Stanford | King Center on Global Development

Isolating the "Tech" from EdTech: Experimental Evidence on Computer Assisted Learning in China

Yue Ma

Robert Fairlie

Prashant Loyalka

Scott Rozelle

March, 2020

Working Paper No. 1065

Isolating the "Tech" from EdTech: Experimental Evidence on Computer Assisted Learning in China

Yue Ma, Robert Fairlie, Prashant Loyalka, Scott Rozelle[1]

March 19, 2020

Abstract

EdTech which includes online education, computer assisted learning (CAL), and remote instruction was expanding rapidly even before the current full-scale substitution for in-person learning at all levels of education around the world because of the coronavirus pandemic. Studies of CAL interventions have consistently found large positive effects, bolstering arguments for the widespread use of EdTech. However CAL programs, often held after school, provide not only computer-based instruction, but often additional non-technology based inputs such as more time on learning and instructional support by facilitators. In this paper, we develop a theoretical model to carefully explore the possible channels by which CAL programs might affect academic outcomes among schoolchildren. We isolate and test the technology-based effects of CAL and additional parameters from the theoretical model, by designing a novel multi-treatment field experiment with more than four thousand schoolchildren in rural China. Although we find evidence of positive overall CAL program effects on academic outcomes, when we isolate the technology-based effect of CAL (over and above traditional pencil-and-paper learning) we generally find small to null effects. Our empirical results suggest that, at times, the "Tech" in EdTech may have relatively small effects on academic outcomes, which has important implications for the continued, rapid expansion of technologies such as CAL throughout the world.

Keywords: Computer-assisted learning, EdTech, ICT, pencil effects, student learning, educational productivity, RCT

JEL Codes: I21, O15

[1] Ma: Stanford University (email: yma@stanford.edu); Fairlie: University of California, Santa Cruz and NBER (email: rfairlie@ucsc.edu); Loyalka: Stanford University (email: loyalka@stanford.edu); Rozelle: Stanford University (email: rozelle@stanford.edu); We would like to thank students at the Center for Experimental Economics in Education (CEEE) at Shaanxi Normal University for exceptional project support as well as Dell Global Giving and the TELOS Initiative at the GSE at Stanford for financing the project. We thank David Card, Thomas Dee, Mark Duggan, Caroline Hoxby, Natalia Lazzatti, Hongbin Li, Bryan Pratt, Jon Robinson, Sean Sylvia, and seminar participants at UC Berkeley, Stanford, UC Merced, and Montana State University for comments and suggestions.

1 Introduction

Computer assisted learning (CAL), online courses, massive open online courses (MOOCs), and other forms of educational technology (EdTech) are revolutionizing the way in which students are educated. Billions of dollars are spent each year in the United States on software for K-12 students, and the global EdTech industry was projected to grow to more than $250 billion by 2020 and $340 billion by 2025 (Escueta et al. 2017). With the large-scale, comprehensive movement of schoolchildren and college students in China, Europe, the United States and most other countries around the world to online platforms in response to the coronavirus outbreak (COVID-19), actual expenditures on EdTech will be substantially higher. Although EdTech is rapidly being deployed throughout the developed and developing world, the relatively limited evidence on whether and how it affects academic outcomes is mixed (Bulman and Fairlie 2016).

Bolstering arguments for the continuing expansion of EdTech in developing countries, however, recent evaluations of supplemental learning CAL programs across a wide range of software types have consistently found large positive effects on academic outcomes (Lai et al. 2013; 2015; Mo et al. 2014; Muralidharan et al. 2019; Bohmer, Burns, and Crowley 2014).[2] In these studies, the effects are identified through randomized experiments in which the treatment group that receives supplemental CAL (and in some cases additional inputs) is compared to a control group that receives no inputs. Although often attributed to the EdTech component, the estimated effects of the supplemental use of CAL, however, include other non-technology based

[2] Earlier studies of computer-based programs find large positive effects (Banerjee et al. 2007; Linden 2008), but the evidence is not always clear (Rouse and Krueger 2004; Rockoff 2015). For the less common use of computer-assisted learning as a direct substitute for regular teacher instruction in the classroom the evidence tends to show null effects (Dynarski et al. 2007, Campuzano et al. 2009; Linden 2008; Barrow et al. 2009; Carillo et al. 2011) but this might depend on how computers are used (Falck, Mang, and Woessmann 2018). Finally, the less structured provision of computers and laptops for home and/or school use among schoolchildren tends to show null effects (e.g. Malamud and Pop-Eleches 2011; Fairlie and Robinson 2013; Beuermann et al. 2015; Cristia et al. 2017; Malamud et al. 2019; Hull 2019). For recent reviews, see Glewwe et al. (2013), Bulman and Fairlie (2016), and Escueta et al. (2017).

inputs in educational production. These inputs include more time learning academic material, additional instructional support by facilitators, more attention to students, and potential crowd out effects on homework time. Thus, a fundamental question for making decisions over investments of educational resources is whether the "Tech" in EdTech is driving the supplemental educational program effects or whether another input is driving the effects. This represents a more general problem in evaluating and interpreting the impacts of any supplemental education intervention program (e.g. after-school tutoring and community technology centers) because those programs also consist of additional inputs such as more learning time.

In this paper, we first create a theoretical model to carefully explore the possible channels by which supplemental CAL might affect academic outcomes among schoolchildren. We then estimate several key parameters from the theoretical model including the technology-based effect of CAL. To generate exogenous variation in CAL and other inputs, we design and conduct a randomized controlled trial (RCT) involving more than four thousand 4th to 6th grade students across 352 math classes in 130 schools in rural China. Of the 185 million schoolchildren in China roughly 75 percent live in rural areas (Chen et al. 2015; UNESCO 2020). The RCT includes three treatment arms: i) supplemental CAL, ii) traditional supplemental learning (i.e. solving problems using pencil and paper workbooks), and iii) a pure control that receives no supplemental learning. The traditional supplemental learning sessions were designed to have identical content and duration as the supplemental CAL sessions so that they could be used to isolate the technology effects of CAL from the overall program effects. To further isolate effects, the RCT was also designed so that the CAL treatment does not provide any additional inputs and is not part of a larger program.[3]

[3] A recent study (Johnson et al. 2018) evaluates a pilot program that involves training teaching assistants to deliver a structured package of literacy materials to groups of 3-4 young children in England. They cross-randomize the TA

The theoretical model demonstrates that there are multiple channels by which a supplemental CAL program can effect educational outcomes, but does not provide a prediction on whether the isolated technology component of the program improves or worsens educational outcomes. Estimates from the field experiment indicate that, for the average student in rural China, the supplemental CAL program increases students' math grades However, when we isolate the technology-based effects of CAL we find point estimates that are small and statistically indistinguishable from zero. The differential between the estimated overall CAL program effect and the estimated CAL technology specific effect is also statistically significant. We find no evidence of positive effects of the CAL Program for math test scores, however, suggesting the program improves non-cognitive more than cognitive skills.

Given well-documented gender differences in computer use and achievement (Eble and Hu 2019; Xu and Li 2018; Algan and Fortin 2018; Hannum and Park 2007), we examine effects for boys and girls separately. Focusing on boys, we find that the CAL program increases math grades by 3.4 percentile points and math test scores by 0.10σ. Isolating the technology-effects from CAL, however, we find point estimates for the CAL technology effect that are notably smaller and statistically indistinguishable from zero. For girls, we do not find positive estimates of the CAL program effect or the isolated CAL technology effect.

Turning to other measures, we find no evidence of time substitution effects from the CAL and workbook sessions: neither type of session crowds out homework time in math. On the other hand, the CAL program and isolated CAL technology effect increase how much students like their math class.

---

teaching with ICT or paper equivalent sessions and find positive effects for both (slightly larger for non-ICT). However, the emphasis on the teaching assistant intervention, small group assignment, and implementation of this program in the classroom make it difficult to isolate the technology-based effects of the ICT session.

In addition to contributing to the literature on whether and how CAL programs work, the findings from our experiment provide novel evidence on whether the isolated technology component of CAL improves academic outcomes. To our knowledge, only one previous study addresses this question.[4] Muralidharan et al. (2019) find large positive effects of after-school Mindspark Center programs in India which include both extensive software use and instructional support. To rule out the effects of the instructional support and extra learning time inputs of the program they compare their impact estimates to those from of an after-school private tutoring program that did not include a technology component but was conducted in the same location and student age group, and for more time (Berry and Mukherjee 2016). The comparison program has no impacts on student outcomes suggesting that additional instructional time and tutoring were not the key drivers of the Mindspark impacts (Muralidharan et al. 2019). Building on the use of this comparison, our experiment provides the first estimate in the literature directly identifying the technology-based effects of CAL on educational outcomes. It is the first experiment to use a second comparison group to remove additional inputs such as more time learning, instructional support from teachers and aides in the sessions, more attention to students, and crowd out of homework time. Isolating the technology-based effects is fundamental to understanding how CAL works, and whether the "Tech" in EdTech positively affects educational outcomes.

Our paper also contributes to the broader literature on the effects of computer technology in education and the labor market by providing a new "Pencil Test." The seminal paper by DiNardo and Pischke (1997) found that workers who use pencils at work experience a wage premium that

[4] The general finding of null effects when CAL substitutes for regular teacher instruction in the classroom provides some indirect evidence on the question (Dynarski et al. 2007, Campuzano et al. 2009; Linden 2008; Barrow et al. 2009; Carillo et al. 2011). However, the implementation of these programs within the classroom and the substitution for several factors (not just learning time) such as teacher lecture time, in-class discussions and small group work make it difficult to isolate the technology-related effects of CAL.

is similar to computer users. "Pencil skills" are not scarce, however, and cannot have a large return in the labor market, raising the concern that the large estimated returns to computer skills in previous studies were due to unobserved worker and job characteristics. Several recent studies evaluating CAL programs rule out concerns about unobserved heterogeneity among students, parents and schools by using RCTs, but ignore a related threat to interpretation. In a similar vein, comparing CAL to pencil and paper workbook estimates provides evidence that the large positive estimates of the effects of CAL programs commonly found in previous studies might be at least partly due to other inputs such as more time devoted to learning material, which could have also been achieved with a pencil and workbook. Taken together, the findings have particular relevance to the questions of whether technology has a distinct advantage in improving student outcomes, and what advantages and disadvantages it has over traditional “pencil and paper” forms of learning.

The remainder of the paper is organized as follows. In Section 2, we create a theoretical model to illustrate the channels by which CAL might affect academic outcomes among schoolchildren relative to more traditional, “pencil and paper” forms of learning. Section 3 describes the design and implementation of the experiment. Section 4 presents our main results and reports estimates of the structural parameters of the theoretical model. Section 5 concludes.

2 Theoretical Model of Investment in EdTech

A theoretical model illustrates the channels by which CAL might affect academic outcomes among schoolchildren. The broad question of interest is whether parents, students and schools are choosing optimal levels of technology inputs for education given constraints on financial resources, information, and in-school and after-school time allocated to learning. Can academic achievement be improved by investing in additional technology use? The answer to this question necessarily involves a trade-off between inputs. Investment in technology likely offsets investment in traditional resources. During after-school hours, the question of whether time spent using computer-based learning offsets traditional learning is especially salient because of the flexibility of this time.

