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

YUE MA
Stanford University

ROBERT FAIRLIE
University of California, Los Angeles

PRASHANT LOYALKA and SCOTT ROZELLE
Stanford University

I. Introduction

Computer-assisted learning (CAL), online courses, massive open online courses, 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 is projected to grow to more than US$340 billion by 2025 (Escueta et al. 2017). With the large-scale, comprehensive movement of schoolchildren and college students in 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. The scale of the substitution of EdTech for classroom learning is remarkable: more than 1.5 billion schoolchildren around the world have moved to online learning because of social-distancing restrictions (UNESCO 2020). The phased return to classroom-based teaching will take time, and it is likely that the pandemic accelerated the longer-term shift from in-person learning to online learning.

Although EdTech is rapidly being deployed throughout the developed and developing world, the relatively limited evidence on whether and how it affects

We would like to thank Dell Global Giving and the TELOS Initiative at the Graduate School of Education at Stanford for financing the project. We thank David Card, Thomas Dee, Mark Duggan, Oliver Falck, Caroline Hoxby, Anna Kerkhof, Natalia Lazzatti, Hongbin Li, Bryan Pratt, Jon Robinson, Abhijeet Singh, Sean Sylvia, Sebastian Wichert, and seminar participants at the University of California, Berkeley, Stanford University, University of California, Merced, Syracuse University, University of Munich/CESifo, and Montana State University for comments and suggestions. We thank Gagandeep Sachdeva for research assistance. Contact the corresponding author, Robert Fairlie, at rfairlie@ucla.edu.

Electronically published June 3, 2024

Economic Development and Cultural Change, volume 72, number 4, July 2024.

https://doi.org/10.1086/726064

academic outcomes is mixed (Glewwe et al. 2013; Bulman and Fairlie 2016; Escueta et al. 2017).[1] The effects are often identified through randomized experiments in which the treatment group that receives supplemental CAL is compared with a control group that receives no inputs. Although usually attributed to the EdTech component, the estimated effects of the supplemental use of CAL, however, include other non-technology-based inputs. 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 program effects or whether another input is driving the effects. This represents a more general problem in evaluating and interpreting the effects of any supplemental education intervention program (e.g., after-school tutoring and community technology centers) because these programs also consist of additional inputs such as more time on learning.[2]

In this paper, we clarify and discuss the channels by which supplemental CAL may affect academic outcomes among schoolchildren, and we then present estimates that clarify the channels by which CAL affects those outcomes. A theoretical model that lays out these channels more formally is presented in appendix A. The technology-based components of CAL include engaging video- and game-based material, rapid feedback on problems through provision of assistance or answers, adjustment of the difficulty of problems for each student, and input through a keyboard instead of writing.[3] To generate exogenous variation in CAL and other inputs, we design and conduct a randomized controlled trial (RCT) involving more than 4,000 fourth-to-sixth-grade students across

[1] Recent evaluations of supplemental learning CAL programs across a wide range of software types find large positive effects on academic outcomes (e.g., Banerjee et al. 2007; Lai et al. 2013, 2015; Böhmer 2014; Mo et al. 2014). For the less common use of CAL as a direct substitute for regular teacher instruction in the classroom, the evidence tends to show null effects (Dynarski et al. 2007; Linden 2008; Barrow, Markman, and Rouse 2009; Campuzano et al. 2009; Carrillo, Onofa, and Ponce 2011), but this might depend on how computers are used (Falck, Mang, and Woessmann 2018) or what levels of CAL are being used (Bettinger et al. 2023). Finally, the less structured provision of computers and laptops for home and/or school use among schoolchildren tends to show null or mixed effects (e.g., Fuchs and Woessmann 2004; Schmitt and Wadsworth 2006; Machin, McNally, and Silva 2007; Fiorini 2010; Malamud and Pop-Eleches 2011; Fairlie and London 2012; Fairlie and Robinson 2013; Beuermann et al. 2015; Cristia et al. 2017; Hull and Duch 2019; Malamud et al. 2019).

[2] Understanding the role of various inputs in improving student learning outcomes is especially important in many developing countries where students have low and stagnant learning outcomes (see Muralidharan and Zieleniak 2013; Pritchett 2013).

[3] See the reviews in Glewwe et al. (2013), Bulman and Fairlie (2016), and Escueta et al. (2017), as well as Ebner and Holzinger (2007), Burguillo (2010), Van der Kleij et al. (2015), Vincent (2016), Van Klaveren, Vonk, and Cornelisz (2017), and Muralidharan, Singh, and Ganimian (2019) for more discussion.

352 math classes in 130 schools in rural China.[4] The RCT includes three treatment arms: (i) supplemental CAL, (ii) traditional supplemental learning (i.e., solving problems by use of pencils and paper workbooks), and (iii) a pure control that receives no program intervention. 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 and used to test whether CAL can substitute for traditional learning. The workbooks, however, do not replicate "tech" components such as engaging video- and game-based material, rapid feedback on problems through provision of assistance or answers, and input through a keyboard instead of writing. 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 intervention.

There are multiple channels by which a supplemental CAL program can affect educational outcomes, but there is no theoretical 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, the supplemental CAL program has no effect on students' math test scores. When we isolate the technology-based effects of CAL, we also do not find evidence of positive effects on test scores. Given well-documented gender differences in computer use, benefits, and achievement, we examine effects for boys and girls separately.[5] Focusing on boys, we find that the CAL program increases math test scores by 0.10σ. Isolating the technology effects from CAL, however, the point estimates for the CAL technology effect are smaller and lose statistical significance. For girls, we do not find positive estimates of the CAL program effect or the isolated CAL technology effect on test scores.

Following previous studies, we also examine CAL program and technology effects on class grades and find that, for the average student, the supplemental CAL program increases students' math grades.[6] However, when we isolate the technology-based effects of CAL, we find point estimates that are small and statistically indistinguishable from zero. On the other hand, the differential between the

[4] Of the 185 million schoolchildren in China, roughly 75% live in rural areas (Chen, Yang, and Ren 2015). Technology might have long-term benefits for rural children in China and help close some of the urban/rural achievement gap. Bianchi, Lu, and Song (2020) find, e.g., that an earlier large-scale program that connected urban teachers to rural schoolchildren using satellite internet and physical CDs improved academic achievement and labor market outcomes.

[5] See Hannum and Park (2007), Xu and Jaggers (2014), Algan and Fortin (2018), Xu and Li (2018), and Eble and Hu (2019).

[6] For a few examples, see Malamud and Pop-Eleches (2011), Fairlie and London (2012), Fairlie and Robinson (2013), and Bergman (2021).

estimated overall CAL program effect and the estimated CAL-technology-specific effect is positive and statistically significant. For boys, we find that the CAL program increases math grades by 3.4 percentile points, but the point estimates for the CAL technology effect are notably smaller and statistically indistinguishable from zero. We do not find positive estimates of the CAL program effect or the isolated CAL technology effect on math grades for girls.

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 the isolated CAL technology effect both increase how much students like their math class.

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. 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 crowding 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. The experimental design also provides new evidence on the substitutability of CAL for traditional learning, which has implications for the full-scale substitution of technology that may sometimes occur (e.g., during the COVID-19 pandemic).

Our paper builds on three previous studies that explore technology-specific effects.[7] In one recent study, Muralidharan, Singh, and Ganimian (2019) find large positive effects of after-school Mindspark Center programs in India, which include both extensive software use and instructional support. To separate out the effects of the instructional support and extra learning time inputs of the program, they compare their effect estimates with those 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 effects on student outcomes, suggesting that additional instructional time and tutoring were not the

[7] 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; Linden 2008; Barrow, Markman, and Rouse 2009; Campuzano et al. 2009; Carrillo, Onofa, and Ponce 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.

key drivers of the Mindspark effects. A second recent study evaluates a pilot program that involves the training of teaching assistants (TAs) to deliver a structured package of literacy materials to groups of three to four young children in England (Johnson et al. 2019). Cross-randomizing the TA teaching with information and communication technology (ICT) or paper equivalent sessions, they find positive effects for both (slightly larger for non-ICT). The emphasis of the experiment on evaluating the TA intervention, small group assignment, and implementation of this program in the classroom, however, make it more difficult to isolate the technology-based effects of the ICT session. In a third recent study, Büchel et al. (2022) conduct an effect evaluation on a broader education program in El Salvador that includes after-school Khan Academy lessons with teachers, Khan Academy lessons with nonteacher supervisors, lessons with teachers that repeat the curriculum, and a control group (grades 3–6). Although there were problems with high absentee rates of both students and teachers and lack of preparation among teachers, the experiment provides evidence that Khan Academy lessons had positive effects on learning of math relative to traditional lessons by teachers.

