Journal of Development Effectiveness

ISSN: 1943-9342 (Print) 1943-9407 (Online) Journal homepage: http://www.tandfonline.com/loi/rjde20

More is not always better: evidence from a randomised experiment of computer-assisted learning in rural minority schools in Qinghai

Fang Lai, Linxiu Zhang, Yu Bai, Chengfang Liu, Yaojiang Shi, Fang Chang & Scott Rozelle

To cite this article: Fang Lai, Linxiu Zhang, Yu Bai, Chengfang Liu, Yaojiang Shi, Fang Chang & Scott Rozelle (2016) More is not always better: evidence from a randomised experiment of computer-assisted learning in rural minority schools in Qinghai, Journal of Development Effectiveness, 8:4, 449-472, DOI: 10.1080/19439342.2016.1220412

To link to this article: http://dx.doi.org/10.1080/19439342.2016.1220412

JOURNAL OF DEVELOPMENT EFFECTIVENESS, 2016
VOL. 8, NO. 4, 449–472
http://dx.doi.org/10.1080/19439342.2016.1220412

Routledge
Taylor & Francis Group

More is not always better: evidence from a randomised experiment of computer-assisted learning in rural minority schools in Qinghai

Fang Lai[a,b], Linxiu Zhang[b], Yu Bai[b], Chengfang Liu[b], Yaojiang Shi[c], Fang Chang[c] and Scott Rozelle[d]

[a]LICOS, Katholic University Leuven, Leuven, Belgium; [b]Center for Chinese Agricultural Policy, Institute for Geographic Sciences and Natural Resource Research, Chinese Academy of Sciences, Beijing, P. R. China; [c]Center for Experimental Economics in Education, Shaanxi Normal University, Xi'an, P. R. China; [d]Rural Education Action Project, Freeman Spogli Institute, Stanford University, Stanford, CA, USA

ABSTRACT

The education of poor and disadvantaged populations has been a long-standing challenge for education systems in both developed and developing countries. Drawing on data from two randomised controlled trials involving two cohorts of grade 3 students in poor rural minority schools in China's Qinghai province, this paper explores the effects of computer-assisted learning (CAL) on student academic and non-academic outcomes for underserved student populations, and how interactions between the CAL programme and existing classroom resources affect the programme effectiveness. Results show that CAL could have significant beneficial effects on both student academic and non-academic outcomes. However, when the scope of the programme expanded to include a second subject (in this case, math – which was added on top of the Mandarin subject matter that was the focus of the first phase of the programme), some schools had to use regular school hours for CAL sessions. As a result, the phase II programme did not generate any (statistically) significant improvement over the first phase.

ARTICLE HISTORY
Received 3 December 2014
Accepted 1 August 2016

KEYWORDS
Education; development; computer-assisted learning; random assignment; test scores; China; rural minorities

Introduction

The education of poor and disadvantaged populations has been a long-standing challenge for education systems in both developed and developing countries. Unfortunately, many studies have found that spending on traditional educational inputs alone are insufficient and do not improve learning (Hanushek 1986, 1995; Glewwe and Kremer 2006; Planty et al. 2008; Jacob and Lefgren 2004). Consequently, researchers are actively exploring alternative methods of delivering educational inputs in order to better improve learning outcomes.

Computer-assisted learning (or CAL) is one such alternative. CAL entails the use of computers and computing technologies to enhance learning through computerised instruction, drills and

CONTACT Chengfang Liu cfliu.ccap@igsnrr.ac.cn Center for Chinese Agricultural Policy, Institute for Geographic Sciences and Natural Resource Research, Chinese Academy of Sciences, No. 11-A Datun Road, Chaoyang District, Beijing 100101, P. R. China
Present address of Fang Lai is Rural Education Action Project, Freeman Spogli Institute, Stanford University, Stanford, CA, USA and IMPAQ International, Oakland, CA, USA.
Present address of Chengfang Liu is China Center for Agricultural Policy, School of Advanced Agricultural Science, Peking University, Beijing, China.

exercises (Kirkpatrick and Cuban 1998; President's Committee of Advisors on Science and Technology, 1997). By integrating regular class materials with interactive interfaces and games, CAL is thought to hold promise for making the learning process more engaging for students (Inal and Cagiltay 2007; Schaefer and Warren 2004). In fact, CAL has been shown to meet many otherwise unmet needs for students in underserved schools (for example, Banerjee et al. 2007; Barrow, Markman, and Rouse 2008; Linden 2008; He, Linden, and MacLeod 2008; Lai et al. 2013a,b; McEwan 2014). For these vulnerable students, CAL may serve as a valuable learning tool when teachers are not available or have too little training and/or motivation to provide adequate instruction during or after school. CAL may also provide remedial tutoring services when commercial services are either not available or not affordable. Finally, CAL may provide help that parents/grandparents who are illiterate or too busy cannot provide. In this sense, in-school CAL programmes may be particularly effective in developing countries, where schools are often plagued with poor facilities and unqualified teachers and computer technologies are relatively new and not affordable for most families.

Despite its promise, the empirical evidence on the effectiveness of CAL in promoting learning is mixed, including studies based on randomised evaluations (for example, Angrist and Lavy 2002; Fuchs and Woosmann 2004; Goolsbee and Guryan 2006). For example, both Dynarski et al. (2007) and Rouse and Krueger (2004) found no significant gain in test scores in either math or reading from CAL programmes for US students. In comparison, Barrow, Markman, and Rouse (2008) found a computer-assisted learning programme improved student math test scores by 0.17 standard deviations in Chicago schools. CAL programmes seem to be particularly effective in developing countries. Both Banerjee et al. (2007) and He, Linden, and MacLeod (2008) found significant positive effects of CAL programmes on student test scores, and these effects increased over time in Banerjee et al. (2007). Lai et al. (2013a), 2013b, 2015) and Mo et al. (2013), 2014) also found that computer-assisted learning programmes can significantly improve the performance of underserved students in both rural-to-urban migrant schools in Beijing and rural schools in Shaanxi, China. Lai et al. (2015) found that a Mandarin language CAL programme improved language skills and had positive spillovers onto math scores among non-Mandarin speaking ethnic minorities.

