

This article was downloaded by: [DI Mo]

On: 25 April 2014, At: 23:23

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Journal of Development Effectiveness

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/rjde20>

### Integrating computer-assisted learning into a regular curriculum: evidence from a randomised experiment in rural schools in Shaanxi

Di Mo<sup>abc</sup>, Linxiu Zhang<sup>c</sup>, Renfu Luo<sup>c</sup>, Qinghe Qu<sup>c</sup>, Weiming Huang<sup>d</sup>, Jiafu Wang<sup>e</sup>, Yajie Qiao<sup>f</sup>, Matthew Boswell<sup>b</sup> & Scott Rozelle<sup>b</sup>

<sup>a</sup> Freeman Spogli Institute, Stanford University, Stanford, USA

<sup>b</sup> LICOS Centre for Institutions and Economic Performance, KU Leuven, Leuven, Belgium

<sup>c</sup> CCAP, IGSNRR, Chinese Academy of Sciences, Beijing, China

<sup>d</sup> Economics Department, Colby College, Waterville, USA

<sup>e</sup> Department of Economics, Indiana University Bloomington, Bloomington, USA

<sup>f</sup> Department of Economics and Management, Ankang University, Ankang, China

Published online: 23 Apr 2014.

To cite this article: Di Mo, Linxiu Zhang, Renfu Luo, Qinghe Qu, Weiming Huang, Jiafu Wang, Yajie Qiao, Matthew Boswell & 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, DOI: [10.1080/19439342.2014.911770](https://doi.org/10.1080/19439342.2014.911770)

To link to this article: <http://dx.doi.org/10.1080/19439342.2014.911770>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or

howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## Integrating computer-assisted learning into a regular curriculum: evidence from a randomised experiment in rural schools in Shaanxi

Di Mo<sup>a,b,c,\*</sup>, Linxiu Zhang<sup>c</sup>, Renfu Luo<sup>c</sup>, Qinghe Qu<sup>c</sup>, Weiming Huang<sup>d</sup>, Jiafu Wang<sup>e</sup>, Yajie Qiao<sup>f</sup>, Matthew Boswell<sup>b</sup> and Scott Rozelle<sup>b</sup>

<sup>a</sup>Freeman Spogli Institute, Stanford University, Stanford, USA; <sup>b</sup>LICOS Centre for Institutions and Economic Performance, KU Leuven, Leuven, Belgium; <sup>c</sup>CCAP, IGSNRR, Chinese Academy of Sciences, Beijing, China; <sup>d</sup>Economics Department, Colby College, Waterville, USA; <sup>e</sup>Department of Economics, Indiana University Bloomington, Bloomington, USA; <sup>f</sup>Department of Economics and Management, Ankang University, Ankang, China

Recent attention has been placed on whether computer assisted learning (CAL) can effectively improve learning outcomes. However, the empirical evidence of its impact is mixed. Previous studies suggest that the lack of an impact in developed countries may be attributable to substitution of effort/time away from productive, in-school activities. However, there is little empirical evidence on how effective an in-school programme may be in developing countries. To explore the impact of an in-school CAL programme, we conducted a clustered randomised experiment involving over 4000 third and fifth grade students in 72 rural schools in China. Our results indicate that the in-school CAL programme has significantly improved the overall math scores by 0.16 standard deviations. Both the third graders and the fifth graders benefited from the programme.

**Keywords:** computer assisted learning; out-of-school programme; in-school programme; substitution effect; test scores

### Introduction

In the last decade, attention has been placed on initiatives that adopt computer technology to confront the long-standing challenge of delivering quality education to poor and disadvantaged populations (for example, Banerjee et al. 2007; Barrow, Markman, and Rouse 2009; Linden 2008; Cristia et al. 2012; Guimarães et al. 2013). These studies aim to evaluate whether educational input, such as computer assisted learning (CAL) programmes, can improve student learning. CAL programmes utilise computers and modern computing technologies (including both software and hardware devices) to enhance learning through computerised instruction, drills and exercises (Present's Committee of Advisors on Science and Technology 1997; Kirkpatrick and Cuban 1998). When integrated with well-designed educational games, CAL programmes can sustain the interest and curiosity of students and lead to gains in student performance. The programme also allows for the delivery of a consistent curriculum regardless of training or expertise of the teachers (Nara and Noda 2003). In some contexts, such a CAL programme may be more cost-effective than using teachers to provide additional instruction (Banerjee et al. 2007).

Despite the popularity of investment in computer technology in education, the empirical evidence on the impact of such programmes has been mixed. Early studies in Israel

---

\*Corresponding author. Email: [dimo@stanford.edu](mailto:dimo@stanford.edu)

Present affiliation for Weiming Huang is Freeman Spogli Institute, Stanford University, Stanford, USA. Present affiliation for Jiafu Wang is Department of Economics, Duke University, Durham, USA.

and the United States found little consistent evidence as to whether the application of computer technology in school instruction has beneficial effects for student academic achievement (for example, Angrist and Lavy 2002; Fuchs and Woosmann 2004; Goolsbee and Guryan 2006). Later studies utilising randomised experiments to evaluate specific CAL programmes also found mixed evidence. For example, both Dynarski (2007) and Krueger and Rouse (2004) found no significant gain in math and reading test scores from CAL programmes for students in the United States. In contrast, Barrow, Markman, and Rouse (2009) found a CAL programme improved student math test scores by 0.17 standard deviations in urban schools in three districts in the United States. This particular CAL programme used computer-aided instructions in algebra to replace traditional classroom teaching. Although relatively few in number, evaluations conducted in developing countries mostly show CAL has had positive effects on student test scores (Banerjee et al. 2007; He, Linden, and MacLeod 2008; Linden 2008).

An important limitation to these studies is that they usually examine CAL as an educational input, and do not consider whether the programme is rolled out as an *in-school programme* or *out-of-school programme*. Miettinen (1999) defines in-school programmes as those that occur during regular school hours and/or those that are organised as formal classroom activities. In contrast, out-of-school programmes take place after school hours and/or are built around less formal group activities.

Out-of-school programmes have several potential advantages over in-school programmes. First, out-of-school programmes can be run without the restrictions of the formal classroom (for example, the limited scope of activities that are allowed to take place under the supervision of formal teachers). Second, out-of-school programmes can be run without taking time away from other regularly scheduled classroom learning activities. For these reasons, remedial education camps and other after-school programmes have been found to lead to higher levels of learning (Hull and Schultz 2002). Third, out-of-school programmes are also potentially easier to design in a way that effectively caters to the needs of individuals, as has been shown with professional training programmes that teach workplace skills (Dias et al. 1999).

Despite their potential benefits, out-of-school programmes also have several disadvantages. First, successfully implementing such programmes often requires schools to make various structural changes regarding curriculum, staff allocation and meal programmes. Second, the success of out-of-school programmes often depends on the extraordinary effort of teachers or volunteers. As a consequence, it is often suggested that for programmes to be sustainable, they ultimately need to be incorporated into the regular school-day curriculum (Underwood et al. 2000). Furthermore, the advantages of out-of-school programmes may disappear if in-school programmes are organised to use educational resources more efficiently (Cole 1996).

Previous research shows that an in-school CAL programme is ineffective in improving student learning in India (Linden 2008). In contrast, the same research team showed that an out-of-school version of the same programme did improve student learning. According to Linden (2008), one of the reasons for the difference in impact is that the in-school programme appears to have created a substitution of effort/time away from other productive, in-class activities (such as formal instruction by the teacher in math and language). The out-of-school programme instead appears to have served as a complement to existing resources.

Despite the interesting findings, the Linden study did not take into account how the in-school CAL programme was incorporated into the regular school curriculum. India is known for short school days. Schools often provide only five to six hours per day of instruction (as reported in the study by Duflo, Hanna, and Ryan (2012)). Incorporating CAL during school hours may have to replace relatively productive teaching periods.

There is also high absenteeism of teachers during regular school hours (Chaudhury et al. 2005), which may have contributed to the ineffectiveness of the in-school CAL programme, particularly if the supervising teachers frequently missed the CAL sessions.

In contrast, rural schools in China have many features that potentially make an in-school CAL programme more effective. In rural China, a school day typically runs from seven to eight hours with an additional one-hour noon break. China is reported to have much lower teacher absenteeism than India (Liu and Kumar 2008). Moreover, all Chinese schools are required to allocate time for computer, art or music classes. However, since many rural schools do not have teachers for these subjects, rural schools typically have multiple time slots in a week that are relatively unproductive. In other words, there may be less of a substitution effect if CAL programmes are run in-school. Thus it is likely that an in-school CAL programme in rural China may be more effective in improving student learning than a similar in-school programme in India.

