents into Portuguese and reached 2.6 million students, which registered in the platform in the period of 2012 to 2017.[6]

The platform may enhance students' math performance through three main channels.

[6] According to information reported o the Lemann Foundation's website *https://fundacaolemann.org.br/materiais/khan-academy-in-brazil*

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. 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.

Elementary education in Brazil is mandatory and goes from 1st to 9th grades, with students ranging from 6 to 14 years old. There are three main groups of schools in terms of the grades they offered: (a) schools that offer only the first 5 grades (Cycle I), (b) schools that offer only the final 4 grades, from 6th to 9th (Cycle II) and (c) the entire elementary level, from 1st to 9th grade (Cycles I and II). Elementary education is in its majority publicly provided. In 2017, among the 183,743 schools offering elementary education, 78.8% of them were public, covering 83.2% of the 27 million enrolled students.[7] 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.

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.

[7] According to the 2017 Schooling Census.

One of their initiatives is to promote the use of Khan Academy 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 Khan Academy and iii) close monitoring of intervention's implementation by Lemann Foundation staff, which acted as promoters of Khan Academy, providing assistance for solving any potential difficulties schools/teachers were facing. The program also allows teachers to have access to a detailed feedback report on students' performance, indicating their strengths and weaknesses.

The implementation of Khan Academy 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.

If this program is scaled, we should expect variation in the implementation across schools. For example, schools with a higher rate of computers per students should be more

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

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 a larger proportion of schools implementing the program in the rotation mode 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. 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 these 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 compose a stratum. This resulted in 157 schools in the final evaluation sample.[9]

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 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;[10] 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. It is worth noticing that all Cycle I and II schools serve as treatment for one grade and control for the

[9]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.

[10]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.

other.[11]

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 grades. 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 major concerns.

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.

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

[11] 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.

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.[12]

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

[12]There is a literature documenting the effects of this index, viewed as one instrument for increasing school accountability. For example, see Firpo et al. (2017).

[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.

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.

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 plot 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

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

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 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.

[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.

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 Khan Academy platform integrated to one math class every week (around 50 minutes per week). The assignment to the treated group also involved frequent visits from Lemann foundation staff, which followed up on treated grades' usage of the platform, solved any potential difficulties and acted as promoters of Khan Academy usage. 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, substituting one traditional math class per week for the Khan Academy 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 avoid 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 Khan Academy platform.

Our ITT estimates are based on the following regression:

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

where y_{igs} is an outcome of interest for individual i, who belongs to grade g in a school s,

[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.

Z_{igs} is an indicator variable that takes value 1 if individual i belongs to a treated school-grade, $\mathbf{X}_{igs}$ is a set of baseline controls, which includes strata fixed effects, and ϵ_{igs} is an error term. 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 (2019).

We consider two main outcomes: math proficiency and attitudes towards math.[19] 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 with Experimental Design

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

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

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 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 Khan Academy 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.[20]

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 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 for the remainder of the class.[21]

Such implementation issues 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.[22]

In columns 1 to 3 of Table 5, we show how the total number of minutes logged in the

[20]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.

[21]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.

[22]greenThe 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.

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.[23] 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. We also find that 5th grade students logged in more weeks than 9th graders, and that 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 margin of usage. Interestingly, Appendix Figure A.1 shows that the infrastructure problems were concentrated in the beginning of the intervention around April-June. In the last months of intervention most of the schools did not present infrastructure problems.

6 Results

6.1 Treatment Effects on Math Proficiency

Columns 1 and 2 of Table 6 shows intent to treat 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

[23]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.

shows the results for the regression on the treatement indicator and strata fixed effect while the second columns 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 (2019) does not detect large problems with the inference procedure.[24]

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 (0.060σ 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.2, 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

[24]The assessment proposed by Ferman (2019) 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 sample should be considered with caution.

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 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.

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. Therefore, it might be that those with worse implementation may be driving the null result on math proficiency.