To illustrate these points, we start by adding computer resources such as CAL to a standard model of education production.[5] In the context of after-school education production by students, the binding constraints in such a model are the amount of after-school time available for learning and the budget for parental or school resources for after-school learning. The focus of the model is on how CAL investment affects various math time inputs, but we also discuss the theoretical implications of how CAL programs, more generally, might provide additional instructional support by teachers or aides and more attention to students during sessions. We consider a value-added model of education and focus on academic performance in math.[6]

$$A_i = f\left(X_i, S_i, T_i^C, T_i^M\right) \; s.t. \tag{2.1}$$

$$T_i^M = T_i^{TR} + T_i^C$$

$$T_i^M + T_i^{Oth} \leq T$$

$$P^{TR} T_i^{TR} + P^C T_i^C \leq B_i,$$

A measure of academic performance in math, A_i, is assumed to depend on the characteristics of a student and his or her family (including prior academic performance), X_i, school and teacher characteristics, S_i, total time allocated to learning math, T_i^M, and time allocated to learning math on the computer, T_i^C. Time allocated to learning math on the computer is essentially entered twice to allow for a direct technology effect and a separate time learning math

[5] See Hanushek (1979, 1986); Rivkin, Hanushek, and Kain (2005); Figlio (1999) for examples.
[6] See Hanushek (1979) for an early discussion of value-added models in the economics of education literature.

effect. Total time allocated to learning math consists of traditional learning, T_i^{TR}, and CAL, T_i^C. The amount of time spent on learning math is constrained by total available after-school learning time T which includes time spent after-school on all other activities, T_i^{Oth}. Investments in traditional and CAL are subject to costs (P^{TR} and P^C) and per student budget B_i for after-school learning expenditures on math.

If students, parents and schools do not make optimal choices of CAL, possibly due to not having access to technology, or other resource and information constraints, then an exogenous reallocation toward CAL could be positive. On the other hand, if students, parents, and schools already optimally allocate time then an exogenous reallocation toward CAL and away from other more productive forms of learning will result in a negative or zero effect on math performance.[7] From Equation (2.1) the total marginal effect of CAL on academic achievement is:

$$\frac{dA}{dT^C} = \frac{\delta A}{\delta T^C} + \frac{\delta A}{\delta T^M}\frac{\delta T^M}{\delta T^C} \tag{2.2}$$

The total effect is comprised of a direct effect of increasing CAL time on math and an indirect effect through increasing total time spent learning math.

CAL might have a direct or "technology" effect on academic achievement independent of more time on math (i.e. $\frac{\delta A}{\delta T^C} \neq 0$). CAL is video-based, and often game-based, and thus might be more engaging than traditional learning. The game-based features of educational software might increase learning interest as well as learning performance (Ebner and Holzinger, 2007; Burguillo, 2010). CAL might also provide faster feedback on problems compared to the feedback associated with traditional modes of learning (Van der Kleij et al., 2015). On the other hand, the game-based nature of CAL might reduce interest in completing traditional homework or learning in class and

[7] Parents and students might limit time on computers for after-school learning because of concerns over distraction, safety, and other issues.

hence decrease achievement. In addition, solving math problems on a computer instead of writing them down on paper with a pencil could commit them less to memory (Vincent, 2016). The net technology effect of these potentially offsetting mechanisms is theoretically ambiguous.

Using the total time on math constraint in (2.1) we can rewrite the total marginal effect of CAL on math achievement.

$$\frac{dA}{dT^C} = \frac{\delta A}{\delta T^C} + \frac{\delta A}{\delta T^M}\left(1 + \frac{\delta T^{TR}}{\delta T^C}\right) \tag{2.3}$$

Here we can view the indirect effect (the second term in equation 2.3) as having two parts. The first part is the effect of a one-to-one increase in math time by increasing CAL time. As discussed in detail below, this part of the effect of introducing educational technology is important and often overlooked in the previous literature. Introducing CAL in a subject implicitly increases time spent learning that subject. The second part of the indirect effect of CAL captures the possibility of crowd out (or crowd in) of traditional learning in math. CAL might displace some of the time a student normally devotes to traditional forms of learning such as homework or independent studying because of the overall time constraint (i.e. $\frac{\delta T^{TR}}{\delta T^C} < 0$). Crowd out of homework time might result because of the time constraint and/or the student viewing traditional learning as less fun or engaging compared to learning math on the computer (which is often game-based). Working in the opposite direction, however, there could be crowd in where CAL might increase a student's interest and confidence in math and ultimately increase independent time studying math.

To make the theoretical model more tractable we approximate with a linear education production function. We modify (2.1) and (2.3) as:

$$A_i = \beta X_i + \gamma S_i + \theta T_i^C + \lambda T_i^M \tag{2.1'}$$

$$\frac{dA}{dT^C} = \theta + \lambda(1 + \eta). \tag{2.3'}$$

Arguably, the parameter of most interest is the direct or technology-based effect of CAL on academic performance, θ. It captures how CAL affects achievement stripped of any mechanical effects through increased hours learning math or any crowd out or crowd in effects on traditional forms of learning math. Policymakers, however, might not be as concerned about removing crowd out or crowd in effects, but want to know the net reduced form effect that captures the relative returns to different investments in math learning. In this case, the budget constraint and relative prices would also play an important role. The crowd out or crowd in parameter, is also of interest because it provides a sense of the behavioral response to different technology investment policies. For example, a better understanding of whether investing in CAL, that makes traditional forms of learning less interesting (crowd out) or builds confidence (crowd in), is useful.

The total effect of implementing a CAL program, captured by (2.3’), captures everything: time learning math on the computer, total time learning math, and potential crowd out effects on homework. Although the focus here has been on time effects, as discussed in more detail below, implementing a CAL program often includes additional inputs such as providing learning new material outside the standard curriculum, additional instructional support by teachers or aides running the sessions, and more attention to schoolchildren in sessions.

2.2 Interpreting Estimates from Previous CAL Program Evaluations

As noted above, several previous studies estimate the effects of supplemental CAL on academic outcomes and find large positive effects. For example, Lai et al. (2013; 2015) and Mo et al. (2014) find large positive effects of supplemental CAL programs for Chinese schoolchildren (0.12 to 0.18σ in math) from 40 minutes of instruction, 2 times a week. Muralidharan et al. (2019) find large positive intent-to-treat effects of after-school Mindspark Center programs in India which include software use and instructional support (0.37σ in math and 0.23σ in Hindi) from 90 minutes per session, six sessions a week. Bohmer, Burns, and Crowley (2014) find large positive effects from an after-school program providing CAL and student coaches in South Africa (0.25σ in math) for 90 minutes, biweekly. These studies essentially estimate (2.3') without identifying the technology parameter, θ. Because (2.3') includes the total effects from an increase in T^M in addition to an increase in T^C, it favors finding positive effects on academic outcomes. Outside of the theoretical model, many of the CAL programs evaluated in the previous literature include additional educational inputs such as coaches and tutoring sessions which further complicate the interpretation of CAL effects on academic outcomes.

Muralidharan et al. (2019) recognize this concern and note that the impact estimates from the after-school Mindspark program intervention include a combination of the computer program, group-based instruction, and extra instructional time. Although their experiment is not designed to distinguish between these different channels of impact, they address the concern by comparing their CAL program estimates to estimates from a contemporaneous experimental study of the impacts of an after-school tutoring program in the same location and with similarly-aged students (Berry and Mukherjee, 2016). Although the tutoring program was longer (3 hours per session, six days per week instead of 1.5 hours per session, six days per week) no impacts on academic outcomes were found. Muralidharan et al. (2019) note that the null impacts from this program suggest that the additional instructional time or group-based tutoring on their own may have had limited impacts without the CAL program provided at the Mindspark Centers.

The markedly different findings from evaluations of the two programs is important. We build on these results by directly estimating the technology parameter, θ, in (2.3') using our experiment. Specifically, we make use of two treatment arms and a control group to isolate the effects of the different inputs.

3 Estimation, Experimental Design, and Data

Estimating the parameters from the theoretical model is complicated for two primary reasons. First, academic performance and CAL use is likely to be correlated with unobservables leading to biased estimates, especially if there is positive selection bias. Second, the multicollinearity of total math time and CAL time makes it difficult to identify the separate effects of math time and CAL time on academic performance. To address both concerns we designed and implemented a field experiment in which students are randomly assigned to either a control group, a treatment group that receives supplemental CAL sessions, or a treatment group that receives supplemental traditional workbook sessions. As discussed in more detail below, the supplemental traditional workbook sessions were designed to provide similar content, time learning math, and other characteristics as the CAL sessions. Production of math achievement in the control, CAL treatment, and workbook treatment groups can be represented, respectively, by:

$$A_i^0 = \beta X_i + \gamma S_i + \lambda T_i^{TR\_0} \tag{3.1}$$

$$A_i^{CAL} = \beta X_i + \gamma S_i + \theta + \lambda\left(T_i^{TR\_0} + 1 + \eta^C\right) \tag{3.2}$$

$$A_i^{WK} = \beta X_i + \gamma S_i + \lambda\left(T_i^{TR\_0} + 1 + \eta^{WK}\right), \tag{3.3}$$

where $T_i^{TR\_0}$ is the base or control level of traditional homework time, and η^C and η^{WK} are the potential crowd out (or crowd in) responses of math homework time to CAL and workbook sessions, respectively. To normalize time units and simplify the notation, the CAL treatment sets

$T_i^C = 1$ and the extra time allocated to learning math to 1. The workbook treatment, which is of the same duration of time, also sets the extra time allocated to learning math to 1.

The parameters of these three equations can be recovered by using adjusted means and the following two equations:

$$\theta = \bar{A}^{CAL} - \bar{A}^{WK} - \lambda(\eta^C - \eta^{WK}) \tag{3.4}$$

$$\lambda = \frac{\bar{A}^{WK} - \bar{A}^0}{1 + \eta^{WK}} \tag{3.5}$$

$\eta^C - \eta^{WK}$ can be estimated from the difference in total hours learning math between the CAL treatment group and the workbook treatment group, η^C can be estimated from the difference in total hours learning math between the CAL treatment group and the control group, and η^{WK} can be estimated from the difference in total hours learning math between the workbook treatment group and the control group. If these hours substitution effects are small then we are essentially identifying θ from the CAL-workbook difference, and λ from the workbook-control difference. The RCT, in expectation, holds constant student, teacher, and school characteristics and the base or control level of traditional learning time on math, $T_i^{TR\_0}$.