Previous studies that employed large-scale randomised controlled trials to evaluate CAL interventions were all run as out-of-school programmes (Lai et al. 2011, 2012, 2013; Mo et al. 2013). These studies have shown that CAL can significantly improve the math and Chinese test scores of rural students in China. However, in these previous efforts to test CAL, there was no danger of the CAL programme substituting for other classroom activities run by teachers during the regular teaching day. The question remains whether a CAL programme in Chinese schools will be equally effective in improving learning when it is implemented as an in-school programme during regular school hours.

The overall goal of this paper is to explore the impact of an in-school CAL programme on the academic outcomes of an underserved student population in a developing country. To achieve this goal, the main question that we seek to answer in this paper is whether an in-school CAL programme increases school performance. To do this, we also address the questions of whether in-school CAL programmes increase the academic performance of grade 3 students and grade 5 students, and how the in-school CAL effect compares with that of a CAL programme implemented outside of the classroom.

The rest of the paper is organised as follows. The next section reviews the study by Lai et al. (2013), which reports on the results of an out-of-school CAL programme in rural China that was also conducted by the current authors. The third section describes the current study's methodology, including the research design and sampling, intervention design, data collection and statistical approach. The remaining sections present the results from the study, discuss the findings and conclude.

### **In-school CAL programme in China**

The question of whether an in-school CAL can improve student learning is of particular relevance to China. As part of a new effort to improve the facilities in rural schools, the government has recently invested in improving the computing infrastructure of rural public schools (Yuan 2012). By 2011, 86 per cent of the rural public schools had set up computer rooms with an average of 17 computers in each school (Yang et al. 2013). China's Ministry of Education, however, has even more ambitious plans. The recently announced 12th Five-Year Plan for Integrating Information Technology into Education aspires to set up a computer room in every rural school by 2020 (Ministry of Education 2012). Since the plan requires such an enormous investment of fiscal resources, it is important to learn how to effectively use the new computing resources.

Unfortunately, China's rural schools face several constraints in providing quality computer-based education. Teachers at rural schools typically do not have the

qualifications or materials necessary to promote learning in computer classes (Lai et al. 2011). Teachers lack the training and/or motivation to adequately instruct students and pique the interest of students in using computers for learning. There is also a shortage of curricula to use during computing class, particularly in poor rural areas (Yang et al. 2013). Although 69 per cent of students in rural public schools have computer classes, research shows that few schools have employed computers and/or educational software for instructional purposes in core academic subjects. Even when they do, computer classes are frequently cancelled due to a lack of teachers and instructional materials.

If CAL classes could occur in periods already assigned for computer classes – which are not being used effectively – it may be that in-school CAL classes in China could bring the proven benefits of the CAL programme without any offsetting effects. In other words, if CAL could be conducted during the regularly scheduled but poorly utilised computer class period, CAL in rural China may be both integrative and supplemental.

## **Sampling, data and methods**

### *Sampling and the process of randomisation*

We conducted a clustered (at the school level) Randomised Controlled Trial of CAL in Shaanxi rural schools during the 2011–2012 academic school year. A total of 5267 students in 72 rural Shaanxi schools were involved in the study. The study covered third grade and fifth grade students.<sup>1</sup>

Choosing the sample consisted of several steps. First, to focus our study on students from poor rural areas, we restricted our sample frame to four counties randomly selected out of the ten counties in Ankang Prefecture, the prefecture that covers one of the poorest areas in the southern region of Shaanxi Province. Shaanxi Province is situated in north-west China, which is one of the poorest regions in the country (Ezroj et al. 2004). Shaanxi ranks the second place among all provinces in China (NBSC 2013) in terms of number of nationally designated poor counties. Ankang prefecture (where our sample counties are selected from) covers one of the poorest areas in the southern part of Shaanxi Province. Eighty per cent of the counties in Ankang are nationally designated counties. The average per capita income of the randomly selected four counties was about 4000RMB (\$650) per year in 2011, which is far below rural China's average per capita income of 6977RMB in the same year (NBSC 2011). Three out of the four sample counties are nationally designated poor counties in China.<sup>2</sup>

After choosing the counties, we obtained a comprehensive list of all *wanxiao* (elementary schools with six full grades – grade one through grade six) in each of the four counties from the Department of Education of Ankang Prefecture.<sup>3</sup> We included all 72 schools that met the above criterion in our sample.

Within the sample schools, we included both third grade and fifth grade students in the 72 schools in our sample. We chose third grade and fifth grade students for two reasons. First, at the time of the launch of the project, we only had remedial tutoring material for students from third to sixth grade. It is for this reason that we did not choose students from first grade or second grade. Second, a subset of the fourth grade and sixth grade students in the school had already participated in a pilot project during the previous academic year. In order to avoid confounding the treatment effect, we chose to focus the intervention on third grade and fifth grade students. Again, none of these students had ever participated in a CAL programme prior to the 2011–2012 academic year.

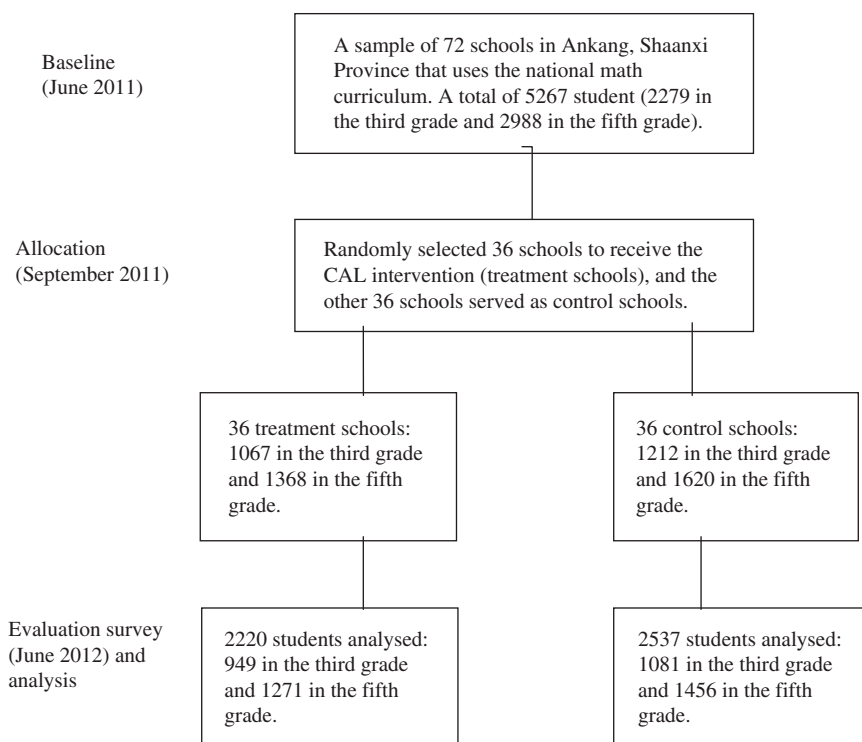


Figure 1. Experiment profile.

All of the third grade and fifth grade students in the 72 sample schools were included. In phase one of this study (Lai et al. 2013), we had only included students who boarded at school. In this study we included students who were boarding at school and students who were living at home. Of the total number of students involved in the study (5267), 2279 were third grade students and 2988 were fifth grade students (Figure 1).

Although at the time of the baseline survey the main sample included a total of 72 schools and 5267 students, for various reasons (mainly because of school transfers and extended absences due to illness or injuries), there was some attrition by the end of the study. By the time of the evaluation survey we were able to follow up with 4757 students in the 72 sample schools (Figure 1, final row). In other words, 4757 out of the initial 5267 students (who took the baseline survey) were included in our evaluation survey and were part of the subsequent statistical analysis; 9.8 per cent of the sample dropped out between the baseline and endline surveys. There were 249 attrited students (10.9%) from the third grade and 261 attrited students (8.7%) from the fifth grade. Fortunately for the study's integrity, there were no variables that were systematically related between the characteristics of students and their attrition status (Table 2).

After choosing the 72 schools for our sample, we randomly assigned them to either the treatment or control group. This assignment was done after the baseline. During the baseline, both the enumerators and the respondents/participants were blind to their eventual group assignment. In order to assure that the treatment and control groups were similar in terms of key characteristics at the time of the baseline, we pre-balanced along several key variables when we randomised. These key variables include the control

variables listed in Appendix 1 (that is, *student gender, student age, boarding student, ever repeated a grade, only child, age of father, age of mother, father has at least junior high school degree, mother has at least junior high school degree, at least one parent lives at home and family wealth*). This method was discussed by Bruhn and McKenzie (2009). In doing so, we re-randomised several times until the key baseline variables that are listed in Appendix 1 in the revised manuscript were balanced between the treatment and control groups.

