

Are Public Schools Ready to Integrate Math Classes with Khan Academy? *

Bruno Ferman[†]

Lucas Finamor[‡]

Lycia Lima[§]

First Draft: June 20th, 2019

This Draft: June 27th, 2019

Abstract

We study the impacts of the program *Khan Academy in Schools* 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 Khan Academy platform, instead of their standard math classes. We find positive effects of the program on measures of attitudes towards math, which were not translated to a positive average treatment effect on students' math proficiency. We also explore treatment heterogeneity by quality of implementation, suggesting that the program can have positive effects 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.

*We would like to thank the staff from the Lemann Foundation and from Khan Academy for their support to the project. This project was pre-registered at AEA RCT Registry under the number AEARCTR-0002456. An initial draft of this paper was part of Lycia Lima's PhD dissertation at the Sao Paulo School of Business Administration - FGV.

[†]Sao Paulo School of Economics - FGV, bruno.ferman@fgv.br

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

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

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; World-Bank, 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 softwares 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).

One of the most popular online platforms focused on delivering educational content tailored at each students' level is the Khan Academy, which offers free instructional videos and personalized exercises both in math as well as in other subject areas, ranging from kindergarten to college levels. Khan Academy stands out for its worldwide popularity, having reached 71 million of 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. In this paper, we present the findings of the first randomized evaluation of the *Khan Academy in Schools* program, an effort to promote the use of the Khan Academy platform in Brazilian public schools.

The program was implemented in Brazil as a partnership between Khan Academy and the nonprofit Lemann Foundation, and its main feature was to integrate the Portuguese version of Khan Academy platform into math classes, once a week, in the the schools' computer lab. Our study measures the impacts of the intervention on math proficiency and attitudes towards math based on 5th and 9th grade students from 157 schools located across three different regions of Brazil. We analyze both average treatment effects, and also perform an exercise to estimate the heterogeneous effect of the program based on whether schools faced technology infrastructure challenges to program's implementation and whether they adopted the implementation modality based on an individual or rotational use of the computer during class.

We first show that students in treated grades report to use Khan Academy in math classes and that this increase did not crowd out the use of computer lab by other subjects.

In terms of outcomes, our findings show that *Khan Academy in Schools* had positive effects on measures of students' attitudes towards math, which were not translated to a positive average treatment effect on math proficiency, measured in a standardized national exam. However, we find suggestive evidence that such null effect on students' test scores hides 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. We are able to carry out this comparison by leveraging the design of the experiment which delivered one treated grade at every participant school.

While this paper is the first randomized evaluation of an effort to integrate the Khan Academy into regular class hours, there has been a series of studies investigating the effects of technology-enhanced instruction interventions in developing countries on learning outcomes. A review by Glewwe and Muralidharan (2016) shows the results are largely varied, with estimates ranging from significantly negative to significantly positive magnitudes. The available evidence suggests the characteristics of the computer-aided learning (*henceforth* CAL) interventions are an important factor to explain the heterogeneity of findings. Positive effects on learning are registered in studies mostly focused on programs that complement traditional teaching with CAL activities, such as Muralidharan et al. (2019), Banerjee et al. (2007), Lai et al. (2015), Linden (2008), Yang et al. (2013) and Mo et al. (2013). One common feature among all of these programs is that they increase the number of hours students are exposed to academic instruction.

However, when we consider the performance of CAL as an alternative for regular teaching, pulling students out of traditional class for classes that integrate CAL sessions, the limited available evidence presents mixed findings (Banerjee et al., 2007; Linden, 2008; Carrillo et al., 2011). In this context, our results are directly relevant as they shed light on potential reasons for the diverging results found in the literature on the effectiveness of CAL as a substitute for standard math classes. We find that details of program implementation are determinant for the performance of CAL programs as an alternative for traditional teaching pedagogy in developing countries. 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 discusses the results and section 6 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, originally created for the United States, 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.

Funded by volunteer contributions and partnerships with private sector foundations, the non profit initiative has greatly expanded over the years and currently reaches millions of students in over 190 countries. Khan Academy resources are available in 36 languages, and there are versions of the website in Spanish, French and Brazilian Portuguese. 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 2017.¹

The platform may enhance students' math performance through three main channels. First, it may increase the quality of math content accessed by students by offering quality material developed by specialists. The second potential channel is by increasing students' learning through offering content and exercises tailored to each students' level, addressing students' heterogeneity within class. A third channel through which the platform may have an impact on a students' performance is by shifting the students' perceptions regarding math, turning the studying experience more attractive. By presenting the math content in an interactive and friendly way, designed to promote a fun and exciting learning experience, the platform may change the students' attitudes towards math, which may be ultimately translated into an increased math performance.

The Lemann Foundation has promoted the use of Khan Academy in Brazilian public schools through the program *Khan Academy in Schools*.² 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

¹According to information reported on the Lemann Foundation's website <https://fundacaolemann.org.br/materiais/khan-academy-in-brazil>

²"*Khan Academy nas Escolas*", later renamed to "*Innovation in Schools*" or "*Inovação nas escolas*"

2017 edition of the program, which we evaluate in this paper, 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. 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. 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. 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.