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 similar to that of 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—the careful choice of the control condition. By comparing CAL with pencil and paper-workbook estimates directly, we provide cautionary 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 a paper workbook. Taken together, the findings have particular relevance to the questions 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.

Finally, the findings contribute to a rapidly growing literature on the effects of school closures on schoolchildren during the pandemic (e.g., Azevedo et al. 2020; Maldonado and De Witte 2020; Clark et al. 2021; Grewenig et al. 2021; Angrist, Bergman, and Matsheng 2022; Li et al. 2023). The findings shed light on whether the extensive substitution of online learning for traditional learning due to COVID-19 is likely to have net negative effects on academic outcomes among schoolchildren. Our finding that EdTech and workbook exercise sessions of equal time and content outside of school hours had the same effect on standardized math test scores and grades in math classes suggests that EdTech might provide at least some substitution for traditional learning.

The remainder of the paper is organized as follows. In section II, we discuss the channels by which CAL might affect academic outcomes among schoolchildren relative to traditional "pencil-and-paper" forms of learning. This section also describes the design and implementation of the experiment. We present our empirical results in section III, and we conclude in section IV.

II. Estimation, Experimental Design, and Data

A 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 whether time spent using computer-based learning offsets traditional learning is especially salient because of the flexibility of this time.

As presented in appendix A, we create a theoretical model to illustrate the channels by which CAL might affect academic outcomes among schoolchildren. In brief, the model starts by adding computer resources such as CAL to a standard model of education production with the binding constraints being 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. The model formalizes an important, and intuitively plausible, insight that the total (marginal) effect of CAL on academic achievement is composed of a direct effect of increasing CAL time on math and an indirect effect through increasing total time spent learning math. The distinction is important and guides the design of our field experiment and estimating equations.

A. Direct and Indirect Effects

CAL might have a direct or "technology" effect on academic achievement independent of more time on math. 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 with the feedback associated with traditional modes of learning (Van der Kleij, Feskens, and Eggen 2015). Another "technology" aspect of CAL that could be important is that it can be adaptive or personalized for students. The adaptive component might be especially advantageous when students are below grade level and have widely differing levels of preparation (Muralidharan, Singh, and Ganimian 2019). In countries where student preparation is stronger, more regimented, and more homogeneous, however, adaptive CAL might not work better than regular CAL (Van Klaveren, Vonk, and Cornelisz 2017).

On the other hand, the game-based nature of CAL might reduce interest in completing traditional homework or learning in class and 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.

Turning to the indirect effect, we can view this effect 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 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 study because of the overall time constraint. 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 with 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.

Arguably, the component of most interest is the direct or technology-based effect of CAL on academic performance.[8] 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

[8] In the theoretical model presented in app. A, the direct or technology-based effect of CAL is θ in eq. (A3′).

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 component 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, which makes traditional forms of learning less interesting (crowd out) or builds confidence (crowd in), is useful.

The total effect of implementing a CAL program 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, implementation of a CAL program often includes additional inputs such as provision of new learning material outside the standard curriculum, additional instructional support by teachers or aides running the sessions, and more attention to schoolchildren in sessions.

A few recent 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 (0.12σ to 0.18σ in math) from 40 minutes of instruction two times a week. Muralidharan, Singh, and Ganimian (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 six 90-minute sessions a week. Böhmer (2014) finds large positive effects from an after-school program providing CAL and student coaches in South Africa (0.25σ in math) over the course of a year. These studies essentially estimate the total effect without identifying the technology component.[9] Because this equation includes the total effects from an increase in time on math in addition to an increase in time on the computer, it might favor finding positive effects on academic outcomes. Outside of the model, many of the CAL programs evaluated in previous literature include additional educational inputs such as coaches and tutoring sessions, which further complicate the interpretation of CAL effects on academic outcomes. To investigate this issue, we directly estimate the technology component using our experiment. Specifically, we make use of two treatment arms and a control group to isolate the effects of the different inputs.

[9] In the theoretical model presented in app. A, the total effect of implementing a CAL program is captured by (A3′), the technology parameter is θ, time on math is T^{M}, and time on the computer is T^{C}.

B. Estimation

Estimating the relationships discussed above is complicated for two primary reasons. First, academic performance and CAL use are 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

$$A_i^0 = \beta X_i + \gamma S_i + \lambda T_i^{TR\_0}, \tag{1}$$

$$A_i^{CAL} = \beta X_i + \gamma S_i + \theta + \lambda\left(T_i^{TR\_0} + 1 + \eta^C\right), \tag{2}$$

$$A_i^{WK} = \beta X_i + \gamma S_i + \lambda\left(T_i^{TR\_0} + 1 + \eta^{WK}\right), \tag{3}$$

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:

$$\theta = \bar{A}^{CAL} - \bar{A}^{WK} - \lambda(\eta^C - \eta^{WK}), \tag{4}$$

$$\lambda = \frac{\bar{A}^{WK} - \bar{A}^0}{1 + \eta^{WK}}, \tag{5}$$

where $\eta^C - \eta^{WK}$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. Student, teacher, and school characteristics and the base or control level of traditional learning time on math, $T_i^{\mathrm{TR}\_0}$, are balanced in expectations because of the RCT.

We estimate these parameters (which are also discussed in app. A) represented in equations (1)–(3) using the following regression:

$$Y_{ij} = \alpha_0 + \alpha_1\left(-D_{1j}\right) + \alpha_2\left(-D_{2j}\right) + \mathbf{X}_{ij}\beta + \mathbf{S}_{ij}\gamma + \tau_c + \varepsilon_{ij}, \tag{6}$$

where Y_{ij} is the academic outcome of interest measured at end line 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; $\mathbf{X}_{ij}$ is a vector of baseline student control variables, $\mathbf{S}_{ij}$ 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, α_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, $\mathbf{X}_{ij}$ 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.

C. *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 discussed in appendix A. The field experiment involves more than 4,000 fourth-to-sixth-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-grade class, one fifth-grade class, and one sixth-grade class that each had at least four boarding students. This resulted in a total of 352 classes or school-grades (instead of 130 schools × 3 grades = 390 school-grades). 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.

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 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 end-line) survey.

Regarding external validity, we focused on boarding students because there was no time at which to provide after-school CAL and workbook sessions to nonboarding students. Boarding students represent 37% of students in our schools. We find that boarding and nonboarding students are similar across numerous characteristics (results are available by request).[10] Although these students do not appear to be substantially different across observable characteristics, we have to be cautious in generalizing our results to the broader population of students in China.[11] Boarding students and their families face different constraints than those face by nonboarding students. But, boarding students are interesting in their own right. There are 32 million primary and junior high boarding students in China, which represents 32% of all students (Ministry of Education 2017). We also focus on China, which even in rural areas tends to have relatively good access to and familiarity with computers and internet compared with poorer parts of the world. On one hand, improved access and familiarity could accentuate the positive effects of EdTech, but on the other hand, they could limit the novelty and excitement of use of EdTech by schoolchildren.

D. Baseline Survey

The baseline survey collected information on students, teachers, and school principals. The student survey collected information about student and household

[10] The summary statistics reported in table A1 pertain to the experimental sample (boarding students only). In regard to line 1 of panel A, the larger sample of students (boarding and nonboarding) was used to standardize baseline math scores; as baseline math score averages and SD are close to 0 and 1, respectively, this indicates the distributions of math scores for boarding students and nonboarding students are quite similar.

[11] Boarding students are similarly likely to be girls, are slightly younger, like math slightly less at baseline, and have slightly less educated parents.

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.