We estimate the parameters of the theoretical model represented in Equations (2.11), (2.12) and (2.13) using the following regression equation:

$$Y_{ij} = \alpha_0 + \alpha_1(-D_{1j}) + \alpha_2(-D_{2j}) + X_{ij}\beta + S_{ij}\gamma + \tau_c + \varepsilon_{ij} \tag{3.6}$$

where Y_{ij} is the academic outcome of interest measured at endline for student *i* in school *j*; D_{1j} is a dummy variable indicating the treatment assignment for the control condition of class *j*, D_{2j} is a dummy variable indicating the class treatment assignment for the workbook condition of class *j*; X_{ij} is a vector of baseline student control variables, S_{ij} is a vector of baseline teacher and classroom control variables, and τ_c is a set of county-grade (strata) fixed effects. Both the control and workbook dummy variables are entered with negative signs to capture relative differences with the CAL treatment (which is the left-out condition in the equation). In this case $\alpha_1$captures the CAL-control difference, which is the overall program effect or the "CAL program" effect, and α_2 captures the CAL-workbook difference, which is the isolated technology-based effect of CAL or the "CAL technology" effect. In all specifications, X_{ij} includes the baseline value of the dependent variable (when available). We also estimate treatment effects with an expanded set of baseline controls including student age, gender, whether each parent finished junior high school or not, teacher gender, teacher experience, whether the teacher attended college, number of boarding students in the class, and total class size. In all regressions, we adjust standard errors for clustering at the class level.

3.2 Experimental Design

We designed the field experiment to generate exogenous variation in both supplemental CAL as well as supplemental traditional learning with the purpose of estimating the parameters from the theoretical model. The field experiment involves more than four thousand 4^{th} to 6^{th} grade students across 352 school-grades (with one math class per school-grade) in 130 schools in rural China. The RCT includes three treatment arms: a supplemental CAL arm, a supplemental traditional learning (pencil and paper workbook) arm, and a pure control arm. The supplemental learning offered by the first two treatment arms is identical in terms of content and duration.

The experiment was conducted among rural primary schools in Northwest China (Shaanxi Province). Specifically, 130 schools from 9 impoverished counties were sampled to participate in the experiment. In each school, we randomly sampled one fourth, fifth, and sixth grade class that had at least 4 boarding students. All students in the sampled classes were surveyed, but the experimental sample includes only boarding students. Altogether, we sampled and surveyed 4,024 boarding students and their 352 math class teachers.[8] We focused on boarding students because there was no time at which to provide after-school CAL and workbook sessions to non-boarding students. Boarding students represent 37 percent of students in our schools. There are 32 million primary and junior high boarding students in China representing 32 percent of all students (Ministry of Education 2017).

The experiment took place in four stages. First, in October 2017 (near the start of the school year), we conducted a baseline survey of students, teachers and principals. Second, after we collected the baseline data, we randomized the 352 classes into the three different treatment conditions. Third, we began conducting the interventions with boarding students in the treated classes in the first half of November 2017. Fourth, in June 2018 (at the end of the school year), we returned to the same classes to conduct a follow up (or endline) survey.

3.3 Baseline Survey

The baseline survey collected information on students, teachers, and school principals. The student survey collected information about student and household characteristics (age, gender, father completed junior high school (yes/no), mother completed junior high school (yes/no), the degree to which they liked math class). Students also took a 35 minute standardized exam in math. The teacher survey collected information on teacher gender, experience, and college attendance. Finally, we collected data on the number of boarding students in the class and class size.

3.4 Randomization

[8] We find that boarding and non-boarding students are similar across numerous characteristics.

We designated each of 27 county-grades (9 counties and 3 grades) in our sample as strata or blocks. We then randomly allocated classes within these strata to one of three different treatment conditions (T1 = Supplemental CAL, T2 = Supplemental Workbook, or C = Control). There were 116 classes in 88 schools for supplemental CAL (T1), 118 classes in 86 schools for supplemental workbook (T2), and 118 classes in 85 schools for the control group.

To ensure adequate sample sizes, power calculations were conducted before the beginning of the trial (Spybrook et al., 2009).[9] We expected to lose a small amount of statistical power due to student attrition. Based on our experience, we assumed an attrition rate of 5%. The actual attrition rate from baseline to endline was only 2.4%.

3.5 Program (Treatment) Administration

The CAL and workbook programs were implemented by a university-based NGO in western China that specializes in after-school programs. Program sessions were held once a week from October 2017 to June 2018. Sessions were held for 40-minutes on Sunday afternoon each week. In the weekly sessions students were asked to complete math exercises taken from the (same) chapter of the standardized math textbook that students were supposed to cover (according to the national curriculum) in class each week. The programs had facilitators who were trained by our research team to organize and supervise the supplemental learning time. The facilitators were instructed to not provide instruction to the students, but rather to make sure that students stayed on task in terms of doing supplemental exercises particular to the week. Facilitators were only allowed to assist students with scheduling, computer hardware issues, software operations, and handing out and

[9] We conservatively used the following parameters to estimate the sample size for the study: (a) intraclass correlation coefficient (adjusted for strata fixed effects): 0.10; (b) average number of boarding students per class: 11; (c) R-squared of 0.40 (controlling e.g. for baseline math achievement). With alpha = 0.05 and beta = 0.8, we estimated that we would need 115 classes per treatment arm for a minimum detectable effect size (MDES) of 0.14 SDs.

collecting workbooks. They were instructed to not answer questions regarding the material. According to our observations, there was little instruction-based communication during the CAL sessions. The facilitators were, for the most part, not the regular math teachers for the students.

The CAL and workbook programs were designed to be as similar as possible. For example, the two programs were implemented by the same NGO, they were held on the same days of the week, for the same amount of time, and number of times during the school year. They also had the same curricular content each week, and same facilitator training and instructions. At endline we found roughly similar attendance rates for CAL and workbook sessions for the two treatment groups.

The software is used in schools in China nationwide. Similar to most CAL software for this age group, the software relied on vivid images and was gamified (see Appendix Figure 1). If students answered an exercise correctly, they received virtual coins with which they could buy virtual gear and outfits. When students did a problem incorrectly they would receive feedback that it was incorrect and solutions if they got stuck. Instead of using computers, students assigned to workbook sessions completed pencil-and-paper math exercises. As with any standard workbook, students could check solutions for the odd-numbered exercises at the back of the workbook.

The supplemental CAL and workbook content was aligned with the standardized, government-mandated curricula for each grade. Unlike less developed countries, students in rural China are much more likely to be on-grade level in terms of achievement outcomes (Li et al. 2018). Students and teachers are rarely absent from class, and students are taught a standardized curriculum at a regular pace.

The CAL program that we evaluate in this study demonstrated positive effects on a range of educational outcomes in previous studies in China (Lai et al. 2013; 2015; Mo et al. 2014, 2015).

Students receiving supplemental CAL increased math and language test scores, the degree to which they liked school, self-efficacy, and interest in learning (Lai et al. 2015; Bai et al. 2018). Although it is difficult to compare estimates across studies, effect sizes for supplemental CAL in China are similar to the effects sizes for supplemental CAL in India once the lower levels of time commitment are factored in.[10] The large positive effects in India may also be due to the CAL program's adaptive component, which is critical for when students are below grade level and have widely differing levels of preparation (Muralidharan et al. 2019). In countries where student preparation is stronger, more regimented and more homogenous, such as China, regular CAL appears to work well for supplemental learning (Mo et al., 2014; 2015). Additionally, in another setting with more homogenous student preparation, Van Klaveren et al. (2017) who conduct an RCT in Dutch secondary schools do not find significant differences between adaptive vs. non-adaptive software and even find negative relative effects for higher ability students (0.08σ). Software differs along many dimensions, however, and thus some caution is needed in generalizing the results to different applications, but the software we evaluate here has been shown to work and is widely used in China.

3.6 Endline Survey and Primary Outcomes

We conducted the endline survey with the students, teachers, and principals. As in the baseline, students took a 35 minute standardized math exam.[11] In the analyses, we convert endline math

[10] The CAL program effect sizes ranged from 0.1 to 0.2σ for sessions of from 40 minutes of instruction, 2 times a week in rural China and the CAL program effect sizes were 0.37σ from 90 minutes per session (for two subjects), six sessions a week in India. The after-school program includes 45 minutes of software use and 45 minutes of instructional support on two subjects each of six days per week. As noted above, Muralidharan et al. (2018) cite evidence in Berry and Mukherjee (2016) showing no effects of a private tutoring program in India (run by Pratham) with similar aged children and time frame suggesting that the effects are primarily driven by CAL.

[11] Like the baseline test, the endline math test was grade-appropriate, tailored to the national and provincial-level mathematics curricula. Although grade-appropriate tests may present a problem in some developing countries (since student learning is, on average, below grade level), this was not the case in our sample schools. Our baseline and

exam scores into z-scores by subtracting the mean endline math score of the control sample and dividing by the standard deviation of the control sample. We also asked math teachers to provide each student's math grades (as distinct from math test/exam/achievement scores). Separate from test scores, grades capture other cognitive and non-cognitive dimensions of human capital and are predictive of later life outcomes (Borghans et al. 2016). [12] Furthermore in the context of China, grades are operationalized as the teacher's independent evaluation of a student's within-class ranking in overall ability and are a less lumpy measure than letter grades provided in the United States. For the analyses, we convert math grade ranks into percentiles based on class size. The correlation between math test scores and math grades is 0.529. Although randomization was at the class level, boarding students represent only a fraction of the students in the class thus providing variation in ranks. We also asked students about the degree to which they liked math class, time on math homework, as well as time on language homework.

3.7 Balance Check

Appendix Table 1 presents tests for balance on baseline observables across the treatment arms. The table presents the results from a total of 36 tests comparing average variable values across the treatment and control arms. These tests were conducted by regressing each baseline variable on a

---

endline math tests, which had anchor items, allowed us to produce vertically scaled scores. The scaled scores show that the sample students, on average, made substantive achievement gains within each grade.
The tests were constructed by trained psychometricians in multiple steps. Mathematics test items tests were first selected from standardized mathematics curricula for each grade (4, 5, and 6). The content validity of these test items was checked by multiple experts. The psychometric properties of the test were then validated using data from extensive pilot testing. The tests had good psychometric properties (Cronbach alphas of approximately 0.8, unidimensionality, and a lack of differential item functioning by gender). An analysis of the pilot, baseline and endline test results also indicated that the tests did not suffer from floor or ceiling effects.

[12] In conversations with teachers we found that grades in math courses were determined by homework, class performance, understanding of material, exams, and final exam.

treatment group indicator and controlling for strata. For tests of student-level variables, standard errors were adjusted for clustering at the class level.

Out of the 36 tests, only one was statistically different from zero at the 10% level and one at the 5% level. The results from Table A1 therefore indicate that balance was achieved across the three arms, especially as a small number of significant differences is to be expected (by random chance). Our key baseline covariates (baseline math test scores and grades) were not statistically different between any of the three treatment arms (even at the 10% level).