Many factors might have contributed to this mixed record. First, the quality of implementation might be important to the effectiveness of the programme. For example, Mo (2015) found that the effect of a CAL programme decreased dramatically when implemented by local governments instead of a research team. However, not many studies in the literature have described the process of programme implementation and how well the programme fits into the (schooling) environment in which the programme was being implemented. Second, when CAL sessions are held during the school day, it is possible that (some of) the positive impacts of the CAL programme are offset by the negative effects of missing regularly scheduled class time (Angrist and Lavy 2002; Rouse and Krueger 2004). Linden (2008) used a randomised experiment to explore this substitution effect by examining two different CAL programmes implemented in India, finding that CAL implemented as an in-school programme (that is, replacing regular school inputs) improved student test scores less than when students were participating in after-school CAL programmes (where substitution was less of an issue).

The overall goal of this paper is to explore the effects of CAL on student academic and non-academic outcomes for underserved student populations in a developing country context. In pursuit of this goal, we have two specific objectives. First, we examine the overall impacts of a two-phase CAL Mandarin and math programme on student academic performance (as measured by standardised test scores). Second, we examine whether the interactions between the CAL programme and existing classroom resources (as well as interactions between the CAL programme and the educational context) matter for a programme's effectiveness. We achieve the second objective by comparing the effects of two CAL regimens: one that delivered Mandarin language instruction after school, and one that delivered instruction in two subjects, Mandarin and math, partly during and partly after school. As part of our analysis, we also report the effects of CAL programmes on non-academic student outcomes.

To meet the goal of this study, we present the results from two randomised field experiments using a CAL remedial tutoring programme as the treatment in 57 elementary schools. The schools are in poor minority rural areas of Qinghai Province, one of the largest and poorest provinces in China. Qinghai also is densely populated by ethnic minority groups. In fact, around 86 per cent of the students in our sample are ethnic minorities. Most students in the study schools are from impoverished rural families. Hence, these students are vulnerable not only because they are from poor rural families and attend rural schools of poor educational quality, but also because of the language barrier they face (Yang et al. 2013). Rural minority students often speak their own languages or dialects at home, and, at most, only use Mandarin at school. Their parents often do not read at all or speak Mandarin well. Most teachers in rural minority schools are minorities themselves and often have only limited competency in Mandarin. Given the fact that almost all text books are in Mandarin, and tests for educational progress (for example, high school and college entrance examinations) are taken in Mandarin, poor Mandarin language competency may be a major obstacle in the educational lives of rural minority students. As such improving the Mandarin language skills of non-Mandarin speakers may be among their most urgent academic needs.

Among the 57 sample schools (used in both phase I and phase II of the study) 26 schools were randomly selected to receive the CAL intervention and the remaining 31 served as controls. As described earlier, two successive cohorts of grade 3 students in the same schools received the CAL programmes. During the 2011–12 academic school year, the first cohort included 1822 grade 3 students (mostly aged nine to eleven). During the 2012–13 academic school year, the second cohort included 1592 grade 3 students (also mostly aged 9–11 years).

The two CAL regimens run in the sample treatment schools – one each for each of the cohorts – differed in two ways. The first difference was in the intensity of the CAL programme. For the first cohort (in phase I of our study), the CAL programme was implemented in the 26 treatment schools and only included materials on Mandarin. The students had two CAL Mandarin sessions per week during the programme period. For the second cohort (in phase II), the CAL programme included both Mandarin and math subjects. Because the math subject was added on (to the original Mandarin-only CAL programme), the number of CAL sessions per week doubled. The students then had four CAL sessions per week, two in Mandarin and two in math. As a consequence, the CAL programme for the second cohort was more intensive than the programme for the first cohort.

The second difference between phase I and phase II of the study involved the approach to programme implementation (or the nature of the classroom context of our CAL interventions). In the case of the first cohort, the schools were asked to use after-school hours for all CAL sessions. Given the lower intensity of the CAL programme for the first cohort, the CAL sessions were, in fact, all held outside of regular school hours. Although we requested that the schools use after-school hours for the second cohort's CAL sessions (as they had done in phase I), some schools found this impossible and ended up using significant numbers of in-school hours for some of the CAL sessions. Because of these differing approaches to implementation, in the case of the first cohort the programme should be thought of as a purely complementary input, whereas in the case of the second cohort it may be that to a certain extent, educators (CAL promotes) have substituted CAL for existing classroom lessons.

Our study contributes to the existing literature in one key way. By comparing the effects of two CAL programmes we demonstrate that the interactions between the CAL programme and classroom context (regular class schedule and time, availability of teachers, student level of performance and attitudes) appear to affect the effectiveness of the CAL programme. Our work demonstrates that CAL has the potential to provide a relatively quick and effective remedy for the severe underperformance of rural minority children in China (and possibly disadvantaged children in other parts of the world) in language and other subjects. In order to maximise the programme effects, however, it is important that educators and policy makers consider interactions between the programme and classroom context (in particular, the substitution away from existing teaching resources that might occur if the programme is implemented during regular classroom time periods).

The rest of this paper is organised as follows. The next section reviews the study's research design and sampling, explains the intervention in more detail, and describes the study's data and statistical approach. The following sections then present the results, discuss the findings and conclude.

Sampling, data and methods

Sampling and the process of randomisation

We conducted a clustered randomised controlled trial of CAL in rural schools beginning in 2011. A total of 2907 students in 57 elementary schools in poor minority areas in China's Qinghai Province participated in our study. The sample consists of two successive cohorts of grade 3 students. The first cohort includes those who were grade 3 students in the academic year of 2011–12. The second cohort includes those who were grade 3 students in the academic year of 2012–13.

We adopted a three-step process to arrive at this sample. First, because we wanted to focus on rural ethnic minority students – among the most disadvantaged student groups in the country – we restricted our samples to ethnic minority counties in Qinghai province, where the population is 60 per cent rural and 46 per cent ethnic minority. More specifically, we chose our sample counties in Haidong Prefecture, which covers one of the poorest areas in the east part of Qinghai Province. The counties in our sample frame were three relatively accessible minority autonomous counties in Haidong Prefecture. Two out of the three sample counties are nationally-designated poverty counties (Xunhua and Hualong); the remaining county is a provincially-designated poverty county (Huzhu).[1] Among these three counties, Xunhua is a Salar minority autonomous county, with 94 per cent of its population ethnic minorities; Hualong is a Hui minority autonomous county (around 79% of its population are ethnic minorities), and Huzhu is a Tu minority autonomous county (25% of its population are ethnic minorities; and it is the only Tu minority autonomous county in China).

