



Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing



Fang Lai^{a,b,*}, Renfu Luo^c, Linxiu Zhang^c, Xinzhe Huang^d, Scott Rozelle^a

^a Freeman Spogli Institute for International Studies, Stanford University, 5th Floor, Encina Hall, 616 Serra Street, Stanford, CA 94305, USA

^b LICOS (Centre for Institutions and Economic Performance), Katholic University of Leuven, De Bériotstraat 34, B-3000 Leuven, Belgium

^c Center for Chinese Agricultural Policy, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, No. 11-A Datun Road, Chaoyang District, Beijing 100101, PR China

^d Department of Agricultural and Resource Economics, 2116 Social Sciences and Humanities, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

ARTICLE INFO

Article history:

Received 14 September 2013

Revised 17 March 2015

Accepted 18 March 2015

Available online 8 April 2015

JEL classification:

I20

I21

I28

O15

Keywords:

Education

Development

Computer-assisted learning

Random assignment

Test scores

China

Migration

ABSTRACT

The education of the disadvantaged population has been a long-standing challenge to education systems in both developed and developing countries. Although computer-assisted learning (CAL) has been considered one alternative to improve learning outcomes in a cost-effective way, the empirical evidence of its impacts on improving learning outcomes is mixed. This paper uses a randomized field experiment to explore the effects of CAL on student academic and non-academic outcomes for students in migrant schools in Beijing. Our results show that a remedial CAL program held out of regular school hours improved the student standardized math scores by 0.15 standard deviations and most of the program effect took place within 2 months after the start of the program. Students with less-educated parents benefited more from the program. Moreover, CAL also significantly increased the students' interest in learning.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The education of poor and disadvantaged population has been a long-standing challenge to education systems in both developed and developing countries (e.g. Glewwe & Kremer, 2006; Planty et al., 2008; World Bank, 2004). In confronting the challenge, efforts have been made to provide adequate

educational inputs such as textbooks and school facilities for disadvantaged populations in both developed and developing countries. Unfortunately these initiatives seem to have been unsuccessful in advancing learning outcomes (e.g., Glewwe & Moulin, 2002; Glewwe & Zitzewitz, 2004; Hanushek, 1986, 1995). As a consequence, researchers are actively exploring other ways of delivering educational inputs in order to better improve learning outcomes.

Computer-assisted learning (CAL) is one such alternative (e.g., Banerjee, Cole, Duflo, & Linden, 2007; Barrow, Markman, & Rouse, 2008; Linden, 2008). Computer-assisted learning

* Corresponding author at: Freeman Spogli Institute for International Studies, 5th Floor, Encina Hall, 616 Serra Street, Stanford, CA 94305-6055, USA. Tel.: +1 510 847 9811; fax: +1 650-725 1992.

involves the use of computers and modern computing technologies, embodied in both software and hardware devices, to enhance learning via computerized instruction, drills and exercises (Kirkpatrick & Cuban, 1998; Present's Committee of Advisors on Science and Technology, 1997). For students that live in vulnerable populations who often have to go to schools that are poor in quality, CAL can be a good substitute for teachers when the teachers are not available or have too little training and/or motivation to provide adequate instruction to the students either during or after school hours. CAL can also provide remedial tutoring services when commercialized services are either not available or not affordable. Finally, CAL can provide help that parents who are illiterate or too busy cannot provide. In these senses, CAL can be particularly effective in developing countries, where schools are plagued with poor facilities and unqualified teachers and computer technologies are relatively new and may be beyond the purchase options for most families.

Despite its promise, the empirical evidence on the effectiveness of CAL in promoting learning is mixed (e.g. Angrist & Lavy, 2002; Fuchs & Woosmann, 2004; Goolsbee & Guryan, 2006). An important limitation of these early studies is that they usually examined CAL as a form of an educational input with little attention to how computers were used in the schools. Later studies improved on these early studies by evaluating well-defined individual CAL programs using randomized experiments also found mixed evidence of the effectiveness of CAL (e.g., Barrow et al., 2008; Dynarski, 2007; Rouse & Krueger, 2004).

The existing literature has several limitations that have contributed to the ambiguous assessment of the effectiveness of CAL programs. First, the majority of CAL evaluations have been done in the context of developed countries, where educational resources may be abundant and computers are not novel to the students. Thus, it may not be surprising that many studies have found no significant beneficial effects of CAL on learning outcomes. However, there are few evaluations of CAL in the context of developing countries—where educational resources are highly constrained and access to technologies is limited. In the rare cases in which CAL programs in developing countries have been evaluated, most show positive effects on student test scores (Banerjee et al., 2007; He, Linden, & MacLeod, 2008; Linden, 2008). Second, instead of being supplementary to regular school time, many of the CAL programs in the existing literature often interfere with the regular school curriculum (as students are pulled out of class for CAL sessions). As a consequence, part of the full impact of CAL may be offset by the negative effects of missing classes, creating a downward bias in the estimation of the genuine impacts of the CAL intervention (Angrist & Lavy, 2002; Linden, 2008; Rouse & Krueger, 2004). Finally, besides academic performance, CAL might also have beneficial effects on non-academic outcomes. These non-cognitive outcomes, to our knowledge, have seldom been examined in the literature.

The overall goal of this paper is to explore the nature of the effects of CAL on not only student academic outcomes but also non-academic outcomes among underserved student populations in a developing country. In pursuit of this goal, we identify four particular objectives. First, we examine the immediate impacts of a CAL math program implemented

out of regular school hours on student academic performance in math (as measured by standardized test scores). Second, we examine how the program effects change over time and across students with different academic and family backgrounds. Third, we examine the spillovers of a math-focused CAL program on student academic performance in other subjects (in our case, language class). Finally, we investigate the impacts of CAL on non-academic student outcomes, an effect that has almost never been reported in the literature. Specifically, we examine the effect of the CAL intervention on the interest that students have in learning; student self-confidence; and self-efficacy in studying.

To meet this goal, in this paper we present the results of a randomized field experiment of a CAL program involving over 4000 third-grade students, mostly aged 9 and 10, from poor migrant families, in 43 migrant schools in Beijing. Migrant schools are private-run for-profit schools specifically serving migrant children. These schools are unregulated by the state and are typically thought to provide low quality education. Urban residents never send their children to migrant schools. Students in migrant schools are vulnerable in the sense that their teachers are busy and underpaid and almost never offer out of class tutoring; parents of the students are often poor and inadequately educated and, thus, can neither afford commercial tutoring nor are they able to effectively tutor their own children.

While our research design for computer-assisted learning is similar in many ways to Banerjee et al. (2007) and Linden (2008), there remain several differences. The most important feature of our research design is that we include an additional set of control schools to detect any program spillovers onto the control group. Second, we used within school (between class) randomization to both eliminate school-level confounding influences and improve the efficiency of the estimation. Third, our CAL program was implemented out of the regular school hours and thus was able to avoid any downward bias of the estimation of the treatment effect due to substitution effects.

Our results show that a remedial CAL program held out of regular school hours improved the student standardized math scores by 0.15 standard deviations and most of the program effect took place within 2 months after the start of the program. Students with less-educated parents benefited more from the program. The CAL program also significantly increased the interest of students in learning. We observe no significant spillovers of CAL program in math onto Chinese language test scores. Our results are also shown to be robust to the potential threats of the program spillovers.

The rest of the paper is organized as follows. The first section briefly lays out the context of the study—migrant schools and migrant education in Beijing. The next section reviews the study's approach, including the research design and sampling, an explanation of the intervention, a description of the data and an explanation of the statistical approach. The following sections present the results, discuss the findings and conclude.

2. Migrant schools and migrant education in Beijing

As China's economy has grown over the past 30 years, the number of rural-to-urban migrants has increased to

150 million, among which, 20 million are children of school age (All-China Women's Federation, 2008; Sa, 2004). As urban public schools can only accommodate a small fraction of migrant children, limited space and steep out-of-district school-access fees unaffordable to many migrant families, compel most migrant children to attend for-profit schools that specifically serve children from poor migrant families (*migrant schools*). The students in these migrant schools are exclusively migrant children—no urban children enroll in these schools. In Beijing it is estimated that 70% of migrant children attend migrant schools (Tao & Yang, 2007).