E. Randomization

We designated each of 27 county-grades (9 counties and 3 grades) in our sample of 130 schools and 352 classes (or school-grades) 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). As a result of the randomization, 116 classes in 88 of the schools were assigned to supplemental CAL (T1), 118 classes in 86 of the schools were assigned to supplemental workbook (T2), and 118 classes in 85 of the schools were assigned to the control group.[12]

To ensure adequate sample sizes, power calculations were conducted before the beginning of the trial (Spybrook et al. 2009).[13] We expected to lose a small amount of statistical power because of student attrition. On the basis of our experience, we assumed an attrition rate of 5%. The actual attrition rate from baseline to end line was only 2.4%.

F. Program (Treatment) Administration

The CAL and workbook programs were implemented by a university-based nongovernmental organization (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

[12] Our sample still consisted of 130 schools and 352 classes (school-grades). However, because we randomized the total 352 classes within county-grades (strata), it was not necessary that the classes in any given school had all three treatment conditions. For example, the grades 4, 5, and 6 classes in school A may have all been assigned to supplemental CAL, while the grades 4, 5, and 6 classes in school B may have been assigned to supplemental workbook, control, and supplemental workbook, respectively.

[13] 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^2 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 of 0.14 SD.

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 allowed to assist students only with scheduling, computer hardware issues, software operations, and handing out and 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 and were held on the same days of the week, for the same amount of time, and for the same number of times during the school year. They also had the same curricular content each week and the same facilitator training and instructions. At end line, 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 fig. A1). 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 students in less developed countries, students in rural China are much more likely to be on grade level in terms of achievement outcomes (Khor et al. 2016). 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. 2023). Software differs along many dimensions, of course, and thus some caution is warranted in generalizing the results to different applications, but the software we evaluate here has been shown to work and is widely used in China.

G. End-Line Survey and Primary Outcomes

We conducted the end-line survey with the students, teachers, and principals. As in the baseline, students took a 35-minute standardized math exam.[14] In the analyses, we convert end-line math exam scores into z-scores by subtracting the mean end-line 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 noncognitive dimensions of human capital and are predictive of later life outcomes (Borghans et al. 2016).[15] 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 on the basis of 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 spent on math homework, and time spent on language homework.[16]

H. Balance Check

Table A1 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 treatment group indicator

[14] Like the baseline test, the end-line 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 (because student learning is, on average, below grade level), this was not the case in our sample schools. Our baseline and end-line 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. Test items for mathematics 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 by using data from extensive pilot testing. The tests had good psychometric properties (Cronbach alpha of approximately 0.8, unidimensionality, and a lack of differential item functioning by gender). An analysis of the pilot, baseline, and end-line test results also indicated that the tests did not suffer from floor or ceiling effects.

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

[16] As stated in our preanalysis plan, math exam scores are the primary outcome for the study while the rest are secondary outcomes (that we do not power for and do not adjust for multiple hypothesis testing).

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).

III. Results

Estimates of equation (6) for math test scores are reported in table 1. For the full sample, we find a positive point estimate but no statistical evidence of a positive effect of the overall CAL program on math test scores. Turning to isolation of the technology effect of CAL, we also find no effect on test scores. Finally, the CAL technology estimate is not statistically different from the CAL program estimate.

As noted above, a 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. This crowd out could be responsible for the null effect finding. Additionally, the next step toward recovering the technology component (θ) defined in equations (4) and (5) is to estimate

TABLE 1
CAL PROGRAM AND TECHNOLOGY EFFECTS ON MATH TEST SCORES

	All Students (N = 3,928)		Boys Only (n = 2,142)		Girls Only (n = 1,785)	
	(1)	(2)	(3)	(4)	(5)	(6)
CAL program (CAL treatment − control)	.033	.032	.099**	.099**	−.044	−.045
	(.039)	(.039)	(.049)	(.049)	(.046)	(.046)
CAL technology (CAL treatment − workbook treatment)	.059	.061	.075	.074	.039	.041
	(.044)	(.044)	(.054)	(.054)	(.054)	(.054)
CAL program – CAL technology (workbook treatment − control)	−.026	−.029	.025	.025	−.084	−.086
	(.046)	(.046)	(.060)	(.060)	(.052)	(.052)
Additional controls	No	Yes	No	Yes	No	Yes
R^2	.432	.436	.442	.445	.432	.437

Note. 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). All columns control for baseline math score. Even-numbered columns also control for the following baseline covariates: liking math (scale 1–100), 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. Cluster (class-level)-robust standard errors are in parentheses.
** $p < .05$.

TABLE 2
CAL PROGRAM AND WORKBOOK TREATMENT EFFECTS ON TIME ON MATH OUTSIDE OF SCHOOL (HOMEWORK HOURS)

	All Students (N = 3,930)		Boys Only (n = 2,145)		Girls Only (n = 1,784)	
	(1)	(2)	(3)	(4)	(5)	(6)
CAL treatment − control (η^{C})	−.149	−.154	.065	.065	−.387*	−.393**
	(.199)	(.199)	(.241)	(.242)	(.198)	(.197)
Workbook treatment − control (η^{WK})	.128	.123	.305	.298	−.068	−.085
	(.201)	(.201)	(.239)	(.240)	(.207)	(.206)
CAL treatment − workbook treatment	−.277	−.277	−.240	−.233	−.319	−.308
	(.198)	(.197)	(.243)	(.242)	(.202)	(.201)
Additional controls	No	Yes	No	Yes	No	Yes
R^2	.099	.099	.096	.098	.121	.125

Note. CAL treatment − control (η^{C}) and workbook treatment − control (η^{WK}) are reported for crowd out (crowd in) estimates of the two treatments (see eqq. [2] and [3]). Math homework time (hours last week): control group mean = 3.36, SD = 2.70. All columns control for baseline math score. Even-numbered columns also control for the following baseline covariates: liking math (scale 1–100), 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. Cluster (class-level)-robust standard errors are in parentheses.
* $p < .10$.
** $p < .05$.

whether homework time is affected by the two treatments. 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 (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.[17] Given these findings, we can interpret the CAL technology effects estimates presented in table 1 as estimates of the parameter θ in equations (A1′) and (A3′).

Boys and girls use computers differently, and there are much higher levels of video-game use among boys (Kaiser Family Foundation 2010; Fairlie 2015; Algan and Fortin 2018).[18] Additionally, boys and girls differ substantially in academic performance in schools in China (Hannum and Park 2007; Xu and Li 2018; Eble and Hu 2019). Thus, we estimate effects of CAL, which is video game based, separately for boys and girls. Table 1 also reports estimates of equation (6)

[17] 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).
[18] Program for International Student Assessment (PISA) data indicate that 47% of boys compared with 16% girls play a computer game every day (Algan and Fortin 2018).

for boys and girls. The patterns for CAL program and isolated technology effects are more apparent for boys.[19] The CAL program has positive and significant effects on boys' test scores (0.10σ). On the other hand, we find no evidence of positive CAL technology effects for boys. For test scores, the point estimates are smaller and statistically indistinguishable from zero. Estimates reported in table 2 for effects 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 parameter θ.

We find no evidence of significant effects of either the CAL program or CAL technology effects on the test scores of girls. Table 1 reports estimates of equation (6) for girls. The CAL program and CAL technology point estimates for test scores are small in magnitude and not statistically significant. The estimated effects for CAL might differ by gender because boys and girls engage differently with technology (Kaiser Family Foundation 2010; Fairlie 2015; Algan and Fortin 2018). Additional analyses do not reveal 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 2 for effects on homework time show some evidence of negative effects for girls.

We are finding smaller CAL program effects on test scores than the large positive estimates in previous studies (e.g., Banerjee et al. 2007; Lai et al. 2013, 2015; Böhmer 2014; Mo et al. 2014). Treatment intensity might or might not be a reason that we do not find strong positive effects on test scores for the total sample. Previously, some China studies' CAL interventions involved two session of 40 minutes a week for 2 or 3 semesters (Lai et al. 2013; Mo et al. 2014), but some studies found a significant effect with intervention of 2 sessions of 40 minutes a week for 1 semester (Lai et al. 2015), which would be similar to our treatment intensity of one 40-minute session a week for 2 semesters. As noted above, the software is similar to software used in previous studies, so this is unlikely to explain differences. The types of tests are also unlikely to explain differences because they were also similar to those of previous studies in China. Admittedly, it is not clear what might explain the differences in findings for test scores.