Second, after choosing the counties, we created a sample of schools. To do so, we obtained a comprehensive list of all schools in the sample counties. After this, we called each school to confirm the number of minority students in that school. We excluded schools with less than 20 per cent minority students. For convenience we also excluded schools that were either too big (more than 150 grade 3 students – such schools would require too many computers) or too small (less than 20 grade 3 students – such schools were in danger of being closed or merged). After these two exclusions, there were 70 schools that met the study's criteria. Due to limitations on the number of computers available for use in the CAL programme, we were able to only implement CAL in 26 treatment schools. Consequently, we then randomly chose 57 schools (26 schools for treatment; 31 schools for controls) for our final sample.[2]

Finally, because the Mandarin CAL software we used was only available at the grade 3 level at the time, we focused our CAL programme on grade 3 students. In total, there were 3701 grade 3 students included in the initial sample (1981 of whom belonged to the first cohort, and 1720 of whom belonged to the second cohort – see Table 1). Among all of the sample grade 3 students, 86 per cent were from minority families.

Table 1. Description of sample and attrition.

		Cohort 1			Cohort 2		
Number of students	Total	Total	Treatment	Control	Total	Treatment	Control
Initial	3701	1981	805	1176	1720	722	998
Follow-up	3164	1723	683	1040	1441	594	847
Attrition	15%	13%	15%	12%	16%	18%	15%

After choosing the 57 schools for our sample, we randomly chose (without stratification) 26 schools from these 57 schools to receive the CAL intervention. The grade 3 students in the 26 treatment schools during the 2011–12 and 2012–13 academic years constituted the treatment students (805 of whom were from the first cohort, and 722 of whom were from the second cohort – see Table 1). The 2174 grade 3 students in the other 31 schools served as the control students (1176 of whom were from the first cohort, and 998 of whom from the second cohort – see Table 1).

Experiment arms/interventions

Our experiment focused fully on two treatment groups: the grade 3 students in the 26 treatment schools during the 2011–12 academic year (the first cohort) and the grade 3 students in the same 26 treatment schools during the 2012–13 academic year (the second cohort). We also used two control groups: the grade 3 students of the 31 control schools in the first cohort (during the 2011–12 academic year) and the grade 3 students of the same 31 control schools in the second cohort (during the 2012–13 academic year). For clarity and simplicity in the rest of the paper we call the intervention received by the first cohort the *phase I programme*, and the intervention received by the second cohort the *phase II programme*.

CAL intervention: phase I programme

The intervention of the phase I programme involved computer-assisted Mandarin remedial tutoring sessions that were designed to complement the regular Mandarin language curriculum.[3] Under the supervision of two local teacher–supervisors 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. No regular class time was used during the phase I programme. The sessions were mandatory. The content of each session emphasised core competencies in the uniform national Mandarin language curriculum.

During each session, two students randomly assigned as a pair shared one computer and played educational games matched to the material covered that week in regular Mandarin class. In a typical session, the students first watched an animated video that reviewed the material and then they were able to play Mandarin games to practice the relevant skills, such as vocabulary, reading comprehension and grammar introduced in class. If a student had a Mandarin-related question, he/she was encouraged to discuss it with his/her teammate, not 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. According to our in-class observations, the sessions were so intense that the attention of the students was fully on the computer and there was almost no communications among the groups or between any of the groups and the teacher–supervisor.

To facilitate the implementation, we also designed a detailed CAL curriculum and implementation protocol. The protocol was targeted at the teacher–supervisors that were charged with implementing the CAL programme in each school. One of the most important jobs of the teacher–supervisor was to make sure the CAL sessions were proceeding on a pace that matched that of the students' regular Mandarin classes. To avoid confounding the effect of the CAL intervention with any influence of additional teaching inputs to the students, none of the teacher–supervisors were Mandarin or math teachers or homeroom teachers of the grade 3 students. An implementation protocol was presented in a manual, which was given to the teacher–supervisor as a bound booklet that contained detailed instructions.

CAL intervention: phase II programme

The implementation of the phase II CAL programme (for the second cohort of the treatment group) was almost the same as that of the phase I programme, except that the math subject was introduced into the programme package. Because of this, the number of CAL sessions doubled.

Besides the two sessions of CAL intervention in Mandarin (which were identical to those provided to students in the phase I programme), the students in the phase II programme also were offered two CAL sessions per week in math. Like the CAL programme packages in Mandarin, the CAL package in math was designed to complement the national uniform math curriculum on a week-by-week basis. The classroom rules and implementation methods for the math sessions were exactly the same as that of the Mandarin sessions.

Importantly, although we requested schools to 'replicate' the phase I programme in terms of the implementation of the phase II Programme, due to the increased number of CAL sessions per week (from two sessions per week during the phase I programme to four sessions per week during the phase II programme), during the phase II programme some of the schools believed they had no option but to use regular school hours for at least some of the CAL sessions. In some schools, students and teachers live far away from their schools. Because of this, school officials want children to leave for home shortly after classes are over. In other schools, it is not convenient for students to spend lunch breaks at school because students must go home for lunch. In such schools, school officials had difficulties arranging all of the CAL sessions during after-school/out-of-school hours and had no option but to schedule some or all of the CAL sessions during in-school hours. As a result, whereas the after-school phase I CAL programme would act as a complement to the regular in-class instruction, the phase II CAL programme – at least in some schools – might be more of a substitute (to the regular in-class instruction).

CAL control group

Grade 3 students in the 31 control schools constituted the CAL control group for both phase I programme and phase II programme. For both cohorts, students in the control group did not receive any CAL intervention. To avoid any type spillovers from the CAL intervention, the principals, teachers and students (and their parents) of the control schools were not informed of the CAL project. The students in the control group took their regular Mandarin and math classes at school.

Data collection

For each cohort, the research group conducted two rounds of surveys in the 57 control and treatment schools. The first-round survey for the phase I programme was a baseline survey conducted with all grade 3 students in the 57 schools during the 2011–12 academic year. The baseline was done before any implementation of the CAL programme had begun (and before the random assignments were made). The second-round data collection effort was a final evaluation survey conducted at the conclusion of the programme at the end of the 2011–12 academic year. The baseline and endline surveys for the phase II programme were conducted before and after the implementation of the phase II CAL programme during the 2012–13 academic year.