After the randomisation, 36 schools were assigned to receive the CAL intervention. As the CAL intervention engaged both third grade and fifth grade students, the 2435 students of the third and fifth grades in the 36 treatment schools constitute the treatment group (Figure 1). Among these students, there were 1067 third grade students and 1368 fifth grade students. The 2832 students (1212 from the third grade and 1620 from the fifth grade) in the other 36 schools served as the control group. Because of the attrition, there were 4757 students left in our final analytic sample, among whom 2220 were in the 36 treatment schools, and 2537 were in the control schools.

### *Experiment arm/intervention*

The main intervention involved computer-assisted math remedial tutoring sessions that were designed to complement the regular in-class math curriculum for the entire school year 2011–2012. Under the monitoring of two teacher-supervisors trained by our research group, the students in the treatment group had two 40-minute CAL sessions per week as regular classes in school.<sup>4,5</sup> The sessions were mandatory and attendance was taken by the teacher-supervisors.

According to our protocol, the CAL sessions were supposed to be given during the normal ‘computer class’ time period. We chose the computer class time periods since typically these are reserved for teaching non-academic material. Based on our surveys, in the computer classes offered in most of China’s rural schools, students were taught basic computer operations, such as using a mouse, typing Chinese and using Microsoft Office software. On average, in 75.6 per cent of the rural public schools in Shaanxi Province students are taught such basic computer operations in computer classes. When the schools do not have computer teachers to teach the class, computer class time is frequently used for students to practice math, Chinese or English questions under teacher supervision.

The instructional videos and games that comprise the content of each CAL session were designed for improving students’ basic competencies in the uniform national math curriculum. The content was exactly the same for all students within the same grade among schools in the treatment group. During each session, two students shared one computer and played math games designed to help students review and practice the basic math material that was being taught in their regular school math classes. In a typical session, the students first watched an animated video that reviewed the material that they were receiving instruction on during their regular math class sessions in that week. The students then played math games with animated characters to practice the skills introduced in the video lecture.<sup>6</sup> If students had a math-related question, they were encouraged to discuss it with their teammate (the student they shared the computer with). The students were not allowed to discuss their questions with other teams or the teacher-supervisor. Our protocol required that the teachers could only help students with scheduling, computer hardware issues and software operations.<sup>7</sup> In fact, according to our observations, the sessions were so intense that the students were almost always exclusively focused on their computers. There was little communication among the groups or between any of the



groups and the teacher-supervisor. The CAL software had enough content and exercise games to cover the math course materials for the entire school year 2011–2012 and the material was sufficient to provide 80 minutes of remedial tutoring per week (two 40-minute sessions).

With both software and hardware ready, we then worked out a detailed CAL curriculum and implementation protocol. The protocol was targeted exclusively at the teacher-supervisors who were responsible for implementing the CAL programme in each school. The CAL curriculum was designed to keep pace with the progress of school instruction on a week-by-week basis. This was done so that our CAL sessions provided a timely review and an opportunity to practice the knowledge and skills that were introduced and covered as part of their regular math class. One of the most important jobs of the teacher-supervisor was to make sure the weekly CAL sessions proceeded on a pace that matched the pace of the students' regular math classes. Because this work was clearly beyond the scope of their normal classroom duties, we compensated the teacher-supervisors with a monthly stipend of 100 yuan (approximately 15 USD), an amount roughly equivalent to 15 per cent of the wage of a typical rural teacher. All teacher-supervisors of the 36 treatment schools also participated in a two-day mandatory training programme.

*CAL control group (the students in the 36 control schools).* The third grade and fifth grade students in the 36 control schools constituted the CAL control group. Students in the control group did not receive any CAL intervention. To avoid any type of the spillover effects of the CAL intervention, the principals, teachers and students (and their parents) of the control schools were not informed of the CAL project. The research team did not visit the control schools except for during the baseline and final evaluation surveys. The students in the control group took their regular math and computer classes at school as usual.

#### *Data collection*

The research group conducted two rounds of surveys in the 72 control and treatment schools. The first-round survey was a baseline survey conducted with all third and fifth graders in the 72 schools in June 2011 at the end of the spring semester and before any implementation of CAL programme had begun. The second-round survey was a final evaluation survey conducted at the end of the programme in June 2012.

In each round of the survey, the enumeration team visited each school (treatment and control schools) and conducted a two-part survey. In the first part students were given a standardised math test, which provided us our main outcome variable. The math test included 29–31 questions (tests in different grades and rounds included slightly different numbers of questions).<sup>8</sup> All the questions were chosen to not repeat the questions that were contained in the exercises in the CAL software. Students were required to finish tests in each subject in 25 minutes. Time limits were strictly enforced.

In the second part, enumerators collected data on the characteristics of students and their families. From this part of the survey we were able to create demographic and socioeconomic variables. The data set includes measures of each student's *gender*, *age* (measured in years), whether the student is a *boarding student*, has the student *ever repeated a grade*, if the student is the *only child* of his or her family, father's education level (*father has at least junior high school degree*), mother's education level (*mother has at least junior high school degree*) and parental care (*at least one parent lives at home*). To create the indicator of family wealth, we documented the ownership of household

appliances to proxy the family asset value. The variable of family wealth equals 1 if the family assets are higher than the median value and 0 otherwise.

### *Statistical methods*

We used both unadjusted and adjusted ordinary least squares (OLS) regression analyses to estimate how the academic outcome changed in the treatment group relative to the control group. Our unadjusted analysis regressed the outcome variable (that is, post-programme math test score<sup>9</sup>) on a dummy variable of the treatment (CAL intervention) status. We used adjusted analysis as well to improve statistical efficiency. In all regressions, we corrected for school-level clustering (relaxing the assumption that disturbance terms are independent and identically distributed within schools).

The model we estimated is as follows:

$$y_{isc} = \alpha + \beta * treatment_s + \theta * y_{0isc} + X_{isc}\gamma + \varepsilon_{isc} \quad (1)$$

where  $y_{isc}$  is the outcome variable after the CAL programme for child  $i$  in school  $s$  and class  $c$ ,  $treatment_s$  is a dummy variable for a student attending a treatment school (equal to one for students in the treatment group and zero otherwise) and  $\varepsilon_{isc}$  is a random disturbance term clustered at the school level. We also included a set of control variables. Specifically, we controlled for  $y_{0isc}$ , the pre-programme math test score and Chinese test score for student  $i$  in school  $s$  and class  $c$ , and  $X_{isc}$ , a vector of additional control variables. The control variables are expected to only improve the precision of the estimates. The variables in  $X_{isc}$  are student and family characteristics (*gender, age, boarding student, ever repeated a grade, only child, age of father, age of mother, father has at least junior high school degree, mother has at least junior high school degree, at least one parent lives at home and family wealth*). By construction, the coefficient of the dummy variable  $treatment_s$ ,  $\beta$  is equal to the unconditional difference in the outcome ( $y_{isc}$ ) between the treatment and control groups over the programme period. In other words,  $\beta$  measures how the treatment group changed in the standardised math test score levels after the CAL programme relative to the control group. We estimate Equation (1) with control variables (adjusted model) and without control variables (unadjusted model).

The attrition pattern does not differ between the treatment and control groups. The results comparing the attrition rates between the treatment group and the control group show that the attrition rates are not affected by the treatment status (Table 1). In conducting the test, we estimate Equation (1) with the attrition status as the dependent variable and without control variables. As the results show, the attrition rates are not correlated with the treatment status when pooling the third grade and fifth grade students (column 1, row 1). Similarly, when testing attrition rates separately for the third and fifth grade students, no significant difference is found between the treatment and control groups (columns 2 and 3, row 1). All the coefficients are insignificant and close to zero.

We used a set of student characteristics to check the validity of the random assignment. We estimate Equation (1) without control variables using the baseline characteristics each at a time as the dependent variable. According to our data, we found that for all student characteristics, none of the differences between the treatment and control groups were statistically significant among the samples before attrition (Appendix 2) or among the samples after attrition (Appendix 3). The assignment of treatment is not correlated with any of the student characteristics for the sample students before attrition (Appendix 2, column 1). In addition, the differences are almost all small in magnitude.

Table 1. Comparisons of attrition between the treatment and control students.