3 Study Design

3.1 Sample Selection

This experiment was conducted in primary public schools of five cities in three different regions of Brazil for 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.

3.2 Experimental design

The study took place in 157 primary education schools.³ 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).

Schools were initially stratified based on four criteria: i) the municipality they were in; ii) type of school in terms of the grades they offered (cycle I, cycle II or both cycles); iii) whether they had ever received the Khan Academy program in the years preceding the experiment;⁴ and iv) 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, conditional on data availability.

Every school in our sample was assigned at least one treatment and one control grade, with the purpose of increasing engagement and reducing attrition. This study is based on students from the 5th and 9th grades, since for these grades there is a national standardized exam every two years and math proficiency data would be available for the 2017 academic year. For Cycle I schools, 3rd (or 4th) and 5th grades were eligible to receive the program, and we randomized treatment in the 5th grade. Schools assigned as controls in the 5th grade automatically received treatment in the 3rd or 4th grade. Similarly, for Cycle II schools, 6th and 9th grades were eligible, and treatment in the 9th grade was randomly assigned. For schools assigned 9th grade as control, the 6th grade received the intervention. Schools with both cycles had only the 5th and 9th grades eligible, and similar procedure was followed. Randomization allocated which grade would receive treatment.

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

³There 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.

⁴In 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 5.1 we check whether control students were even exposed to the platform.

study, our sample is composed of a total of 217 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 survey data collected 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). We collected data for an instrument that measured students' attitudes towards mathematics (Brito, 1998). 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*").⁵ 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.⁶ We also collected data on students' demographic characteristics, students' self reported access and usage of computer and internet both at home and at school as well as their preference in relation to school subjects. On the follow-up survey, information on the knowledge and usage of Khan Academy was also collected to assess program compliance and contamination in the control group. Survey data is not available for 7 out of the 157 schools, which left the study after treatment assignment.

Our second data source is 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. Data on teachers' characteristics is also collected, including age

⁵This measure was originally developed by Aiken Jr and Dreger (1961) and translated and validated to Portuguese by Brito (1998). See the original papers for the full list of questions.

⁶An 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 paper.

and educational level. Although this exam is implemented in all public schools in Brazil with more than 20 enrolled students, the Ministry of Education only releases proficiency data for those school grades that had at least 80 percent of enrolled students taking the test. We have administrative data for all schools in our sample (including those that left the study after treatment assignment), with the exception of those school grades that did not meet the minimum attendance requirement. Unfortunately, we are not able to link individual level administrative data with survey data because the SAEB dataset is de-identified.

We also use information extracted from the Khan Academy platform on the usage of Khan Academy 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 Survey

Table 1 presents survey student level baseline characteristics for the pooled sample and for the samples of the 5th and 9th grades separately. For each group, the table displays 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, with standard errors clustered at the school level. 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 0.695, 0.453 and 0.720 respectively for the three samples considered).

There are two potential sources of attrition in the survey, school-level and student-level attrition. Our first source of attrition is associated with schools that left the program after treatment assignment. Seven schools out of our sample of 157 schools - both in treatment and control groups - left the study after randomization took place for various reasons, mostly unrelated with treatment assignment. The small number of school dropouts and the different reasons associated with the withdraw minimize our concerns with differential selective attrition. Two out of seven schools left the program after randomization and previously to the 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. Student-level attrition in the survey 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 Table 2 we show attrition results for our different measures of attrition. 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.⁷ In Panel A, we show that survey attrition rate (attrition defined by the absence of data on attitudes towards math) was relatively high, at almost 40% for the pooled sample in the control group. High survey attrition is relatively common in studies that collect data in Brazilian public schools at the end of school year, as it is not atypical for school attendance in Brazil to drop significantly during the last month of classes. Attrition in treatment group is 2.5 percentage points lower than that in the control group (p -value=0.027). In Appendix Table A.1, however, 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. This suggests that the significant differences in attrition rates are unlikely to generate differential selective attrition that could threaten the validity of our results.

4.2.2 SAEB data

Table 3 shows covariates are also balanced for characteristics reported in the SAEB data set, confirming there are no significant differences between treatment arms in none 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 B 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. There are no significant differences in attrition rates between treatment and control groups for the math proficiency outcome, for the pooled sample, and for the 5th and 9th grades separately. The results show that the intervention is not correlated with the likelihood of the schools having SAEB data reported. In Panel C, we

⁷The dependent variable is an indicator whether there is no outcome data available.

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).

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.⁸

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 5.1 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:

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

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

⁸There was some variation on the start date of the intervention in the different cities. Pelotas, Barueri and Mogi had 24 weeks of exposure, while Sao Bernardo had 16 weeks and Manaus had 20 weeks

grade), and separately for each grade. Standard errors are clustered at the school level.