In each round of survey, the enumeration team visited each school and conducted a two-part survey. In the first part students were given a standardised Mandarin test and a standardised math test. The Mandarin test included 40–45 questions (tests in different rounds included slightly different numbers of questions). The math test included 25 questions. Students were required to finish tests in each subject in 25 min. All students took the Mandarin test before taking the math test. Our enumeration team strictly enforced time limits and proctored the examinations. We use the scores of the students on the Mandarin and math tests as our measures of student academic performance.

In the second part of the survey, 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 whether the students were *male*, their ethnic group (*Han, Hui, Salar, Tu and Tibetan*), county, whether they were the *only child* in the family, whether their father was illiterate (*father illiterate*), whether their mother was illiterate (*mother illiterate*), whether their parents were still farmers and worked on the family farm or if their parents worked off-farm (*family off-farm*) and whether students had *ever used a computer* or had ever had *access to other modern technologies*. To create indicators of parental care, during the survey the students were also asked whether their parents had migrated to some other location outside of his/her home town or whether their parents stayed at home for most of the time during the semester (*live with both parents*).

In the second part of the survey (during the surveys in both phase I and phase II), students were also asked to answer questions that could help us measure their noncognitive traits. To create indicators for student attitudes towards schooling (*like school*), the students were asked to rate their attitude towards school on a 0–100 scale, where '0' indicates 'extremely hate school' and '100' indicates 'extremely enjoy school.' The indicator of self-efficacy of studying Mandarin was created from the responses of students to a seven-item psychological scale for the self-efficacy of studying Mandarin. To measure the self-efficacy of Mandarin studying, a professor in psychometrics at Beijing Normal University chose among the 12 indicators of studying attitudes used in TIMSS 2003 and developed a seven-item scale of self-efficacy of Mandarin studying that is appropriate to use under the context of elementary schools in China. Self-efficacy of studying is a psychological concept that measures one's belief in one's ability to succeed in learning and problem-solving in a certain subject. One's sense of efficacy of studying can play a major role in how one approaches goals, tasks, and challenges related in the study of a subject. Individuals with higher levels of self-efficacy in studying typically take control over their own learning experience and are more likely to participate in class and prefer hands-on learning experiences.

Statistical methods

We used ordinary least squares (OLS) regression analysis (both with and without control variables) to estimate how the academic and noncognitive outcomes changed in the treatment group relative to the control group. Our basic OLS analysis regressed the endline outcome variables (that is, postprogramme outcome variables) on the value of outcome variables at baseline, a dummy variable measuring each student's treatment (CAL intervention) status and a cohort dummy. First, the basic OLS model is

$$y_{isc} = \alpha + \beta treatment_s + \theta y_{0isc} + \gamma c + \epsilon_{isc}, \tag{1}$$

where y_{isc} is the endline outcome variable for child i in school s and cohort c, y_{0isc} measures the outcome variable of the same child at the baseline, *treatment* is a dummy variable for a grade 3 student attending a treatment school (equal to 1 for students in the treatment group and 0 otherwise), c is a cohort dummy variable which is equal to 1 for the second cohort and 0 for the first cohort, and ϵ_{isc} is a random disturbance term clustered at the school level.

We used several dependent variables to measure the student academic and noncognitive outcomes (y_{isc}). The primary outcome variable of our analysis is the student academic outcome, measured by the student standardised test scores in Mandarin and math. Importantly, besides variables measuring academic outcomes, we also included two noncognitive outcome variables, namely, *like school* and self-efficacy in Mandarin studying.

By construction, the coefficient of the dummy variable *treatment*, β, measures the difference in the value of the outcome variable between the treatment and control groups over the programme period conditional on the baseline value of the outcome variable for each group. In other words, βmeasures how the treatment group changed in the outcome levels during the programme period relative to the control group.

In order to improve the efficiency of the estimation, we also estimated the same model controlling a set of additional control variables including student individual and family characteristics (*male*, ethnic group (*Han, Tibetan, Tu, Hui and Salar*), *only child, father illiterate, mother illiterate, family off-farm, live with both parents, ever used a computer, and access to other modern* *technology*). By doing so, β is an unbiased, efficient estimate of the CAL treatment effect. In all regressions, we accounted for the clustered nature of our sample by constructing Huber-White standard errors corrected for school-level clustering.

To test for the difference in the CAL treatment effects between phase I treatment (which was done for the first cohort) and phase II treatment (for the second cohort), we ran the same model as in equation (1) (with control variables) but included one additional interaction term. The interaction term was created by multiplying the treatment dummy by the cohort dummy. We then can use the coefficient of the interaction term to test for the significance of the difference of the two treatments. If the coefficient on the interaction term is positive (negative) and significant, it indicates that the phase II CAL programme was more (less) effective than the phase I CAL programme.

Attrition and validity of randomisation

Although at the time of the baseline survey the main sample included 3701 grade 3 students, there was some attrition by the end of the study for each cohort. For various reasons (mainly school transfers and absences due to illness or injuries) at the time of the evaluation survey for the first cohort, we were only be able to follow up with a total of 1723 grade 3 students (both treatment and controls) (Table 1). Among the 805 treatment students in the first cohort, 683 remained in our sample at the time of the final evaluation. This means that the attrition rate was 15 per cent. Among the 1176 control students in the first cohort, 1040 remained in our sample at the time of the final evaluation, indicating an attrition rate of 12 per cent. Among the 1720 students in the second cohort (722 in the treatment group and 998 in the control group), we were only able to follow up with 1441 students (594 in the treatment group and 847 in the control group).

To understand the nature of those that attrited and assess if the attrition affected the nature of the sample or the validity of the randomisation, we regressed each variable from a set of student characteristics (see later for more details on the definition of the variables) on a constant, a dummy for attrition, a dummy for treatment and an interaction of the attrition and treatment dummies. The student individual characteristics were those that were included as control variables in the full model (defined earlier). According to our findings (Table 2), students who attired from the sample did differ from those who did not attrite. For the first cohort (row 4), we see that students who attrited had lower Math scores (column 2), a higher share of mothers that were illiterate (column 10), were less likely to live with both parents (column 14) and were less likely to be Tu and more likely to be Hui (columns 5 and 6). For the second cohort (row 7), students who attrited had lower Mandarin and Math test scores (columns 1 and 2) and more access to modern technologies, and were less likely to be Tu and Salar and more likely to be Hui (columns 5–7). However, as seen from the fact that none of the coefficients of the interaction variables were statistically significant (rows 3, 6 and 9), there is no evidence that the attrition affected the validity of our randomisation for either cohort.