Dependent variable: attrition (1 = students attrited; 0 = students remained)				
		All (Third grade and fifth grade) (1)	Third grade (2)	Fifth grade (3)
[1]	Treatment (1 = treatment group; 0 = control group)	-0.02 (0.02)	0.00 (0.03)	-0.03 (0.02)
[4]	Observations	5267	2279	2988
[5]	R <sup>2</sup>	0.001	0.000	0.003

Notes: Robust standard errors in parentheses clustered at school level.

The test aims to show whether attrition rates are different between the treatment and control groups. The test regresses attrition status on the treatment variable.

Consistently, none of the student characteristics are significantly different between the treatment and control groups when testing the validity of random assignment separately for third and fifth grade students before attrition (Appendix 2, columns 3 and 5). We found the same results on the sample students after attrition (Appendix 3). In other words, our results show that student characteristics are well balanced between the treatment and control groups both before and after attrition.

## Results

According to our analysis, the in-school CAL programme significantly improved the academic performance of the students in the sample treatment schools (Table 2). The multivariate regression analyses (adjusted and unadjusted) show that by pooling the third grade and fifth grade students, the programme impact is estimated to be 0.18 standard deviations (although not significant at the 10% level) using the unadjusted model (column 1, row 1). By controlling for student baseline scores and other characteristics, the standard error is largely reduced and the estimate is significant at the 1 per cent level. The programme impact slightly decreased to 0.16 standard deviations using the adjusted model in Equation (2) (column 2, row 1).

Using only third grade students or only fifth grade students, in the unadjusted model the estimated CAL treatment effects on math test scores are equal to 0.20 standard deviations for third grade students (Table 3, column 1, row 1), and 0.17 standard deviations for fifth grade students (Table 3, column 2, row 1). Both of the coefficients of interest in the unadjusted model are not significant (Table 3, columns 1 and 2, row 1).

Next we estimate the treatment impact by including control variables to improve the precision of the estimates. When we add the control variables (using the adjusted model), the more efficient estimates show that the CAL programme had a positive and significant impact on the standardised math scores of the students. The estimated treatment effect for third grade students is 0.17 standard deviations (Table 3, column 3, row 1) and is significant at the 10 per cent level. The estimated treatment effect for fifth grade students remains at 0.17 standard deviations and is significant at the 5 per cent level (Table 3, column 4, row 1).

So how big an effect size is 0.17 standard deviations? According to McEwan (2013) and Schagen and Hoden (2009), educators are often interested in promoting new

Table 2. Ordinary least squares estimators of the CAL impact on standardised math test scores on all students (third grade and fifth grade).

Dependent variable: standardised post-CAL math test score (standard deviation)		(1)	(2)
[1]	Treatment (1 = treatment group; 0 = control group)	0.18 (0.11)	0.16*** (0.06)
[2]	Baseline math score (units of standard deviation) <sup>a</sup>		0.39*** (0.02)
[3]	Baseline Chinese score (units of standard deviation) <sup>b</sup>		0.18*** (0.02)
[4]	Gender (1 = boy; 0 = girl)		0.07** (0.03)
[5]	Age (years)		-0.03 (0.02)
[6]	Boarding student (1 = yes; 0 = no)		0.06 (0.04)
[7]	Only child (1 = yes; 0 = no)		-0.13*** (0.04)
[8]	Ever repeated grade (1 = yes; 0 = no)		-0.03 (0.03)
[9]	Age of father (years)		0.00 (0.00)
[10]	Age of mother (years)		-0.00 (0.00)
[11]	Father has at least junior high school degree (1 = yes; 0 = no)		-0.01 (0.03)
[12]	Mother has at least junior high school degree (1 = yes; 0 = no)		0.01 (0.03)
[13]	At least one parent lives at home (1 = yes; 0 = no)		-0.01 (0.02)
[14]	Family wealth (1 = higher than the median; 0 = otherwise)		0.02 (0.03)
[15]	Observations	4757	4757
[16]	R <sup>2</sup>	0.008	0.317

Notes: \*\* significant at 5%; \*\*\* significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show the impact of the in-school CAL treatment on student math test scores of both the third and fifth grade students. The test regresses standardised post-CAL math test scores on the treatment variable and a set of control variables.

<sup>a,b</sup>The baseline math/Chinese score is the normalised score on the math/Chinese test that is given to all sample students before the CAL programme.

programmes if they have a 0.2 standard deviations effect on test scores. In a majority of the papers that were reviewed by a team (Krishnaratne, White, and Carpenter 2013), the most common effect size used for power calculations is 0.2 standard deviations. The implicit assumption is that 0.2 is big enough of an effect to be relevant. As seen in the paper, the in-school CAL programme managed to improve student performance by 0.17 standard deviations. According to a paper by Lai et al. (forthcoming), which calculates an urban-rural academic achievement gap, our CAL programme's effect size (0.17) could reduce the urban-rural gap by almost 20 per cent. In another study, we found that a similar

Table 3. Ordinary least squares estimators of the CAL impact on standardised math test scores of the third grade and fifth grade students.

Dependent variable: standardised post-CAL math test score (standard deviation)		Third grade (1)	Fifth grade (2)	Third grade (3)	Fifth grade (4)
[1]	Treatment (1 = treatment group; 0 = control group)	0.20 (0.15)	0.17 (0.13)	0.17* (0.09)	0.17** (0.07)
[2]	Baseline math score (units of standard deviation) <sup>a</sup>			0.31*** (0.03)	0.48*** (0.02)
[3]	Baseline Chinese score (units of standard deviation) <sup>b</sup>			0.17*** (0.04)	0.17*** (0.02)
[4]	Gender (1 = boy; 0 = girl)			0.00 (0.04)	0.12*** (0.03)
[5]	Boarding student (1 = yes; 0 = no)			-0.06*** (0.02)	-0.08*** (0.03)
[6]	Age(years)			0.13 (0.08)	0.01 (0.04)
[7]	Only child (1 = yes; 0 = no)			-0.14** (0.05)	-0.10*** (0.03)
[8]	Ever repeated grade (1 = yes; 0 = no)			-0.08* (0.04)	-0.01 (0.03)
[9]	Age of father (years)			0.01 (0.00)	-0.00 (0.01)
[10]	Age of mother (years)			-0.00 (0.00)	0.00 (0.01)
[11]	Father has at least junior high school degree (1 = yes; 0 = no)			0.01 (0.04)	-0.02 (0.03)
[12]	Mother has at least junior high school degree (1 = yes; 0 = no)			0.00 (0.05)	0.02 (0.03)
[13]	At least one parent lives at home (1 = yes; 0 = no)			-0.05 (0.04)	0.04 (0.03)
[14]	Family wealth (1 = higher than the median; 0 = otherwise)			0.01 (0.04)	0.02 (0.03)
[15]	Observations	2030	2727	2030	2727
[16]	R-squared	0.010	0.007	0.299	0.381

Notes: \* Significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show the impact of the in-school CAL treatment on student math test scores separately for the third and fifth grade students. The test regresses standardised post-CAL math test scores on the treatment variable with or without a set of control variables.

<sup>a,b</sup>The baseline math/Chinese score is the normalised score on the math/Chinese test that is given to all sample students before the CAL Programme.

rise in standardised test scores was associated with a jump in intraschool district school rankings of up to five places (out of around 30 places). These studies, together with the interest policymakers in our study area have shown about CAL (they are now sponsoring an upscaling project), we believe that our effect size of 0.17 standard deviations is sufficiently large to attract the interest of policymakers.

The results testing the heterogeneous effects show that the effect of the CAL treatment varies by boarding status only for the third grade students (Table 4).<sup>10</sup> When we interact

Table 4. Ordinary least squares analysis of the heterogeneous impact of CAL Programme on student standardised math test scores of boarding and non-boarding students.

Dependent variable: standardised post-CAL math test score (standard deviation)			
		Third grade (1)	Fifth grade (2)
[1]	Treatment (1 = treatment group; 0 = control group)	0.30*** (0.09)	0.17** (0.08)
[2]	Interaction: Treatment * Boarding student	-0.21** (0.10)	-0.01 (0.08)
[3]	Boarding student (1 = yes; 0 = no)	0.17** (0.08)	0.01 (0.04)
[4]	Control variables <sup>a</sup>	Yes	Yes
[5]	Observations	2030	2727
[6]	R <sup>2</sup>	0.318	0.381

Notes: \* Significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show the heterogeneous effects of the in-school CAL treatment by student boarding status. The test regresses standardised post-CAL math test scores on the treatment variable, the interaction term between boarding and treatment status, boarding status and a set of control variables.