In this paper, we consider two main outcomes: math proficiency and attitudes towards math.⁹ 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 Results

5.1 Program Implementation and Compliance with Experimental Design

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

Our results show that around 97% of the students in treated grades report using Khan Academy (around 82% report using it in school). In the control group, only 6.3% of the students report using the platform (4.4% report using in school), so contamination does not raise major concerns. Considering the 5th and 9th grades separately, we observe that proportion of students reporting use of Khan Academy is slightly lower for the 5th grade (96% in the 5th grade as opposed to 98% in the 9th grade).

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

⁹Math proficiency and attitudes towards math were the main outcomes registered in the paper’s pre-analysis plan. AEA RCT Registry: AEARCTR-0002456.

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 (-0.055pp) 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.

While virtually all treated students were exposed to platform, many schools experienced some implementation problems during the program. 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. In about 31% of those visits, they reported that the implementation was inadequate. In 71% of those cases with 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. 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. 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.¹⁰ In only 1% of the cases, more than one student shared the same computer. Teachers were advised not to let that happen, because this would undermine the effectiveness of one of Khan Academy’s main feature, which is its adaptive learning nature that tailors the content according to each student’s needs.

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 600 minutes of use for the duration of the study, roughly 25 minutes per week, while in the modality of one computer per

¹⁰There 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.

student the expectation was for students to have approximately twice this exposure. In columns 1 to 3 of Table 5, we show how the total number of minutes logged in the platform correlates with infrastructure problems and with the type of implementation. In schools that implemented the program with rotation and had infrastructure problems, 5th graders spent 540 minutes logged in the platform from April to October.¹¹ 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. Interestingly, even in schools with one computer per student, the total number of minutes for 9th graders is still only about the same as the total number of minutes for 5th graders in schools with infrastructure problems and rotation. 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.

5.2 Treatment Effects on Main Outcomes

Table 6 shows intent to treat estimates of the program on math proficiency (columns 1-2) and attitudes towards math (columns 3-4) for the pooled sample (Panel A), and for the 5th and 9th grades separately (Panels B and C). The first column for each outcome omits the covariates from the regression 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.

Our results also 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). Our initial hypothesis was that one of the channels through which the program could foster math proficiency was by improving the students' math learning

¹¹We 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.

experience. This hypothesis was based on the assumption that, by learning math in a more exciting and interactive manner, students would have better attitudes regarding math, potentially paying more attention on the exposed content or even spending longer hours studying it, which could ultimately impact proficiency. While we confirm that the intervention has a positive impact on the attitudes towards math, the effects were very small, and our findings suggest the modest gains in attitudes were not translated into higher math proficiency on average.

There are a few factors that may have prevented positive average treatment effects from arising. First, one important aspect to note about the intervention is that, although it exposes students to a potentially more engaging learning experience, it does so by integrating Khan Academy into one of the weekly math classes, so students' total exposure to traditional methods of teaching is reduced. Also, the Khan Academy class was carried out at the schools' computer lab, and there is anecdotal evidence that a significant proportion of class time was wasted moving the students to a different location. Second, the implementation of the program faced some challenges, as 51% of the schools reported infrastructure problems in at least one month of implementation. Lastly, the different types of implementation (individual *vs* rotational use of the computer) may have played an important role. Our data shows that implementation was based in rotation in 59% of the treated schools in the 5th grade and in 55% of the schools treated in the 9th grade. Overall, it may be that students' total hours of exposure to math materials remained constant or even decreased.

5.3 Treatment Heterogeneity

If the null effect we estimated for students' test scores comes from infrastructure problems and/or from a implementation modality based on rotation of students, then we should expect to find positive effects in schools that had a better implementation. 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 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 3rd, 4th, or 9th grades in these schools) to this information for 5th grade treated schools (so this information comes from implementation in the 5th grade). 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. 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 zero for all 14 schools in this group. 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, providing further evidence that the type of implementation should not be dependent on the grade that received the program in a given school.

Table 8 presents the results for the 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 show that the integration with Khan Academy may be an effective alternative to traditional curriculum if adequately implemented. Students assigned in treated grades that did not face infrastructure problems had higher math scores (0.058σ), although not significant ($p\text{-value}=0.158$), but significant gains were registered when the modality of implementation was one computer per student (0.081σ). On the other hand, students assigned to grades that implemented the rotational modality of the program performed worse in the SAEB exam (-0.076σ), which may not be surprising if this type of implementation ultimately leads to a reduction in students' total math

exposure.

Our results 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) and for students assigned to the individual use of the computer modality (0.127σ). In the 5th grade, negative effects on math scores were registered for students in the poorer implementation group, but only statistically significant for the group that implemented with rotational use (-0.082σ). 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 0.478 for the heterogeneity with respect to no infrastructure problems and 0.723 for type of implementation).