While we do not have experimental variation on whether schools experienced infrastructure problems, or on whether they implemented the program with one student per computer, we take advantage of the fact that all schools implemented Khan Academy 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, within each school, grades that were not assigned to receive treatment would have had the same quality and modality of implementation as grades that were treated. 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. 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 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. 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.[25] 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

[25] 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.

evidence that we should expect that the computer lab of a given school would comport the same modality of treatment (rotation versus one computer per student) regardless of the treated grade. Finally, in Appendix Tables A.6, A.7, A.8 and A.9, 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. Columns 1-2 show the heterogeneity results for math proficiency, while columns 3-4 display the results for attitudes towards math. Our results provides suggestive evidence that integration with Khan Academy may be an effective alternative to traditional curriculum if adequately implemented. Students assigned to treated grades that did not face infrastructure problems had marginally higher math scores (0.058σ, p-value=0.220), and gains were registered when the modality of implementation was one computer per student (0.081σ, p-value=0.121). On the other hand, treated students in schools with infrastructure problems and students assigned to grades that implemented the rotational modality of the program performed worse in the SAEB exam. The p-values for the test that the coefficients on the good and bad implementation is the same is 0.09 for infrastructure problems and 0.02 for implementation modality.

The positive estimates for the samples with better implementation are mostly driven by the 5th grade subsample, which experienced larger than the average gains both for students assigned to treated grades that faced no infrastructure problems (0.093σ, p-value=0.110) and for students assigned to the individual use of the computer modality (0.127σ, p-value=0.016). In the 5th grade, negative effects on math scores were registered for students in the poorer implementation group, with statistically significant effects for the group that implemented with rotational use (-0.082σ, p-value = 0.044). For the 9th graders, no significant differences are found, and all estimated coefficients are negative. These findings are consistent with results from Table 5, where we show 9th grades did not have a large exposure to the platform, even in schools with good implementation.

Columns 3-4 of Table 8 present the heterogeneous effects on students' attitudes towards math. In all three panels, standard errors are relatively large, and we cannot reject the null hypothesis that the effects are the same for schools with better and worse implementation (for the pooled sample, p-values equal to 0.948 for the heterogeneity with respect to no infrastructure problems and 0.726 for type of implementation).

Overall, 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 presented in Sections 6.1 and 6.2 (see Duflo et al. (2020) for a discussion on the potential benefits of presenting analyses there 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 experienced infrastructure problems would have had the same expected effect of a school with better infrastructure if it had not have infrastructure problems. For example, it may be that there are other variables, such as motivation of the school principal, that explains both the infrastructure problems 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 (2019) 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 Discussion

Our experimental results point in the direction of a zero overall treatment effect of the platform on math proficiency and a positive effect on students' attitudes towards math. Taking advantage of the fact that control schools also implemented the program in non-evaluated grades, we extrapolate the infrastructure measure and implementation modality to the control grades in those schools and we find suggestive evidence that schools with better implementation had gains in math proficiency, while attitudes towards math was similar in both groups.

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 quality and type of implementation. Figure A.1 shows that infrastructure problems were concentrated in the first months of the experiment. In the last months of the experiment, even the schools labeled as with infrastructure problems reported good use of the platform. Also, students in the rotation implementation, despite having to split one of their weekly classes 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 good and worse implementation) on attitudes towards math. The Khan Academy platform can therefore be seen as an effective way to change attitudes towards math, even in the short-run and with short exposition.

However, even though virtually all students experimented the platform, those schools with inadequate infrastructure or with rotation modality had on average fewer hours of usage. As we showed in Table 7, infrastructure problems can represent a reduction of 30-40% of average usage, and for the rotation modality, a reduction of 40-60%. If there are returns to scale in spending more time in one activity, these math activities are not as effective as standard math classes, and/or there is relevant time wasted in the transition from one activity to the other, then the implementation of the program in these schools may have actually reduced the total amount of math content that these students were exposed to, relative to a setting with no intervention. Moreover, it is conceivable that some classes were wasted trying to connect to the internet without success, which again could have reduced the total amount of math content that these students were exposed to.

Overall, these heterogeneous patterns can be rationalized in a model in which perceptions about math can be affected by exposing students to a more attractive way to present math content, regardless of whether such exposure comes at the expense of a reduction in standard math classes. However, to achieve proficiency gains, the platform may require consistent and longer usages.

These results indicate that further research on the use of Khan Academy is warranted, and provide guidance on how such studies should be implemented. To the extent that improvements in attitudes may eventually lead to improvements in proficiency, a longer-term exposition could lead to improvements in proficiency. Moreover, our results indicate that the type and quality of implementation matter substantially.