4 Results

Estimates of Equation (3.6) are reported in Table 1. For the full sample, we find a positive and statistically significant effect of the overall CAL program on the student's math grade although no statistical evidence of a positive effect of the overall CAL program on math test scores. The coefficient estimate on math grade indicates that the CAL program increased a student's ranking in the class by 1.8 percentiles. After excluding very small classes, which create a high level of variance because movements in grade-class rankings are amplified, we find that the CAL program increased a student's ranking in the class by 1.9 percentiles.[13] Turning to isolating the technology effect of CAL we find no effect on test scores and no effect on math grades. For math grades, even in face of the positive CAL program estimate, we do not find a technology-based effect of CAL that is statistically distinguishable from zero. Furthermore, the CAL technology estimate is statistically different from the CAL program estimate (1.53 to 1.70 percentile points).

[13] The total sample size is only 2 percent smaller than the total sample size. The median class size in the sample is 36.

The next step towards recovering the technology parameter, θ, from the theoretical model defined in Equations (3.4) and (3.5) is to estimate whether homework time is affected by the two treatments. As noted above, an additional complication regarding the interpretation of the overall CAL program effect estimates is that they include the potential crowd out (or crowd in) of homework time on the subject. Table 2 reports estimates of CAL and workbook treatment effects for time spent on math homework (i.e. not during school and not during the CAL or workbook sessions as part of the experiment). From Equation (3.4), η^{C} can be estimated from CAL treatment – control, and η^{WK} can be estimated from workbook treatment - control. All of the point estimates on homework time are small and statistically insignificant. We do not find evidence that students' homework time is altered by either the CAL or workbook treatments. This is reasonable given that teachers continued to require regular homework, and the CAL and workbook sessions were run independently of the classroom.[14] Given these findings we can interpret the CAL technology effects estimates presented in Table 1 as estimates of the theoretical parameter, θ, in Equations (2.1') and (2.3').

Taken together, the results suggest that even though the "EdTech" program may positively influence student learning outcomes for the average student, part of the effect is due to additional inputs such as time on instruction that supplemental workbook sessions (the "Ed" without the "Tech") also offer. In fact, our estimates for performance in math class suggest that the entire effect is due to additional inputs and that the isolated technology-based CAL effect is zero.

Boys and girls use computers differently with much higher levels of video game use among boys (Kaiser Family Foundation 2010; U.S. Department of Education 2011; Fairlie 2017; Algan

[14] We also find that the CAL or workbook sessions do not crowd out time on other subjects (in our case, the main other subject students took in primary school – language).

and Fortin 2018).[15] Additionally, boys and girls differ substantially in academic performance in schools in China (Eble and Hu 2019; Xu and Li 2018; Hannum and Park 2007). Thus, we estimate impacts of CAL, which is video game based, separately for boys and girls. Tables 3 and 4 report estimates of Equation (3.6) for boys and girls, respectively. The patterns for CAL program and isolated technology effects are more apparent for boys. The CAL program has positive and significant effects on both boys' test scores (0.10σ) and performance in math class (3.4 percentile points). On the other hand, we find no evidence of positive CAL technology effects for boys. For both test scores and grades, the point estimates are smaller and statistically indistinguishable from zero. The estimates for math grades are also precise enough to show a statistically significant difference between the CAL program and CAL technology estimates essentially at the 10 percent level (p-value = 0.103). The results for boys provide additional evidence that the isolated CAL technology effect might be small and that part of the positive CAL program estimate is due to additional program inputs such as more time learning math. Estimates reported in Table 5 for impacts on homework time show null effects similar to the results for the total sample. Thus, we do not find evidence of substitutability for homework time for boys implying that the CAL technology estimate can be interpreted as the theoretical parameter, θ.

We find no evidence of significant effects of either the CAL program or CAL technology effects on the learning outcomes of girls. Table 4 reports estimates of Equation (3.6) for girls. The CAL program and CAL technology point estimates are small in magnitude, inconsistent in sign, and not statistically significant. The estimated effects for CAL might differ by gender because boys and girls engage differently with technology (U.S. Department of Education 2011; Kaiser Family Foundation 2010; Fairlie 2017; Algan and Fortin 2018). Additional analyses do not reveal

[15] International PISA data indicate that 47 percent of boys compared with 16 percent girls play a computer game every day (Algan and Fortin 2018).

any clear explanations for why our results differ, however. One possibility is there might have been a small amount of substitutability away from homework time for girls. Estimates reported in Table 5 for impacts on homework time show some evidence of negative effects for girls.

A common argument for how CAL, or EdTech more generally, works is that it increases engagement in subject material. If students enjoy learning math through CAL that enjoyment could spill over to their math classes. Table 6 reports estimates of Equation (4.1) for whether students report liking their math class.[16]

The results differ for liking math class. For all students, both the CAL program effect and the CAL technology effect are positive and statistically significant. The CAL technology effect is roughly 2.7 percentile points. Another key finding here is that the CAL program and CAL technology effects sizes are essentially the same. Spending more time on math is not the underlying cause of why the CAL program treatment has a positive effect on liking math and instead the vivid images, gamification and other technology-based attributes of CAL might have increased overall enjoyment of math. For boys, the CAL program effect is positive and statistically significant, but the CAL technology effect is statistically insignificant. The difference in point estimates, however, is small. For girls, the CAL technology effect is positive and statistically significant, but the CAL program effect is not significant. The CAL program versus technology difference is larger than for boys, but also not statistically significant. Overall, we find some evidence that the technology component of CAL has a positive spillover effect on students liking their math class. This is consistent with the argument that the use of technology can increase interest in subject material. This increased interest may or may not translate into higher academic performance over the long-run.

[16] The endline survey question was worded carefully to refer to the student's math class and not to the CAL or workbook sessions.

4.2 Distributional Effects

The results from the treatment regressions provide some evidence of CAL program effects and smaller or null CAL technology effects at the mean. Turning the focus to other parts of the distribution, we first estimate models in which we create dependent variables indicating that the student is above the median of the test score or grade distribution (Appendix Table 2). For test scores (above the median), we find little evidence of significant effects for either the CAL program or CAL technology. For grades (above the median), we find a positive and statistically significant coefficient on the CAL program effect (0.036, se=0.018), but a small and statistically insignificant coefficient on the CAL technology effect (0.009, se=0.018). These results are consistent with the main regression results.

We also estimate quantile treatment effects regressions to test for differential treatment effects across the post-treatment outcome distribution. Appendix Figures 2 and 3 display estimates and 95 percent confidence intervals for each percentile for the CAL Program and CAL technology effects for math test scores and math grades, respectively. For test scores we find some evidence of positive CAL technology effects at the bottom of the distribution. For most of the distribution we find null estimates of CAL program and CAL technology effects. For math grades, the patterns are consistent with the findings for mean treatment effects – larger positive CAL program effects throughout the distribution, but essentially zero CAL technology effects throughout the distribution. Although the quantile treatment estimates are not precisely measured they do not change the conclusion from the mean impacts reported in Table 1. Mean impact estimates do not appear to be hiding differential effects at different parts of the distribution. We thus focus on mean impacts.

4.3 Heterogeneity on Initial Math Ability

We estimate CAL program and CAL technology effects by baseline math ability terciles. Appendix Table 3 reports estimates of (4.1) separately by tercile defined by baseline math test scores. Teachers do not assign class ranks at the beginning of the school year. For the bottom and top terciles we find similar results as for the results for all students. We find positive CAL program effects on endline math class rank, and CAL technology point estimates that are notably smaller and are not statistically distinguishable from zero. We find no discernable effects on endline math test scores. For the middle tercile we do not find statistically significant coefficients for either test scores or grades. The main findings thus hold for both the lowest ability and highest ability students.

4.4 Robustness Checks

We conduct several robustness checks of our main results. First, we examine whether the lack of evidence of a CAL technology effect is due to students not having any experience working on computers in school. In contrast, we find that all of the schools in our sample have computer time at school, and self-reported use by schoolchildren indicates that 87 percent have used computers in school at baseline. Nevertheless, we estimate the test score and grade regressions with only students who self-report using a computer at school as a check. We find similar results to those reported in Table 1 (results not shown for the sake of brevity). The null finding for the CAL technology effect is not due to schoolchildren not being familiar with using computers at school.

Second, we examine whether the estimates for grade are sensitive to having a high percentage of boarding students in the classroom. In classes with a high percentage of boarding students and treatment being assigned at the class level, there could be an attenuated treatment effect. To address this issue, we estimate the grade regression (reported in columns 3 and 4 of Table 1) excluding classes with 90 percent boarding students, 80 percent boarding students and 70 percent boarding students. We find that the CAL program coefficients remain positive and roughly similar in magnitude although they lose some statistical power (Appendix Tables 4-6). The CAL technology coefficients remain small (often negative) and not close to statistical significance. The robustness of results is consistent with boarding students representing a small share of students in the classroom.

5 Conclusions

Although EdTech is rapidly expanding around the world and accelerating in response to recent global health developments, relatively little is known about the advantages and disadvantages of using technology in education. Is EdTech, as proponents argue, revolutionizing the way in which students learn? Our theoretical model illustrates that the answer to this question is not straightforward because there are several possible inputs to educational production that are often entangled with the technology provided in CAL programs, making it difficult to isolate effects. To estimate the technology effect of CAL and other key parameters from the theoretical model, we design a field experiment in rural China that includes a novel "pencil and paper" workbook treatment in addition to a regular CAL program treatment and a control group. Estimates from the experiment indicate that, for the average student in rural China, the overall CAL program improves math grades whereas the isolated technology component of CAL has no discernable effect on math grades. The difference between the two estimated effects is statistically significant. We do not find evidence of a CAL program or CAL technology effect for math test scores, which may be due to the program having greater effects on non-cognitive than cognitive skills.

Given gender differences in computer use, we examined effects for boys and girls separately. For boys, we find that the CAL program increases math grades by 3.5 percentile points and math test scores by 0.10σ. But, when we isolate the CAL technology effect, the point estimates become noticeably smaller and statistically indistinguishable from zero. We also find no evidence of substitution effects of the CAL and workbook sessions on homework time in math. For girls we do not find positive effects of the CAL program or CAL technology component. On the other hand, we find evidence suggesting that the both the CAL program and CAL technology affect how much students report that they like their math class, which might or might not have longer-term effects.

Our study provides a second generation "pencil test" (DiNardo and Pischke 1997). If similarly timed, content and structured "pencil and paper" workbook sessions show roughly similar effects on academic performance as the CAL program then it raises concerns that another factor common to both is driving the results. In particular, the "pencil and paper" workbook sessions, by construction, provide more time learning subject material, which might be the key educational input that increases academic performance and not the new computer technology in CAL programs. The technology-based effect of CAL might be relatively small and might not be the primary driver of the estimated large positive effects of CAL programs found in many previous studies.

Another argument for the rapid adoption of EdTech around the world is that it has low marginal costs. Once developed, copying software or providing it online is nearly costless to provide access to the additional student. In our experiment, however, we find that the marginal costs of pencil and paper workbooks are also low, and in fact, are lower. The costs of photocopying workbooks are inexpensive. Furthermore, workbooks do not require the high fixed costs and maintenance costs of computers, Internet connections, and extra space to house computers. Back-of-the-envelope calculations indicate that the workbook program has roughly 22 percent lower costs (see Appendix B).