It is possible to rationalize the heterogeneous effects on students' math proficiency and the (lack of) heterogeneous effects on attitudes towards math if we consider that virtually all treated students were exposed to the platform, regardless of the quality and type of implementation. However, students in the rotation implementation had to split one of their weekly classes between studying in the platform and doing other math activities. 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, students in schools with the rotation system spent significantly less time in the platform. Likewise, students in schools with infrastructure problems were also exposed to the platform. However, they spent significantly less time in the platform relative to schools with no infrastructure problems. 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. Therefore, 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. Moreover, the extensive margin with respect to exposure to the platform may be more relevant in shaping such views about math relative to the

intensive margin of usage. This may explain the lack of heterogeneous effects on attitudes towards math. When we consider the effects on students' math proficiency, however, then this reduction in standard math classes and/or the intensive margin of exposure to the platform may be more relevant, so we find heterogeneous effects depending on the quality and type of implementation.

5.4 Discussion

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)).

When we consider the evidence on CAL programs as substitutes for standard math classes, our results help rationalize the mixed evidence found in the literature. We show that the quality and type of implementation are important determinants of whether such programs should have positive or negative effects. Importantly, since in this case the impact of the program depends on the net effectiveness of the CAL program relative to a standard math class, it is possible that the impact of the program is negative when the implementation is inadequate. In our study, we show that this can be the case when students have to rotate between the CAL activity and other math activities, and when infrastructure problems in the school prevents a more extensive usage of the platform. In contrast, CAL programs implemented as complements should be less likely to generate negative results when there are implementation problems.

Overall, these results point out that the external validity of experimental results on CAL programs should be considered with caution. 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.

6 Conclusion and Policy Implications

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 the program does not have an impact on average over students' math scores, although we find small but significant effects on attitudes towards math. We also explore treatment heterogeneity by quality of implementation, showing that the program has positive effects when there are no infrastructure problems and when the implementation modality is based on one computer per student. However, it can have negative effects in settings with implementation problems, or in which the implementation modality is based on rotation.

The available evidence points out that computer assisted learning (CAL) programs are very beneficial when they are delivered supplementing the traditional school curriculum. As highlighted by Muralidharan et al. (2019), 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. Our conclusion is that details of program implementation matter, and these must be taken into account when considering scaling up of CAL programs as an alternative for traditional teaching pedagogy in developing countries.

References

- 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.
- Brito, M. R. F. (1998). Adaptação e validação de uma escala de atitudes em relação à matemática). *Zetetiké: Revista de Educação Matemática*, 6(9).
- 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.
- 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.
- 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.
- Linden, L. L. (2008). *Complement or substitute?: The effect of technology on student achievement in India*. InfoDev Working Paper, Columbia University.
- 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.
- 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.

Figures and Tables

Table 1: 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 [1.000]	0.004 [0.030]	11422	0.000 [1.000]	-0.007 [0.035]	7203	0.000 [1.000]	0.024 [0.059]	4219
Male	0.505 [0.500]	-0.005 [0.009]	12369	0.513 [0.500]	-0.015 [0.010]	7871	0.488 [0.500]	0.012 [0.016]	4498
Year of Birth	2004.614 [2.298]	-0.010 [0.027]	12381	2005.911 [1.396]	-0.053 [0.040]	7872	2001.820 [1.013]	0.066 [0.043]	4509
White	0.327 [0.469]	-0.014 [0.009]	10703	0.364 [0.481]	-0.028 [0.014]	6540	0.256 [0.437]	0.008 [0.015]	4163
Black	0.107 [0.309]	-0.013 [0.006]	10703	0.111 [0.314]	-0.010 [0.008]	6540	0.100 [0.300]	-0.017 [0.012]	4163
Indian	0.038 [0.192]	0.002 [0.004]	10703	0.041 [0.198]	0.004 [0.005]	6540	0.033 [0.180]	0.000 [0.006]	4163
Mixed	0.488 [0.500]	0.026 [0.010]	10703	0.450 [0.498]	0.034 [0.015]	6540	0.563 [0.496]	0.012 [0.014]	4163
Asian	0.039 [0.194]	-0.001 [0.004]	10703	0.034 [0.182]	0.001 [0.005]	6540	0.048 [0.214]	-0.004 [0.007]	4163
Has computer at home	0.580 [0.494]	-0.007 [0.010]	12396	0.572 [0.495]	-0.014 [0.016]	7892	0.596 [0.491]	0.005 [0.018]	4504
Frequently uses computer at home	0.455 [0.498]	-0.003 [0.009]	12380	0.454 [0.498]	-0.007 [0.013]	7884	0.457 [0.498]	0.006 [0.018]	4496
Has internet at home	0.736 [0.441]	-0.008 [0.012]	12360	0.741 [0.438]	-0.022 [0.016]	7867	0.726 [0.446]	0.017 [0.018]	4493
Uses computer at home for school activities	0.520 [0.500]	-0.006 [0.010]	12365	0.518 [0.500]	-0.018 [0.014]	7872	0.526 [0.499]	0.016 [0.017]	4493
Uses computer lab at school	0.367 [0.482]	-0.011 [0.033]	12374	0.419 [0.493]	-0.013 [0.044]	7879	0.255 [0.436]	-0.008 [0.050]	4495
Uses computer lab at school during portuguese classes	0.237 [0.426]	0.023 [0.031]	12403	0.290 [0.454]	0.019 [0.043]	7896	0.123 [0.329]	0.031 [0.052]	4507
Uses computer lab at school during math classes	0.255 [0.436]	0.048 [0.033]	12368	0.318 [0.466]	0.035 [0.041]	7873	0.119 [0.323]	0.071 [0.050]	4495
Uses computer lab at school during other classes	0.332 [0.471]	-0.052 [0.029]	12334	0.335 [0.472]	-0.018 [0.037]	7852	0.327 [0.469]	-0.112 [0.061]	4482
Uses computer lab at school not during class	0.144 [0.351]	-0.013 [0.009]	12377	0.148 [0.355]	-0.018 [0.011]	7878	0.135 [0.342]	-0.005 [0.022]	4499