<sup>a</sup>Control variables include all the variables in Appendix 1.

the treatment variable with students' boarding status, the coefficient on the interaction term is negative but insignificant on the full sample (column 1, row 2). These results suggest that, on average, boarding students do not seem to benefit differently from the programme than the non-boarding students. For third grade students, the coefficient on the interaction term is negative and significant (column 2, row 2). The results suggest that the programme effect on third grade non-boarding students is 0.3 standard deviations, which is 0.21 standard deviations higher than the impact on third grade boarding students (column 2, rows 1–2). For fifth grade students, the results show that the programme does not differ by boarding status (column 3, row 2). The coefficient of the interaction term is close to zero and insignificant. Both the boarding and non-boarding students improved by 0.17 standard deviations after the CAL programme relative to the control group (column 3, rows 1–2).<sup>11</sup>

Using our data on the computer class activities, we also conducted a test on whether the treatment effect differs for schools where students learn basic computer skills in computer classes and the schools where students do not learn these skills. We included an interaction term between the treatment variable and a variable indicating whether the students learn basic computer operations in the regression that estimates the treatment effect (using Equation (1)). We can only run such a test among the fifth grade students because there are too few third grade students who have these activities. Our results show that we cannot reject the hypothesis that there is no significant difference between having CAL replace the learning activities of basic computer operations and other activities in computer classes. The coefficient on the interaction term is not significant. The result table is available upon request.

### **In-school programme versus out-of-school programme**

Before we conducted the in-school programme, we conducted an out-of-school programme as an efficacy trial to test whether CAL could be made to work in rural China. As is detailed in Lai et al. (2013), the programme was conducted to investigate the impact that CAL had on

the poor. After the success of the efficacy trial, we designed another CAL programme to be incorporated into the regular school day (the current in-school CAL study). In other words, these two programmes were designed to have two different study goals.

The in-school and out-of-school CAL programmes constitute an interesting comparison. First, as stated above, the two programmes are different in the way they were integrated into the school day (either in-school or out-of-school). At the same time, the actual content of the two CAL programmes remained the same. Second, the two studies were based on samples from the same student population: they were conducted in the same schools among students who were in the same grades in different school years.<sup>12</sup> The third and fifth grade students in 2010 participated in the first phase and the third and fifth grade students in 2011 participated in the second phase. Therefore, we have decided to not only report the evaluation result of phase two but also link it with phase one for comparison.

To test the effectiveness of out-of-school CAL (the efficacy trial), we conducted a cluster-RCT in Shaanxi Province during the 2010–2011 school year. A total of 5967 students in 72 rural Shaanxi schools were involved in the study. Among the students, 2726 students were boarding students and the other 3074 students were non-boarders. The boarding students constituted the sample for the study; the non-boarders only were used as additional controls. In other words, only the boarding students in the 36 treatment schools received the CAL programme.

According to the analysis in Lai et al. (2013), which is replicated in Appendix 4 in this paper, the out-of-school CAL programme had a positive and significant effect on the math test scores of students in the treatment schools. Appendix 4 includes the results of the regressions when using Equation (1) with control variables. Overall, as seen in Appendix 4, scores went up by 0.13 standard deviations (column 1). The impact on third grade students was 0.18 standard deviations (column 2) and the fifth grade students was 0.10 standard deviations (column 3). Although the result for fifth grade students is not significant, Lai et al. (2013) found significant impact on the poorer students.

By contrasting the in-school and out-of-school programmes, we find that both programmes improved the performance of most of the students who participated in the CAL treatment programmes by non-trivial magnitudes. Specifically, by integrating the CAL programme into the course of the regular school day, the programme improved performance of third and fifth grade students by 0.17 standard deviations. These estimates are comparable to the out-of-school programme effect of 0.18 standard deviations on the third grade students and 0.10 standard deviations on the fifth grade students (for students up to the 70th percentile – using their pretest scores).

Although the in-school programme is proved to have successfully improved students' performance, there may have been substitution of the in-school programme on the third grade boarding students. Table 4 shows that third grade boarding students benefit less from the in-school programme than the non-boarding students by 0.21 standard deviations. One possible explanation is that the substitution effect for the boarding students is larger than the non-boarding students in the third grade. The third grade students were likely to receive teacher-supervised/assisted exercise sessions in computer classes that were replaced by CAL (only two schools had computer competency activities in third grade computer classes). As boarding students were less likely to get any assistance at home, it could be that these computer classes are the only chance they have to get assistance in math learning besides the regular math teaching. As a result, the substitution effect of taking away these exercise sessions is larger for the boarding students than for the non-boarding students. This may explain why the out-of-school programme works better than the in-school programme in helping the third grade boarding students.

In order to test for the different substitution effects by boarding status, we run a regression by including the interaction term of three variables, including the treatment status, boarding status and whether the students were having computer competency activities during computer classes. The results show that boarding students benefit more if the CAL classes replaced computer competency activities instead of the supervised exercise sessions. The point estimate in the difference in treatment effect on boarding students is 0.14 standard deviations, although the estimate is not significant as a result of low power (too few third grade students had computer competency activities).<sup>13,14</sup>

## Conclusion

In this paper we present the results from a randomised field experiment of a CAL programme in 72 rural public schools in Ankang, Shaanxi. The study involves around 5267 third grade and fifth grade students. To evaluate the effectiveness of the programme we randomly chose 36 schools from the entire sample as treatment schools and third and fifth grade students in these schools received the CAL intervention. The remaining 36 schools served as control schools. The main intervention was designed to be a math CAL programme held during regular school hours (during a regularly scheduled computer class). The students were offered 40 minutes of shared computer time after school, twice a week. During the sessions students played computer-based games that required them to practice using their knowledge of math and relatively simple problem-solving skills. The CAL programme was tailored to the regular school math curriculum and was remedial in nature, providing the students with drills and exercises with different levels of difficulty.

Our results indicate that the in-school CAL programme significantly improved student academic outcomes. Two 40-minute CAL math sessions per week increased the student standardised math scores by 0.17 standard deviations for third grade students and by 0.17 standard deviations for fifth grade students. Although out-of-school programmes have typically been considered superior to in-school programmes, the gains from this in-school programme do not vary much from the overall impact of the out-of-school pilot programme reported by Lai et al. (2013).

These results suggest that given the possibility of substitution, the in-school programme still improves student learning. The reason that our results differ from the Linden study (2008) is that by integrating the CAL programme during a relatively unproductive period of time with low teacher absenteeism, the substitution effect may have been minimised. In order to investigate the substitution effect, we examined the differential impacts of CAL in schools that were teaching computer competencies (and other non-academic material) and those that (sometimes) used the classes as review sessions for math, Chinese and English.<sup>15</sup> Our results show that we cannot reject the hypothesis that there is no significant difference between having CAL replace learning activities of basic computer operations and other activities in computer classes. The coefficient on the interaction term is not significant. This seems to indicate that the conditions in China make in-school programmes a viable (and effective) means for introducing CAL to rural students.

There are limitations to our paper. For example, we are interested mostly in the academic outcomes of children (for example, their math scores) rather than competencies in other areas (for example, a student's ability to use a mouse or operate Microsoft Office software). We also do realise that our sample covers only one prefecture in southern Shaanxi. Although our study is restricted to one prefecture in Northwest China, we



believe that many aspects of our study environment apply to not only our study schools but also schools in other parts of rural China. Rural schools in China are homogeneous in many aspects. For example, almost all schools are public. The Ministry of Education requires that all primary schools cover first through sixth grade. The primary school curriculum used by almost all rural provinces is called 'renjiaoban'. Renjiaoban is produced and distributed by the centrally administered People's Education Publishing House, which is part of the Ministry of Education. Teaching credentials are uniform across provinces and salaries of nearly all teachers are paid by the national government (especially in the case of poor rural areas). Although there is potential external validity, we know there are limits to how representative our study areas are.

Given China's current effort to have computer rooms in all schools in poor rural areas of western China, these findings are timely, policy-relevant and immediately actionable. As more computers are installed in rural schools, policymakers and school officials will need to explore various options and ultimately decide which type of CAL programme to implement. As the results of our study indicate that both in-school and out-of-school CAL programmes can produce positive results in rural China, policymakers and school officials can thus select from a broader range of choices when making their CAL programme decisions.<sup>16</sup>

In the end, the final decision on whether to implement out-of-school or in-school programmes may be different for different schools and different counties. While out-of-school programmes may be more effective in some cases (the out-of-school programme benefited the third grade boarding students more than the in-school programme), in rural China running an after-school programme means that some students (such as non-boarding students with long daily commutes) will be less likely to participate. As such, schools need to balance the potentially greater effectiveness of out-of-school programmes with the greater inclusiveness and broader reach of in-school programmes. Perhaps if school officials are given the flexibility and the proper training, they can decide what degree of integration will give their particular institution the best mix of effectiveness and equity.