A. Effects on Grades

Estimates of equation (6) for math grades are reported in table 3. For the full sample, we find a positive and statistically significant effect of the overall CAL

[19] We note that the results presented in table 1 should be interpreted with some caution because of the nine reported coefficients, only one is statistically significant.

TABLE 3
CAL PROGRAM AND TECHNOLOGY EFFECTS ON MATH GRADES: ALL STUDENTS

	Grade (Rank; $N = 3{,}829$)		Grade Rank, Class $N \geq 10$ ($N = 3{,}750$)	
	(1)	(2)	(3)	(4)
CAL Program (CAL treatment − control)	1.743*	1.758*	1.866**	1.876**
	(.919)	(.922)	(.925)	(.929)
CAL technology (CAL treatment − workbook treatment)	.212	.155	.234	.178
	(.996)	(.999)	(1.013)	(1.017)
CAL program − CAL technology (workbook treatment − control)	1.531*	1.603*	1.632*	1.697*
	(.877)	(.876)	(.895)	(.894)
Additional controls	No	Yes	No	Yes
R^2	.300	.308	.299	.308

Note. 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). All columns control for baseline class rank in math. Even-numbered columns also control for the following baseline covariates: liking math (scale 1–100), 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. Cluster (class-level)-robust standard errors are in parentheses.
* $p < .10$.
** $p < .05$.

program on the student's math grade. 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.[20] Turning to isolating the technology effect of CAL, we find no effect on math grades. For math grades, even in the 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 (1.53–1.70 percentile points). As noted above, by the finding of null effects on homework time by either the CAL or workbook treatments, we can interpret the CAL technology effects estimates presented in table 1 as estimates of the parameter θ in equations (A1′) and (A3′).

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

[20] The resulting sample size is only 2% smaller than the total sample size. The median class size in the sample is 36.

is due to additional inputs and that the isolated technology-based CAL effect is zero.

We also estimate effects of CAL on math grades separately for boys and girls. Tables 4 and 5 report estimates of equation (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 boys' test performance in math class (3.4 percentile points). On the other hand, we find no evidence of positive CAL technology effects for boys. For 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% level. 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 spent learning math. Estimates reported in table 2 for effects 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 parameter θ.

We find no evidence of significant effects of either the CAL program or CAL technology effects on math grades of girls. Table 5 reports estimates of equation (6) for girls. The CAL program and CAL technology point estimates are small in magnitude, inconsistent in sign, and not statistically significant.

TABLE 4

CAL PROGRAM AND TECHNOLOGY EFFECTS ON MATH GRADES: BOYS ONLY

	Grade (Rank; n = 2,095)		Grade, Class $N \geq 10$ (n = 2,053)	
	(1)	(2)	(3)	(4)
CAL program (CAL treatment − control)	3.414***	3.430***	3.488***	3.506***
	(1.280)	(1.277)	(1.291)	(1.288)
CAL technology (CAL treatment − workbook treatment)	1.540	1.530	1.482	1.455
	(1.456)	(1.449)	(1.480)	(1.472)
CAL program − CAL technology (workbook treatment − control)	1.874	1.900	2.006	2.051
	(1.218)	(1.223)	(1.249)	(1.253)
Additional controls	No	Yes	No	Yes
R^2	.307	.311	.307	.312

Note. 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). All columns control for baseline class rank in math. Even-numbered columns also control for the following baseline covariates: liking math (scale 1–100), 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. Cluster (class-level)-robust standard errors are in parentheses.

*** $p < .01$.

TABLE 5
CAL PROGRAM AND TECHNOLOGY EFFECTS ON MATH GRADES: GIRLS ONLY

	Grade (Rank; n = 1,733)		Grade, Class $N \geq 10$ (n = 1,696)	
	(1)	(2)	(3)	(4)
CAL program (CAL treatment − control)	−.526	−.590	−.345	−.434
	(1.371)	(1.371)	(1.374)	(1.373)
CAL technology (CAL treatment − workbook treatment)	−1.451	−1.350	−1.365	−1.268
	(1.519)	(1.526)	(1.540)	(1.549)
CAL program − CAL technology (workbook treatment − control)	.925	.759	1.019	.834
	(1.375)	(1.396)	(1.394)	(1.416)
Additional controls	No	Yes	No	Yes
R^2	.302	.308	.299	.305

Note. 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). All columns control for baseline class rank in math. Even-numbered columns also control for the following baseline covariates: liking math (scale 1–100), 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. Cluster (class-level)-robust standard errors are in parentheses.

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 (6) for whether students report liking their math class.[21]

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 that 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

[21] The end-line survey question was worded carefully to refer to the student's math class and not to the CAL or workbook sessions.

TABLE 6
CAL PROGRAM AND TECHNOLOGY EFFECTS ON LIKING MATH CLASS (SCALE 1–100)

	All Students (N = 3,931)		Boys (n = 2,145)		Girls (n = 1,785)	
	(1)	(2)	(3)	(4)	(5)	(6)
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 − technology)	−.094	−.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
R^2	.170	.172	.163	.165	.195	.201

Note. 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). All columns control for baseline liking math class (scale 1–100), control group mean = 87.2. 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. Cluster (class-level)-robust standard errors are in parentheses.
* $p < .10$.
** $p < .05$.

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.

B. *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 (table A2). For having a test score above the median, we find little evidence of significant effects for either the CAL program or CAL technology. For receiving a grade 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 posttreatment outcome distribution (Parente and Santos Silva 2016). Figures A2 and A3 display estimates and 95% 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 effects reported in tables 1 and 3. Mean effect estimates do not appear to be hiding differential effects at different parts of the distribution. We thus focus on mean effects.

C. *Heterogeneity on Initial Math Ability*

We estimate CAL program and CAL technology effects by baseline math ability terciles. Table A3 reports estimates of equation (6) 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 end-line math class rank and CAL technology point estimates that are notably smaller and are not statistically distinguishable from zero. We find no discernible effects on end-line 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.

D. *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% have used computers in school at baseline. Nevertheless, we estimate the test score and grade regressions with only students who self-report use of 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 use of computers at school.

Second, we examine whether the estimates of effects on student grades 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 (which is reported in cols. 1 and 2 of table 3) excluding classes with 90% or more boarding students, 80% or more boarding students, and 70% or more boarding students. We find that the CAL program coefficients remain positive and roughly similar in magnitude although they lose some statistical power (tables A4–A6). 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.

IV. Conclusions

Although EdTech is rapidly expanding around the world and accelerated in response to the COVID-19 pandemic, relatively little is known about the advantages and disadvantages of use of technology in education. Is EdTech, as proponents argue, revolutionizing the way in which students learn? 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 theoretical parameters, 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, the overall CAL program and the isolated technology component of CAL have no effect on math test scores. Given gender differences in computer use, we examined effects for boys and girls separately. For boys, we find that the CAL program increases 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. For girls, we do not find positive effects of the CAL program or CAL technology component on test scores.

Turning to math grades, estimates from the experiment indicate that for the full sample the CAL program improves grades, whereas the isolated technology component of CAL has no discernible effect on grades. The difference between the two estimated effects is statistically significant. For boys, we find that the CAL program increases math grades by 3.5 percentile points, but the point estimates become noticeably smaller and statistically indistinguishable from zero for the CAL technology effect. The stronger effects on grades than on test scores might be due to the program having greater effects on noncognitive than on cognitive skills. We also find no evidence of substitution effects of the CAL and workbook sessions on homework time in math. On the other hand, we find evidence suggesting that 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 pencil and paper-workbook sessions and the CAL program are similarly timed, have similar content, and are similarly structured, and the pencil and paper-workbook sessions show effects on academic performance that are roughly similar to those of 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 spent 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 pencils and paper workbooks are also low and, in fact, are lower. The costs of photocopying workbooks are small. 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% lower costs with the conservative assumption of zero costs for computers and internet (see app. B).