After excluding those that attrited, we used the same set of student characteristics to check the validity of the random assignment for our final analytic sample by regressing each variable on the treatment dummy variable. We found the differences between the treatment and control groups for all of the variables were statistically insignificant (Table 3). The *F*-test of joint significance of the student characteristics was also insignificant.

Finally, we also used the same set of student characteristics to check the validity of the random assignment for the whole sample (that is, the sample that included both the students in our final analytic sample and those who attrited). When doing so, we found that the differences between the treatment and control groups were also statistically insignificant for all student characteristics (see the Appendix).

Table 2. 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.

Variables	(1) Baseline Mandarin score (units of standard deviation)[a]	(2) Baseline math score (units of standard bdeviation)[b]	(3) Female (1 = yes; 0 = no)	(4) Tibetan	(5) Tu	(6) Hui	(7) Salar	(8) Only child (1 = yes; 0 = no)	(9) Father illiterate (1 = yes; 0 = no)	(10) Mother illiterate (1 = yes; 0 = no)	(11) Ever used a computer (1 = yes; 0 = no)	(12) Access to other modern technologies[b]	(13) Family off-farm	(14) Live with both parents
Both cohort jointly														
(1) Attrition	−0.37***	−0.40**	−0.01	−0.04	−0.13***	0.18***	−0.05	−0.03	0.03	0.07**	0.06	0.04**	0.02	−0.16***
	(0.13)	(0.16)	(0.03)	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
(2) Treatment	0.00	−0.08	−0.02	0.02	0.03	−0.03	0.01	−0.01	−0.00	0.02	−0.04	−0.00	−0.00	−0.02
	(0.14)	(0.12)	(0.02)	(0.06)	(0.10)	(0.12)	(0.06)	(0.02)	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.04)
(3) Treatment*attrition	0.07	0.18	0.00	0.02	−0.02	0.00	−0.00	0.02	0.01	−0.02	0.03	−0.02	0.03	0.03
	(0.21)	(0.22)	(0.04)	(0.05)	(0.06)	(0.07)	(0.04)	(0.03)	(0.05)	(0.05)	(0.04)	(0.03)	(0.03)	(0.05)
First cohort														
(4) Attrition	−0.7	−1.14***	0.03	−0.02	−0.10**	0.18***	−0.00	0.00	0.00	0.10**	0.09	0.00	0.01	−0.29***
	(0.36)	(0.33)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.04)	(0.06)	(0.03)	(0.02)	(0.04)
(5) Treatment	0.00	−0.11	−0.03	0.03	0.01	−0.04	0.06	−0.01	−0.02	0.03	−0.06	0.01	−0.00	−0.03
	(0.16)	(0.15)	(0.03)	(0.07)	(0.10)	(0.12)	(0.12)	(0.02)	(0.03)	(0.04)	(0.07)	(0.03)	(0.02)	(0.05)
(6) Treatment*attrition	−0.68	0.08	−0.01	0.03	−0.03	0.05	−0.07	−0.01	0.04	−0.01	0.06	0.04	0.01	0.1
	(0.38)	(0.35)	(0.07)	(0.06)	(0.05)	(0.07)	(0.07)	(0.04)	(0.06)	(0.07)	(0.09)	(0.04)	(0.04)	(0.06)
Second cohort														
(7) Attrition	−0.35***	−0.34**	−0.05	−0.05	−0.17***	0.19***	−0.09***	−0.05	0.05	0.04	0.04	0.06***	0.03	−0.04
	(0.12)	(0.15)	(0.05)	(0.03)	(0.05)	(0.07)	(0.03)	(0.03)	(0.05)	(0.04)	(0.02)	(0.02)	(0.03)	(0.05)
(8) Treatment	0.00	−0.05	−0.00	0.01	0.05	−0.02	−0.06	−0.02	0.02	0.00	−0.02	−0.01	−0.01	0.01
	(0.16)	(0.14)	(0.03)	(0.06)	(0.10)	(0.13)	(0.05)	(0.03)	(0.04)	(0.05)	(0.04)	(0.02)	(0.02)	(0.05)
(9) Treatment*attrition	0.06	0.11	0.01	0.02	−0.02	−0.05	0.07	0.04	−0.02	−0.03	−0.01	−0.05	0.06	−0.04
	(0.20)	(0.21)	(0.07)	(0.04)	(0.08)	(0.10)	(0.04)	(0.04)	(0.06)	(0.08)	(0.04)	(0.03)	(0.05)	(0.07)

Robust standard errors clustered at the school level are shown in parentheses. For each cohort, each column reports the results of one regression of the corresponding student characteristics in columns (1)–(14) on the treatment dummy variable, attrition dummy variable and the interaction between these two dummy variables. Rows (1)–(3) report the results for both cohorts jointly, rows (4)–(6) report results for the first cohort and rows (7)–(9) report results for the second cohort.

*** $p < 0.01$, ** $p < 0.05$.

[a,b] The baseline math/Mandarin score is the score on the standardised math/Mandarin test that was given to all students in the sample before the CAL programme.

[c] 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.

Table 3. Comparison of student and family characteristics and student access to computer and other modern technologies between the treatment and control groups for students who did not attrite the sample.