More research is clearly needed on the effectiveness of EdTech and what underlies these effects. In settings where students are substantially behind grade level or there is substantial heterogeneity, the technology effects of *adaptive* CAL might be larger because technology can personalize education. This is consistent with the large positive effects of an after-school program in India that includes adaptive computer instruction (Muralidharan et al., 2019).[17] In educational systems in which students are generally at grade level and there is less heterogeneity, however, experimental evidence shows that adaptive CAL does not have an advantage over non-adaptive CAL and might even be less effective for high-ability students (Van Klaveren et al. 2017). Moreover, in comparison with the widely-used software we evaluate, adaptive software is much more costly to develop and maintain. There may be no one-size adaptive algorithm, making it difficult to generalize the benefits of adaptive software even across contexts where teaching at the right level is important.[18] More research is needed on the tradeoffs between adaptive vs non-adaptive software.

Another area of promise is that we find evidence of a positive effect of CAL technology on student interest in math whereas no effect on math interest from extra time learning math. More research is needed on whether the technology in EdTech can spark an interest in math among young children and generate longer-term interest and success in math. More research is also needed on separating the effects of various inputs in educational production, especially the mechanical effects extra time learning, in supplemental educational programs. The results of this study raise

[17] As compared to the null effects of a similar after-school program in India without adaptive computer instruction found in Berry and Mukherjee (2016).
[18] Other components of EdTech could also be evaluated. For example, CAL might be improved with a component that regularly informs teachers and parents of student progress (Bergman 2020).

concerns about the attribution of the effectiveness of key inputs in these programs and have broader implications for evaluations of any supplemental educational program.

Finally, more research is needed on whether and to what degree EdTech can substitute for traditional learning. This is especially pertinent today in light of the full-scale, comprehensive, global movement to EdTech at all levels of education in response to the coronavirus pandemic. How much human capital accumulation will be lost or will CAL, online classes, remote learning, and other forms of EdTech be able to substitute adequately for traditional teaching and learning methods? These are important questions as we move forward with education in which limited person-to-person contact is essential for health reasons.

References

Algan, Yann, and Nicole M. Fortin. 2018. "Computer Gaming and the Gender Math Gap: Cross-Country Evidence among Teenagers." In *Transitions through the Labor Market: Work, Occupation, Earnings and Retirement*, pp. 183-228. Emerald Publishing Limited.

Bai, Yu, Bin Tang, Boya Wang, Di Mo, Linxiu Zhang, Scott Rozelle, Emma Auden and Blake Mandell. 2018. "Impact of Online Computer Assisted Learning on Education: Evidence from a Randomized Controlled Trial in China." REAP Working Paper.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122 (3): 1235-1264.

Barrow, Lisa, Lisa Markman, and Cecilia Elena Rouse. 2009 "Technology's Edge: The Educational Benefits of Computer-aided Instruction." *American Economic Journal: Economic Policy* 1 (1): 52-74.

Bergman, Peter. 2020. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment," *Journal of Political Economy* (forthcoming).

Beuermann, Diether W., Julian Cristia, Santiago Cueto, Ofer Malamud, and Yyannu Cruz-Aguayo. 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.

Böhmer, Bianca. 2014. "Testing Numeric: Evidence from a Randomized Controlled Trial of a Computer Based Mathematics Intervention in Cape Town High Schools." University of Cape Town Working Paper.

Bulman, George, and Robert W. Fairlie. "Technology and education: Computers, software, and the internet." *Handbook of the Economics of Education*. Vol. 5. Elsevier, 2016. 239-280.

Burguillo, Juan C. 2010. "Using Game Theory and Competition-based Learning to Stimulate Student Motivation and Performance." *Computers & Education* 55 (2): 566-575.

Campuzano, Larissa, Mark Dynarski, Roberto Agodini, and Kristina Rall. 2009. "Effectiveness of Reading and Mathematics Software Products: Findings from Two Student Cohorts. https://eric.ed.gov/?id=ED504657.

Carrillo, Paul E., Mercedes Onofa, and Juan Ponce. 2011. "Information Technology and Student Achievement: Evidence from a Randomized Experiment in Ecuador." IDB Working Paper 78.

Chen, L.J., Yang, D.L., & Ren, Q. (2015). Report on the state of children in China. Chicago, IL: Chapin Hall at the University of Chicago.

Cristia, Julian, Pablo Ibarrarán, Santiago Cueto, Ana Santiago, and Eugenio Severín. 2017. "Technology and child development: Evidence from the one laptop per child program." *American Economic Journal: Applied Economics* 9, no. 3: 295-320.

Dynarski, Mark, Roberto Agodini, Sheila Heaviside, Timothy Novak, Nancy Carey, Larissa Campuzano, Barbara Means, Robert Murphy, William Penuel, Hal Javitz, Deborah Emery, Willow Sussex. 2007. "Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort." https://telearn.archives-ouvertes.fr/hal-00190019/.

Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. 2017. “Education Technology: an Evidence-based Review”. NBER Working Paper w23744.

DiNardo, John, and Jörn-Steffen Pischke. 1997. "The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?" *Quarterly Journal of Economics* 112 (1): 291-303.

Ebner, Martin, and Andreas Holzinger. 2007. "Successful Implementation of User-centered Game Based Learning in Higher Education: An Example from Civil Engineering." *Computers & education* 49 (3): 873-890.

Eble, Alex, & Hu, Feng. 2019. “How important are beliefs about gender differences in math ability? Transmission across generations and impacts on child outcomes.” Columbia University Working Paper.

Falck, Oliver, Constantin Mang, and Ludger Woessmann. 2018. "Virtually No Effect? Different Uses of Classroom Computers and their Effect on Student Achievement." *Oxford Bulletin of Economics and Statistics* 80 (1): 1-38.

Fairlie, Robert W. 2016. "Do Boys and Girls Use Computers Differently, and Does it Contribute to Why Boys Do Worse in School than Girls?." *The BE Journal of Economic Analysis & Policy* 16.1: 59-96.

Fairlie, Robert W., and Jonathan Robinson. 2013. "Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren," *American Economic Journal: Applied Economics* 5(3): 211-240.

Glewwe, Paul W., Eric A. Hanushek, Sarah D. Humpage, and Renato Ravina. 2013. “School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010,” in *Education Policy in Developing Countries* (ed. Paul Glewwe): University of Chicago Press: Chicago.

Hull, Marie, and Katherine Duch. 2019. "One-to-One Technology and Student Outcomes: Evidence From Mooresville’s Digital Conversion Initiative." *Educational Evaluation and Policy Analysis* 41.1: 79-97.

Kaiser Family Foundation. 2010. *Generation M^2: Media in the Lives of 8- to 18-Year Olds*. Kaiser Family Foundation Study.

Lai, Fang, Linxiu Zhang, Xiao Hu, Qinghe Qu, Yaojiang Shi, Yajie Qiao, Matthew Boswell, and Scott Rozelle. 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, Fang, Renfu Luo, Linxiu Zhang, Xinzhe Huang, and Scott Rozelle. 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.

Levin, Henry M., and Clive Belfield. 2015. "Guiding the Development and Use of Cost-effectiveness Analysis in Education." *Journal of Research on Educational Effectiveness* 8 (3): 400-418.

Levin, Henry M., Patrick J. McEwan, Clive Belfield, A. Brooks Bowden, and Robert Shand. 2017. *Economic Evaluation in Education: Cost-effectiveness and Benefit-cost Analysis*. SAGE Publications.

Li, Yanyan, Prashant Loyalka, Guirong Li, Chengfang Liu, and Scott Rozelle. 2019. Learning Trajectories among Middle School Students in Developing Contexts: Evidence from Rural China." REAP Working Paper.

Linden, Leigh L. 2008. "Complement or Substitute? The Effect of Technology on Student Achievement in India," http://www.leighlinden.com/Gyan_Shala_CAL_2008-06-03.pdf.

Malamud, Ofer, and Cristian Pop-Eleches. "Home computer use and the development of human capital." *The Quarterly journal of economics* 126.2 (2011): 987-1027.

Malamud, Ofer, Santiago Cueto, Julian Cristia, and Diether W. Beuermann. "Do children benefit from internet access? Experimental evidence from Peru." *Journal of Development Economics* 138 (2019): 41-56.

McEwan, Patrick J. 2015. "Improving Learning in Primary Schools of Developing Countries: A Meta-analysis of Randomized Experiments." *Review of Educational Research* 85 (3): 353-394.

Ministry of Education. 2011. 2010 National Statistical Report on Education Development. *China Geology Education*, 2011(3): 93-96.

Ministry of Education. 2017. The Yearbook of Education Statistics in China. National Statistics Press, Beijing, China. http://www.stats.gov.cn/tjsj/ndsj/2017/indexeh.htm (accessed March 19, 2019)

Mo, D., Swinnen, J., Zhang, L., Yi, H., Qu, Q., Boswell, M. and Rozelle, S., 2013. Can One-to-one Computing Narrow the Digital Divide and the Educational Gap in China? The Case of Beijing Migrant Schools. *World Development*, 46: 14-29.

Mo, Di, Linxiu Zhang, Renfu Luo, Qinghe Qu, Weiming Huang, Jiafu Wang, Yajie Qiao, Matthew Boswell, and Scott Rozelle. 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, Di, Weiming Huang, Yaojiang Shi, Linxiu Zhang, Matthew Boswell, and Scott Rozelle. 2015. "Computer Technology in Education: Evidence from a Pooled Study of Computer Assisted Learning Programs among Rural Students in China." *China Economic Review* 36: 131-145.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2016. “Disrupting Education? Experimental Evidence on Technology-aided Instruction in India.” *American Economic Review* 109.4 (2019): 1426-1460.

Parente, Paulo MDC, and João MC Santos Silva. 2016. "Quantile Regression with Clustered Data." *Journal of Econometric Methods* 5 (1): 1-15.

Rockoff, Jonah E. 2015. “Evaluation Report on the School of One i3 Expansion." https://bit.ly/2OfPDXs.

Rouse, Cecilia Elena, and Alan B. Krueger. 2004. Putting Computerized Instruction to the Test: A Randomized Evaluation of a “Scientifically Based” Reading Program," *Economics of Education Review* 23(4): 323–338.

UNESCO 2020. “Indicators: Primary Education, pupils; Secondary Education, pupils,” Institute for Statistics, https://data.worldbank.org/indicator.

Van der Kleij, Fabienne M., Remco CW Feskens, and Theo JHM Eggen. 2015. "Effects of Feedback in a Computer-based Learning Environment on Students’ Learning Outcomes: A Meta-analysis." *Review of Educational Research* 85 (4): 475-511.

Van Klaveren, Chris, Sebastiaan Vonk, and Ilja Cornelisz. 2017. "The Effect of Adaptive Versus Static Practicing on Student Learning—Evidence from a Randomized Field Experiment." *Economics of Education Review* 58: 175-187.

Vincent, Jane. 2016. "Students’ Use of Paper and Pen versus Digital Media in University Environments for Writing and Reading—A Cross-cultural Exploration." *Journal of Print Media and Media Technology Research* 5 (2): 97-106.