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. 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 (for a review of the literature see, for instance, 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 heterogeneity 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.

Given this discussion, we stress that the results we present on the effects of the Khan Academy platform should be viewed as the effects of this platform integrated to math classes, with a specific type and a given quality of implementation. Given the available evidence, we should expect different results if we considered different types of implementation of the Khan Academy platform, or if we considered a setting with better infrastructure. Still, our results help clarify the conditions in which we should expect to find positive effects from these kind of programs, and provide guidance for further research.

7 Conclusion

In this paper, we present novel experimental evidence on the impacts of the Khan Academy platform, through the program *Khan Academy in Schools*, implemented across five cities in three different regions of Brazil. The program aimed at integrating one weekly math class (50 minutes) with a Khan Academy session in the computer lab. 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 there are no infrastructure problems and when the implementation modality is based on one computer per student. However, it may have negative effects in settings with implementation problems, or in which the implementation modality is based on rotation.

The available evidence points out that CAL programs are very beneficial when they are delivered supplementing the traditional school curriculum. As highlighted by Muralidharan et al. (2019) and Mo et al. (2020), mode of delivery is important, and effectiveness of CAL programs may vary depending on whether these are implemented in substitute or supplementary manners, in-school or out-of-school. Evidence on the effectiveness of CAL programs as substitutes for teacher delivered curriculum is limited, and the available evidence is not conclusive. Our results contribute to the debate on this issue. We show that implementation challenges may prevent positive treatment effects from arising and that, when adequately implemented, CAL programs may be effective even when it does not increase the total number of hours of exposure to math content. We stress that we see our results depending on program implementation as suggestive, and that they should induce further research for more conclusive answers. Our conclusion is that details of program implementation matter, and these must be taken into account when considering scaling up CAL programs as an alternative for traditional teaching pedagogy in developing countries.

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.

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.

Barrera-Osorio, F. and Linden, L. L. (2009). *The use and misuse of computers in education: evidence from a randomized experiment in Colombia*. The World Bank.

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.

Beuermann, D. W., Cristia, J., Cueto, S., Malamud, O., and Cruz-Aguayo, Y. (2015). One laptop per child at home: Short-term impacts from a randomized experiment in peru. *American Economic Journal: Applied Economics*, 7(2):53–80.

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.

Carrillo, P. E., Onofa, M., and Ponce, J. (2011). Information technology and student achievement: Evidence from a randomized experiment in ecuador.

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

Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., and Severín, E. (2017). Technology and child development: Evidence from the one laptop per child program. *American Economic Journal: Applied Economics*, 9(3):295–320.

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.

Ferman, B. (2019). A simple way to assess inference methods. *arXiv e-prints*, page arXiv:1912.08772.

Firpo, S., Pieri, R., and Souza, A. P. (2017). Electoral impacts of uncovering public school quality: Evidence from brazilian municipalities. *EconomiA*, 18(1):1–17.

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.

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.

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

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.

Mo, D., Zhang, L., Wang, J., Huang, W., Shi, Y., Boswell, M., and Rozelle, S. (2013). The persistence of gains in learning from computer assisted learning: Evidence from a randomized experiment in rural schools in shaanxi province. Technical report, REAP working paper.

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.

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

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)						
5th grade		540.0			13.407	
		[64.8]			[1.221]	
9th grade		386.3			11.359	
		[34.4]			[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	0.060
s.e.	[0.035]	[0.024]	[0.033]	[0.022]
p-value	(0.513)	(0.515)	(0.090)	(0.008)
Inference assessment	*0.068*	*0.078*	*0.068*	*0.068*
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)
Inference assessment	*0.061*	*0.069*	*0.066*	*0.069*
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)
Inference assessment	*0.084*	*0.087*	*0.071*	*0.092*
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 13, 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 is based on the assessment proposed by Ferman (2019) using 1000 draws of iid normal random variables.

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.