	(1)	(2)
Dependent variable: the treatment dummy variable	First cohort	Second cohort
(1) Baseline Mandarin score (units of standard deviation)[a]	0.00	0.00
	(0.16)	(0.16)
(2) Baseline math score (units of standard deviation)[b]	−0.11	−0.05
	(0.15)	(0.15)
(3) Female (1 = yes; 0 = no)	−0.03	−0.00
	(0.03)	(0.03)
(4) Tibetan	0.03	0.01
	(0.07)	(0.06)
(5) Tu	0.01	0.05
	(0.10)	(0.10)
(6) Hui	−0.04	−0.02
	(0.12)	(0.13)
(7) Salar	0.06	−0.06
	(0.12)	(0.05)
(8) Only child (1 = yes; 0 = no)	−0.01	−0.02
	(0.02)	(0.03)
(9) Father illiterate (1 = yes; 0 = no)	−0.02	0.02
	(0.03)	(0.04)
(10) Mother illiterate (1 = yes; 0 = no)	0.03	0.00
	(0.04)	(0.05)
(11) Ever used a computer (1 = yes; 0 = no)	−0.06	−0.02
	(0.07)	(0.04)
(12) Access to other modern technologies[c]	0.01	−0.01
	(0.03)	(0.02)
(13) Family off-farm	−0.00	−0.01
	(0.02)	(0.02)
(14) Live with both parents	−0.03	0.01
	(0.05)	(0.05)
(15) Observations	1723	1,441

Robust standard errors clustered at the school level in parentheses.
*** $p < 0.01$, ** $p < 0.05$.
[a] The baseline Mandarin score is the score on the standardised Mandarin test that was given to all students in the sample before the CAL programme.
[b] The baseline math score is the score on the standardised math test that was given to all students in the sample before the CAL programme.
[c] 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.
The difference between the treatment and control groups conditional on the other variables is calculated by regressing each of the row variable on the treatment dummy variable.

Results

The impacts of the CAL programme on the student test scores

The data show that overall the CAL programme has significant effects on student academic performance. Over the whole period (including both phases I and II of the programme), students in the treatment group improved their standard test scores by 0.15 standard deviations more than students in the control school on both Chinese and math tests (Table 4, row 1, columns 3–4). The results were consistent whether we estimated the model with or without control variables (columns 1–4).

To test the significance of the differences in programme effects between the phase I and II programmes, we used the pooled sample and added the interaction term (which was the product of the treatment dummy and cohort dummy) into model (1) with all control variables. As discussed earlier, we then tested for the different in programme effects by examining the significance of the coefficient of the interaction term. According to our findings, the relevant coefficients was negative

Table 4. Ordinary least squares estimators of the impacts of CAL programme on student academic outcomes.

			Both	Cohorts	Jointly		
Dependent variable: student test scores of the final evaluation	Mandarin (1)	Math (2)	Mandarin (3)	Math (4)	Mandarin (5)	Math (6)
(1) Treatment (1 = the treatment group; 0 = the control group)	0.11 (0.06)	0.15** (0.07)	0.15*** (0.05)	0.15** (0.06)	0.20*** (0.07)	0.22*** (0.08)
(2) Baseline Mandarin score (units of standard deviation)[a]	0.69*** (0.03)		0.50*** (0.03)	0.21*** (0.03)	0.50*** (0.03)	0.21*** (0.03)
(3) Baseline math score (units of standard deviation)[b]		0.60*** (0.02)	0.24*** (0.02)	0.46*** (0.02)	0.24*** (0.02)	0.46*** (0.02)
(4) Female (1 = yes; 0 = no)			0.12*** (0.02)	0.01 (0.03)	0.12*** (0.02)	0.01 (0.03)
(5) Tibetan			−0.07 (0.07)	−0.02 (0.07)	−0.07 (0.07)	−0.03 (0.07)
(6) Tu			−0.10 (0.06)	−0.05 (0.07)	−0.11 (0.06)	−0.05 (0.07)
(7) Hui			0.04 (0.09)	−0.14 (0.08)	0.04 (0.09)	−0.14 (0.08)
(8) Salar			−0.13 (0.08)	−0.05 (0.09)	−0.14 (0.08)	−0.07 (0.09)
(9) Only child (1 = yes; 0 = no)			−0.08** (0.04)	−0.04 (0.04)	−0.08** (0.04)	−0.04 (0.04)
(10) Father illiterate (1 = yes; 0 = no)			−0.08 (0.04)	−0.09** (0.04)	−0.08 (0.04)	−0.09** (0.04)
(11) Mother illiterate (1 = yes; 0 = no)			−0.04 (0.03)	−0.01 (0.03)	−0.04 (0.03)	−0.01 (0.03)
(12) Ever used a computer (1 = yes; 0 = no)			0.08** (0.04)	0.05 (0.04)	0.08** (0.04)	0.05 (0.04)
(13) Access to other modern technologies[c]			0.22*** (0.07)	0.16** (0.08)	0.22*** (0.07)	0.16** (0.07)
(14) Family off-farm			0.07 (0.05)	0.00 (0.04)	0.06 (0.05)	−0.00 (0.04)
(15) Live with both parents			0.02 (0.03)	−0.01 (0.03)	0.02 (0.03)	−0.01 (0.03)
(16) County dummies			Y	Y	Y	Y
(17) Cohort dummies	Y	Y	Y	Y	Y	Y
(18) Interactions of treatment dummy and cohort dummy					−0.11 (0.09)	−0.16 (0.11)
(19) Number of observations	3164	3164	3164	3164	3164	3164

Robust standard errors clustered at the school level are shown in parentheses. Each column reports the results of one regression of the student standardised Mandarin or math test scores over the programme period on the corresponding variables in rows (1)–(17).

*** $p < 0.01$, ** $p < 0.05$.

[a] The baseline Mandarin score is the score on the standardised Mandarin test that was given to all students in the sample before the CAL programme.

[b] The baseline math score is the score on the standardised math test that was given to all students in the sample before the CAL programme.

[c] 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.

and insignificant (Table 4, row 18, columns 5 and 6). This result indicates that the difference in programme effects between the two CAL regimens was not statistically significant.

Based on these findings, therefore, the empirical results suggest that on the one hand, the CAL programme (as a whole package including both regimens) did significantly improve student academic performance in both Mandarin and Math. On the other hand, the CAL regimen in the phase II of the experiment, while having a broader scope (Mandarin + math), did not produce improvement in either subject (neither Mandarin nor math) when compared to the results of the phase I programme.