Migrant schools often have poor facilities, under-qualified and unmotivated teachers with high job turnover and fragmented curricula (CCAP, 2009; Ding, 2004; Han, 2004; Human Rights Watch, 2006; Kwong, 2004; Liu, 2002; Ma, Bai, Liu, Pang, & Rozelle, 2008). Due to their private nature, migrant schools are often unacknowledged by the municipal government and rarely receive any support (Han, 2004; Kwong, 2004). Worse still, these schools are often transient, with sudden closings due to anything from having their leases pulled because of rebuilding projects to local regulation violations. This transience discourages long-term investment in these schools. The poor quality of facilities and under-qualified teachers have almost certainly created an environment in which migrant children are underperforming educationally (e.g.; Lai et al., 2011; Song, Loyalak, & Wei, 2010). For example, Lai et al. (2014) found that after controlling for individual and family backgrounds, students in public schools still outperformed on academic tests those in migrant schools by around 12 points on a 0–100 scale. Moreover, the longer migrant children stayed in migrant schools, the worse their performance was.

Improving educational quality and student performance in migrant schools has important implications for the children of China's 150 million migrant workers now and in the future as the migrant population continues to grow. When children of well-off families fall behind in study, their parents can either send them to commercial remedial tutoring programs, or the parents themselves almost always make efforts to assist their children in their coursework. Their schools and teachers also can help them catch up via various school-sponsored, remedial tutoring programs. However, migrant students do not have the above privileges. According to our survey data, only 13% of the migrant parents have a high school diploma, and they are too poor to afford commercial remedial tutoring services. Moreover, migrant school teachers are rarely competent and often lack the motivation and dedication that is needed to help their students.

3. Sampling, data and methods

3.1. Sampling and the process of randomization

We conducted a clustered (at the class level) RCT of CAL in Beijing migrant schools in the fall semester of 2010. A total of 4103 students in 98 classes of 43 Beijing migrant schools are involved in our study. Among them, 2514 students in 55 classes of 24 schools constitute the main sample. The other schools/classes serve as additional controls to check for intra-school spillovers and program spillovers.

To focus our study on districts where migrant schools are most concentrated, we restricted our sample to the three districts in Beijing. These three districts are the areas of Beijing most densely populated by migrants and migrant schools. Of the 230 migrant schools in Beijing, 69 schools were in these three districts. Therefore, our sample can be considered to be representative of the children of migrant families—most of whom are relatively poor, compared to the average urban household—who attend migrant schools. Other migrant families have put their children into urban public schools. Our sample is not representative of the class of families/students that are in urban public schools. However, it should be noted that the majority of migrant children in China attend migrant schools (Lai et al., 2011). We believe our sample is reasonably representative of migrant schools across China, which suffer from poor facilities and teachers, and whose students often do not receive remedial care at home. What is more, given that similar shortcomings appear to be common in poor schools outside of China, such as those in rural India (Muralidharan & Kremer, 2006) and rural Indonesia (Newhouse & Beegle, 2006) our sample may also serve as a useful frame of reference for struggling schools in other LDCs or underserved neighborhoods in developed countries. In summary, we believe our study is useful in gauging the effectiveness of computer-assisted learning in schools where poor students lack high quality teachers and facilities, and have little assistance with schoolwork from their parents at home.

Having identified our sample frame, we then proceeded to exclude schools that had only one class per grade at the grade 3 level. In total, 43 schools met the criteria of having two or more grade 3 classes. Our power calculations, however, required that we only needed 24 classes in each of the two groups (the treatment and control groups) in the 24 schools.¹ Therefore, of the 43 schools in the sample frame, we randomly chose 24 schools as our sample schools (Fig. 1). The rest of the schools—19 of them—while excluded from being part of the CAL intervention, were kept as an additional control group which we use to check the robustness of our empirical results regarding possible program spillovers (described more below).

Because the math CAL software we used was only available at the third grade level, we only examined third-grade students in our sample schools. All of the third-grade students in the 24 sample schools were included in the study (although only one class in each school received the CAL intervention). In total, there were 2514 third-grade students in 54 classes, approximately 47 students per class in the 24 core sample schools (Fig. 1). There were more than 48 classes in the 24 core sample schools because five schools had three classes and one school had four.

Although the core sample at the baseline survey included a total of 24 schools and 2514 students, there was some

¹ In the multi-site cluster randomized trial, within each site (school) we randomly assigned one class to treatment group and the other classes to control group. We calculated that we require 45 individuals per class and 24 classes per group to detect a standardized effect size for the outcome variable of 0.25 with 85% power at the 5% significance level. We assumed an intra-cluster correlation of 0.25, a pre- and post-intervention correlation of 0.5, 60% reduction in the between-class variation by blocking on school affiliations, the variability in effect sizes equal to 0.01 and a 15% loss to follow-up.

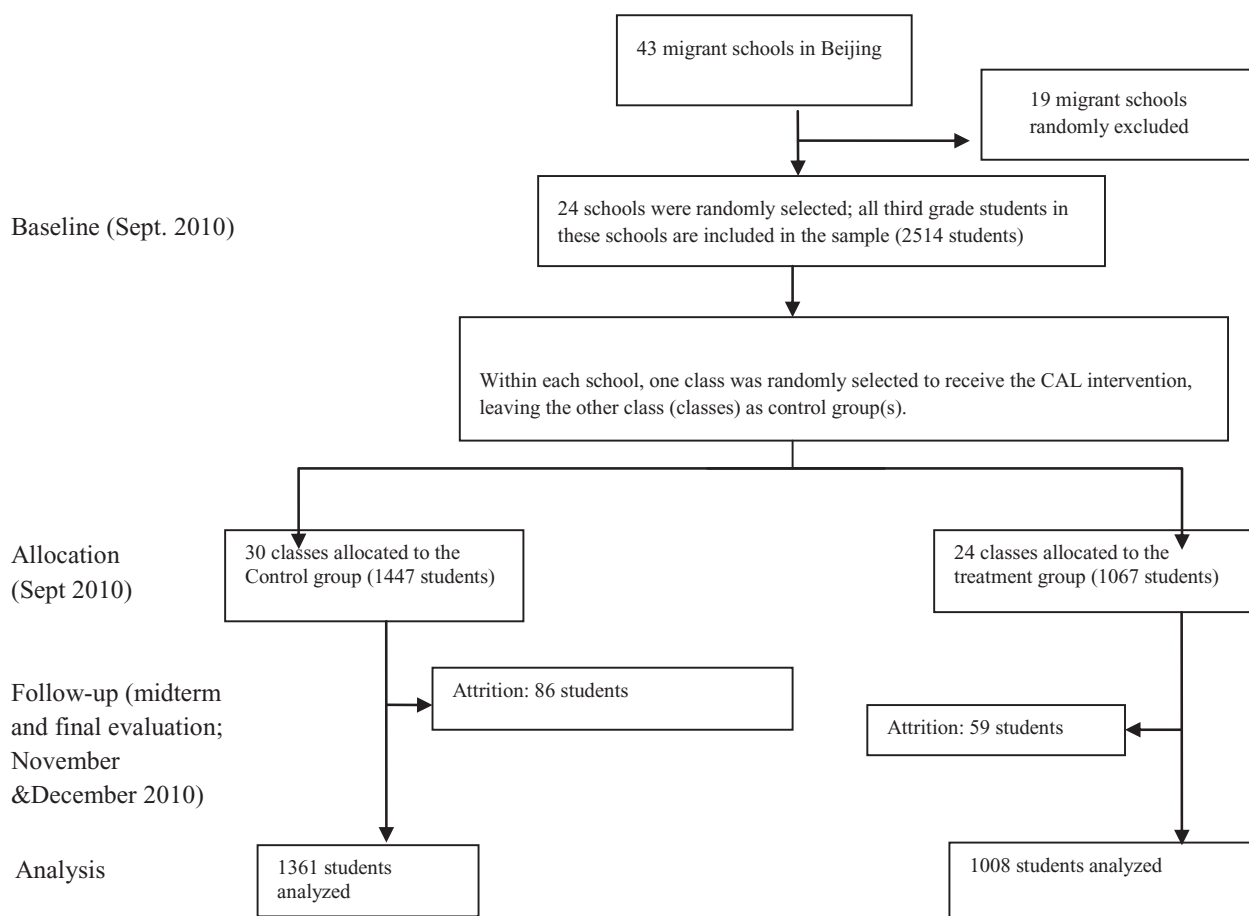


Fig. 1. Experiment profile.

attrition by the end of the study. For various reasons (mainly school transfers and extended absences due to illness or injuries), by the time of the evaluation survey we were only able to follow up with 1361 in the control classes and 1008 students in the treatment classes in the 24 sample schools (Fig. 1, final row). In other words, 2369 out of the initial 2514 students were included in our evaluation survey and were part of the subsequent statistical analysis, implying a moderate attrition rate of 6%. Boys, older students and students with lower Chinese baseline test scores attrited at higher rates than other students (Table 1, column 1). Compared with students that did not attrite, attrition students had more access to some other types of electronic technology, such as a mobile phone.