(cont)

Table 1 Cont. - Baseline Covariates Balance - Survey

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
<i>(cont)</i>									
Has mobile phone	0.715 [0.452]	-0.001 [0.010]	12265	0.683 [0.466]	0.000 [0.014]	7808	0.783 [0.412]	-0.001 [0.014]	4457
Has internet on mobile phone	0.706 [0.455]	-0.003 [0.010]	11286	0.680 [0.467]	-0.004 [0.015]	6925	0.759 [0.428]	-0.003 [0.015]	4361
Lives with mother	0.893 [0.309]	0.005 [0.005]	12362	0.902 [0.298]	0.007 [0.006]	7864	0.874 [0.332]	0.001 [0.009]	4498
Lives with father	0.617 [0.486]	0.003 [0.009]	12360	0.640 [0.480]	-0.002 [0.012]	7861	0.569 [0.495]	0.013 [0.014]	4499
Has books at home	0.767 [0.422]	-0.009 [0.010]	12394	0.740 [0.439]	-0.021 [0.014]	7890	0.826 [0.379]	0.013 [0.013]	4504
Parents talk about school	0.844 [0.363]	-0.001 [0.007]	12394	0.867 [0.339]	-0.012 [0.008]	7891	0.795 [0.404]	0.019 [0.012]	4503
Works outside home	0.082 [0.274]	0.000 [0.005]	12388	0.080 [0.272]	-0.004 [0.007]	7882	0.084 [0.278]	0.008 [0.008]	4506
Has ever repeated a grade	0.238 [0.426]	-0.006 [0.010]	12304	0.186 [0.389]	0.011 [0.013]	7830	0.349 [0.477]	-0.036 [0.016]	4474
Math is the preferred subject	0.428 [0.495]	0.008 [0.013]	12389	0.506 [0.500]	0.007 [0.015]	7894	0.260 [0.439]	0.009 [0.023]	4495
Portuguese is the preferred subject	0.249 [0.432]	0.008 [0.013]	12389	0.267 [0.443]	0.007 [0.014]	7894	0.208 [0.406]	0.010 [0.024]	4495
Other subject is preferred	0.323 [0.468]	-0.016 [0.012]	12389	0.226 [0.418]	-0.014 [0.012]	7894	0.532 [0.499]	-0.018 [0.027]	4495
Participated in Math Olympics	0.192 [0.394]	0.000 [0.009]	11340	0.074 [0.262]	0.005 [0.010]	7192	0.444 [0.497]	-0.009 [0.021]	4148
P value joint	0.695			.453			.720		

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 school 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 Clusters	Mean (control)	Diff	N Obs.	N Clusters	Mean (control)	Diff	N Obs.	N Clusters
Panel A: Student-level Attrition in the Survey											
0.393	-0.025 [0.011] (0.027)	18065	150	0.377	-0.030 [0.015] (0.048)	12220	136	0.433	-0.015 [0.027] (0.585)	5845	72
Panel B: School-grade-level Attrition in the SAEB exam											
0.142	-0.008 [0.038] (0.830)	217	157	0.099	-0.002 [0.045] (0.964)	143	143	0.229	-0.020 [0.089] (0.819)	74	74
Panel C: Student-level Attrition in the SAEB exam											
0.132	0.005 [0.007] (0.482)	17151	143	0.123	0.006 [0.008] (0.468)	11906	129	0.156	0.002 [0.011] (0.849)	5245	58

Notes: This table reports differences in attrition between treatment and control groups in the follow-up survey (Panel A) and in the SAEB exam (school-grade-level in Panel B and student-level in Panel C). 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 school level. P-values are in parenthesis.