When we split the sample and compare the results of each CAL regimen on each cohort, we find even stronger evidence of the above pattern (that the more intensive and broader phase II CAL programme did not generate any additional benefits when compared with the phase I programme, the programme with a narrower scope). Specifically, during the phase I programme period, grade 3 students in the treatment group improved significantly more in their academic performance in Mandarin than students in the control group (Table 5, row 1, column 1). Using the sample including all grade 3 students in the 57 sample schools, the estimated CAL treatment effect on Mandarin test scores is equal to 0.19 and significant at the 1 per cent level using the model controlling for the various student characteristics.[4] Moreover, even though the CAL programme during phase I only focused on Mandarin, there was a positive and significant impact on math test scores (Table 5, row 1, column 2). Compared to their counterparts in the control schools, grade 3 students in the treatment schools improved their math test scores by 0.23 standard deviations more than students

Table 5. Ordinary least squares estimators of the impacts of CAL programme on student academic outcomes by cohort.

	First	cohort	Second	cohort
Dependent variable: student test scores of the final evaluation	Mandarin (1)	Math (2)	Mandarin (3)	Math (4)
(1) Treatment (1 = the treatment group; 0 = the control group)	0.19***	0.23***	0.1	0.06
	(0.06)	(0.07)	(0.07)	(0.09)
(2) Baseline Mandarin score (units of standard deviation)[a]	0.53***	0.16***	0.46***	0.26***
	(0.03)	(0.03)	(0.04)	(0.03)
(3) Baseline math score (units of standard deviation)[b]	0.21***	0.50***	0.29***	0.39***
	(0.03)	(0.03)	(0.04)	(0.03)
(4) Female (1 = yes; 0 = no)	0.12***	0.08	0.11***	−0.07
	(0.03)	(0.04)	(0.04)	(0.04)
(5) Tibetan	−0.04	0.01	−0.07	0.04
	(0.07)	(0.07)	(0.23)	(0.14)
(6) Tu	−0.09	−0.11	−0.07	0.13
	(0.07)	(0.06)	(0.22)	(0.17)
(7) Hui	0.01	−0.19	0.09	0.06
	(0.12)	(0.10)	(0.11)	(0.12)
(8) Salar	−0.20	0.07	−0.08	0.03
	(0.22)	(0.13)	(0.22)	(0.15)
(9) Only child (1 = yes; 0 = no)	−0.05	0.03	−0.13**	−0.11
	(0.06)	(0.06)	(0.05)	(0.06)
(10) Father illiterate (1 = yes; 0 = no)	−0.03	−0.05	−0.12	−0.13
	(0.05)	(0.04)	(0.06)	(0.07)
(11) Mother illiterate (1 = yes; 0 = no)	−0.07**	0.02	0.00	−0.07**
	(0.03)	(0.04)	(0.04)	(0.04)
(12) Ever used a computer (1 = yes; 0 = no)	0.05	0.02	0.16**	0.13
	(0.05)	(0.05)	(0.07)	(0.07)
(13) Access to other modern technologies[c]	0.28**	0.35***	0.14	−0.12
	(0.11)	(0.09)	(0.11)	(0.11)
(14) Family off-farm	−0.04	−0.11	0.15**	0.13
	(0.06)	(0.07)	(0.06)	(0.07)
(15) Live with both parents	0.03	0.05	0.01	−0.08
	(0.04)	(0.05)	(0.04)	(0.05)
(16) County dummies	Y	Y	Y	Y
(17) Cohort dummies				
(18) Number of observations	1583	1583	1324	1324

Robust standard errors clustered at the school level are shown in parentheses. Each column reports the results of one regression of the student standardised Mandarin or math test scores over the programme period on the corresponding variables in rows (1)–(17). The results are similar with or without controlling variables in rows (2)–(17). We report the results with controls for simplicity.

*** $p < 0.01$, ** $p < 0.05$.

[a] The baseline Mandarin score is the score on the standardised Mandarin test that was given to all students in the sample before the CAL programme.

[b] The baseline math score is the score on the standardised math test that was given to all students in the sample before the CAL programme.

[c] 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.

in the control school during the programme period. These results suggest that the CAL intervention did not improve the performance of students in Mandarin at the expense of their performance in math. Instead, the academic benefits from the CAL intervention also appear to have had strong spillovers on mathematics.

There are several possible reasons for the positive spillover effect of the Mandarin CAL intervention onto the math performance of students during the phase I CAL programme. First, the CAL intervention might have increased the general learning ability of the students, which helped improve their performance in other subjects including math. Second, the game-based CAL software may have made learning a more fun and engaging process and thus stimulated student interest in learning Mandarin as well as other subjects (including math). Perhaps most importantly, as discussed earlier, all academic subjects, including math, in the sample schools were taught in Mandarin and the textbooks were written in Mandarin. Therefore, improved Mandarin language skills through CAL may also have helped the students better understand the math course materials.[5]

With such positive impacts of the phase I programme on Mandarin performance as well as significant positive spillovers on student math performance, ex ante, one might expect to observe even larger impacts of the CAL programme on academic performance in phase II, when a math component was explicitly introduced into the CAL programme. With a CAL package augmented by two sessions in math, the students not only were given more CAL sessions each week, but also received remedial tutoring in math directly through the CAL math programmes. Because of this, it was reasonable to expect that, at least in terms of math tests scores, student should improve even more relative to the control group than they did during the phase I programme.

However, during the phase II programme, for the entire sample of treatment schools there was no significant improvement in the student academic performance in either Mandarin or math. The point estimates of the CAL treatment effects during the phase II programme for Mandarin and math were 0.1 and 0.06 standard deviations, respectively (Table 5, row 1, columns 3 and 4). Although positive, the measured impacts are smaller than those during the phase I programme in magnitude and are statistically insignificant.

At first glance the findings are puzzling for two reasons. Of course, the most puzzling result relates to the outcomes on the math tests. During the phase I CAL, the spillover from the Mandarin programme was enough to yield a statistically significant gain in student math performance. It is puzzling why adding a set of CAL math sessions to the intervention had no added effect on student math performance. The other puzzle is with both the math and Mandarin results. During the phase II programme, the overall programme intensity increased compared to that of the phase I programme. However, more CAL sessions seem to have not brought about larger programme effects in either math or Mandarin.

What might explain this result? There are several possible reasons that may be related to the fact that the phase II regimen of the CAL programme became broader in scope. First, the programme could have increased the burden on the staff managing the programme. If the more intense programme was burdensome enough for the staff, this might have decreased the overall quality of the implementation. However, considering that the teacher supervisors of the programme were not there for instruction, but simply to keep the sessions in order and answer occasional questions about the computer and software operations, it seems unlikely that any increased burden on management could have accounted for the lack of improvement of the phase II programme over the phase I programme.