World Bank. 2020. “Indicators: Primary Education, pupils; Secondary Education, pupils,” data.worldbank.org/indicator/

Xu, Di, and Qiujie Li. 2018. "Gender achievement gaps among Chinese middle school students and the role of teachers’ gender." *Economics of Education Review* 67: 82-93.

Table 1: CAL Program and Technology Effects on Math Test Scores and Grades

	(1)	(2)	(3)	(4)	(5)	(6)
	Math Test Score		Grade (Rank)		Grade Rank, Class N >= 10	
CAL Program	0.033	0.032	1.743*	1.758*	1.866**	1.876**
	(0.039)	(0.039)	(0.919)	(0.922)	(0.925)	(0.929)
CAL Technology	0.059	0.061	0.212	0.155	0.234	0.178
	(0.044)	(0.044)	(0.996)	(0.999)	(1.013)	(1.017)
Difference (Program - Tech)	-0.026	-0.029	1.531*	1.603*	1.632*	1.697*
	(0.046)	(0.046)	(0.877)	(0.876)	(0.895)	(0.894)
Additional Controls	No	Yes	No	Yes	No	Yes
N	3,928	3,928	3,829	3,829	3,750	3,750
R-squared	0.432	0.436	0.300	0.308	0.299	0.308

Notes:

1) CAL program is the overall program effect (i.e. CAL treatment relative to control), and CAL technology is the isolated technology-based effect of CAL (i.e. CAL treatment relative to workbook session treatment).
2) All columns control for baseline counterpart of dependent variable (baseline math score or baseline class rank in math).
3) Even-numbered columns also control for the following baseline covariates: liking math (scale 1 to 100), student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number of boarding students in the class, class size.
4) Cluster (class-level)-robust standard errors in parentheses.
5) *** p<0.01, ** p<0.05, * p<0.1.

Table 2: CAL Program and Workbook Treatment Effects on Time on Math Outside of School (Homework Hours)

	(1)	(2)
	All Students	
CAL Treatment – Control (η^{C})	-0.149	-0.154
	(0.199)	(0.199)
Workbook Treatment – Control (η^{WK})	0.128	0.123
	(0.201)	(0.201)
Additional Controls	No	Yes
N	3,930	3,930
R-squared	0.099	0.099

Notes:

1) CAL treatment – Control and Workbook treatment – Control are reported for crowd out (in) estimates of the two treatments (see Equations 3.2 and 3.3).
2) Math homework time (hours last week): control group mean = 3.36, SD = 2.70.
3) All columns control for baseline math score.
4) Even-numbered columns also control for the following baseline covariates: student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number of boarding students in the class, class size.
5) Cluster (class-level)-robust standard errors in parentheses.
6) *** p<0.01, ** p<0.05, * p<0.1.

Table 3: CAL Program and Technology Effects on Math Test Scores and Grades– Boys Only

	(1)	(2)	(3)	(4)	(5)	(6)
	Math Test Score		Grade (Rank)		Grade, Class N >= 10	
CAL Program	0.099**	0.099**	3.414***	3.430***	3.488***	3.506***
	(0.049)	(0.049)	(1.280)	(1.277)	(1.291)	(1.288)
CAL Technology	0.075	0.074	1.540	1.530	1.482	1.455
	(0.054)	(0.054)	(1.456)	(1.449)	(1.480)	(1.472)
Difference (Program - Tech)	0.025	0.025	1.874	1.900	2.006	2.051
	(0.060)	(0.060)	(1.218)	(1.223)	(1.249)	(1.253)
Additional Controls	No	Yes	No	Yes	No	Yes
N	2,142	2,142	2,095	2,095	2,053	2,053
R-squared	0.442	0.445	0.307	0.311	0.307	0.312

Notes:

1) CAL program is the overall program effect (i.e. CAL treatment relative to control), and CAL technology is the isolated technology-based effect of CAL (i.e. CAL treatment relative to workbook session treatment).
2) All columns control for baseline counterpart of dependent variable (baseline math score or baseline class rank in math).
3) Even-numbered columns also control for the following baseline covariates: liking math (scale 1 to 100), student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number of boarding students in the class, class size.
4) Cluster (class-level)-robust standard errors in parentheses.
5) *** p<0.01, ** p<0.05, * p<0.1.

Table 4: CAL Program and Technology Effects on Math Test Scores and Grades – Girls Only

	(1)	(2)	(3)	(4)	(5)	(6)
	Math Test Score		Grade (Rank)		Grade, Class N >= 10	
CAL Program	-0.044	-0.045	-0.526	-0.590	-0.345	-0.434
	(0.046)	(0.046)	(1.371)	(1.371)	(1.374)	(1.373)
CAL Technology	0.039	0.041	-1.451	-1.350	-1.365	-1.268
	(0.054)	(0.054)	(1.519)	(1.526)	(1.540)	(1.549)
Difference (Program - Tech)	-0.084	-0.086	0.925	0.759	1.019	0.834
	(0.052)	(0.052)	(1.375)	(1.396)	(1.394)	(1.416)
Additional Controls	No	Yes	No	Yes	No	Yes
N	1,785	1,785	1,733	1,733	1,696	1,696
R-squared	0.432	0.437	0.302	0.308	0.299	0.305

Notes:

1) CAL program is the overall program effect (i.e. CAL treatment relative to control), and CAL technology is the isolated technology-based effect of CAL (i.e. CAL treatment relative to workbook session treatment).
2) All columns control for baseline counterpart of dependent variable (baseline math score or baseline class rank in math).
3) Even-numbered columns also control for the following baseline covariates: liking math (scale 1 to 100), student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number of boarding students in the class, class size.
4) Cluster (class-level)-robust standard errors in parentheses.
5) *** p<0.01, ** p<0.05, * p<0.1.

Table 5: CAL and Workbook Treatment Effects on Math Outside of School (Homework Hours) – Boys and Girls

	(1)	(2)	(3)	(4)
	Boys		Girls	
CAL Treatment – Control (η^{C})	0.065	0.065	-0.387*	-0.393**
	(0.241)	(0.242)	(0.198)	(0.197)
Workbook Treat. – Control (η^{WK})	0.305	0.298	-0.068	-0.085
	(0.239)	(0.240)	(0.207)	(0.206)
Additional Controls	No	Yes	No	Yes
N	2,145	2,145	1,784	1,784
R-squared	0.096	0.098	0.121	0.125

Notes:

1) CAL treatment – Control and Workbook treatment – Control are reported for crowd out (in) estimates of the two treatments (see Equations 3.2 and 3.3).
2) Math homework time (hours last week): control group mean = 3.36, SD = 2.70.
3) All columns control for baseline math score.
4) Even-numbered columns also control for the following baseline covariates: student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number of boarding students in the class, class size.
5) Cluster (class-level)-robust standard errors in parentheses.
6) *** p<0.01, ** p<0.05, * p<0.1.

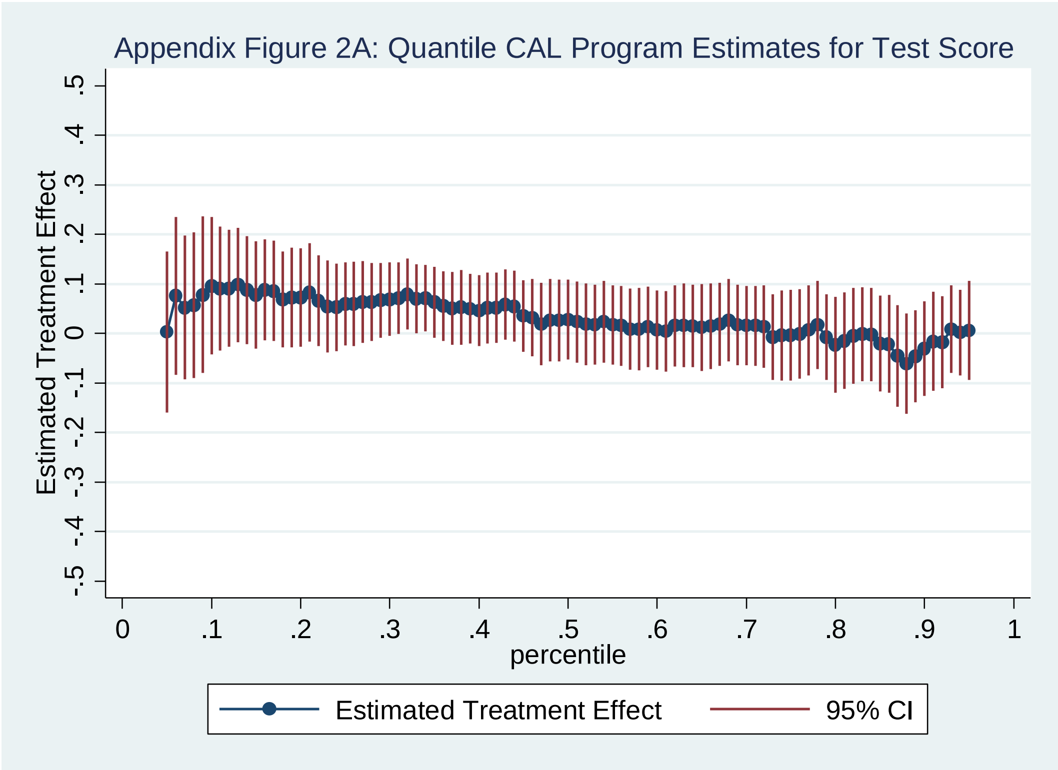
Table 6: CAL Program and Technology Effects on Liking Math Class (Scale 1-100)