Fortunately, not only is the attrition rate low, but it also appears to be independent of the assignment of the CAL intervention, and thus unlikely to either reduce the validity of our research design or bias our results. As shown in the final row of Fig. 1, the attrition is evenly distributed between the treatment and control groups. The attrition rate is around 6% for both the control group and the treatment group. The difference in the attrition rate between these two groups is not statistically significant. In other words, students did not leave the sample because they were or were not assigned to the treatment/control groups. Furthermore, when comparing the attrition students in the treatment group to those in the

control group, we found they had similar characteristics. This suggests that, in general, the factors leading to attrition were largely the same for both groups.

After choosing the 24 schools for the core sample, the first step of our study was to randomly select one class in each sample school to receive the CAL intervention. Hence, our research team randomly selected one third grade class in each of the 24 program schools to receive the CAL intervention (the treatment group), leaving the rest of the third grade classes in each school to serve as the control group. In this way, the baseline student characteristics were balanced between students in the control classes and those in the treatment classes. The raw and within-school differences between the treatment and control groups were not only statistically insignificant for all student characteristics but also small in magnitude in most cases (Table 2, columns 7 and 8).

3.2. Experiment arms/interventions

Excluding the 19 control schools, whose principals were never informed of the CAL intervention in the other 24 CAL sample schools, our experiment focused fully on one group of treatment classes and one group of control classes in the same school.

Table 1

Comparisons of the student characteristics between the attrited students and those remaining in the sample and the characteristics of attrited students between the treatment and control groups in the 24 program schools.

		Sample: sample + attrition observations ^c Within-school difference (nonattrition–attrition) ^d (2)	Sample: attrition observations ^e Within-school difference (treatment–control) ^f (1)
(1)	Baseline math score (units of standard deviation) ^a	0.25	0.01
(2)	Baseline Chinese score (units of standard deviation) ^a	0.28***	0.09
(3)	Age (number of years)	–0.28***	–0.26
(4)	Female (1 = yes; 0 = no)	0.15***	0.08
(5)	Ever used a computer (1 = yes; 0 = no)	0.06	–0.11
(6)	Access to other modern technologies ^b	–0.05**	–0.03
(7)	Living with father (1 = yes; 0 = no)	0.04	–0.02
(8)	Living with mother (1 = yes; 0 = no)	0.03	–0.06
(9)	# of siblings	–0.21	0.24
(10)	Father has high school or college degree (1 = yes; 0 = no)	0.00	0.03
(11)	Mother has high school or college degree (1 = yes; 0 = no)	–0.02	–0.05
(12)	School dummies	Y	Y
(13)	Observations	2369	145

Source: Authors' survey.

Robust standard errors clustered at the class level.

^a The baseline math/Chinese score is the score on the standardized math/Chinese test that was given to all students in the sample before the CAL program.

^b Access to other modern technologies is the mean value of a set of 0/1 dummy variables including whether the student has used cell phone, Internet, game console, CAL software, and videos for learning assistance.

^c The sample include both the sample observations (non-attrition) and the attrition observations.

^d The within-school differences between the attrition and non-attrition observations are calculated by regressions of the row variables on the dummy variable of non-attrition, controlling for school dummy variables.

^e The sample is limited to the attrited observations.

^f The within-school difference between the treatment and control group is calculated by regressions of the row variables on the treatment dummy variable, controlling for school dummy variables.

** Significant at 5%.

*** Significant at 1%.

3.2.1. CAL intervention group (the treatment group)

The main intervention involved computer-assisted math remedial tutoring sessions which were designed to complement the regular in-class math curriculum for the Fall 2010 semester. Under the supervision of one teacher–supervisor trained by our research group, the students in the treatment group had two 40-min CAL sessions per week during lunch break or after school, thus serving as a supplement to the regular in-school instruction hours. The sessions were mandatory and attendance was taken by the teacher–supervisors. To avoid influence from teachers that might confound our estimation of the effect of CAL intervention, none of the teacher–supervisors were math teachers or the home-room teachers of the third-grade students. The content–instructional videos and games—of each session was exactly the same for all students in all 24 treatment classes and emphasized basic competencies in the uniform national math curriculum. In short, the material was remedial in nature; was based on the material that was in student textbooks; and was material taught the same week. The teacher had no discretion on what was taught in the CAL lessons.

During each session, two students shared one computer and played animation-based math games designed to help students review and practice the basic math material that was being taught in their regular school math classes. When playing the games, the students first worked out the solutions with pencils/pens on scratch paper and then submitted the

answers using the keyboards and mice of their computers. If a student had a math-related question, he/she was encouraged to discuss with his/her teammate (the one with whom he/she shared the computer). The students were not supposed to consult the other teams or the teacher–supervisor. According to our protocol, the teachers were only allowed to help students with scheduling, computer hardware issues and software operations. In fact, in our observations, the sessions were so intense that the attention of the students were fully on the computer and, while there was a lot of interaction between the members of the two-person teams, there was little communications among the groups or between any of the groups and the teacher–supervisor.

3.2.2. CAL control group (the control group in the core sample)

Students in the classes that were assigned to the control group (all classes in the 24 program schools, except for the classes which received the CAL intervention) did not receive any CAL intervention. Following the protocol, they were not allowed to access the computers. They were also not allowed to use the software. To our knowledge, no one except for the students in the CAL treatment group and their teacher–supervisors used the computers for any purpose.

To eliminate confounding influences from school visits paid only to the treatment group during the CAL sessions, as Muralidharan and Sundararaman (2011) did in their RCT on teacher performance pay, every time our volunteers visited

Table 2

Comparison of student and family characteristics and student access to computer and other modern technologies between the treatment and control groups in the 24 program schools.

	Treatment (1008 obs)		Control (1361 obs)		The raw difference (treatment–control)		Within-school difference ^c (treatment–control)	
	Mean (1)	SD (2)	Mean (3)	SD (4)	Mean (5)	SD (6)	Mean (7)	SD (8)
(1) Age (number of years)	8.53	0.87	8.49	0.88	0.04	0.04	0.03	0.03
(2) Female (1 = yes; 0 = no)	0.46	0.50	0.45	0.50	0.01	0.02	0.01	0.02
(3) Baseline math score (units of standard deviation) ^a	0.03	0.97	0.00	1.00	0.03	0.06	0.04	0.04
(4) Baseline Chinese score (units of standard deviation) ^a	–0.03	1.00	0.00	1.00	–0.03	0.09	0.00	0.05
(5) Access to other modern technologies ^b	0.50	0.27	0.48	0.26	0.01	0.02	0.01	0.02
(6) Living with father (1 = yes; 0 = no)	0.88	0.33	0.88	0.33	0.00	0.02	–0.00	0.01
(7) Living with mother (1 = yes; 0 = no)	0.93	0.26	0.92	0.27	0.01	0.01	0.01	0.01
(8) # of siblings	1.65	1.72	1.77	2.08	–0.12	0.10	–0.09	0.08
(9) Ever used a computer (1 = yes; 0 = no)	0.49	0.50	0.54	0.50	–0.04	0.04	–0.04	0.03
(10) Father has a college degree (1 = yes; 0 = no)	0.02	0.14	0.02	0.13	0.00	0.01	0.00	0.00
(11) Father has a high school degree (1 = yes; 0 = no)	0.12	0.33	0.12	0.33	0.00	0.01	0.00	0.01
(12) Mother has a college degree (1 = yes; 0 = no)	0.01	0.11	0.01	0.10	0.00	0.00	0.00	0.00
(13) Mother has a high school degree (1 = yes; 0 = no)	0.08	0.27	0.08	0.27	–0.01	0.01	–0.00	0.01

Source: Authors' survey.

** Significant at 5%. *** Significant at 1%.

^a The baseline math/Chinese score is the score on the standardized math/Chinese test that was given to all students in the sample before the CAL program.

^b Access to other modern technologies is the mean value of a set of 0/1 dummy variables including whether the student has used cell phone, Internet, game console, CAL software, and videos for learning assistance.

^c The within-school difference between the treatment and control groups is calculated by regressions of each of the row variables on the treatment dummy variable, controlling for school dummy variables, with robust standard errors clustered at the class level.

the program school and audited the CAL session, they also audited classes in the control group. Therefore, the control groups were visited the same number of times as the treatment groups.