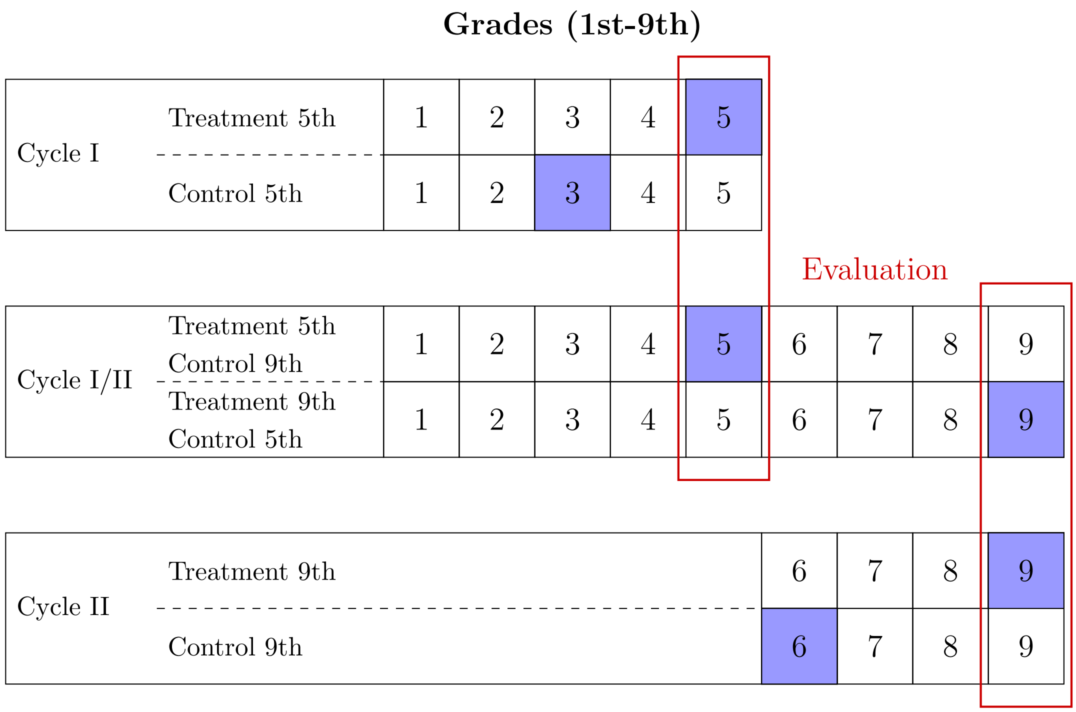
Table 8: ITT Heterogeneity

	Math test score		Attitudes towards math	
	No infrastructure problem	One computer per student	No infrastructure problem	One computer per student
	(1)	(2)	(3)	(4)
Panel A: Full sample				
$T \times X$ (β_1)	0.058	0.081	0.052	0.036
s.e.	[0.048]	[0.052]	[0.049]	[0.047]
p-value	(0.220)	(0.121)	(0.290)	(0.438)
Inference assessment	*0.098*	*0.110*	*0.089*	*0.101*
$T \times (1-X)$ (β_2)	-0.056	-0.076	0.056	0.053
s.e.	[0.040]	[0.032]	[0.035]	[0.020]
p-value	(0.166)	(0.017)	(0.105)	(0.009)
Inference assessment	*0.082*	*0.073*	*0.096*	*0.067*
p-value ($\beta_1 = \beta_2$)	(0.092)	(0.021)	(0.948)	(0.726)
Inference assessment	*0.072*	*0.078*	*0.083*	*0.066*
N	13825	13231	11135	10710
Panel B: 5th grade				
$T \times X$ (β_1)	0.093	0.127	0.066	0.070
s.e.	[0.058]	[0.053]	[0.048]	[0.052]
p-value	(0.110)	(0.016)	(0.167)	(0.179)
Inference assessment	*0.091*	*0.097*	*0.084*	*0.077*
$T \times (1-X)$ (β_2)	-0.062	-0.082	0.039	0.035
s.e.	[0.058]	[0.041]	[0.045]	[0.028]
p-value	(0.287)	(0.044)	(0.385)	(0.207)
Inference assessment	*0.095*	*0.065*	*0.087*	*0.074*
p-value ($\beta_1 = \beta_2$)	(0.085)	(0.005)	(0.717)	(0.531)
Inference assessment	*0.072*	*0.066*	*0.068*	*0.059*
N	9682	9088	7784	7359
Panel C: 9th grade				
$T \times X$ (β_1)	-0.064	-0.102	-0.023	-0.031
s.e.	[0.068]	[0.052]	[0.109]	[0.072]
p-value	(0.350)	(0.048)	(0.830)	(0.661)
Inference assessment	*0.136*	*0.200*	*0.134*	*0.146*
$T \times (1-X)$ (β_2)	-0.009	-0.075	0.076	0.108
s.e.	[0.096]	[0.068]	[0.028]	[0.018]
p-value	(0.926)	(0.271)	(0.007)	(0.000)
Inference assessment	*0.122*	*0.143*	*0.111*	*0.102*
p-value ($\beta_1 = \beta_2$)	(0.693)	(0.781)	(0.437)	(0.085)
Inference assessment	*0.091*	*0.111*	*0.085*	*0.091*
N	4143	4143	3351	3351