In addition to these findings, this study makes another point very clear: as more computer rooms are established in rural schools across western China, it is essential that effective educational-based software is also available and that teachers are trained well in using it.

## Notes

1. The study underwent and successfully passed ethical review by Stanford University's Internal Review Board for non-medical human subjects research with an IRB number of 19341.
2. In terms of educational achievement, Shaanxi Province is at about the national average. However, there is huge inequality within the province (NBSC 2011). For example, in relatively rich areas such as Guanzhong (in the central part of the province), 14.4 per cent of its population received a college education (higher than the national average of 8.9%). In contrast, in Ankang Prefecture only 4.8 per cent of the population holds a college degree.
3. We only included *wanxiao* (or 'complete schools') with six full grades in our sample because the programme requires that third grade and fifth grade students stay in the same school during the programme period (one year and a half). In rural China, there are other elementary schools with only two, three and four grades. These are often small schools (several students per grade) in remote rural villages. In Chinese these are called *jiaoxuedian*, or 'teaching point schools'. The schools that were not complete schools could not be included in the programme because students often transfer to other schools when they reach the third or fourth grades. Teaching point schools also are being shut down and merged into larger complete schools. In either case, it would be impossible for students to continue to attend the CAL sessions. It

would also be difficult for us to follow the students. Therefore, non-wanxiao schools were excluded from the sampling frame.

4. In selecting the teacher-supervisors, we were guided by two principles. First: we wanted to choose the teacher-supervisor rather than the school principal. We also did not want to select a teacher-supervisor who was also a math teacher. With these principles in mind, we excluded from our selection the math teachers or homeroom teachers of the third and fifth grade students. We then created a list of teachers that were available. We then randomly chose the teacher.
5. In terms of teacher training, all teacher-supervisors of the 36 treatment schools participated in a two-day mandatory training programme. During the training, the teachers were introduced to our programme protocol and the two pieces of software. The teachers also underwent hands-on session where they practiced with the software and asked questions. At the end of the training session, randomly selected teachers gave mock classes to all the other teachers who pretended to be students.
6. Both the third and fifth grade CAL software packages consisted of two individual pieces of software. The first piece of software was a commercial, game-based math-learning software programme that we obtained through donation. The software provided remedial tutoring material (both animated reviews and remedial questions) in math for the third and fifth grade students in keeping with the national uniform math curriculum. We developed the second piece of software ourselves. The package (henceforth, the CAL software) was designed to provide the students with a large number of remedial questions.
7. The students were not allowed to discuss math questions with the supervising teacher because the goal of the study is to test whether a CAL programme can improve learning of the underserved students in rural schools. We are interested in knowing whether the programme can benefit students in the poorest schools with little teaching resources. Therefore, we would like to isolate the programme impact from teacher instruction. In other words, teachers were not allowed to intervene in the classes to affect the programme impact. In fact, this is policy relevant given a scenario in which the CAL sessions were run in-school during computer class sessions. The computer teacher would not be an expert in the field and would likely be busy managing the technology and curriculum and not focused on teaching students the material that other teachers were supposed to be teaching. Likewise, the students were not allowed to discuss with other teams (students using a different computer) to limit the influence of student interaction on programme impact. It also makes it easier for teachers to manage the classes without having to organise the group discussion or other activities. According to our observation, students typically had no time to discuss with each other because the sessions were so intense that the students were almost always exclusively focused on their computers.
8. The test questions for the standardised math exam were chosen from the TIMSS test data bank. Drafts of the tests were screened by a set of rural elementary teachers in Shaanxi province. We then rigorously tested the questions in a pilot survey. We then made adjustment to the test by eliminating the questions that were too difficult (almost no one got them right) and the questions that were too easy (almost everyone got them right).
9. The standardised test scores are normalised using the distribution of test scores of the control group students within the same grade and on the same subject.
10. This is important to show because the original analysis for the out-of-school treatment effects for CAL was conducted for boarding school students only. If we show that the effects are the same for boarding and non-boarding students (as we do), then the analysis can really focus on differences between in-school effects (as reported from this study) and out-of-school effects (as shown in Lai et al. 2011).
11. We have also conducted an analysis of heterogeneous effects by student baseline math score, student gender, family wealth and their starting grade. The questions that the tests are intended to address are whether poorer performing students benefit more from the programme than better performing students, whether boys or girls benefit more from the programme, whether poorer or richer students improve more after the programme and whether starting grade (third or fifth grade) makes a difference in how much student learning can be improved. The results show that none of the tests detect any significant difference in programme impact among the subgroups. By following the Bonferroni approach to adjust multiple hypotheses, we divide the significance level of all the correlated outcomes of heterogeneous effects, 0.1, by the number of hypotheses we tested (that is, 4 different types of heterogeneous effects of the CAL

programme). By doing this, we obtain the adjusted p-values for each individual null hypothesis of heterogeneous effects:  $0.1/4 = 0.025$ . Since none of the heterogeneous effects are significant at the 0.1 level, they do not meet the 0.025 adjusted significance requirement either. In other words, with or without adjusting for multiple hypotheses testing, we cannot reject the null hypotheses that there are no heterogeneous effects between the poorer and better performing students, between girls and boys, between the richer and poorer students, and between the third grade and fifth grade students.

12. Although no re-randomisation was done to reassign treatment and control schools in the in-school programme, we did conduct a balance test before the start of the programme to ensure the students in the two groups were balanced. As shown in Appendix 2, the key variables of the treatment and control groups are balanced at the baseline.
13. The result table is available upon request.
14. We also tested whether the programme had any crowding-out effect on Chinese learning. Based on the regression results using Equation (1), the out-of-school programme does not seem to have crowded out student learning in Chinese (Appendix 5). The coefficient of the treatment variable is not significant for either the whole sample (third and fifth grades) or each grade separately. The magnitudes of the coefficients are small and positive.
15. Using our data on the computer class activities, we conducted a test on whether the treatment effect differs for schools where students learn basic computer skills in computer classes and the schools where students do not learn these skills. We included an interaction term between the treatment variable and a variable indicating whether the students learn basic computer operations in the regression that estimates the treatment effect (using Equation (2)). We can only run such a test among fifth grade students because there are too few third grade students who have these activities. The result table is available upon request.
16. Cost-effectiveness analysis suggests that the programme has low cost per unit of improvement in student learning. From the perspective of China's policymakers considering to upscale the programme, computer hardware itself is already a sunk cost given that the government is installing computer labs in every rural elementary school as part of its Twelfth Five Year Plan. The marginal costs that are needed to execute the programme include teacher training, administration costs and allowance for CAL teacher-supervisors. Using the method suggested by Dhaliwal et al. (2012), we calculate the total cost of the programme in our project area to be 9439 USD (in 2011, the project year) and 10,035 USD (in 2014, after taking inflation into account). We then divide the total cost by total impact (total impact = average programme effect multiplied by the total number of students attending CAL sessions):  $10,035 \text{ USD} / (0.17 \text{ SD} * 2435 \text{ students}) = 24.2 \text{ USD/SD}$ . The cost-effectiveness of our programme is comparable to the CAL programme conducted in India. According to the estimates provided by Banerjee et al. (2007), the CAL programme in India costs 21.4 USD/SD (in 2002) and 28.2 USD/SD (in 2014) – also excluding the costs of computers.

## References

- Angrist, J., and V. Lavy. 2002. "New Evidence on Classroom Computers and Pupil Learning." *The Economic Journal* 112 (482): 735–765. doi:10.1111/1468-0297.00068.
- Banerjee, A., S. Cole, E. Duflo, and L. Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122 (3): 1235–1264. doi:10.1162/qjec.122.3.1235.
- Barrow, L., L. Markman, and C. Rouse. 2009. "Technology's Edge: The Educational Benefits of Computer-Aided Instruction." *American Economic Journal: Economic Policy* 1 (1): 52–74.
- Bruhn, M., and D. McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics* 1 (4): 200–232.
- Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers. 2005. "Teacher Absence in India: A Snapshot." *Journal of the European Economic Association* 3 (2–3): 658–667. doi:10.1162/jeea.2005.3.2–3.658.
- Cole, M. 1996. *Cultural Psychology: A Once and Future Discipline*. Cambridge, MA: Harvard University Press.
- Cristia, J., P. Ibararán, S. Cueto, A. Santiago, and E. Severín. 2012. "Technology and Child Development: Evidence from the One Laptop Per Child Program." IDB Working Paper No. 304. Washington, DC: Inter-American Development Bank.