3.2.3. Additional control group

There is one more additional control group in our study design. It is possible that we cannot rely on the control group (control classes in the same school as the treatment classes) alone because the randomization is within each program school and students (in the control classes) might be negatively affected by the treatment. To be more specific with an example of what might happen, the students in the control classes might observe students of the treatment classes in the same grade having CAL sessions twice a week while they just had regular school sessions as usual. This observation alone might affect their efforts/motivation and thus their academic outcomes. In fact, there could have been positive spillovers, too. By interacting with students in the treatment classes, the students in the control classes might also indirectly learn from the CAL sessions. With this positive spillover, the test scores of students in the control group might also increase, generating a downward bias of the estimate of the CAL program effect. To disentangle these spillovers from the true effect of the CAL program, we included a set of pure control schools that did not know of the CAL program, and thus were unaffected by the program spillovers into our analytic sample.

To measure if there was any possible bias from the program spillovers or if there were any spillovers, we also included all third-grade students in the 19 schools that were randomly excluded from the core sample as an additional control group. To minimize their awareness of being part of our study, for the students in this group, we only conducted the standardized math and Chinese tests and collected data on student and family characteristics during the baseline and final evaluation surveys. We did not pay any other visits to these students or to their schools. None of the teachers or students in these 19 schools were informed of the CAL program and these schools did not have any contact with the 24 program schools in the core sample of our study. Consequently, the CAL intervention in the program schools was unlikely to affect these additional control schools via any form of spillover.² Comparisons of the changes in the academic performance over the program period among the treatment

² One might be concerned about possible "contamination" of the pure control schools. For several reasons, we do not believe that this happened. First, the schools are geographically dispersed. Second, all migrant schools are independent—there is no association connecting them together. Third, since migrant schools are private, for-profit entities, they are also competitors. Thus, program schools had no incentive to tell non-program schools about the CAL program. Finally, we directed principals and teachers of the program schools to not talk about the program outside the school, or else be at risk of elimination from the program. We believe under such circumstances, the possibility of spillover of information about the program across schools is minimal.

group, control group (the control classes in the 24 program schools), and the additional control group can inform us of the significance and size of the program spillovers.

3.3. Data collection

The research group conducted a total of three rounds of a survey of each third-grade student in the 24 schools in our main sample, and two rounds of a survey for all third-graders in the 19 additional control schools. The first-round survey was a baseline survey conducted with all third graders in the 43 schools in early September 2010 at the beginning of the Fall semester and before any implementation of CAL program had begun. The second-round survey was a mid-line evaluation survey conducted only with third graders in the 24 program schools in the middle of the semester in early November. The third-round survey was a final evaluation survey conducted at the end of the program in late December, a time that coincided with the end of the Fall semester.

In each round of survey, the enumeration team visited each school and conducted a two-block survey. In the first block students were given a standardized math test and a standardized Chinese test.³

In the second block enumerators collected data on the characteristics of students and their families. From this part of the survey we are able to create demographic and socio-economic variables. The dataset includes measures of each student's *age* (measured in years), gender (described by an indicator *female*, which is equal to one for girls, and zero for boys), how many siblings they had (or *number of siblings*), father's education level (*father has a college degree* and *father has a high school degree*), and mother's education level (*mother has a college degree* and *mother has a high school degree*). To create indicators of parental care, during the survey the students were also asked whether they lived with either of their parents for most of the time during the semester (*living with father* and *living with mother*).

Importantly, in the second block students were also asked to answer questions that could help us measure their non-cognitive traits, such as their attitudes toward schooling, the levels of self-confidence and self-efficacy of math studying (i.e. one's belief in their ability to excel in math and problem solving). To create the indicator depicting the student's attitude toward schooling (*like school*), the students were asked to rate their attitude toward school on a 0–10 scale, where "0" indicates "extremely hate school" and "10" indicates "extremely enjoy school." The indicators of self-confidence and self-efficacy of math studying were created from the responses of students in a ten-item psychological scale measur-

ing self-confidence⁴; and a seven-item psychological scale for self-efficacy of math studying.⁵

For the baseline survey only, information about the access of students to computers and other modern technologies (such as mobile phones) was collected. The information collected in this sub-block of the survey allowed us to create variables that include whether the students had *ever used a computer* and whether they had *access to other modern technologies*.

3.4. Statistical methods

To estimate how the academic and non-academic outcomes changed in the treatment group relative to the control group, we run an Ordinary Least Squares (OLS) model, regressing the outcome variables (i.e. post-program outcome value) on the baseline value of the outcome variable and a dummy variable of the treatment (CAL intervention) status, controlling for the school fixed effects and a set of control variables. As randomization was conducted within schools across classes, we controlled for the school fixed effects to disentangle the systematic within-school differences between the treatment and control classes in y_{isc} and obtain an unbiased estimate of the genuine treatment effect of the CAL intervention. We included other control variables to improve the efficiency of the estimation. In all regressions, we accounted for the clustered nature of our sample by constructing Huber–White standard errors corrected for class-level clustering (relaxing the assumption that disturbance terms are independent and identically distributed within classes). To be more specifically, our model is:

$$y_{isc} = \alpha + \beta \cdot treatment_c + s_s + \theta \cdot y_{0isc} + X_{isc}\gamma + \varepsilon_{isc} \quad (1)$$

where y_{isc} is the endline (or midterm in some cases) outcome variable for child i in school s and class c , $treatment_c$ is a dummy variable for a student attending a class assigned to the treatment group (equal to one for students in the treatment group and zero otherwise), s_s is the vector of school fixed effects, y_{0isc} is the pre-program (baseline) outcome value for student i in school s and class c , X_{isc} is a vector of additional control variables, and ε_{isc} is a random disturbance term (clustered at the class level).

We used several variables to measure the student academic and non-academic outcomes (y_{isc}). The primary outcome variable of our analysis is the student academic outcome, measured by the student standardized math test score.

³ Some experts from the Center for Examination of Beijing helped us pick questions for the tests from official examination books and exercise books. The math test included 29–32 questions (tests in different rounds included slightly different numbers of questions). The Chinese test included 30–35 questions. Students were required to finish test in each subject in 25 min. All students took the math test first and then they took the Chinese test. Our enumeration team monitored the test and strictly enforced the time limits and tried to make sure there was no cheating. We use the scores of the students on the math and Chinese tests as our measures of student academic performance.

⁴ Self-confidence (also called self-efficacy in the psychological literature) is a person's perception of their ability to plan and take action to reach a particular goal. "The construct of perceived self-efficacy reflects an optimistic self-belief" (Schwarzer, 1992). Perceived self-efficacy is an operative construct, i.e., it is related to subsequent behavior and, therefore, is relevant for clinical practice and behavior change. Jerusalem and Schwartz developed the General Self-Efficacy Scale (GSE) in 1979, which was then widely employed in measuring self-efficacy. GSE has 10 items. Each item refers to successful coping and implies an internal-stable attribution of success. In our study, we adopted the Chinese adaption of the GSE developed in Zhang and Schwarzer (1995).

⁵ To measure the self-efficacy of math studying, a professor in psychometrics and measurement in Beijing Normal University helped us choose among the 12 indicators of math attitudes used in TIMSS 2003 and developed a seven-item scale of self-efficacy of math studying that is appropriate to use under the context of elementary schools in China.

We also included the student standardized Chinese test score as an additional academic outcome measure. By doing so, we could examine if there were any positive or negative spillovers of the CAL intervention to the student academic performance in Chinese, the other major subject in China's elementary schools besides math.⁶ Importantly, besides variables measuring academic outcomes, we also included three non-academic outcome variables, namely, the student's level of self-confidence, *like school* and self-efficacy of math studying.

The variables in X_{isc} are student and family characteristics (*female, age, number of siblings, father has a college/high school degree, mother has a college/high school degree, living with father and living with mother*) and the student access to computer and other technical device before the program started (*ever used a computer and access to other modern technologies*).

By construction, β equals the average within-school difference in y_{isc} between the treatment and control groups conditional on y_{ois} (or β measures the within-school difference in changes in the outcome variable between the treatment and control groups over the program period). As the CAL intervention was randomly assigned within schools, β is an unbiased estimate of the effect of being assigned to the treatment group (i.e., the effect of the CAL intervention, or the CAL treatment effect). By including X_{isc} as additional control variables, β is an unbiased, efficient estimate of the CAL treatment effect.