Notes: This table reports results for student-level regressions of math proficiency (columns 1-2) and attitudes towards math (columns 3-4) on interaction terms between the treatment dummy and the heterogeneity variable. In columns (1) and (3), X is an indicator variable which takes value one if there were no infrastructure problems; in columns (2) and (4), X is an indicator variable which takes value one if the implementation modality was based on one computer per student. Specifications in columns 1 and 2 include strata fixed effects, the X variable in level, and the covariates reported in Table 1. Specifications in columns 3 and 4 include strata fixed effects, the X variable in level, and the covariates reported in Table 3. Standard errors are clustered at the strata level. The inference assessment is based on the assessment proposed by Ferman (2019) using 1000 draws of iid normal random variables.

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 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	28.936	27.328
	[0.649]	[0.649]	[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	-0.089	11422	Uses computer lab at school during other classes	0.312	0.003	12334
	[1.001]	[0.025]			[0.464]	[0.008]	
Male	0.504	-0.003	12369	Uses computer lab at school not during class	0.132	0.012	12377
	[0.500]	[0.009]			[0.338]	[0.005]	
Year of Birth	2,004.5	-0.184	12381	Has mobile phone	0.713	0.004	12265
	[2.292]	[0.026]			[0.452]	[0.009]	
White	0.314	-0.009	10703	Has internet on mobile phone	0.710	-0.010	11286
	[0.464]	[0.008]			[0.454]	[0.009]	
Black	0.095	0.013	10703	Lives with mother	0.901	-0.016	12362
	[0.294]	[0.006]			[0.299]	[0.007]	
Native	0.040	-0.002	10703	Lives with father	0.636	-0.050	12360
	[0.196]	[0.003]			[0.481]	[0.009]	
Mixed	0.513	-0.005	10703	Has books at home	0.774	-0.029	12394
	[0.500]	[0.009]			[0.419]	[0.010]	
Asian	0.037	0.003	10703	Parents talk about school	0.840	0.007	12394
	[0.190]	[0.004]			[0.366]	[0.008]	
Has computer at home	0.586	-0.042	12396	Works outside home	0.071	0.031	12388
	[0.493]	[0.010]			[0.257]	[0.005]	
Frequently uses computer at home	0.459	-0.030	12380	Has ever repeated a grade	0.216	0.060	12304
	[0.498]	[0.009]			[0.411]	[0.009]	
Has internet at home	0.731	-0.013	12360	Math is the preferred subject	0.433	-0.023	12389
	[0.443]	[0.011]			[0.495]	[0.007]	
Uses computer at home for school activities	0.520	-0.028	12365	Portuguese is the preferred subject	0.243	0.034	12389
	[0.500]	[0.010]			[0.429]	[0.011]	
Uses computer lab at school	0.365	-0.003	12374	Other subject is preferred	0.325	-0.011	12389
	[0.481]	[0.009]			[0.468]	[0.008]	
Uses computer lab at school during portuguese classes	0.251	0.004	12403	Participated in Math Olympics	0.199	0.009	11340
	[0.434]	[0.012]			[0.399]	[0.008]	
Uses computer lab at school during math classes	0.285	-0.009	12368				
	[0.451]	[0.010]					