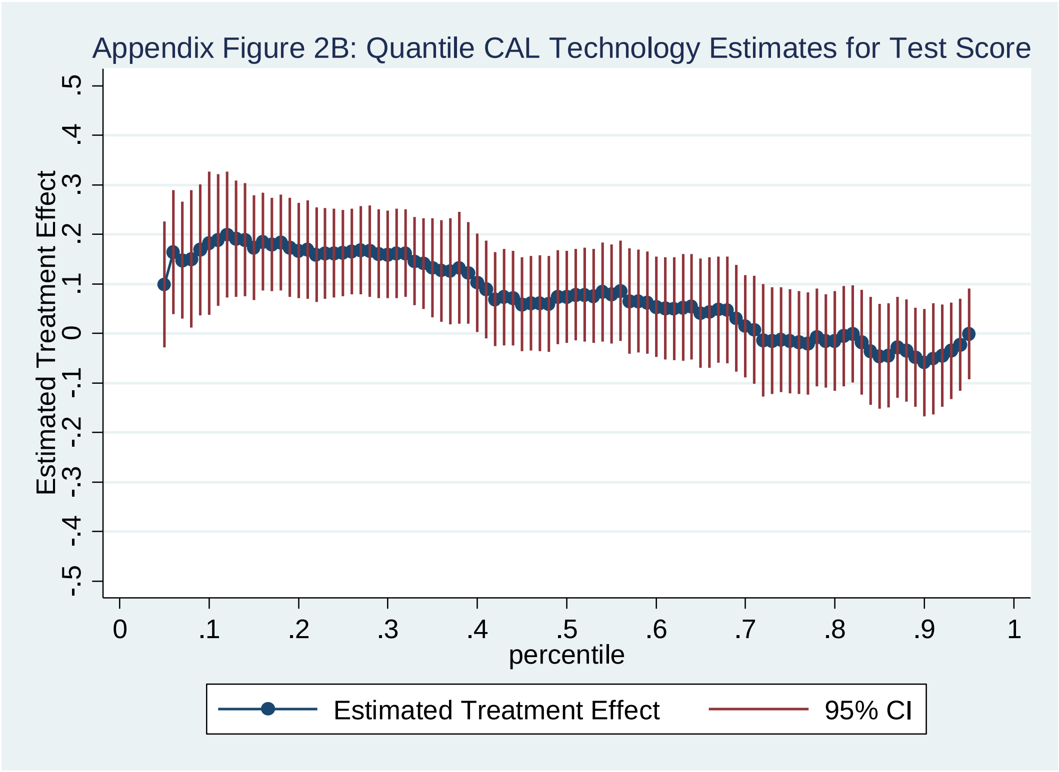
	(1)	(2)	(3)	(4)	(5)	(6)
	All Students		Boys		Girls	
CAL Program	2.580*	2.595*	3.648**	3.675**	1.405	1.476
	(1.336)	(1.327)	(1.644)	(1.639)	(1.557)	(1.540)
CAL Technology	2.675**	2.714**	2.467	2.551	2.813*	2.855*
	(1.357)	(1.359)	(1.624)	(1.628)	(1.672)	(1.679)
Difference (Program - Tech)	-0.094	-0.119	1.181	1.124	-1.408	-1.379
	(1.491)	(1.485)	(1.780)	(1.776)	(1.844)	(1.824)
Additional Controls	No	Yes	No	Yes	No	Yes
N	3,931	3,931	2,145	2,145	1,785	1,785
R-squared	0.170	0.172	0.163	0.165	0.195	0.201

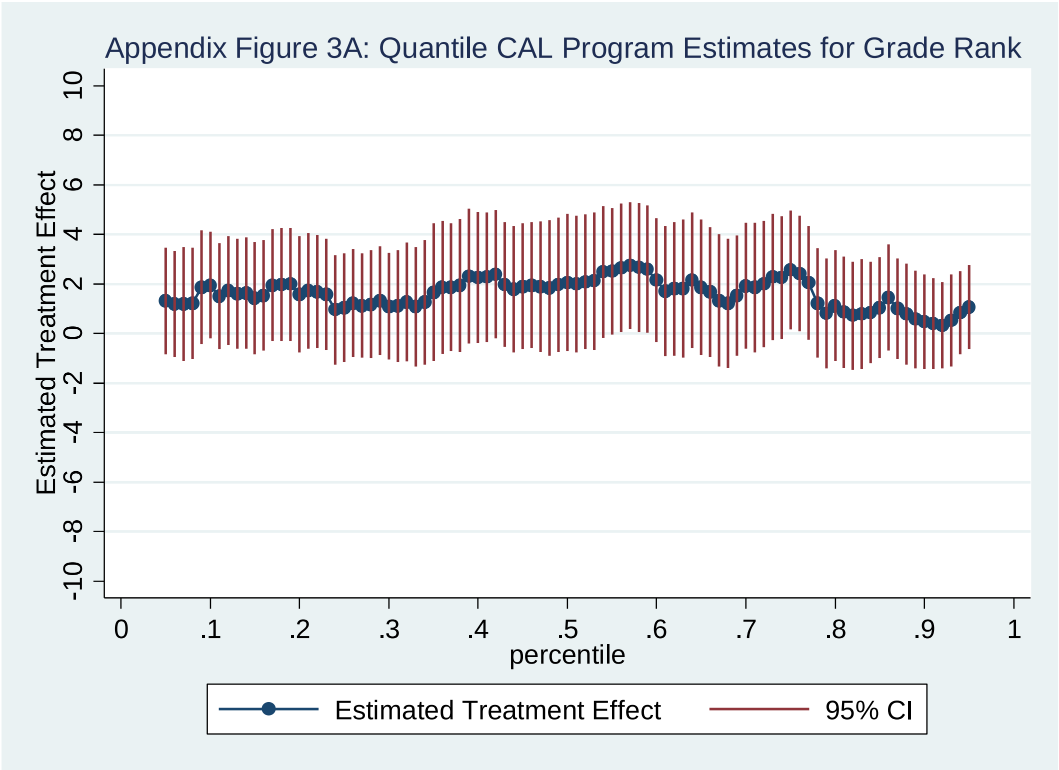
Notes:

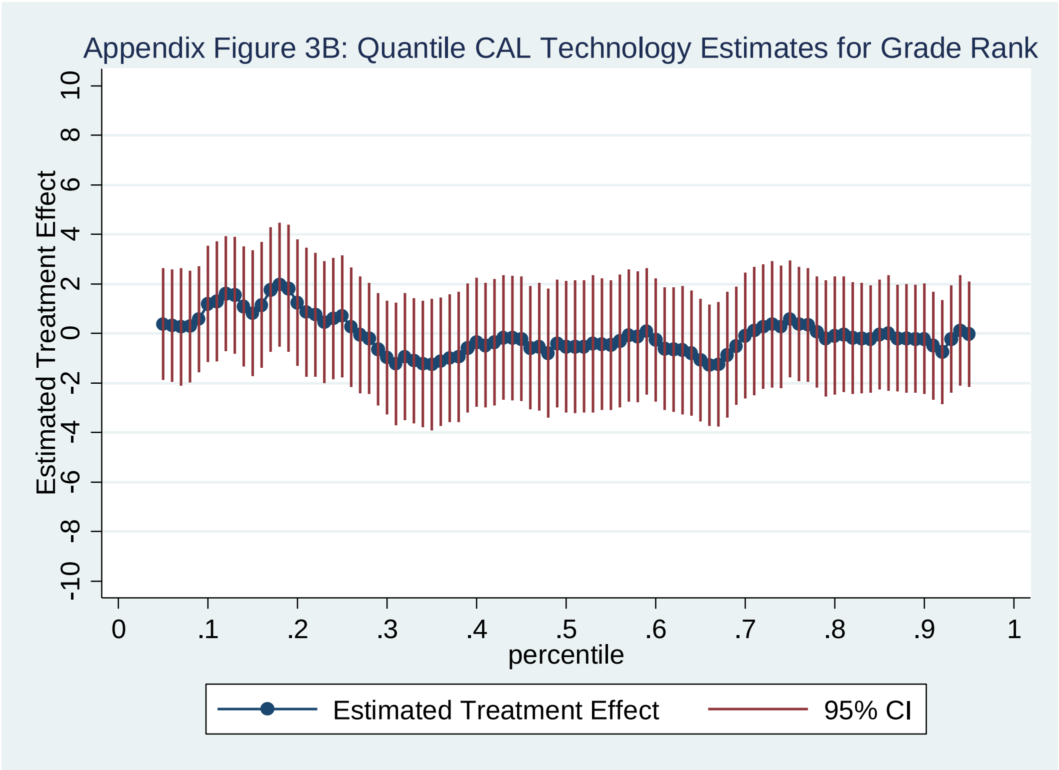
1) CAL program is the overall program effect (i.e. CAL treatment relative to control), and CAL technology is the isolated technology-based effect of CAL (i.e. CAL treatment relative to workbook session treatment).
2) All columns control for baseline liking math class (scale 1 to 100), control group mean = 87.2.
3) Even-numbered columns also control for the following baseline covariates: student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number of boarding students in the class, class size.
4) Cluster (class-level)-robust standard errors in parentheses.
5) *** p<0.01, ** p<0.05, * p<0.1.

Appendix Figure 1


Appendix Figure 2A: Quantile CAL Program Estimates for Test Score
Estimated Treatment Effect
-.5
-.4
-.3
-.2
-.1
0
.1
.2
.3
.4
.5
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
percentile
Estimated Treatment Effect
95% CI


Appendix Figure 2B: Quantile CAL Technology Estimates for Test Score
Estimated Treatment Effect
-.5
-.4
-.3
-.2
-.1
0
.1
.2
.3
.4
.5
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
percentile
Estimated Treatment Effect
95% CI


Appendix Figure 3A: Quantile CAL Program Estimates for Grade Rank
Estimated Treatment Effect
-10
-8
-6
-4
-2
0
2
4
6
8
10
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
percentile
Estimated Treatment Effect
95% CI


Appendix Figure 3B: Quantile CAL Technology Estimates for Grade Rank
Estimated Treatment Effect
-10
-8
-6
-4
-2
0
2
4
6
8
10
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
percentile
Estimated Treatment Effect
95% CI

Appendix Table 1: Summary Statistics and Balance Check

	Baseline Characteristics	Control	CAL	Workbook	P-value	P-value	P-value
		（1）	（2）	（3）	(4)=(2)-(1)	(5)=(3)-(1)	(6)=(2)-(3)
Panel A. Student Characteristics							
(1)	Standardized math score	-0.026	-0.036	-0.032	[0.669]	[0.881]	[0.776]
		(0.987)	(1.062)	(1.017)			
(2)	Within class rank (using math score)	53.906	53.799	55.008	[0.860]	[0.130]	[0.191]
		(29.498)	(30.539)	(29.453)			
(3)	Female (0/1)	0.458	0.459	0.435	[0.762]	[0.264]	[0.163]
		(0.498)	(0.499)	(0.496)			
(4)	Age (years)	11.095	11.017	11.048	[0.874]	[0.738]	[0.612]
		(1.069)	(1.115)	(1.109)			
(5)	Father education 9 years or less (0/1)	0.441	0.400	0.424	[0.081]	[0.554]	[0.237]
		(0.497)	(0.490)	(0.494)			
(6)	Mother education 9 years or less (0/1)	0.390	0.356	0.365	[0.180]	[0.447]	[0.474]
		(0.488)	(0.479)	(0.482)			
(7)	Liking math	-0.052	0.006	-0.117	[0.419]	[0.399]	[0.124]
		(1.050)	(1.015)	(1.087)			
	N	1390	1345	1289			
Panel B. Teacher and Class Characteristics							
(1)	Female (0/1)	0.445	0.391	0.460	[0.334]	[0.909]	[0.417]
		(0.497)	(0.488)	(0.499)			
(2)	Experience (years)	16.239	13.425	15.424	[0.148]	[0.843]	[0.121]
		(11.886)	(11.050)	(11.384)			
(3)	College degree (0/1)	0.560	0.569	0.574	[0.667]	[0.371]	[0.670]
		(0.497)	(0.496)	(0.495)			
(4)	# boarding students	15.447	14.517	16.290	[0.162]	[0.949]	[0.270]
		(6.776)	(6.632)	(79.547)			
(5)	# of total students	35.426	32.717	35.322	[0.019]	[0.223]	[0.313]
		(13.965)	(14.217)	(15.339)			
(6)	N	118	116	118			

Notes: means and SDs (in parentheses) in columns 1-3. P-values in Columns (4-6) are calculated using the estimated coefficient and standard error on an indicator for the treatment group in a regression of each baseline characteristic on the treatment indicator and controlling for randomization strata with robust standard errors accounting for clustering within classes. Joint tests of all student/teacher baseline covariates simultaneously shows no significant difference between T1 and C (p-value: 0.860/0.124), T2 and C (p-value: 0.790/0.862) or T1 and T2 (p-value = 0.184/0.840).