Table 3: 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 [0.500]	-0.008 [0.009]	14411	0.512 [0.500]	-0.010 [0.010]	10072	0.485 [0.500]	-0.001 [0.019]	4339
White	0.283 [0.450]	-0.009 [0.009]	14423	0.293 [0.455]	-0.013 [0.013]	10047	0.255 [0.436]	0.002 [0.015]	4376
Black	0.073 [0.261]	-0.005 [0.005]	14423	0.070 [0.255]	-0.007 [0.006]	10047	0.082 [0.274]	0.000 [0.008]	4376
Mixed	0.527 [0.499]	0.007 [0.010]	14423	0.517 [0.500]	0.015 [0.014]	10047	0.551 [0.497]	-0.014 [0.018]	4376
Asian	0.028 [0.166]	0.004 [0.003]	14423	0.023 [0.151]	0.002 [0.003]	10047	0.041 [0.198]	0.007 [0.007]	4376
Indigenous	0.025 [0.157]	-0.001 [0.002]	14423	0.025 [0.157]	0.000 [0.003]	10047	0.026 [0.158]	-0.001 [0.004]	4376
Race not declared	0.064 [0.244]	0.004 [0.005]	14423	0.071 [0.257]	0.004 [0.007]	10047	0.045 [0.207]	0.006 [0.006]	4376
Age	12.007 [2.087]	-0.005 [0.018]	14625	10.821 [0.795]	0.018 [0.020]	10220	15.099 [0.916]	-0.063 [0.044]	4405
Mother has completed at least high school	0.625 [0.484]	0.025 [0.013]	9606	0.636 [0.481]	0.019 [0.021]	6034	0.606 [0.489]	0.037 [0.023]	3572
Mother literate	0.985 [0.120]	-0.002 [0.002]	14564	0.989 [0.106]	-0.005 [0.002]	10173	0.976 [0.152]	0.006 [0.004]	4391
Father has completed at least high school	0.571 [0.495]	0.017 [0.013]	8006	0.565 [0.496]	0.007 [0.020]	4990	0.582 [0.493]	0.034 [0.023]	3016
Father literate	0.958 [0.201]	0.001 [0.004]	14373	0.962 [0.192]	0.001 [0.004]	10007	0.948 [0.222]	0.001 [0.008]	4366
Teacher younger than 50 years old	0.760 [0.427]	0.008 [0.045]	12805	0.761 [0.426]	0.012 [0.049]	10530	0.752 [0.432]	-0.017 [0.134]	2275
2015 Prova Brasil math grade	0.095 [1.023]	0.029 [0.063]	16820	0.090 [0.934]	-0.066 [0.091]	11654	0.107 [1.216]	0.266 [0.203]	5166
P value joint	.679			.470			.865		

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 school level are in brackets. P-values for a test that all covariates are balanced are reported at the bottom of the table for each of the three samples considered.

Table 4: Follow-up Survey

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
Has computer at home	0.622 [0.485]	-0.012 [0.013] (0.357)	12816	0.631 [0.483]	-0.013 [0.017] (0.422)	9004	0.595 [0.491]	-0.008 [0.024] (0.723)	3812
Frequently uses computer at home	0.472 [0.499]	0.015 [0.011] (0.182)	12808	0.484 [0.500]	0.020 [0.015] (0.193)	9004	0.438 [0.496]	0.004 [0.020] (0.855)	3804
Has internet at home	0.795 [0.404]	-0.002 [0.011] (0.865)	12745	0.804 [0.397]	-0.002 [0.014] (0.911)	8953	0.770 [0.421]	-0.002 [0.018] (0.893)	3792
Uses computer at home for school activities	0.519 [0.500]	0.004 [0.013] (0.763)	12764	0.526 [0.499]	0.001 [0.018] (0.953)	8962	0.502 [0.500]	0.011 [0.021] (0.593)	3802
Uses computer lab at school	0.488 [0.500]	0.285 [0.033] (0.000)	12820	0.555 [0.497]	0.192 [0.036] (0.000)	9010	0.300 [0.458]	0.513 [0.049] (0.000)	3810
Uses computer lab at school during portuguese classes	0.317 [0.465]	-0.039 [0.029] (0.179)	12801	0.370 [0.483]	-0.057 [0.039] (0.143)	8994	0.167 [0.373]	0.003 [0.042] (0.944)	3807
Uses computer lab at school during math classes	0.340 [0.474]	0.445 [0.035] (0.000)	12743	0.398 [0.490]	0.330 [0.039] (0.000)	8951	0.175 [0.380]	0.728 [0.042] (0.000)	3792
Uses computer lab at school during other classes	0.368 [0.482]	-0.055 [0.026] (0.035)	12703	0.386 [0.487]	-0.066 [0.030] (0.028)	8923	0.316 [0.465]	-0.027 [0.060] (0.647)	3780
Uses computer lab at school not during class	0.151 [0.358]	0.051 [0.014] (0.000)	12791	0.140 [0.347]	0.037 [0.014] (0.008)	8985	0.181 [0.385]	0.084 [0.038] (0.025)	3806
Uses Khan Academy	0.063 [0.244]	0.903 [0.017] (0.000)	12673	0.078 [0.268]	0.882 [0.023] (0.000)	8924	0.022 [0.145]	0.956 [0.005] (0.000)	3749
Uses Khan Academy during during	0.044 [0.204]	0.782 [0.022] (0.000)	12549	0.055 [0.228]	0.707 [0.028] (0.000)	8833	0.010 [0.100]	0.967 [0.005] (0.000)	3716