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: 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.7: 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 [0.074]	0.028 [0.034]	-0.025 [0.072]	0.030 [0.033]	-0.001 [0.094]	0.026 [0.079]
Male	-0.017 [0.015]	-0.008 [0.011]	-0.020 [0.014]	-0.023 [0.013]	-0.012 [0.031]	0.016 [0.018]
Year of Birth	0.1 [0.046]	0.014 [0.045]	0.1 [0.105]	-0.087 [0.047]	0.0 [0.080]	0.180 [0.054]
White	0.006 [0.023]	-0.009 [0.011]	0.019 [0.038]	-0.027 [0.016]	-0.017 [0.025]	0.026 [0.012]
Black	0.000 [0.011]	-0.004 [0.007]	0.007 [0.028]	-0.004 [0.010]	-0.016 [0.028]	-0.005 [0.012]
Native	0.004 [0.005]	-0.001 [0.005]	0.005 [0.008]	0.002 [0.006]	0.002 [0.010]	-0.006 [0.007]
Mixed	-0.018 [0.023]	0.014 [0.015]	-0.042 [0.030]	0.029 [0.022]	0.033 [0.020]	-0.011 [0.016]
Asian	0.007 [0.010]	0.000 [0.005]	0.011 [0.011]	0.001 [0.006]	-0.003 [0.012]	-0.004 [0.009]
Has computer at home	-0.019 [0.019]	-0.006 [0.016]	-0.030 [0.033]	-0.019 [0.020]	-0.008 [0.024]	0.012 [0.037]
Frequently uses computer at home	-0.005 [0.017]	-0.001 [0.012]	-0.014 [0.021]	-0.005 [0.016]	0.005 [0.032]	0.005 [0.025]
Has internet at home	-0.008 [0.009]	-0.008 [0.020]	-0.023 [0.027]	-0.025 [0.026]	0.011 [0.022]	0.022 [0.027]
Uses computer at home for school activities	-0.009 [0.014]	-0.003 [0.014]	-0.035 [0.024]	-0.015 [0.017]	0.021 [0.026]	0.014 [0.034]
Uses computer lab at school	-0.068 [0.099]	0.010 [0.043]	-0.130 [0.178]	0.034 [0.065]	0.054 [0.081]	0.000 [0.047]
Uses computer lab at school during portuguese classes	-0.001 [0.086]	0.018 [0.031]	-0.044 [0.135]	0.031 [0.041]	0.101 [0.073]	0.030 [0.036]
Uses computer lab at school during math classes	0.097 [0.136]	0.026 [0.037]	0.088 [0.107]	0.023 [0.045]	0.144 [0.180]	0.059 [0.057]
Uses computer lab at school during other classes	-0.187 [0.051]	-0.008 [0.029]	-0.176 [0.125]	0.036 [0.036]	-0.175 [0.175]	-0.061 [0.043]
Uses computer lab at school not during class	-0.039 [0.015]	-0.002 [0.014]	-0.063 [0.032]	0.000 [0.014]	-0.004 [0.023]	-0.004 [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.8: 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	