- Dhaliwal, I., E. Duflo, R. Glennerster, and C. Tulloch. 2012. "Comparative Cost-effectiveness to Inform Policy in Developing Countries." Working paper. Cambridge: Abdul Latif Jameel Poverty Action Lab, MIT.
- Dias, P., A. Freedman, P. Medway, and A. Paré. 1999. *Worlds Apart: Acting and Writing in Academic and Workplace Contexts*. Mahwah, NJ: Lawrence Erlbaum.
- Duflo, E., R. Hanna, and S. P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–1278.
- Dynarski, M. 2007. "Effectiveness of Reading and Mathematics Software Products: Findings from the 1st Student Cohort." Mathematica Report. [http://www.mathematica-mpr.com/publications/redirect\\_pubsdb.asp?strSite=PDFs/effectread.pdf](http://www.mathematica-mpr.com/publications/redirect_pubsdb.asp?strSite=PDFs/effectread.pdf)
- Ezroj, A., D. Mullin, G. Pyle, and S. Sylvia. 2004. "The Development of Northwest China: Impacts of Developmental Programs and Policy Reforms on Rural China." Working Paper. Pittsburgh, PA: University of Pittsburgh.
- Fuchs, T., and L. Woosmann. 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.
- Goolsbee, A., and J. Guryan. 2006. "The Impact of Internet Subsidies in Public Schools." *The Review of Economics and Statistics* 88 (2): 336–347. doi:10.1162/rest.88.2.336.
- Guimarães, L. B. D. M., J. L. D. Ribeiro, M. E. Echeveste, and J. J. De Jacques. 2013. "A Study of the Use of the Laptop XO in Brazilian Pilot Schools." *Computers & Education* 69: 263–273. doi:10.1016/j.compedu.2013.07.020.
- He, F., L. Linden, and M. MacLeod. 2008. "How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program." Working Paper. New York: Columbia University.
- Hull, G., and K. Schultz, eds. 2002. *School's Out! Bridging Out-of-school Literacies with Classroom Practice*. New York: Teachers College Press.
- Kirkpatrick, H., and L. Cuban. 1998. "Computers Make Kids Smarter – Right?" *Technos Quarterly for Education and Technology* 7 (2): 26–31.
- Krishnaratne, S., H. White, and E. Carpenter. 2013. "Quality Education for All Children? What Works in Education in Developing Countries." Working Paper No. 20. London: International Initiative for Impact Evaluation.
- Lai, F., C. Liu, R. Luo, L. Zhang, X. Ma, Y. Bai, B. Sharbono, and S. Rozelle. forthcoming. "The Education of China's Migrant Children: The Missing Link in China's Education System." *International Journal of Educational Development*.
- Lai, F., R. Luo, L. Zhang, X. Huang, and S. Rozelle. 2011. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing." REAP Working Paper. Stanford, CA: Rural Education Action Program.
- Lai, F., L. Zhang, X. Hu, Q. Qu, Y. Shi, Y. Qiao, M. Boswell, and S. Rozelle. 2013. "Computer Assisted Learning as Extracurricular Tutor? Evidence from a Randomised Experiment in Rural Boarding Schools in Shaanxi." *Journal of Development Effectiveness* 5 (2): 208–231. doi:10.1080/19439342.2013.780089.
- Lai, F., L. Zhang, Q. Qu, X. Hu, Y. Shi, M. Boswell, and S. Rozelle. 2012. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai, China." REAP Working Paper. Stanford, CA: Rural Education Action Program.
- Linden, L. 2008. "Computer or Substitute? The Effect of Technology on Student Achievement in India." Working Paper. New York: Columbia University.
- Liu, Y. and K. Kumar. 2008. "China and India: Different Educational Paths Toward Prosperity." *Policy insight* 2 (1): 1–4.
- McEwan, P. 2013. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." Working paper. Wellesley: Wellesley College.
- Miettinen, R. 1999. "Transcending Traditional School Learning: Teachers' Work and Networks of Learning." In *Perspectives on Activity Theory*, edited by Y. Engeström, R. Miettinen, and R.-L. Punamäki, 325–344. Cambridge: Cambridge University Press.
- Ministry of Education. 2012. "12th Five-Year Plan for Integrating Information Technology into Education." [http://www.edu.cn/zong\\_he\\_870/20120723/t20120723\\_813704.shtml](http://www.edu.cn/zong_he_870/20120723/t20120723_813704.shtml)
- Mo, D., L. Zhang, J. Wang, W. Huang, Y. Shi, M. Boswell, and S. Rozelle. 2013. "The Persistence of Gains in Learning from Computer Assisted Learning: Evidence from a Randomized

- Experiment in Rural Schools in Shaanxi Province.” REAP Working Paper. New York: Columbia University.
- Nara, H., and M. Noda. 2003. *Acts of Reading: Exploring Connections in Pedagogy of Japanese*. Honolulu: University of Hawaii Press.
- NBSC (National Bureau of Statistics of China). 2011. *China Statistical Yearbook*. Beijing: China Statistics Press.
- NBSC (National Bureau of Statistics of China). 2013. *China Statistical Yearbook*. Beijing: China Statistics Press.
- Planty, M., W. Hussar, T. Snyder, S. Provasnik, G. Kena, R. Dinkes, A. Kewal Ramani, and J. Kemp. 2008. *The Condition of Education 2008 (NCES 2008–031)*. National Center for Education Statistics, Institute of Education Sciences, US Department of Education. Washington, DC.
- President’s Committee of Advisors on Science and Technology. 1997. “Panel on Educational Technology.” 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. E., and A. B. Krueger. 2004. “Putting Computerized Instruction to the Test: A Randomized Evaluation of a ‘Scientifically-based’ Reading Program.” *Economics of Education Review* 23: 323–338. doi:10.1016/j.econedurev.2003.10.005.
- Schagen, I., and E. Hodgen. 2009. “How Much Difference Does It Make? Notes on Understanding, Using, and Calculating Effect Sizes for Schools.” <http://www.nzcer.org.nz/system/files/16886.pdf>
- Underwood, C., M. Welsh, M. Gauvain, and S. Duffy. 2000. “Learning at the Edges: Challenges to the Sustainability of Service-Learning in Higher Education.” *Language and Learning across the Disciplines* 4 (3): 7–26.
- Yang, Y., X. Hu, Q. Qu, F. Lai, Y. Shi, M. Boswell, and S. Rozelle. 2013. “The Roots of Tomorrow’s Digital Divide: Documenting Computer Use and Internet Access in China’s Elementary Schools Today.” *China & World Economy* 21 (3): 61–79.
- Yuan, X. 2012. “Strategy for Improving Public Service: Push Forward the Integration of ICT in Rural Areas in China.” (in Chinese) [http://www.stdaily.com/stdaily/content/2012-02/20/content\\_429399.htm](http://www.stdaily.com/stdaily/content/2012-02/20/content_429399.htm)

**Appendix 1. Descriptive statistics of baseline characteristics of the treatment group and the control group of the third grade and fifth grade students after attrition**

		Students after attrition			
		Treatment group		Control group	
		Mean	Standard deviation	Mean	Standard deviation
[1]	Baseline math score (units of standard deviation) <sup>a</sup>	0.00	1.02	0.00	1.00
[2]	Baseline Chinese score (units of standard deviation) <sup>b</sup>	0.00	1.00	0.00	1.00
[3]	Gender (1 = boy; 0 = girl)	0.54	0.50	0.54	0.50
[4]	Age (years)	9.74	1.27	9.76	1.21
[5]	Boarding student (1 = yes; 0 = no)	0.37	0.48	0.35	0.48
[6]	Ever repeated grade (1 = yes; 0 = no)	0.29	0.46	0.28	0.45
[7]	Only child (1 = yes; 0 = no)	0.29	0.45	0.28	0.45
[8]	Father has at least junior high school degree (1 = yes; 0 = no)	0.55	0.50	0.56	0.50
[9]	Mother has at least junior high school degree (1 = yes; 0 = no)	0.40	0.49	0.41	0.49
[10]	At least one parent lives at home (1 = yes; 0 = no)	0.42	0.49	0.43	0.49
[11]	Family wealth (1 = higher than the median; 0 = otherwise)	0.50	0.50	0.50	0.50
[12]	Observations	2220	2537		

<sup>a,b</sup>The baseline math/Chinese score is the normalised score on the math/Chinese test that is given to all sample students before the CAL Programme.