An area of promise is that we find evidence of a positive effect of CAL technology on student interest in math but no effect on math interest from extra time spent 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 of extra time spent learning, in supplemental educational programs.[22] The results of this study raise 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, the findings also provide new evidence on the substitutability of CAL for traditional learning. More research is clearly needed on whether and to what degree EdTech can substitute for traditional learning both as a pedagogical tool and as a delivery platform. This is especially pertinent today in light of the full-scale, comprehensive, global movement to EdTech at all levels of education in response to COVID-19. 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? Our finding that EdTech and workbook exercise sessions of equal time and content

[22] 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 2021).

outside of school hours had the same effect on standardized math test scores and grades in math classes suggests that EdTech might provide at least some substitution for traditional learning. EdTech was certainly widely used to substitute as a platform for delivering content during the pandemic, but how much information was lost (or gained)? These are important questions as we continue to develop and expand EdTech programs.

Appendix A

Theoretical Model of Investment in EdTech

We sketch out a theoretical model that illustrates the channels by which CAL might affect academic outcomes among schoolchildren. Computer resources such as CAL are added to a standard model of education production.[23] 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:[24]

$$\begin{aligned} A_i &= f(X_i, S_i, T_i^{\mathrm{C}}, T_i^{\mathrm{M}}) \text{ subject to} \\ T_i^{\mathrm{M}} &= T_i^{\mathrm{TR}} + T_i^{\mathrm{C}}, \\ T_i^{\mathrm{M}} &+ T_i^{\mathrm{Oth}} \leq T, \\ P^{\mathrm{TR}} T_i^{\mathrm{TR}} &+ P^{\mathrm{C}} T_i^{\mathrm{C}} \leq B_i. \end{aligned} \tag{A1}$$

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 spent on learning math effect. Total time allocated to learning math consists of traditional learning,

[23] For examples, see Hanushek (1979, 1986), Figlio (1999), and Rivkin, Hanushek, and Kain (2005).

[24] See Hanushek (1979) for an early discussion of value-added models in the economics of education literature.

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 learning 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.[25] From equation (A1), the total marginal effect of CAL on academic achievement is

$$\frac{\mathrm{d}A}{\mathrm{d}T^{\mathrm{C}}} = \frac{\delta A}{\delta T^{\mathrm{C}}} + \frac{\delta A}{\delta T^{\mathrm{M}}}\frac{\delta T^{\mathrm{M}}}{\delta T^{\mathrm{C}}}. \tag{A2}$$

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

CAL might have a direct or "technology" effect on academic achievement independent of more time spent on math (i.e., $\delta A/\delta T^{\mathrm{C}} \neq 0$). 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 with 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 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 (A1), we can rewrite the total marginal effect of CAL on math achievement:

$$\frac{\mathrm{d}A}{\mathrm{d}T^{C}} = \frac{\delta A}{\delta T^{\mathrm{C}}} + \frac{\delta A}{\delta T^{\mathrm{M}}}\left(1 + \frac{\delta T^{\mathrm{TR}}}{\delta T^{\mathrm{C}}}\right). \tag{A3}$$

Here we can view the indirect effect (the second term in eq. [A3]) as having two parts. The first part is the effect of a one-to-one increase in math time by

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

increasing CAL time. As discussed in detail below, this part of the effect of introducing educational technology is important and often overlooked in 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 study because of the overall time constraint (i.e., $\delta T^{\mathrm{TR}}/\delta T^{\mathrm{C}} < 0$). The 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 with 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 (A1) and (A3) as

$$A_i = \beta X_i + \gamma S_i + \theta T_i^{\mathrm{C}} + \lambda T_i^{\mathrm{M}}, \tag{A1$'$}$$

$$\frac{\mathrm{d}A}{\mathrm{d}T^{\mathrm{C}}} = \theta + \lambda(1 + \eta)\,. \tag{A3$'$}$$

In this production function, θ captures the direct or technology-based effect of CAL on academic performance. 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. In contrast, the full equation represented by (A3′) captures the total effect of implementing a CAL program, which includes time learning math on the computer, total time learning math, and potential crowd-out effects on homework.

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).

CAL Program Costs

Facilitator training.—The cost to train facilitators includes communication costs (3 training sessions × 10 renminbi [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.—Zero.

Computer and internet costs.—Zero (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 US$18).

Workbook Session Costs

Facilitator training.—The cost to train facilitators includes communication costs (3 training sessions $\times$ 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 class/week $\times$ 17 weeks $\times$ 50 RMB/class = 850 RMB/teacher. This comes out to $850/12 = 78.33$ RMB/student.

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/student (roughly US$14).

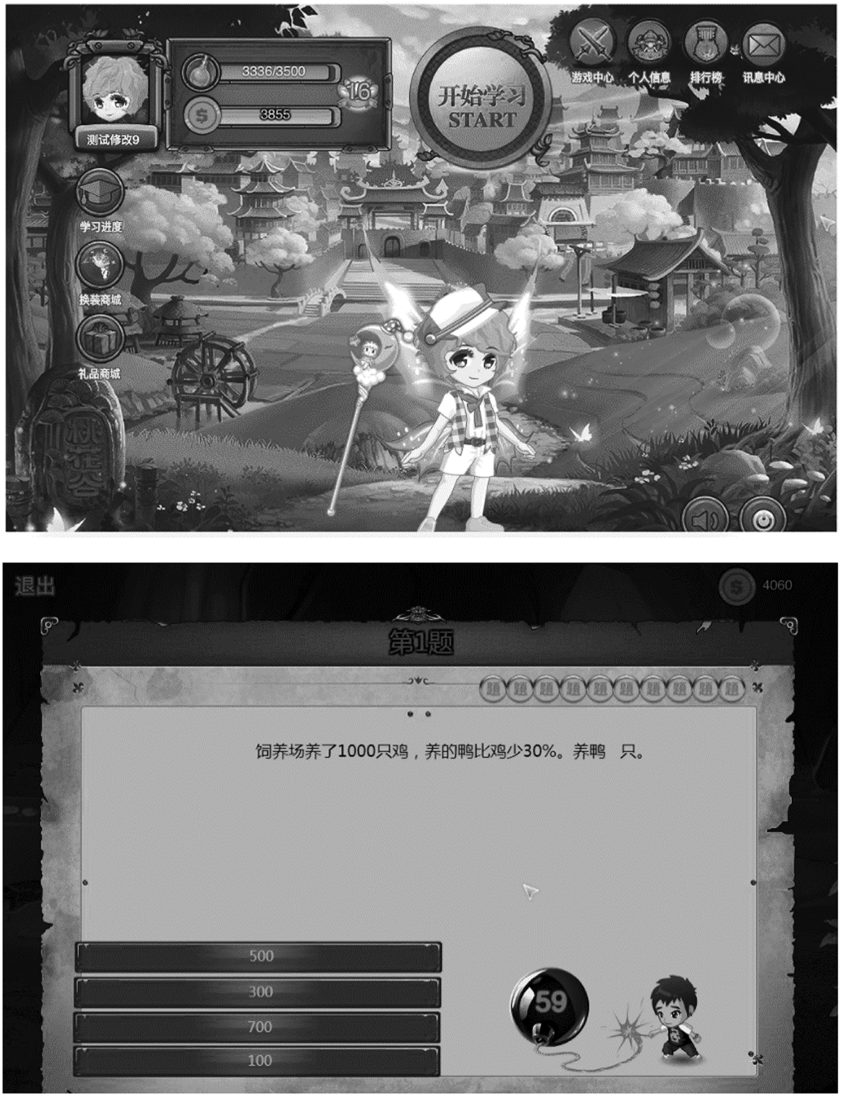

Figure A1. Example graphics from the CAL software.