4. Results

4.1. Main effect of the CAL intervention on student performance

The multivariate regression analyses show that the estimated CAL treatment effect on math test scores is equal to 0.15 standard deviations (Table 3, row 1, column 1) and is significant at the 0.01 level. Considering that the program only ran for one semester, the size of the CAL program effect is comparable to the findings in other CAL evaluations that observed beneficial effects of CAL on student performance (e.g., Banerjee et al., 2007; Barrow et al., 2008; Linden, 2008).^{7,8}

⁶ For example, the CAL program might have improved the student's general learning ability and thus the student Chinese test score might also increase. The CAL program might have also taken up so much of the student's time and energy in learning math that the student had less time and energy to spend on Chinese. In this case, the CAL program in math might negatively affect the student academic performance in Chinese.

⁷ It is possible that teacher effort responded to the CAL program. For example, teachers may have taken advantage of the program and shirked their duties. If this were the case, the lower effort of the teacher would have lowered the observed impact of the CAL program. Unfortunately, in our data set there is no reliable data of the level of teacher effort. Because of the absence of data, we cannot rule out this possibility. However, even if this were the case, our estimate of the CAL impact would be a "lower-bound" estimate. Hence, we still would have to a highly significant finding even after accounting for any confounding influence from the effect of teacher free riding. In the end we would still be able to say that CAL could provide an effective remedy to the underperformance of students in migrant schools.

⁸ One issue of potential interest is whether our results are being driven by strategic switching (that is, are better students jumping into the treatment classes, perhaps at the expense of poor performing student?). We do not believe so for several reasons. First, we do not observe any switching of classes after the intervention. In other words, no students switched from the control

Most of the positive effect of the CAL intervention on standardized math test scores took place within the first 2 months of the program implementation. Since the CAL program was launched in early September, its effect quickly reached 0.15 standard deviations by midterm (early November; Appendix A, row 1, column 2). This pattern of results is similar to those found in Banerjee et al. (2007) who find that the large impacts came early in the program's implementation period. Although the impact did not continue to rise during the second period of the study, this effect persisted over the whole semester (by comparing Table 3, row 1, column 1 and Appendix A, row 1, column 2). In this way, our findings are different from the findings in Banerjee et al. (2007), which found even larger impact of CAL intervention during the second phase of CAL program.

One possible explanation of the absence of further gains during the second period is that there was a fall in the level of excitement about the program. To examine this question, we can look at changes in non-cognitive outcomes over the course of the project (during first and second period). The three non-cognitive variables are: *like school*; self-confidence; and self-efficacy. Using our data, we ran regressions that examined the change of these outcomes over time. The results are in Appendix B. There is no evidence of decreasing "excitement" when measured using any of the three measurable non-cognitive outcomes. The measures of self-confidence and self-efficacy do not fall; when we measure "like school," the rate actually significantly rises after the midterm. We did not observe significant increase in "like school" between baseline and midterm (Appendix B, column 1), but observed significant and large increase in "like school" between baseline and endline (Appendix B, column 2).

Another possible reason for the absence of further gains from CAL program during the second period is household substitution. Das et al. (2011) found that unanticipated school grants led to significant improvements in student test scores but anticipated grants had no impact on test scores. They suggested that over time households would adjust their educational inputs for their children in response to changes in school inputs, thus offsetting part of the impacts of the (anticipated) school inputs. It is possible that this might also apply to our CAL program. For the first half of the semester, the CAL program was unanticipated by the parents and thus parents did not respond immediately (that is, they did not adjust their educational input for their children). By the middle point of the intervention, however, parents may have had time to understand what was going on and realized that the CAL program would continue until the end of the semester. If they believed that the CAL program was doing some of things that they were doing in the evenings with their children, they may have adjusted their time input accordingly. In other words, they might have believed that their children had enough

class to treatment class or vice versa. Second, we do not include students who switched into the program schools after the start of the semester into our analytic sample as we do not have baseline data for these students. Finally, in most cases, migrant students switched schools for reasons irrelevant to CAL (e.g. their parents changed jobs so they had to move to another community). Thus, there was no evidence of strategic switching of classes/schools in response to the CAL intervention.

Table 3

Ordinary Least Squares estimators of the impacts of CAL program on student outcomes in the 24 program schools.

		Standardized math test score (1)	Standardized Chinese test score (2)	Like-school (0–10) (3)	Self-confidence (1–4) (4)	Self-efficacy in math studying (5)
(1)	Treatment (1 = the treatment group; 0 = the control group)	0.15*** (0.04)	0.01 (0.05)	0.31** (0.15)	0.07 (0.05)	–0.00 (0.03)
(2)	Baseline value of the outcome variable	Y	Y	Y	Y	Y
(3)	School dummy variables and other control variables	Y	Y	Y	Y	Y
(4)	Observations	2369	2369	2369	2369	2369
(5)	R-squared	0.39	0.39	0.10	0.20	0.14

Source: Authors' survey.

Robust standard errors in brackets, clustered at the class level.

Each column reports the results of one regression of the student outcome variable (corresponding to the column title) on the treatment dummy variable, controlling for the baseline value of the outcome variable, the school dummy variables and other control variables.

Other control variables include the student age, female (1 = yes; 0 = no), baseline math/Chinese score, ever used a computer (1 = yes; 0 = no), access to other modern technologies, living with father (1 = yes; 0 = no), living with mother (1 = yes; 0 = no), # of siblings, father has college degree (1 = yes; 0 = no), father has high school degree (1 = yes; 0 = no), mother has college degree (1 = yes; 0 = no), mother has high school degree (1 = yes; 0 = no).

** Significant at 5%.

*** Significant at 1%.

tutoring via the CAL program at school and thus reduced the time and effort they spent on tutoring their children at home.

If the household substitution is part of the reason for the leveling off of the impact, we might expect to observe two patterns in the change of student test scores over time. First, we should expect to observe stronger household substitution effects for students whose parents had higher levels of education and were in a better position to be providing in-home tutoring help for their children in the absence of the CAL program.⁹ Second, the first pattern would be expected to emerge only after the midterm exam because parents were only able to effectively respond to the unexpected CAL program sometime after its implementation (for example, about the time of the midline evaluation when they observed improvement in the student midterm test scores).

The results of our regression analysis are consistent with the assumption of household substitution. Before the midline evaluation survey, students whose fathers had higher levels of education did not gain less from the treatment than those who did not (Appendix A, rows 2 and 3, columns 3 and 4). However, after the midline evaluation survey (by comparing student test scores between endline and baseline), students whose father had higher levels of education gained less from the treatment than those who did not (0.20 standard deviations less for students whose father had at least a high school degree and 0.35 standard deviations less for those whose father had a college degree; row 2, columns 8 and 9, Table 4). This result implies that there is a possible household substitution occurring after the onset of the CAL program, when parents were well aware of the CAL program and was stronger among parents who were well educated and more able to tutor their children in the absence of the CAL program.

Unfortunately, we do not have any direct empirical evidence that substitution of hired tutoring services or parental

effort at home is occurring. From responses to survey questions about after-school tutoring activities,¹⁰ we do not observe any significant differences in the amount of after-school tutoring/help with study at home between treatment and control groups (results not shown for brevity). We also do not observe any difference in the amount of tutoring received outside of school (or amount of help received at home) during the evaluation part of the survey between treatment and control classes for students with fathers that have different levels of education. In regressions that examine whether parents with higher levels of education reduced efforts in helping students at home in treatment classes, we see that none of the coefficients on the relevant variables—that is, those made up of an interaction between treatment and fathers' education (high school/college)—are significant (results not shown for brevity). Nonetheless, part of the reason for lack of direct evidence of household substitution using student survey responses of parental tutoring might be that the true level of parental tutoring/attention might not be properly measured by survey questions to students.

The benefits of CAL program come at costs in several dimensions. One major cost is the laptops; although the program laptops were donated, each computer would cost \$100–\$150 if we were to buy them on the market. On average, one computer serves six students. Assuming that the laptops depreciate over 3 years, the per-student cost of computers is \$1.9–\$2.8 during the program period. Another major source of the program cost is compensation for the teacher-supervisors. On average, one teacher-supervisor was in charge of 47 students. We paid each teacher 400 yuan (\$59)

⁹ If the parents were poorly educated, they might not be able to input their time to tutor their children, even if they wanted to. Because of this, the substitution of inputs effect in response to the change in school input—in our case, the CAL program, might even be irrelevant for these types of parents.