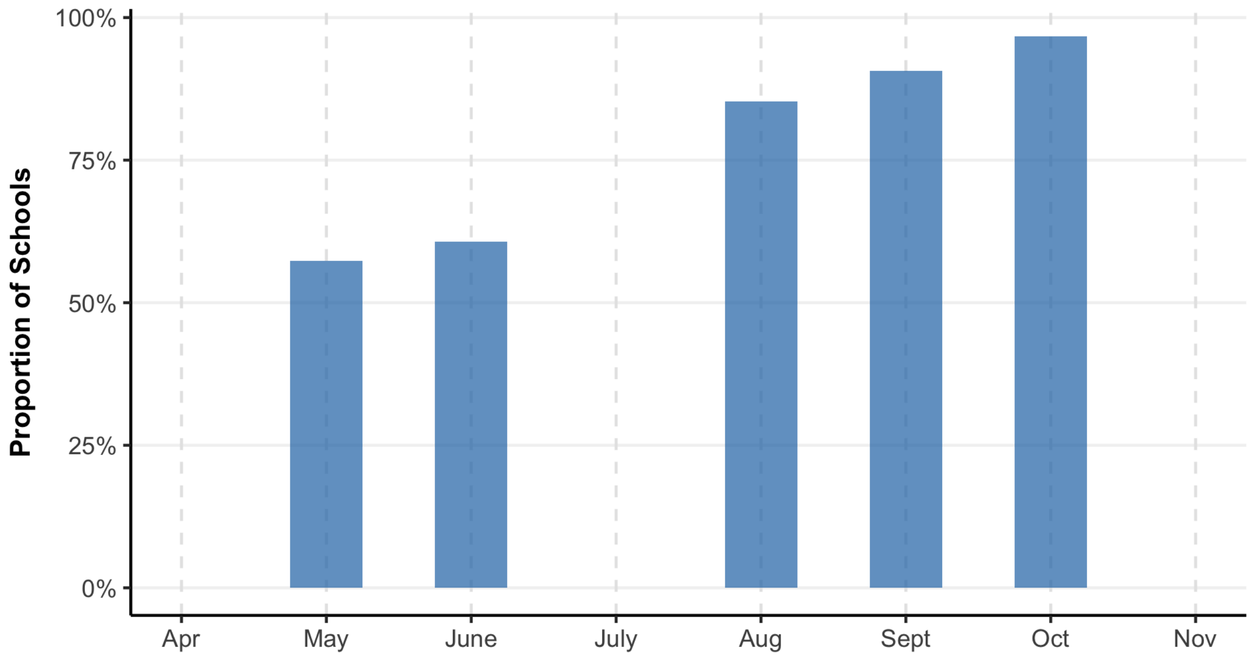
Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the results of student-level regressions of covariates collected in the baseline survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment interacted 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.9: 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 high school	-0.002	0.041	-0.010	0.031	0.030	0.067
	[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 high school	-0.026	0.045	-0.043	0.038	0.032	0.067
	[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	

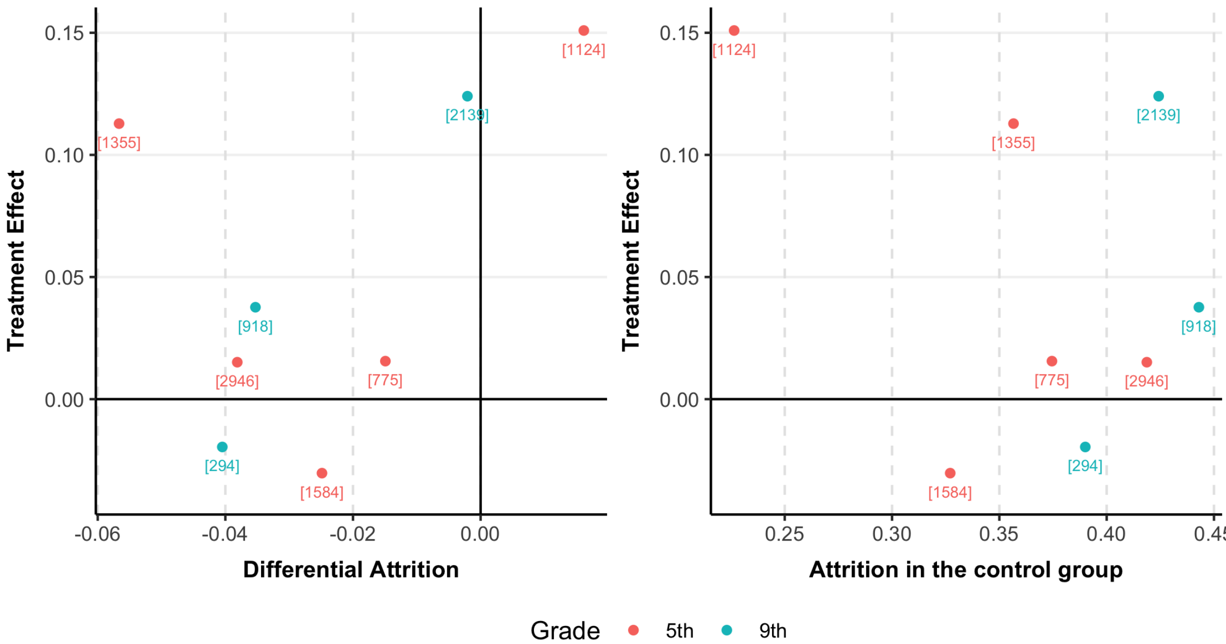
Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the results of student-level regressions of covariates collected in the baseline survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment interacted 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.

Figure A.1: 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.2: 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.