Appendix Table 2: CAL Program and Technology Effects on Whether Math Test Scores/Grades are Above the Median

	(1)	(2)	(3)	(4)	(5)	(6)
	Math Test Score Above Median (Y/N)		Grade (Rank) Above Median (Y/N)		Grade (Rank) Above Median (Y/N) for Class N >= 10	
CAL Program	0.016	0.015	0.036**	0.036**	0.038**	0.038**
	(0.020)	(0.020)	(0.018)	(0.018)	(0.018)	(0.018)
CAL Technology	0.034	0.035*	0.010	0.009	0.012	0.011
	(0.021)	(0.021)	(0.018)	(0.018)	(0.019)	(0.019)
Difference (Program - Tech)	-0.018	-0.020	0.026	0.027	0.026	0.027
	(0.022)	(0.022)	(0.017)	(0.017)	(0.017)	(0.018)
Additional Controls	No	Yes	No	Yes	No	Yes
N	3,928	3,928	3,829	3,829	3,722	3,722
R-squared	0.247	0.253	0.165	0.173	0.165	0.175

Notes:

1) CAL program is the overall program effect (i.e. CAL treatment relative to control), and CAL technology is the isolated technology-based effect of CAL (i.e. CAL treatment relative to workbook session treatment).
2) All columns control for baseline counterpart of dependent variable (baseline math score or baseline class rank in math above median (y/n)).
3) Columns 2 and 4 also control for the following baseline covariates: liking math (1 to 100), student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number boarding students in the class, class size.
4) Cluster (class-level)-robust standard errors in parentheses.
5) *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 3: CAL Program and Technology Effects on Math Test Scores and Grades – By Baseline Achievement Terciles

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	lowest 1/3 achievers				middle 1/3 achievers				top 1/3 achievers			
	Math score		Math grade (rank)		Math score		Math grade (rank)		Math score		Math grade (rank)	
CAL Program	0.059	0.058	4.012**	3.842**	0.006	-0.002	-1.805	-1.681	0.019	0.021	3.168**	3.090**
	(0.068)	(0.068)	(1.822)	(1.802)	(0.056)	(0.057)	(1.582)	(1.609)	(0.043)	(0.043)	(1.556)	(1.568)
CAL Technology	0.049	0.058	1.873	1.691	0.067	0.064	-2.879	-2.945	0.037	0.040	1.060	1.040
	(0.070)	(0.070)	(1.937)	(1.951)	(0.064)	(0.064)	(1.930)	(1.928)	(0.050)	(0.049)	(1.543)	(1.547)
Program - Tech	-0.026	-0.029	1.531*	1.603*	-0.061	-0.066	1.074	1.265	-0.017	-0.019	2.108	2.050
	(0.046)	(0.046)	(0.877)	(0.876)	(0.068)	(0.068)	(1.736)	(1.735)	(0.051)	(0.050)	(1.570)	(1.560)
Add. Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
N	1,316	1,316	1,294	1,294	1,310	1,310	1,280	1,280	1,302	1,302	1,255	1,255
R-squared	0.192	0.199	0.164	0.182	0.118	0.125	0.121	0.129	0.154	0.170	0.089	0.109

Notes:

1) CAL program is the overall program effect (i.e. CAL treatment relative to control), and CAL technology is the isolated technology-based effect of CAL (i.e. CAL treatment relative to workbook session treatment).
2) All columns control for baseline counterpart of dependent variable (baseline math score or baseline class rank in math).
3) Even-numbered columns also control for the following baseline covariates: student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number of boarding students in the class, class size.
4) Cluster (class-level)-robust standard errors in parentheses.
5) *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 4: CAL Program and Technology Effects on Math Grades, Excluding Classes with Certain Percentages of Boarding Students

	(1)	(2)	(3)	(4)	(5)	(6)
	Boarding Students > 90%		Students > 80%		Students > 70%	
CAL Program	1.306	1.318	1.562	1.616	1.356	1.387
	(0.912)	(0.912)	(0.993)	(0.996)	(1.029)	(1.029)
CAL Technology	-0.085	-0.081	-0.174	-0.146	-0.617	-0.560
	(1.060)	(1.061)	(1.110)	(1.108)	(1.224)	(1.225)
Difference (Program - Tech)	1.392	1.399	1.737*	1.762*	1.973*	1.947*
	(0.928)	(0.927)	(0.968)	(0.971)	(1.073)	(1.077)
Additional Controls	No	Yes	No	Yes	No	Yes
N	3,928	3,928	3,829	3,829	3,750	3,750
R-squared	0.432	0.436	0.300	0.308	0.299	0.308

Notes:

1) CAL program is the overall program effect (i.e. CAL treatment relative to control), and CAL technology is the isolated technology-based effect of CAL (i.e. CAL treatment relative to workbook session treatment).
2) All columns control for baseline counterpart of dependent variable (baseline class rank in math test score).
3) Even-numbered columns also control for the following baseline covariates: liking math (scale 1 to 100), student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number of boarding students in the class, class size.
4) Cluster (class-level)-robust standard errors in parentheses.
5) *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 5: CAL Program and Technology Effects on Math Grades, Excluding Classes with Certain Percentages of Boarding Students – Boys Only

	(1)	(2)	(3)	(4)	(5)	(6)
	Boarding Students > 90%		Students > 80%		Students > 70%	
CAL Program	2.836**	2.836**	3.042**	3.089**	2.332	2.426*
	(1.289)	(1.285)	(1.385)	(1.382)	(1.428)	(1.425)
CAL Technology	1.072	1.020	0.749	0.692	-0.123	-0.068
	(1.533)	(1.525)	(1.591)	(1.577)	(1.657)	(1.647)
Difference (Program - Tech)	1.765	1.816	2.294*	2.397*	2.454*	2.494*
	(1.309)	(1.313)	(1.360)	(1.365)	(1.425)	(1.438)
Additional Controls	No	Yes	No	Yes	No	Yes
N	1,940	1,940	1,814	1,814	1,684	1,684
R-squared	0.302	0.306	0.299	0.303	0.307	0.310

Notes:

1) CAL program is the overall program effect (i.e. CAL treatment relative to control), and CAL technology is the isolated technology-based effect of CAL (i.e. CAL treatment relative to workbook session treatment).
2) All columns control for baseline counterpart of dependent variable (baseline class rank in math test score).
3) Even-numbered columns also control for the following baseline covariates: liking math (scale 1 to 100), student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number of boarding students in the class, class size.
4) Cluster (class-level)-robust standard errors in parentheses.
5) *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 6: CAL Program and Technology Effects on Math Grades, Excluding Classes with Certain Percentages of Boarding Students – Girls Only

	(1)	(2)	(3)	(4)	(5)	(6)
	Boarding Students > 90%		Students > 80%		Students > 70%	
CAL Program	-0.715	-0.734	-0.383	-0.443	-0.051	-0.168
	(1.405)	(1.396)	(1.550)	(1.544)	(1.597)	(1.589)
CAL Technology	-1.394	-1.176	-1.281	-1.061	-1.289	-1.068
	(1.615)	(1.619)	(1.681)	(1.680)	(1.775)	(1.787)
Difference (Program - Tech)	1.392	1.399	1.737*	1.762*	1.973*	1.947*
	(0.928)	(0.927)	(0.968)	(0.971)	(1.073)	(1.077)
Additional Controls	No	Yes	No	Yes	No	Yes
N	1,630	1,630	1,500	1,500	1,398	1,398
R-squared	0.295	0.303	0.289	0.297	0.287	0.294

Notes:

1) CAL program is the overall program effect (i.e. CAL treatment relative to control), and CAL technology is the isolated technology-based effect of CAL (i.e. CAL treatment relative to workbook session treatment).
2) All columns control for baseline counterpart of dependent variable (baseline class rank in math test score).
3) Even-numbered columns also control for the following baseline covariates: liking math (scale 1 to 100), student age (years), gender, father graduated junior high, mother graduated junior high, teacher experience (years), teacher gender, teacher attended college, number of boarding students in the class, class size.
4) Cluster (class-level)-robust standard errors in parentheses.
5) *** p<0.01, ** p<0.05, * p<0.1.

Appendix B: Cost Comparison

The main costs of the CAL program and workbook treatment sessions are for training facilitators, paying facilitators to run the sessions, developing the software or workbook, duplicating the software or workbook, and computer and Internet costs for the software. We assume that both the CAL software and workbooks have a limited shelf life. We use the ingredient approach to measure costs (Levin and Belfield 2015; Levin, et al. 2017).

<u>CAL Program Costs</u>

Facilitator training: The cost to train facilitators includes communication costs (3 training sessions ∗ 10 RMB/training session = 30 RMB), training materials (20 RMB), and trainer remuneration (30 RMB). The teacher training subtotal is 80 RMB/teacher, which is equivalent to 6.67 RMB/student (assuming that the number of participants is 12).

Facilitator stipends: Class subsidies are given to program teachers for implementing the CAL sessions; this costs 850 RMB/teacher (for 17 weekly sessions at 50 RMB per session). This comes out to 850 / 12 = 78.33 RMB/student.

Software development: The cost to design and develop the software is a one-time expenditure. Assuming that the software will last for 5 years, its per-student unit cost is 200,000 RMB / 5 years / 88 classes / (12 students/class) = 37.88 RMB/student.
Reproduction costs: 0
Computer and Internet costs: 0 (conservatively assuming that these already exist for regular classes and no extra wear and tear costs from CAL sessions)

Total cost: Based on the above, the total cost for the supplemental CAL intervention is: 6.67+78.33+37.88=122.80 RMB per student (roughly $18 USD).

<u>Workbook Sessions Costs</u>

Facilitator training: The cost to train facilitators includes communication costs (3 training sessions ∗ 10 RMB/training session = 30 RMB), training materials (20 RMB), and trainer remuneration (30 RMB). The teacher training subtotal is 80 RMB/teacher, which is equivalent to 6.67 RMB/student (assuming that the number of participants is 12).

Facilitator stipends: Class subsidies are given to program teachers for implementing the workbook sessions, and cost 1 classes/week ∗ 17 weeks ∗ 50 RMB/class = 850 RMB/teacher. This comes out to 850 / 12 = 78.33 RMB/student (roughly $18 USD).

Workbook development: The cost to design and develop the workbook is a one-time expenditure. Assuming that the workbook content will last for 5 years, its per-student unit cost is 5,300 RMB / 5 years / 88 classes / (12 students/class) = 1 RMB/student.
Reproduction costs: The cost to photocopy and ship the workbook per student is 11 RMB.

Total cost: Based on the above, the total cost for the supplemental workbook intervention is: 6.67+78.33+1+11=97 RMB per student (roughly $14 USD).