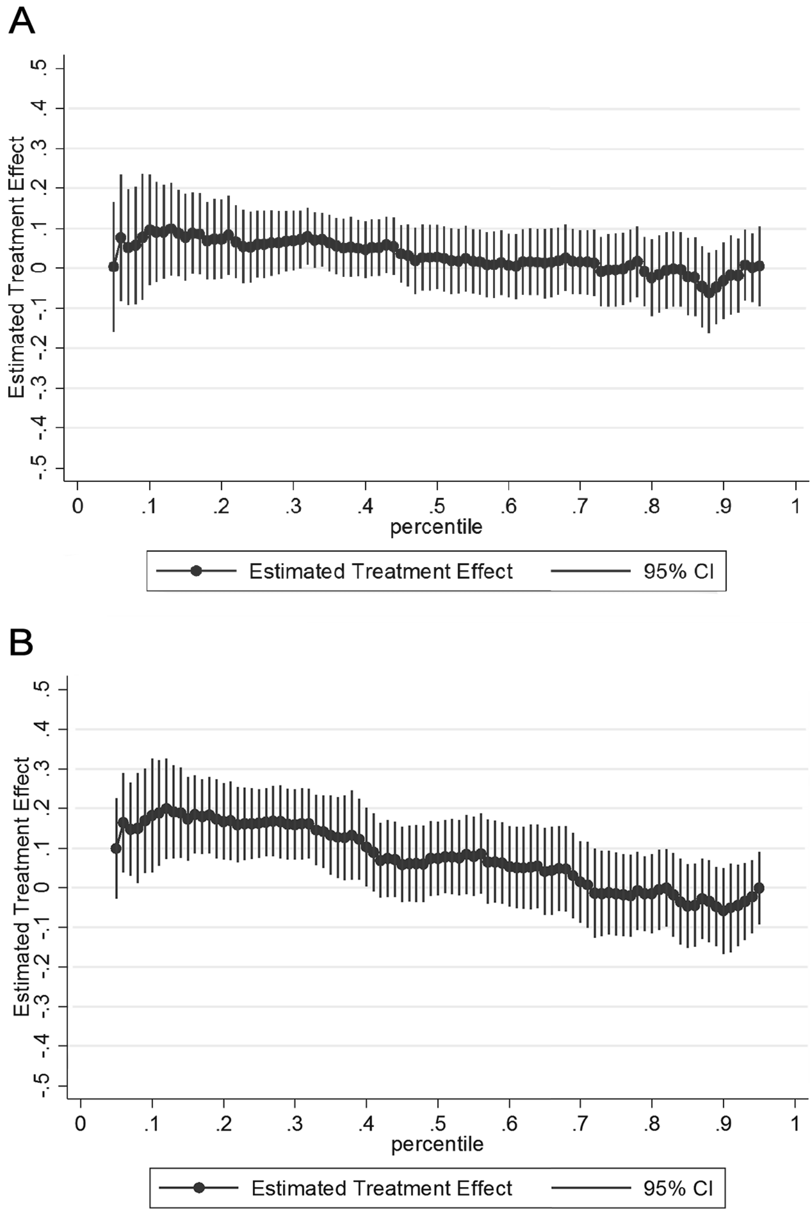

Figure A2. *A*, Quantile CAL program estimates for test score. *B*, Quantile CAL technology estimates for test score. CI = confidence interval.

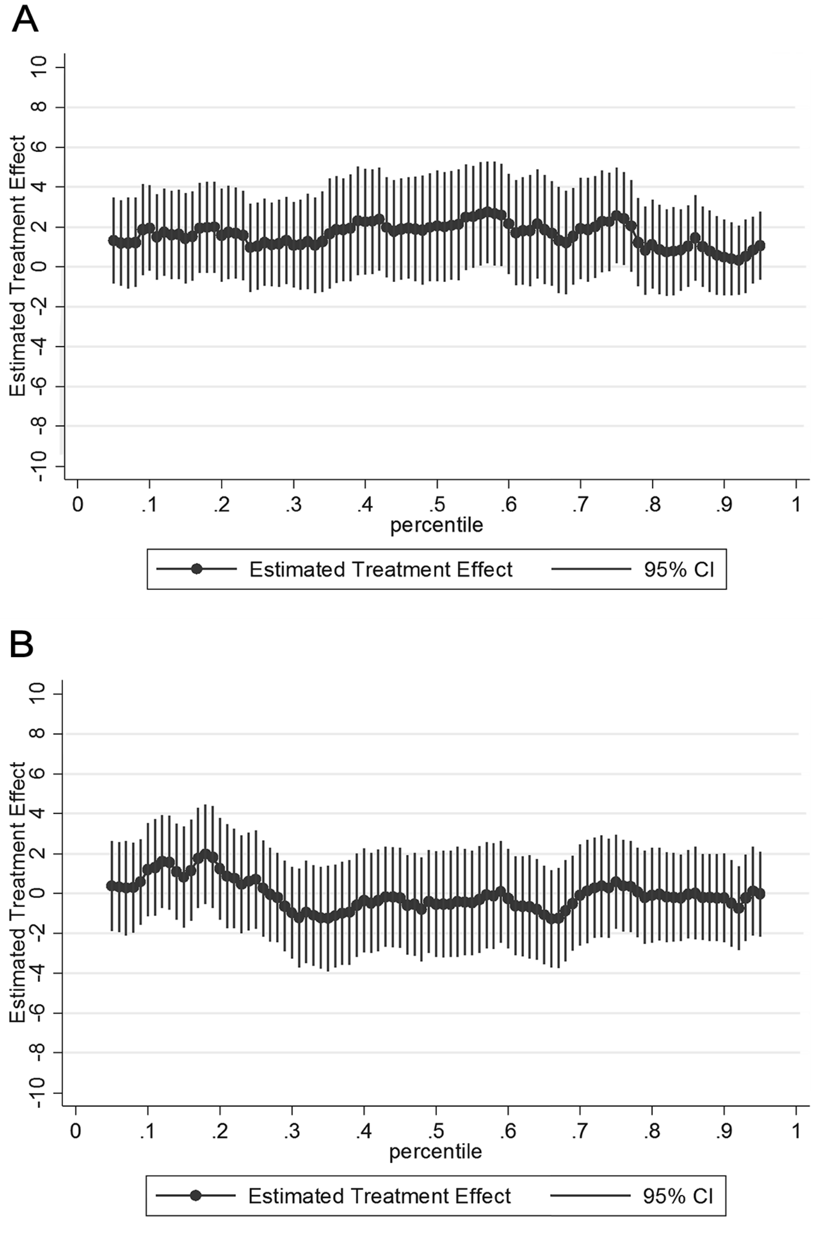

Figure A3. *A*, Quantile CAL program estimates for grade rank. *B*, Quantile CAL technology estimates for grade rank.

TABLE A1
SUMMARY STATISTICS AND BALANCE CHECK

	Control (1)	CAL (2)	Workbook (3)	*p*-Value (Col. 2 – Col. 1) (4)	*p*-Value (Col. 3 − Col. 1) (5)	*p*-Value (Col. 2 − Col. 3) (6)
			A. Student Characteristics			
1. Standardized math score	−.026	−.036	−.032	.669	.881	.776
	(.988)	(1.062)	(1.017)			
2. Within-class rank (using math score)	53.906	53.799	55.008	.860	.130	.191
	(29.498)	(30.539)	(29.453)			
3. Female (0/1)	.458	.459	.435	.762	.264	.163
	(.498)	(.499)	(.496)			
4. Age (years)	11.095	11.017	11.048	.874	.738	.612
	(1.069)	(1.115)	(1.109)			
5. Father education 9 years or less (0/1)	.441	.400	.424	.081	.554	.237
	(.497)	(.490)	(.494)			
6. Mother education 9 years or less (0/1)	.390	.356	.365	.180	.447	.474
	(.488)	(.479)	(.482)			
7. Liking math	−.052	.006	−.117	.419	.399	.124
	(1.050)	(1.015)	(1.087)			
8. Observations	1,390	1,345	1,289			
			B. Teacher and Class Characteristics			
1. Female (0/1)	.445	.391	.460	.334	.909	.417
	(.497)	(.488)	(.499)			
2. Experience (years)	16.239	13.425	15.424	.148	.843	.121
	(11.886)	(11.050)	(11.384)			
3. College degree (0/1)	.560	.569	.574	.667	.371	.670
	(.497)	(.496)	(.495)			
4. Number of boarding students	15.447	14.517	16.290	.162	.949	.270
	(6.776)	(6.632)	(79.547)			
5. Number of total students	35.426	32.717	35.322	.019	.223	.313
	(13.965)	(14.217)	(15.339)			
6. Observations	118	116	118			

Note. Means and SD (in parentheses) in cols. 1–3. The *p*-values in cols. 4–6 are calculated by 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 show no significant difference between T1 and C (*p*-value: .860/.124), T2 and C (*p*-value: .790/.862), or T1 and T2 (*p*-value = .184/.840). The above summary statistics pertain to the experimental sample (boarding students only), In regard to line 1 of panel A, the larger sample of students (boarding and nonboarding) was used to standardize baseline math scores.