¹⁰ We asked two questions during the evaluation surveys: (a) How many times per week do you have after-school tutoring sessions (remedial or advanced—typically in a commercial venue or in a teacher's home, where the teacher is paid)? (b) How many times per week do you receive help with your study after school inside your home (e.g., from your parents, siblings, relatives or private tutors)? Overall, only 18% of the students in our sample reported to have after-school tutoring sessions at least once per week; 34% of the students in our sample reported to have received help with study after school from parents (or other people inside the household) once per week.

Table 4

The Ordinary Least Squares estimators of the heterogeneous program effect on students with different characteristics in the 24 program schools.

Dependent variable: standardized post CAL math test score		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)	Treatment (1 = the treatment group; 0 = the control group)	0.15*** (0.04)	0.15*** (0.04)	0.11 (0.41)	0.12** (0.05)	0.22* (0.13)	0.16 (0.13)	0.11** (0.05)	0.15*** (0.04)	0.17*** (0.05)	0.15*** (0.04)	0.14*** (0.05)	0.20*** (0.06)	0.21*** (0.07)
(2)	Interactions: treatment*	Baseline math score (units of standard deviation) ^a	Baseline Chinese score (units of standard deviation) ^a	Age (number of years)	Female (1 = yes; 0 = no)	Living with father (1 = yes; 0 = no)	Living with mother (1 = yes; 0 = no)	# of siblings	Father has a college degree (1 = yes; 0 = no)	Father has a high school degree (1 = yes; 0 = no)	Mother has a college degree (1 = yes; 0 = no)	Mother has a high school degree (1 = yes; 0 = no)	Ever used a computer (1 = yes; 0 = no)	Access to other modern technolo- gies ^b
(3)	Control variables	–0.03 (0.05)	–0.05 (0.04)	0.00 (0.05)	0.05 (0.07)	–0.09 (0.12)	–0.01 (0.12)	0.02 (0.02)	–0.35* (0.20)	–0.20* (0.11)	–0.41 (0.26)	0.06 (0.12)	–0.10 (0.07)	–0.12 (0.14)
(4)	School dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
(5)	Observations	2369	2369	2369	2369	2369	2369	2369	2369	2369	2369	2369	2369	2369
(6)	R-squared	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39

Source: Authors' survey.

Robust standard errors in brackets, clustered at the class level.

Each column reports the coefficients of the treatment dummy variable (in row (1)) and the interaction of the treatment dummy variable and the corresponding student characteristics in row (2) from one regression. Each regression regresses the student post-CAL standardized math score on the treatment dummy variable and the interaction of the treatment dummy variable and the corresponding student characteristics in row (2) in that column, controlling for all the student characteristics in row (2) and school fixed effects (school dummies).

^a The baseline math/Chinese score is the score on the standardized math/Chinese test that was given to all students in the sample before the CAL program.

^b Access to other modern technologies is the mean value of a set of 0/1 dummy variables including whether the student has used cell phone, Internet, game console, CAL software, and videos for learning assistance.

* Significant at 10%.

** Significant at 5%.

*** Significant at 1%.

per month. Therefore, the cost in teacher compensation was \$5 per student. In addition to the cost related to computers and teacher–supervisor compensation, the maintenance cost, can be assumed to amount to \$1 per student. The software cost is negligible when averaged by the number of students. Therefore, the total per-student cost of our CAL program is \$7.9–\$8.8. Given that the average program effect is 0.14 standard deviations, the CAL program costs \$5.6–\$6.3 per tenth of a standard deviation improvement in the student test score.

Although the CAL program might be costly when compared with other programs (e.g. Banerjee et al., 2007), the cost effectiveness of CAL could be substantially improved if it were implemented on a large-scale, real-world basis. China's government already has committed to make vast investments in the computing facilities in rural and urban public schools. If these investments are not considered to be part of the CAL costs, the cost effectiveness of CAL would increase. By productively using the existing computers, the cost of CAL could be reduced to as low as \$4.3 per tenth of a standard deviation. Moreover, one computer can easily serve eight students, and one teacher can readily manage 64 students when working in a rotation shifts, which might further reduce the cost of the CAL program.

4.2. Heterogeneous effects of the CAL intervention on student academic performance

The estimation results using Eq. (1), but including interaction terms between the treatment variable and certain student characteristics show that the CAL intervention had some heterogeneous program effects on the math test scores of students with different family backgrounds (Table 4, row 2). Compared to the students in the control group, students in the treatment group whose fathers had never received a high school diploma and those whose father had never received a college diploma improved 0.20 standard deviations and 0.35 standard deviations more in their math scores than those whose fathers did, respectively (row 2, columns 8 and 9). Although this is understandable, it is important since it is one of the goals of the CAL program to improve the learning outcomes of those students that do not have anyone at home able to help them.¹¹ At the same time, we find no significant evidence of CAL intervention heterogeneous program effects for other student demographic and family characteristics (*gender, age, number of siblings, mother has a college/high school degree, living with father and living with mother*), or for the student baseline test scores, or for the students' access to computers and other technical devices before the program began (row 2, columns 1–7 and 10–13).

4.3. Spillovers in Chinese language test score

In principle, the CAL program in math might generate either positive or negative spillovers on student academic performance in Chinese, the other major subject in China's

elementary schools. Positive spillover might result from increased interest in learning promoted by the CAL program; negative spillover might result from the crowding out effects of the CAL program in math.

Over the program period, we observe no significant improvements in the standardized Chinese test scores for the treatment group in comparison to the control group. The results of our multivariate regression analysis using Eq. (1) also show that the estimated effect of the CAL intervention on Chinese test scores is 0.01 standard deviations and insignificant (Table 3, column 2).¹² All these results suggest that the CAL intervention in math did not create significantly positive spillovers in the Chinese test scores of the students. What is perhaps more important is that there is no evidence that the CAL program improved math performance at the expense of performance in Chinese.

4.4. Impact on non-academic outcomes

CAL intervention not only improved the student academic performance in math, the analysis demonstrates that there are increases in nonacademic outcomes for students in the treatment group (Table 3). Compared to students in the control group, the students in the treatment group “liked school” significantly more (column 3). On a scale of 0–10, the difference in improvement in the score on the indicator *like school* between the treatment and control groups was equal to 0.31 points and was significant at the 5% level. Moreover, students in the treatment group also improved 0.07 points more on a self-confidence assessment with a scale of 1–4 than those in the control group (column 4; and this difference is weakly significant at the 10% level).

On the contrary, we do not find any significant effects of the CAL intervention on self-reported math study efficacy (Table 3, column 5). One possible reason for CAL's lack of effect on the student's self-reported math study efficacy is the remedial nature¹³ of our CAL program. Due to its remedial nature, our CAL program focused more on repeated exercise rather than creative math learning and problem solving. Consequently, although the CAL program made the students like school better and develop stronger self-confidence in general (possibly via the interesting game-based CAL exercises which led to improved test scores), it may not make the students believe that they are more capable in math problem solving in general.

4.5. Robustness check

To examine whether our estimates of the CAL intervention were contaminated by the potential spillovers of the CAL

¹¹ As mentioned in Section 4.1, this pattern might also be due to the highly possible household substitution effect. Yet whatever the reason for this pattern is, overall, students with less educated parents benefit more from the CAL program.

¹² Given that the point estimate of the treatment effect is 0.01 standard deviations, and given that the standard error of the point estimate is 0.05 standard deviations (with a normal approximation), the upper bound of the treatment effect would be 0.01 plus 2 times 0.05 (the standard error). This means that the upper bound of the estimate is equal to 0.11 standard deviations. In other words, according to our estimates, we can rule out improvements larger than 0.11 standard deviations from the CAL intervention for the Chinese language subject.

¹³ By “remedial”, we mean material that was in the students' textbook and material that was being taught the same week (not material that was taught in the distant past).

Table 5

Ordinary Least Squares analysis of the difference in change in academic performance over the program period between the control group (the control classes in the 24 program schools) or treatment group in the core sample and the additional control group (the 19 schools excluded from our core sample).