Notes: This table reports, for the pooled, 5h grade and 9th grades samples separately: i) the control group mean, ii) the results of a student-level regression of different measures collected in the follow-up survey on a dummy variable indicating whether student belongs to a grade-level that was randomly assigned to receive treatment and strata fixed effects and iii) Number of observations. Standard errors are in brackets and p-values in parenthesis. Standard errors are clustered at the school 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.	[58.0]	[70.3]	[58.4]	[1.330]	[1.498]	[1.475]
p-value	(0.011)	(0.016)	(0.754)	(0.030)	(0.008)	(0.110)
One computer per student	195.0	224.2	183.9	1.669	2.082	1.741
s.e.	[73.3]	[91.0]	[98.0]	[1.402]	[1.586]	[2.190]
p-value	(0.008)	(0.014)	(0.061)	(0.234)	(0.189)	(0.426)
9th grade	-178.3	-	-	-3.206	-	-
	[45.6]			[0.787]		
	(0.000)			(0.000)		
Municipality fixed effects	Y	Y	Y	Y	Y	Y
Mean (with infrastructure problem and rotation)						
5th grade		540.0			13.407	
		[61.2]			[1.154]	
9th grade		386.3			11.359	
		[36.5]			[0.824]	
Number of students	8302	5325	2977	8302	5325	2977
Number of schools	103	65	38	103	65	38

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 school 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.	[0.033]	[0.028]	[0.028]	[0.020]
p-value	(0.490)	(0.566)	(0.043)	(0.003)
N	14846	14846	11157	11157
Panel B: 5th grade				
Treatment	-0.036	-0.002	0.044	0.062
s.e.	[0.049]	[0.037]	[0.031]	[0.024]
p-value	(0.462)	(0.954)	(0.159)	(0.010)
N	10388	10388	7806	7806
Panel C: 9th grade				
Treatment	0.011	-0.051	0.086	0.057
s.e.	[0.075]	[0.034]	[0.049]	[0.033]
p-value	(0.883)	(0.129)	(0.080)	(0.086)
N	4458	4458	3351	3351
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 3, while the specifications reported in column 3 include the covariates presented in Table 1. Standard errors are clustered at the school level.

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 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*X	0.058 [0.041] (0.158)	0.081 [0.047] (0.084)	0.029 [0.043] (0.500)	0.036 [0.041] (0.377)
T*(1-X)	-0.056 [0.039] (0.153)	-0.076 [0.034] (0.028)	0.072 [0.040] (0.071)	0.053 [0.023] (0.022)
p-value	0.047	0.011	0.478	0.723
N	13825	13231	11135	10710
Panel B: 5th grade				
T*X	0.093 [0.051] (0.067)	0.127 [0.059] (0.032)	0.066 [0.034] (0.052)	0.070 [0.045] (0.122)
T*(1-X)	-0.062 [0.054] (0.253)	-0.082 [0.046] (0.074)	0.039 [0.040] (0.329)	0.035 [0.030] (0.243)
p-value	0.037	0.009	0.619	0.516
N	9682	9088	7784	7359
Panel C: 9th grade				
T*X	-0.064 [0.060] (0.291)	-0.102 [0.065] (0.115)	-0.023 [0.079] (0.767)	-0.031 [0.073] (0.666)
T*(1-X)	-0.009 [0.075] (0.904)	-0.075 [0.048] (0.116)	0.076 [0.034] (0.025)	0.108 [0.031] (0.000)
p-value	0.606	0.743	0.263	0.076
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 3. Specifications in columns 3 and 4 include strata fixed effects, the X variable in level, and the covariates reported in Table 1. Standard errors are clustered at the school level.