TABLE A2

CAL PROGRAM AND TECHNOLOGY EFFECTS ON WHETHER MATH TEST SCORES/GRADES ARE ABOVE THE MEDIAN

	Math Test Score Above Median (N = 3,928)		Grade (Rank) Above Median (N = 3,829)		Grade (Rank) Above Median for Class $N \geq 10$ (N = 3,722)	
	(1)	(2)	(3)	(4)	(5)	(6)
CAL program	.016	.015	.036**	.036**	.038**	.038**
	(.020)	(.020)	(.018)	(.018)	(.018)	(.018)
CAL technology	.034	.035*	.010	.009	.012	.011
	(.021)	(.021)	(.018)	(.018)	(.019)	(.019)
Difference (program − technology)	−.018	−.020	.026	.027	.026	.027
	(.022)	(.022)	(.017)	(.017)	(.017)	(.018)
Additional controls	No	Yes	No	Yes	No	Yes
R^2	.247	.253	.165	.173	.165	.175

Note. 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). All columns control for baseline counterpart of dependent variable (baseline math score or baseline class rank in math above median [Y/N]). Columns 2 and 4 also control for the following baseline covariates: liking math (1–100), 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. Cluster (class-level)-robust standard errors are in parentheses.

* $p < .10$.

** $p < .05$.

TABLE A3

CAL PROGRAM AND TECHNOLOGY EFFECTS ON MATH TEST SCORES AND GRADES BY BASELINE ACHIEVEMENT TERCILES

	Lowest One-Third Achievers				Middle One-Third Achievers				Top One-Third Achievers			
	Math Score (n = 1,316)		Math Grade (Rank; n = 1,294)		Math Score (n = 1,310)		Math Grade (Rank; n = 1,280)		Math Score (n = 1,302)		Math Grade (Rank; n = 1,255)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CAL program	.059	.058	4.012**	3.842**	.006	−.002	−1.805	−1.681	.019	.021	3.168**	3.090**
	(.068)	(.068)	(1.822)	(1.802)	(.056)	(.057)	(1.582)	(1.609)	(.043)	(.043)	(1.556)	(1.568)
CAL technology	.049	.058	1.873	1.691	.067	.064	−2.879	−2.945	.037	.040	1.060	1.040
	(.070)	(.070)	(1.937)	(1.951)	(.064)	(.064)	(1.930)	(1.928)	(.050)	(.049)	(1.543)	(1.547)
Program − technology	.009	.00	2.139	2.151	−.061	−.066	1.074	1.265	−.017	−.019	2.108	2.050
	(.071)	(.072)	(1.621)	(1.661)	(.068)	(.068)	(1.736)	(1.735)	(.051)	(.050)	(1.570)	(1.560)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R^2	.192	.199	.164	.182	.118	.125	.121	.129	.154	.170	.089	.109

Note. 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). All columns control for baseline counterpart of dependent variable (baseline math score or baseline class rank in math). 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. Cluster (class-level)-robust standard errors are in parentheses.

** $p < .05$.

TABLE A4

CAL PROGRAM AND TECHNOLOGY EFFECTS ON MATH GRADES, EXCLUDING CLASSES WITH CERTAIN PERCENTAGES OF BOARDING STUDENTS

	Exclusion of Classes with Boarding Students > 90% (n = 3,571)		Exclusion of Classes with Boarding Students > 80% (n = 3,314)		Exclusion of Classes with Boarding Students > 70% (n = 3,082)	
	(1)	(2)	(3)	(4)	(5)	(6)
CAL program	1.306	1.318	1.562	1.616	1.356	1.387
	(.912)	(.912)	(.993)	(.996)	(1.029)	(1.029)
CAL technology	−.085	−.081	−.174	−.146	−.617	−.560
	(1.060)	(1.061)	(1.110)	(1.108)	(1.224)	(1.225)
Difference (program − technology)	1.392	1.399	1.737*	1.762*	1.973*	1.947*
	(.928)	(.927)	(.968)	(.971)	(1.073)	(1.077)
Additional controls	No	Yes	No	Yes	No	Yes
R^2	.295	.304	.291	.300	.294	.302

Note. 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). All columns control for baseline counterpart of dependent variable (baseline class rank in math test score). Even-numbered columns also control for the following baseline covariates: liking math (scale 1–100), 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. Cluster (class-level)-robust standard errors are in parentheses.

* $p < .10$.

TABLE A5

CAL PROGRAM AND TECHNOLOGY EFFECTS ON MATH GRADES, EXCLUDING CLASSES WITH CERTAIN PERCENTAGES OF BOARDING STUDENTS: BOYS ONLY

	Exclusion of Classes with Boarding Students > 90% (n = 1,940)		Exclusion of Classes with Boarding Students > 80% (n = 1,814)		Exclusion of Classes with Boarding Students > 70% (n = 1,684)	
	(1)	(2)	(3)	(4)	(5)	(6)
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	.749	.692	−.123	−.068
	(1.533)	(1.525)	(1.591)	(1.577)	(1.657)	(1.647)
Difference (program − technology)	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
R^2	.302	.306	.299	.303	.307	.310

Note. 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). All columns control for baseline counterpart of dependent variable (baseline class rank in math test score). Even-numbered columns also control for the following baseline covariates: liking math (scale 1–100), 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. Cluster (class-level)-robust standard errors are in parentheses.

* $p < .10$.

** $p < .05$.

TABLE A6

CAL PROGRAM AND TECHNOLOGY EFFECTS ON MATH GRADES, EXCLUDING CLASSES WITH CERTAIN PERCENTAGES OF BOARDING STUDENTS: GIRLS ONLY

	Exclusion of Classes with Boarding Students > 90% (n = 1,630)		Exclusion of Classes with Boarding Students > 80% (n = 1,500)		Exclusion of Classes with Boarding Students > 70% (n = 1,398)	
	(1)	(2)	(3)	(4)	(5)	(6)
CAL program	−.715	−.734	−.383	−.443	−.051	−.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 − technology)	1.392	1.399	1.737*	1.762*	1.973*	1.947*
	(.928)	(.927)	(.968)	(.971)	(1.073)	(1.077)
Additional controls	No	Yes	No	Yes	No	Yes
R^2	.295	.303	.289	.297	.287	.294

Note. 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). All columns control for baseline counterpart of dependent variable (baseline class rank in math test score). Even-numbered columns also control for the following baseline covariates: liking math (scale 1–100), 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. Cluster (class-level)-robust standard errors are in parentheses.
* $p < .10$.

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*, ed. Solomon W. Polachek and Konstantinos Tatsiramos, 183–228. Bingley, UK: Emerald.

Angrist, N., P. Bergman, and M. Matsheng. 2022. "Experimental Evidence on Learning Using Low-Tech When School Is Out." *Nature Human Behaviour* 6, no. 7:941–50.

Azevedo, João Pedro, Amer Hasan, Diana Goldemberg, Syedah Aroob Iqbal, and Koen Geven. 2020. *Simulating the Potential Impacts of COVID-19 School Closures on Schooling and Learning Outcomes: A Set of Global Estimates*. Washington, DC: World Bank.

Bai, Y., B. Tang, B. Wang, D. Mo, L. Zhang, S. Rozelle, E. Auden, and B. Mandell. 2023. "Impact of Online Computer Assisted Learning on Education: Experimental Evidence from Economically Vulnerable Areas of China." *Economics of Education Review* 94:102385.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122, no. 3:1235–64.

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, no. 1:52–74.

Bergman, P. 2021. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment." *Journal of Political Economy* 129, no. 1: 286–322.

Berry, J., and P. Mukherjee. 2016. "Pricing of Private Education in Urban India: Demand, Use and Impact." Unpublished manuscript, Cornell University.

Bettinger, Eric, Robert Fairlie, Anastasia Kapuza, Elena Kardanova, Prashant Loyalka, and Andrey Zakharov. 2023. "Diminishing Marginal Returns to Computer-Assisted Learning." *Journal of Policy Analysis and Management* 42, no. 2:552–70.

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, no. 2:53–80.