Dependent variables: standardized post CAL score	Math (1)	Chinese (2)	Math (3)	Chinese (4)
(1) "Treatment" dummy (1 = "treatment": core sample control group for columns 1 and 2/core sample treatment group for columns 3 and 4; 0 = additional control group)	0.08 (0.06)	0.08 (0.07)	0.21*** (0.07)	0.11 (0.07)
Control variables				
(2) Baseline math score (units of standard deviation) ^a	0.41*** (0.02)	0.17*** (0.03)	0.42*** (0.03)	0.16*** (0.03)
(3) Baseline Chinese score (units of standard deviation) ^a	0.25*** (0.02)	0.42*** (0.03)	0.23*** (0.03)	0.46*** (0.03)
(4) Age (number of years)	-0.07*** (0.02)	-0.10*** (0.02)	-0.07*** (0.02)	-0.11*** (0.02)
(5) Female (1 = yes; 0 = no)	-0.06 (0.03)	0.28*** (0.03)	-0.05 (0.04)	0.25*** (0.04)
(6) Ever used a computer (1 = yes; 0 = no)	0.13*** (0.03)	0.15*** (0.04)	0.09** (0.04)	0.14*** (0.04)
(7) Access to other modern technologies (1 = yes; 0 = no) ^b	-0.03 (0.06)	-0.19*** (0.05)	-0.07 (0.07)	-0.18*** (0.06)
(8) Father has a college degree (1 = yes; 0 = no)	0.21** (0.09)	0.06 (0.11)	0.06 (0.11)	-0.06 (0.11)
(9) Father has a high school degree (1 = yes; 0 = no)	0.08 (0.05)	0.02 (0.05)	-0.02 (0.06)	-0.00 (0.05)
(10) Mother has a college degree (1 = yes; 0 = no)	0.09 (0.13)	-0.14 (0.14)	0.08 (0.15)	-0.17 (0.16)
(11) Mother has a high school degree (1 = yes; 0 = no)	-0.07 (0.05)	0.00 (0.06)	0.01 (0.07)	0.07 (0.07)
(12) Observations	2778	2778	2425	2425
(13) R-squared	0.39	0.39	0.36	0.40

Source: Authors' survey.

Robust standard errors in brackets, clustered at the class level.

Each column reports the results of one regression of the dependent variable (the standardized post-CAL test score on the respective subject) on the corresponding variables in rows (1)–(11).

^a The baseline math/Chinese score is the score on the standardized math/Chinese test that was given to all students in the sample before the CAL program.

^b Access to other modern technologies is the mean value of a set of 0/1 dummy variables including whether the student has used cell phone, Internet, game console, CAL software, and videos for learning assistance.

** Significant at 5%.

*** Significant at 1%.

intervention, we included the 19 additional control schools in our sample and conducted two steps of analyses. We first examined whether there was any significant difference in change in test scores between the control group (30 control classes in the 24 program schools) and the 19 additional control schools over the program period using Eq. (1), excluding the school fixed effects and replacing the treatment dummy with an indicator of whether the student belonged to the control group or additional control group. In principle, the change in test scores over the program period should be the same for the control and additional control groups, as neither of these two groups received the CAL intervention. A significant difference in the change in test scores between these two groups would imply a significant spillover of the CAL intervention. We then estimated the treatment effects using the 19 additional control schools instead of the original control group. If there was no significant program spillover, the estimated program effects would be consistent with those using the original control group.

The results show that the program spillovers are unlikely to have confounded the estimates of the program effects. Over the program period the difference in change in test

scores between the additional control group and the control group in our core sample was not significantly different from zero (Table 5, row 1, columns 1 and 2). The estimated CAL intervention effects on student math performance also did not significantly differ whether we used the control group in our core sample or the additional control group, implying no significant confounding influences from either the program spillovers of the CAL intervention (row 1, columns 3 and 4).¹⁴

5. Conclusion

In this paper we present the results from a randomized field experiment of a CAL program involving over 4000 third-grade students. The main intervention was a math CAL remedial tutoring program that was held outside of regular school hours. Our results indicate that CAL has significant benefi-

¹⁴ The point estimates of the CAL program effects are larger than those using the control group in the core sample, yet the standard errors were also larger.

cial effects on both student academic and non-academic outcomes. Two 40-min CAL math sessions per week increased the student standardized math scores by 0.15 standard deviations. Students with less-educated parents benefited more from the program. CAL also significantly increased the interest in learning of the students.

This paper contributes to the understanding of the effect of CAL on learning outcomes for underserved populations in developing countries in several respects. First, we took care in preparing software and hardware for the program and designing our CAL program implementation and evaluation protocol in order to prevent potentially confounding influences from affecting the measured impacts. Many previous studies reported various shortcomings in program implementation (e.g., schools in the treatment group used program computers for other purposes, such as in the case reported in [Banerjee et al., 2007](#)) that might potentially have biased the evaluation results. Our protocol took various measures to prevent such interferences. By implementing the CAL program as a full supplementary program, we also eliminated any substitution effects that might have diminished the program effects. Further, unlike many studies in the literature that use only one control group, our evaluation design included an additional control group—one that had never heard about the program. This further allowed for rigorous measurement (and thus elimination) of possible program spillovers on to the control group and disentangle such spillovers from the genuine program effects.

This paper also contributes to the understanding of the effects of CAL by its focus on outcomes beyond test scores. Besides the impacts of CAL intervention on student test scores, this paper also examined the impacts on non-academic outcomes such as self-confidence and interest in schooling.

Given the significant impact of CAL on student academic and nonacademic outcomes found in this paper, educational policy makers in China (and in other developing countries, as well as underserved communities in developed countries) should consider upscaling CAL programs, especially in public schools serving disadvantaged students (e.g., rural public schools in China). Of course, private migrant schools in Bei-

jing might not be representative of public schools in China or in other developing countries. Nonetheless, migrant schools in Beijing and public schools (especially public schools in poor rural areas) that serve disadvantaged students do share some common problems: low teacher quality, poor school resources, lack of remedial tutoring and the resulting persistent underperformance of the students.

The importance of CAL and CAL-like programs will likely grow as China's government, and other governments in developing countries, have committed to making large investments in the computing facilities in rural public schools. However, in many rural schools, after the investment in computing facilities has been completed, the computer rooms are locked and the computers are frequently unused because the schools do not know how to properly use them to facilitate student learning. This paper demonstrates that a CAL program can be used as a complementary input to existing computing resources and has the potential to promote learning outcomes for underserved students by productively using these computer technologies. Therefore, we believe that the government—in China and elsewhere—might consider extending CAL programs on a larger scale in China and then rigorously evaluate these new initiatives to inform policies that intend to provide better educational service to the poor.

Acknowledgements

The authors would like to thank *Quanta Computer* for their support for the project. They not only generously provided the computers and other hardware, they also provided suggestions and took an interest in the project from start to finish. Additional acknowledgement goes to Tianhua Shidai for supplying part of the software. The REAP software team, especially Nicholas Cheng, deserves much credit. We also would like to thank the China Children and Teenagers' Fund for their gracious collaboration. Target Foundation, TAG Foundation, Adobe and two anonymous private donors provided funding for the evaluation and implementation. We thank all of the REAP students, interns and volunteers for their tireless efforts.

Appendix A. Ordinary Least Squares analysis of the main and heterogeneous impact of CAL program on student academic outcomes before the midterm for the students in the 24 program schools.

Dependent variable: standardized midterm math score	(1)	(2)	(3)	(4)
(1) Treatment (1 = the treatment group; 0 = the control group)	0.14*** (0.04)	0.15*** (0.04)	0.16*** (0.04)	0.15*** (0.04)
(2) Treatment*father has at least a high school degree			-0.05 (0.12)	
(3) Treatment*father has a college degree				0.24 (0.20)
Control variables				
(4) Baseline math score (units of standard deviation) ^a	0.54*** (0.02)	0.41*** (0.03)	0.41*** (0.03)	0.41*** (0.03)
(5) Baseline Chinese score (units of standard deviation) ^a		0.22*** (0.03)	0.22*** (0.03)	0.22*** (0.03)
(6) Age (number of years)		-0.03 (0.03)	-0.03 (0.03)	-0.03 (0.03)
(7) Female (1 = yes; 0 = no)		-0.00 (0.04)	-0.00 (0.04)	-0.00 (0.04)
(8) Ever used a computer (1 = yes; 0 = no)		0.12*** (0.04)	0.12*** (0.04)	0.12*** (0.04)
(9) Access to other modern technologies ^b		-0.07 (0.07)	-0.07 (0.07)	-0.07 (0.07)
(10) Living with father (1 = yes; 0 = no)		-0.03 (0.06)	-0.03 (0.06)	-0.03 (0.06)
(11) Living with mother (1 = yes; 0 = no)		0.06 (0.07)	0.06 (0.07)	0.06 (0.07)
(12) # of siblings		-0.02*** (0.01)	-0.02*** (0.01)	-0.02*** (0.01)
(13) Father has a college degree (1 = yes; 0 = no)		-0.02 (0.13)		-0.12 (0.16)
(14) Father has at least a high school degree (1 = yes; 0 = no)		-0.03 (0.06)	-0.00 (0.12)	-0.03 (0.06)
(15) Mother has a college degree (1 = yes; 0 = no)		0.07 (0.14)	0.07 (0.14)	0.05 (0.14)
(16) Mother has a high school degree (1 = yes; 0 = no)		-0.00 (0.07)	-0.00 (0.07)	-0.00 (0.07)
(17) School dummy variables	Y	Y	Y	Y
(18) Observations	2369	2369	2369	2369
(19) R-squared	0.34	0.38	0.38	0.38

Source: Authors' survey.