Appendix A Appendix Tables

Table A.1: Balance conditional on non-attriters

	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 [0.031]	7243	0.049 [1.006]	-0.023 [0.037]	4688	-0.012 [0.998]	0.071 [0.059]	2555
Male	0.502 [0.500]	0.004 [0.011]	7761	0.501 [0.500]	-0.001 [0.013]	5056	0.504 [0.500]	0.014 [0.017]	2705
Year of Birth	2004.723 [2.232]	-0.022 [0.032]	7764	2005.995 [1.266]	-0.083 [0.043]	5054	2001.894 [0.949]	0.093 [0.047]	2710
White	0.336 [0.472]	-0.015 [0.011]	6692	0.369 [0.483]	-0.017 [0.015]	4194	0.269 [0.444]	-0.011 [0.016]	2498
Black	0.099 [0.299]	-0.005 [0.007]	6692	0.104 [0.305]	-0.009 [0.009]	4194	0.090 [0.286]	0.001 [0.011]	2498
Indian	0.040 [0.196]	0.003 [0.005]	6692	0.043 [0.203]	0.004 [0.006]	4194	0.034 [0.181]	0.002 [0.007]	2498
Mixed	0.486 [0.500]	0.019 [0.012]	6692	0.447 [0.497]	0.021 [0.016]	4194	0.563 [0.496]	0.014 [0.016]	2498
Asian	0.039 [0.194]	-0.002 [0.005]	6692	0.037 [0.188]	0.001 [0.005]	4194	0.044 [0.205]	-0.007 [0.007]	2498
Has computer at home	0.602 [0.490]	-0.016 [0.013]	7772	0.597 [0.491]	-0.020 [0.017]	5065	0.613 [0.487]	-0.009 [0.024]	2707
Frequently uses computer at home	0.468 [0.499]	-0.004 [0.012]	7765	0.465 [0.499]	-0.003 [0.015]	5062	0.476 [0.500]	-0.008 [0.023]	2703
Has internet at home	0.740 [0.439]	-0.005 [0.013]	7749	0.751 [0.433]	-0.024 [0.017]	5049	0.716 [0.451]	0.030 [0.021]	2700
Uses computer at home for school activities	0.531 [0.499]	-0.010 [0.013]	7750	0.528 [0.499]	-0.018 [0.016]	5050	0.539 [0.499]	0.005 [0.024]	2700
Uses computer lab at school	0.372 [0.483]	-0.016 [0.031]	7751	0.419 [0.494]	-0.005 [0.046]	5051	0.266 [0.442]	-0.035 [0.045]	2700
Uses computer lab at school during portuguese classes	0.245 [0.430]	0.010 [0.034]	7773	0.301 [0.459]	0.002 [0.046]	5065	0.122 [0.328]	0.024 [0.055]	2708
Uses computer lab at school during math classes	0.263 [0.440]	0.047 [0.034]	7758	0.333 [0.471]	0.035 [0.043]	5055	0.108 [0.311]	0.070 [0.054]	2703
Uses computer lab at school during other classes	0.337 [0.473]	-0.055 [0.029]	7732	0.337 [0.473]	-0.020 [0.038]	5039	0.337 [0.473]	-0.123 [0.061]	2693
Uses computer lab at school not during class	0.138 [0.345]	-0.014 [0.010]	7760	0.142 [0.349]	-0.021 [0.012]	5057	0.130 [0.337]	-0.002 [0.024]	2703

(cont)

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

	Pooled Sample			5th grade			9th grade		
	Mean (control)	Diff	N	Mean (control)	Diff	N	Mean (control)	Diff	N
	<i>(cont)</i>								
Has mobile phone	0.711 [0.454]	0.010 [0.012]	7699	0.680 [0.467]	0.008 [0.016]	5018	0.779 [0.415]	0.013 [0.016]	2681
Has internet on mobile phone	0.710 [0.454]	0.007 [0.012]	7026	0.689 [0.463]	0.004 [0.017]	4401	0.752 [0.432]	0.013 [0.016]	2625
Lives with mother	0.902 [0.297]	0.001 [0.006]	7752	0.908 [0.289]	0.007 [0.008]	5048	0.888 [0.315]	-0.010 [0.012]	2704
Lives with father	0.639 [0.480]	0.001 [0.012]	7748	0.658 [0.474]	-0.010 [0.015]	5047	0.595 [0.491]	0.021 [0.020]	2701
Has books at home	0.777 [0.416]	-0.009 [0.010]	7771	0.748 [0.434]	-0.013 [0.014]	5064	0.841 [0.366]	0.000 [0.015]	2707
Parents talk about school	0.837 [0.370]	0.009 [0.009]	7772	0.859 [0.348]	-0.002 [0.010]	5066	0.787 [0.410]	0.030 [0.016]	2706
Works outside home	0.067 [0.251]	0.004 [0.005]	7772	0.064 [0.245]	0.006 [0.007]	5063	0.075 [0.263]	0.000 [0.010]	2709
Has ever repeated a grade	0.211 [0.408]	-0.001 [0.011]	7724	0.163 [0.369]	0.011 [0.014]	5033	0.319 [0.466]	-0.025 [0.021]	2691
Math is the preferred subject	0.440 [0.496]	0.007 [0.015]	7769	0.521 [0.500]	0.003 [0.017]	5064	0.260 [0.439]	0.015 [0.026]	2705
Portuguese is the preferred subject	0.238 [0.426]	-0.001 [0.014]	7769	0.250 [0.433]	0.002 [0.014]	5064	0.212 [0.409]	-0.007 [0.026]	2705
Other subject is preferred	0.321 [0.467]	-0.006 [0.014]	7769	0.229 [0.420]	-0.005 [0.015]	5064	0.528 [0.499]	-0.008 [0.028]	2705
Participated in Math Olympics	0.182 [0.386]	-0.001 [0.010]	7086	0.063 [0.243]	0.007 [0.012]	4606	0.446 [0.497]	-0.018 [0.024]	2480
P value joint	0.845			0.801			0.578		

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-attriters, individuals for which there is follow-up data available. Standard errors clustered at the school level are in brackets. 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. Standard errors clustered at the school level are presented in brackets. P-values are presented in parenthesis.

Table A.2: Number of Students Enrolled per Classroom

	Cycle I schools	Cycle II schools	Two cycle schools
	(1)	(2)	(3)
3rd grade	0,526 [0,429] (0,220)		
4th grade	-0,588 [0,450] (0,192)		
6th grade		2,357 [1,474] (0,110)	
9th grade			0,190 [0,743] (0,799)
Mean (omitted group)	28,936 [0,649]	28,936 [0,649]	27,328 [0,949]
Omitted group	5th grade	9th grade	5th grade
Number of schools	78	14	58

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