**Appendix 2. Comparison of characteristics between the treatment group and the control group of the third grade and fifth grade students before attrition**

		Students before attrition					
		Third grade and fifth grade		Third grade only		Fifth grade only	
		Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
[1]	Baseline math score (units of standard deviation) <sup>a</sup>	0.01	0.03	0.02	0.03	0.01	0.04
[2]	Baseline Chinese score (units of standard deviation) <sup>b</sup>	0.02	0.03	0.01	0.03	0.02	0.03
[3]	Gender (1 = boy; 0 = girl)	0.01	0.01	0.00	0.02	0.01	0.02
[4]	Age (years)	-0.01	0.01	-0.03	0.02	0.01	0.03
[5]	Boarding student (1 = yes; 0 = no)	0.02	0.07	0.01	0.09	0.02	0.07
[6]	Ever repeated grade (1 = yes; 0 = no)	0.01	0.03	-0.02	0.05	0.03	0.04
[7]	Only child (1 = yes; 0 = no)	0.01	0.04	0.03	0.05	0.00	0.04
[8]	Father has at least junior high school degree (1 = yes; 0 = no)	0.00	0.03	0.01	0.03	-0.01	0.04
[9]	Mother has at least junior high school degree (1 = yes; 0 = no)	-0.01	0.03	-0.01	0.03	-0.02	0.04
[10]	At least one parent lives at home (1 = yes; 0 = no)	-0.01	0.03	0.00	0.03	-0.03	0.03
[11]	Family wealth (1 = higher than the median; 0 = otherwise)	0.00	0.04	0.00	0.04	0.00	0.04
[12]	Observations	5267		2279		2988	

Robust standard errors in parentheses clustered at school level.

The test aims to show whether the samples are well balanced in the treatment and control groups before attrition for the third and fifth grade students (combined and separately). These tests regress the student and family characteristics on the treatment status one at a time.

<sup>a,b</sup>The baseline math/Chinese score is the normalised score on the math/Chinese test that is given to all sample students before the CAL programme.

**Appendix 3. Comparison of characteristics between the treatment group and the control group of the third grade and fifth grade students after attrition**

		Students after attrition					
		Third grade and fifth grade		Third grade only		Fifth grade only	
		Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
[1]	Baseline math score (units of standard deviation) <sup>a</sup>	0.01	0.03	0.02	0.03	0.01	0.04
[2]	Baseline Chinese score (units of standard deviation) <sup>b</sup>	0.02	0.03	0.01	0.03	0.02	0.03
[3]	Gender (1 = boy; 0 = girl)	0.00	0.01	-0.01	0.02	0.01	0.02
[4]	Age (years)	0.00	0.01	-0.03	0.02	0.01	0.03
[5]	Boarding student (1 = yes; 0 = no)	0.02	0.08	0.03	0.09	0.01	0.07
[6]	Ever repeated grade (1 = yes; 0 = no)	0.01	0.04	-0.02	0.06	0.04	0.04
[7]	Only child (1 = yes; 0 = no)	0.01	0.05	0.03	0.05	0.00	0.05
[8]	Father has at least junior high school degree (1 = yes; 0 = no)	-0.01	0.03	0.01	0.03	-0.02	0.04
[9]	Mother has at least junior high school degree (1 = yes; 0 = no)	-0.01	0.03	0.00	0.03	-0.02	0.04
[10]	At least one parent lives at home (1 = yes; 0 = no)	-0.01	0.03	0.01	0.04	-0.02	0.03
[11]	Family wealth (1 = higher than the median; 0 = otherwise)	-0.01	0.04	-0.01	0.04	-0.01	0.04
[12]	Observations	4757		2030		2727	

Robust standard errors in parentheses clustered at school level.

The test aims to show whether the samples are well balanced in the treatment and control groups after attrition for the third and fifth grade students (combined and separately). These tests regress the student and family characteristics on the treatment status one at a time.

<sup>a,b</sup>The baseline math/Chinese score is the normalised score on the math/Chinese test that is given to all sample students before the CAL programme.



#### Appendix 4. Ordinary Least Squares analysis of the out-of-school CAL Programme on student standardised math test scores of boarding students of the third and fifth grade students who participated in the CAL Programme during March and June, 2011

Dependent variable: standardised post-CAL math test score (standard deviation)			
	All (Third grade & Fifth grade) (1)	Third grade (2)	Fifth grade (3)
[1] Treatment (1 = treatment group; 0 = control group)	0.13** (0.06)	0.18** (0.09)	0.10 (0.07)
[3] Control variables	Yes	Yes	Yes
[4] Observations	2426	1038	1388
[5] R <sup>2</sup>	0.427	0.416	0.453

Notes: \*\* significant at 5%. Robust standard errors in parentheses clustered at school level.

This is a replication of the analysis in Lai et al. (2013). The test aims to show the impact of the out-of-school CAL treatment on student math test scores. The test regresses standardised post-CAL math test scores on the treatment variable and a set of control variables that are listed in Appendix 1.

#### Appendix 5. Ordinary Least Squares analysis of the in-school CAL Programme on student standardised Chinese test scores

Dependent variable: standardised post-CAL Chinese test score (standard deviation)			
	All (Third grade & Fifth grade) (1)	Third grade (2)	Fifth grade (3)
[1] Treatment (1 = treatment group; 0 = control group)	0.05 (0.06)	0.03 (0.06)	0.07 (0.10)
[3] Control variables <sup>a</sup>	Yes	Yes	Yes
[4] Observations	2426	1038	1388
[5] R <sup>2</sup>	0.345	0.430	0.319

Notes: Robust standard errors in parentheses clustered at school level.

The test aims to show the spillover effect of the in-school CAL treatment on standardised post-CAL Chinese test scores. The test regresses standardised post-CAL Chinese test scores on the treatment variable and a set of control variables.

<sup>a</sup>Control variables include all the variables in Appendix 3.

#### Appendix 6. CAL hardware and software

The intervention team spent considerable time to prepare the necessary hardware, software, CAL curriculum and programme implementation protocol in a way that would both facilitate smooth implementation of the CAL programme and avoid confounding influences that might bias our results. As the first step, to meet the hardware requirements of the CAL programme, we acquired (by way of donation from Dell, Inc.) 640 brand-new identical desktop computers and installed our CAL software package on these desktops. We then removed all preinstalled software that would not be used during the CAL

intervention (such as Windows built-in games and Microsoft Office) and disabled the Internet and USB functions on all of the computers. In this way, not only could we prevent students or teachers from using the programme computers for other purposes that might affect the operation of the regular CAL programme but we could also avoid the interruptions that might otherwise be caused by accidental deletion of the CAL software or the introduction of viruses. ‘Sealing the computers’ also ensured the quality of our evaluation of the programme effects without capturing any other confounding influences (spillovers) if students used the computers to gain knowledge by having access to other sources of information such as the Internet. It also prevented teachers and students from the control schools from copying our CAL software onto their computers.

The educational programme contained in the computer-assisted learning software has two parts. The first piece of software was a commercial, game-based math-learning software programme that was obtained through donation. We adopted this package because it was uniquely suited to the CAL programme. The software provided remedial tutoring material (both animated reviews and remedial questions) in math for the third and fifth grade students following the national uniform math curriculum. The designers of the programme also developed their software so it could be used in conjunction with the material that students were learning in their math class on a week-by-week basis.

We also developed the second piece of software by ourselves. Our software package (henceforth, the CAL software) was developed to provide the students with a large number of practice questions. Students answered the questions in game-based exercises. In choosing the math questions to include in the CAL software, we consulted experienced elementary school math teachers in both public schools in cities and rural areas, as well as expert committee members of the Center for Examination of Beijing, an institute that designs citywide uniform tests for elementary schools in Beijing. With their direction and assistance, we chose questions for the CAL software from several commercially available books of practice questions. By combining the commercial software and the CAL software, we had enough content and exercise games to provide 80 minutes of remedial tutoring (two weekly sessions of 40 minutes each) that cover the math curriculum for the spring 2011 semester.

We also developed the second piece of software by ourselves. Our software package (henceforth, the CAL software) was developed to provide the students with a large number of practice questions. Students answered the questions in game-based exercises. In choosing the math questions to include in the CAL software, we consulted experienced elementary school math teachers in both public schools in cities and rural areas, as well as expert committee members of the Center for Examination of Beijing, an institute that designs city-wide uniform tests for elementary schools in Beijing. With their direction and assistance, we chose questions for the CAL software from several commercially available books of practice questions. By combining the commercial software and the CAL software, we had enough content and exercise games to provide 80 minutes of remedial tutoring (two weekly sessions of 40 minutes each) that cover the math curriculum for the spring 2011 semester.