Bianchi, Nicola, Yi Lu, and Hong Song. 2020. "The Effect of Computer-Assisted Learning on Students' Long-Term Development." NBER Working Paper no. 28180 (December), National Bureau of Economic Research, Cambridge, MA.

Böhmer, B. 2014. "Testing Numeric: Evidence from a Randomized Controlled Trial of a Computer-Based Mathematics Intervention in Cape Town High Schools." MS thesis, University of Cape Town.

Borghans, L., B. H. Golsteyn, J. J. Heckman, and J. E. Humphries. 2016. "What Grades and Achievement Tests Measure." *Proceedings of the National Academy of Sciences of the USA* 113, no. 47:13354–59.

Büchel, K., M. Jakob, C. Kühnhanss, D. Steffen, and A. Brunetti. 2022. "The Relative Effectiveness of Teachers and Learning Software: Evidence from a Field Experiment in El Salvador." *Journal of Labor Economics* 40, no. 3:737–77.

Bulman, George, and Robert W. Fairlie. 2016. "Technology and Education: Computers, Software, and the Internet." In *Handbook of the Economics of Education*, vol. 5, ed. Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 239–80. Amsterdam: North-Holland.

Burguillo, Juan C. 2010. "Using Game Theory and Competition-Based Learning to Stimulate Student Motivation and Performance." *Computers and Education* 55, no. 2:566–75.

Campuzano, Larissa, Mark Dynarski, Roberto Agodini, and Kristina Rall. 2009. "Effectiveness of Reading and Mathematics Software Products: Findings from Two Student Cohorts." NCEE 2009-4041, National Center for Education Evaluation and Regional Assistance. 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 no. 78, Inter-American Development Bank, Washington, DC.

Chen, L. J., D. L. Yang, and Q. Ren. 2015. "Report on the State of Children in China." Chicago: Chapin Hall at the University of Chicago.

Clark, Andrew E., Huifu Nong, Hongjia Zhu, and Rong Zhu. 2021. "Compensating for Academic Loss: Online Learning and Student Performance During the COVID-19 Pandemic." *China Economic Review* 68:101629.

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.

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, no. 1:291–303.

Dynarski, Mark, Roberto Agodini, Sheila Heaviside, Timothy Novak, Nancy Carey, Larissa Campuzano, Barbara Means, et al. 2007. "Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort." TeLearn, hal-00190019. https://telearn.archives-ouvertes.fr/hal-00190019/.

Eble, Alex, and Feng Hu. 2019. "How Important Are Beliefs about Gender Differences in Math Ability? Transmission across Generations and Impacts on Child Outcomes." CDEP-CGEG Working Paper no. 53, Center for Development Economics and Policy, Columbia University.

Ebner, Martin, and Andreas Holzinger. 2007. "Successful Implementation of User-Centered Game Based Learning in Higher Education: An Example from Civil Engineering." *Computers and Education* 49, no. 3:873–90.

Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. 2017. "Education Technology: An Evidence-Based Review." NBER Working Paper no. 23744 (August), National Bureau of Economic Research, Cambridge, MA.

Fairlie, Robert W. 2015. "Do Boys and Girls Use Computers Differently, and Does It Contribute to Why Boys Do Worse in School than Girls?" *BE Journal of Economic Analysis and Policy* 16, no. 1:59–96.

Fairlie, Robert W., and Rebecca A. London. 2012. "The Effects of Home Computers on Educational Outcomes: Evidence from a Field Experiment with Community College Students." *Economic Journal* 122, no. 561:727–53.

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, no. 3:211–40.

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, no. 1:1–38.

Figlio, D. N. 1999. "Functional Form and the Estimated Effects of School Resources." *Economics of Education Review* 18, no. 2:241–52.

Fiorini, Mario. 2010. "The Effect of Home Computer Use on Children's Cognitive and Non-cognitive Skills." *Economics of Education Review* 29, no. 1:55–72.

Fuchs, Thomas, and Ludger Woessmann. 2004. "Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at School." CESifo Working Paper no. 1321, CESifo, Munich.

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, chap. 2. Chicago: University of Chicago Press.

Grewenig, Elisabeth, Philipp Lergetporer, Katharina Werner, Ludger Woessmann, and Larissa Zierow. 2021. "COVID-19 and Educational Inequality: How School Closures Affect Low- and High-Achieving Students." *European Economic Review* 140:103920.

Hannum, E., and A. Park, eds. 2007. *Education and Reform in China*. London: Routledge.

Hanushek, E. A. 1979. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions." *Journal of Human Resources* 14, no. 3:351–88.

———. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24, no. 3:1141–77.

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, no. 1:79–97.

Johnson, Helen, Sandra McNally, Heather Rolfe, Jenifer Ruiz-Valenzuela, Robert Savage, Janet Vousden, and Clare Wood. 2019. "Teaching Assistants, Computers and Classroom Management." *Labour Economics* 58:21–36.

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

Khor, N., L. Pang, C. Liu, F. Chang, D. Mo, P. Loyalka, and S. Rozelle. 2016. "China's Looming Human Capital Crisis: Upper Secondary Educational Attainment Rates and the Middle-Income Trap." *China Quarterly* 228:905–26.

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.

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, no. 2:208–31.

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, no. 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*. Thousand Oaks, CA: Sage.

Li, G., X. Zhang, D. Liu, H. Xue, D. Hu, O. Lee, C. Rilling, et al. 2023. "Education and EdTech during COVID-19: Evidence from a Large-Scale Survey during School Closures in China." *Comparative Education Review* 67 no. 1:53–77.

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.

Machin, Stephen, Sandra McNally, and Olmo Silva. 2007. "New Technology in Schools: Is There a Payoff?" *Economic Journal* 117, no. 522:1145–67.

Malamud, Ofer, Santiago Cueto, Julian Cristia, and Diether W. Beuermann. 2019. "Do Children Benefit from Internet Access? Experimental Evidence from Peru." *Journal of Development Economics* 138:41–56.

Malamud, Ofer, and Cristian Pop-Eleches. 2011. "Home Computer Use and the Development of Human Capital." *Quarterly Journal of Economics* 126, no. 2:987–1027.

Maldonado, Joana Elisa, and Kristof De Witte. 2020. "The Effect of School Closures on Standardised Student Test Outcomes." Discussion Paper no. DPS20.17, Department of Economics, Katholieke Universiteit Leuven.

Ministry of Education. 2017. *The Yearbook of Education Statistics in China*. Beijing: National Statistics Press. http://www.stats.gov.cn/tjsj/ndsj/2017/indexeh.htm.

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* 36:131–45.

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, no. 3:300–323.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109, no. 4:1426–60.

Muralidharan, Karthik, and Yendrick Zieleniak. 2013. "Measuring Learning Trajectories with Longitudinal Data and Item Response Theory." Lecture presented at the Young Lives Conference, "Inequalities in Children's Outcomes in Developing Countries," Oxford, July 8–9.

Parente, Paulo M. D. C., and João M. C. Santos Silva. 2016. "Quantile Regression with Clustered Data." *Journal of Econometric Methods* 5, no. 1:1–15.

Pritchett, Lant. 2013. *The Rebirth of Education: Schooling Ain't Learning*. Washington, DC: Center for Global Development.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73, no. 2:417–58.

Schmitt, John, and Jonathan Wadsworth. 2006. "Is There an Impact of Household Computer Ownership on Children's Educational Attainment in Britain? *Economics of Education Review* 25:659–73.

Spybrook, J., S. W. Raudenbush, R. Congdon, and A. Martinez. 2009. "Optimal Design for Longitudinal and Multilevel Research." Unpublished manuscript, University of Michigan.

UNESCO. 2020. "Education: From Disruption to Recovery." https://en.unesco.org/covid19/educationresponse.

Van der Kleij, Fabienne M., Remco C. W. Feskens, and Theo J. H. M. Eggen. 2015. "Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-analysis." *Review of Educational Research* 85, no. 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* 58:175–87.

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, no. 2:97–106.

Xu, Di, and Shanna S. Jaggars. 2014. "Performance Gaps between Online and Face-to-Face Courses: Differences across Types of Students and Academic Subject Areas." *Journal of Higher Education* 85, no. 5:633–59.

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.