Robust standard errors in brackets, clustered at the class level.

Each column reports the results of one regression of the dependent variable on the corresponding variables in rows (1)–(17), controlling for the school dummy variables.

^a The baseline math/Chinese score is the score on the standardized math/Chinese test that was given to all students in the sample before the CAL program.

^b Access to other modern technologies is the mean value of a set of 0/1 dummy variables including whether the student has used cell phone, Internet, game console, CAL software, and videos for learning assistance.

** Significant at 5%.

*** Significant at 1%.

Appendix B. The impact of the CAL program on non-cognitive outcome at midterm and endline periods

	Like school midterm (1)	Like school endline (2)	Self-efficacy midterm (3)	Self-efficacy endline (4)	Self-confidence midterm (5)	Self-confidence endline (6)
(1) Treatment (1 = the treatment group; 0 = the control group)	0.01 (0.11)	0.31** (0.15)	0.04 (0.03)	-0.00 (0.03)	0.06 (0.03)	0.07 (0.04)
(2) Baseline value of the outcome variable	Y	Y	Y	Y	Y	Y
(3) School dummy variables and other control variables	Y	Y	Y	Y	Y	Y
(4) Observations	2369	2369	2369	2369	2369	2369

Source: Authors' survey.

Robust standard errors in brackets, clustered at the class level.

Each column reports the results of one regression of the student outcome variable (corresponding to the column title) on the treatment dummy variable, controlling for the baseline value of the outcome variable, the school dummy variables and other control variables.

Other control variables include the student age, female (1 = yes; 0 = no), baseline math/Chinese score, ever used a computer (1 = yes; 0 = no), access to other modern technologies, living with father (1 = yes; 0 = no), living with mother (1 = yes; 0 = no), # of siblings, father has college degree (1 = yes; 0 = no), father has high school degree (1 = yes; 0 = no), mother has college degree (1 = yes; 0 = no), mother has high school degree (1 = yes; 0 = no).

** Significant at 5%. *** Significant at 1%.

References

- All-China Women's Federation. (2008). *Quan Guo Nong Cun Liu Shou Er Tong Zhuang Kuang Yan Jiu Bao Gao* (Report of the status of rural left-behind children in China).
- Angrist, J., & Lavy, V. (2002). New evidence on classroom computers and pupil learning. *The Economic Journal*, 112(October), 735–765.
- Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122(3), 1235–1264.
- Barrow, L., Markman, L., & Rouse, C. (2008). Technology's edge: The educational benefits of computer-aided instruction. *National Bureau of Economic Research Working Paper No. 14240*.
- CCAP (Center for Chinese Agricultural Policy Chinese Academy of Sciences) (2009). *Policy Brief, 9*. Beijing, China.
- Das, Jishnu, Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2011). School inputs, household substitution, and test scores. *NBER Working Paper No. 16830*.
- Ding, J. (2004). Overhauling and regulating schools set up by migrants. *Chinese Education and Society*, 37(5), 10–28.
- Dynarski, M. (2007). Effectiveness of reading and mathematics software products: Findings from the 1st student cohort. *Mathematica Report*.
- Fuchs, T., & Woosmann, L. (2004). Computers and student learning: Bivariate and multivariate evidence on the availability and use of computers at home and at schools. *Brussels Economic Review*, 47, 359–585.
- Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook on the Economics of Education: vol. 2* (pp. 945–1017). New York, NY: Elsevier.
- Glewwe, P., & Moulin, S. (2002). Textbooks and test scores: Evidence from a prospective evaluation in Kenya. *BREAD Working Paper*, Cambridge, MA.
- Glewwe, P., & Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: The case of flip charts in Kenya. *Journal of Development Economics*, 74, 251–268.
- Goolsbee, A., & Guryan, J. (2006). The impact of Internet subsidies in public schools. *The review of economics and statistics*, MIT Press, 88(2), 336–347.
- Han, J. (2004). Survey report on the state of compulsory education among migrant children in Beijing. *Chinese Education and Society*, 37(5), 29–55.
- Hanushek, E. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24, 1141–1177.
- Hanushek, E. (1995). Interpreting recent research on schooling in developing countries. *World Bank Research Observer*, 10, 227–246.
- He, F., Linden, L., & MacLeod, M. (2008). How to teach English in India: Testing the relative productivity of instruction methods within the Pratham English language education program. *Working Paper*. Columbia University Department of Economics.
- Human Rights Watch. (2006). China: Beijing closes schools for migrant children in pre-Olympic clean-up. September 26. Accessed online March 17, 2011: <http://www.hrw.org/en/news/2006/09/25/china-beijing-closes-schools-migrant-children-pre-olympic-clean>.
- Kirkpatrick, H., & Cuban, L. (1998). Computers make kids smarter – Right? *Technos Quarterly for Education and Technology*, 7(2), 26–31.
- Kwong, J. (2004). Educating migrant children: Negotiations between the state and civil society. Research Report. *The China Quarterly*, 180, 1073–1088.
- Lai, F., Liu, C., Luo, R., Zhang, L., Ma, X., Bai, Y., et al. (2011). Private migrant schools or rural/urban public schools: Where should China educate its migrant children? *REAP Working Paper*.
- Lai, F., Liu, C., Luo, R., Zhang, L., Ma, X., Bai, Y., Sharbono, B., & Rozelle, S. (2014). The Education of China's Migrant Children: The Missing Link in China's Education System? *International Journal of Educational Development*, 37, 68–77.
- Linden, L. (2008). Complement or substitute? The effect of technology on student achievement in India. *Working Paper*. Columbia University Department of Economics.
- Liu, X. (2002). *Overview of educational needs & philanthropic opportunities in China*. San Diego, CA: International Community Foundation.
- Ma, X., Bai, Y., Liu, C., Pang, X., & Rozelle, S. (2008). Migrant schools in China's capital city: A descriptive review of a representative migrant survey. *Working Paper*. Center for Chinese Agricultural Policy, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China.
- Muralidharan, K., & Kremer, M. (2006). Public and private schools in rural India. *Working paper*. Harvard University, Graduate School of Education.
- Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119, 39–77.
- Newhouse, D., & Beegle, K. (2006). The effect of school type on academic achievement: Evidence from Indonesia. *Journal of Human Resources*, 41(3), 529–557.
- Planty, M., Hussar, W., Snyder, T., Provasnik, S., Kena, G., Dinkes, R., et al. (2008). *The condition of education 2008* (NCES 2008-031). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Present's Committee of Advisors on Science and Technology. (1997). Panel on educational technology. *In Report to the President on the use of technology to strengthen K-12 education in the United States*. Washington DC: Office of the President.
- Rouse, C., & Krueger, A. (2004). Putting Computerized instruction to the test: A randomized evaluation of a 'scientifically-based' reading program. *Economics of Education Review*, 23, 323–338.
- Sa, Z. (2004). A case study of rural-urban migrant children's education barriers in China. In *American Sociological Association Annual Meeting*. San Francisco, CA.
- Schwarzer, R. (Ed.). (1992). *Self-efficacy: Thought control of action*. Washington, DC: Hemisphere.
- Song, Y., Loyalak, P., & Wei, J. (2010). *Does going to public schools matter for migrant children's academic achievement in China*. China Institute for Educational Finance Research, Peking University.
- Tao, H., & Yang, D. (2007). Problems of "moving children" of Beijing city in education and its countermeasures. *Jiangxi Educational Research*, 1, 61–63.
- World Bank. (2004). *World development report 2004: Making services work for the poor*. New York, NY: Oxford University Press.
- Zhang, J. X., & Schwarzer, R. (1995). Measuring optimistic self-beliefs: A Chinese adaptation of the general self-efficacy scale. *Psychologia: An International Journal of Psychology in the Orient*, 38(3), 174–181.