Second, the students may simply have been burned out by the additional CAL session time. When there were multiple sessions (Mandarin and math) rather than only one (Mandarin only), students (to the extent that any part of the CAL programme was given during out of school hours) had to stay longer in school for the phase II CAL programme. However, the student burnout hypothesis is unlikely to be one of the primary reasons for the ineffectiveness of the phase II CAL programme, since during the endline surveys CAL treatment school students reported that their interest in schooling increased over time.[6]

Third, language difficulties during the implementation of the CAL math programme may have been one of the reasons for the absence of an effect for the phase II CAL programme. The CAL math programme materials (that is, the game-based, animated lectures and exercises) were all in Mandarin. The minority students, most of whom speak their own language in daily life and Mandarin only in class (some teachers even used their own language instead of Mandarin to deliver part of their lecture for clarity and convenience), might not have understood materials in some of the sessions of the phase II CAL math programme due to language problems.[7] Because of these language difficulties it is possible that they did not benefit as much. Hence, it is possible that such difficulties in language may have confused or frustrated the students more than helped them in learning math. In other words, it might have been too early to introduce the phase II CAL math programme in Mandarin – especially in the case of grade 3 minority students.[8]

Finally, substitution effects might have played an important role in this finding. There are several reasons to think this might be the case. First, it is possible that the additional CAL session time might have come at the expense of other activities (such as homework help sessions). In other words, students may not have been able to become engaged in learning the subjects that were being taught/remediated after school. Second, the phase II CAL programme in our study might have substituted for existing classroom inputs in a way that offset at least part of the CAL treatment effect that was observed during phase I.

According to the existing literature, as well as according to our field records from the CAL enumerators, we found that substitution effects might be one possible reason for this finding. Linden (2008) tried to isolate possible negative substitution effects of a CAL programme run during schools hours by comparing two programmes, one that was run after school and one during. He concluded that when implemented in place of existing classroom inputs, a substitution effect appears to have attenuated the positive effects of CAL programme on student performance. According to our data in the China phase II CAL programme, due to the increased number of CAL sessions, many schools had to use regular school hours for at least some of the CAL sessions. Specifically, 49 per cent, 62 per cent and 79 per cent of the treatment students in Huzhu, Xunhua and Hualong counties, respectively, were in schools that used school hours to implement some of the programme (Table 6, column 2). On average, during the phase II CAL programme around 63 per cent of the grade 3 treatment students were in schools that used regular school hours for the CAL sessions. In the case of the phase I CAL programme schools, our records do not show any schools used regular hours. Therefore, it is plausible that the CAL sessions might have crowded out regular school instruction (or other in-classroom learning activities) and, thus, offset part of the positive effects of the CAL programme.

To further examine whether the substitution effect played a role in the lack of improvement of the broader phase II CAL programme over the phase I CAL programme, we executed an empirical exercise. We restricted the observations in our analytical sample to students in schools that did not use any regular school hours for the CAL sessions (and thus should not have been affected by the

Table 6. The percentage of students in schools that take up regular school hours for CAL in each of the three programme counties.

County	(1) Total number of students	(2) Percentage of students in schools that use regular school hours for at least some of their CAL sessions
(1) Huzhu	504	49
(2) Xunhua	301	62
(3) Hualong	472	79
Total	1277	63

substitution effects). When doing so, the phase II CAL treatment effects – similar to the phase I CAL treatment effects – were positive and statistically significant for both subjects. Specifically, the phase II CAL treatment raised the Mandarin test scores of grade 3 treatment students over control students by 0.20 standard deviations (Table 7, row 1, column 3). The coefficient was significant at the 1 per cent level. Similarly, the phase II CAL treatment raised math test scores by 0.35 standard deviations in math (statistically significant at the 1 per cent level – Table 7, column 4). This effect is larger than the impact of phase I CAL treatment on student math test scores (0.23 standard deviations), which is just a spillover effect from the phase I CAL Mandarin programme.

The most serious concern for this subsample approach is that there is no guarantee that the differences between the measured effects of the phase II CAL when including all schools and when only including schools that did not use regular hours is solely due to a substitution effect. It is also possible that other inherent differences between schools that implemented the CAL sessions as designed (using only out of regular school hour time periods) and those that chose to use regular in-school hours for CAL sessions were driving the results. If there were such differences, we say that the measured differences were in part due to selection bias.

Table 7. The estimators of CAL treatment effects restricting the sample to schools without substitution.

Dependent variable: student test scores of the final evaluation	(1) Mandarin	(2) Math	(3) Mandarin	(4) Math
(1) Treatment (1 = the treatment group; 0 = the control group)	0.22***	0.33***	0.20***	0.35***
	(0.08)	(0.12)	(0.07)	(0.11)
(2) Baseline Mandarin score (units of standard deviation)[a]	0.67***		0.47***	0.29***
	(0.04)		(0.05)	(0.04)
(3) Baseline math score (units of standard deviation)[b]		0.57***	0.26***	0.34***
		(0.04)	(0.04)	(0.03)
(4) Female (1 = yes; 0 = no)			0.11**	−0.09**
			(0.04)	(0.04)
(5) Tibetan			0.15	−0.05
			(0.12)	(0.19)
(6) Tu			0.11	0.19
			(0.15)	(0.21)
(7) Hui			0.10	−0.03
			(0.10)	(0.19)
(8) Salar			0.08	−0.04
			(0.16)	(0.20)
(9) Only child (1 = yes; 0 = no)			−0.08	−0.17**
			(0.06)	(0.07)
(10) Father illiterate (1 = yes; 0 = no)			−0.15***	−0.11
			(0.05)	(0.07)
(11) Mother illiterate (1 = yes; 0 = no)			−0.00	−0.06
			(0.04)	(0.04)
(12) Ever used a computer (1 = yes; 0 = no)			0.20***	0.12
			(0.07)	(0.08)
(13) Access to other modern technologies[c]			0.13	−0.00
			(0.12)	(0.11)
(14) Family off-farm			0.10	0.16
			(0.07)	(0.09)
(15) Live with both parents			−0.01	−0.09
			(0.04)	(0.05)
(16) Observations	1065	1065	1065	1065

Robust standard errors are shown in parentheses. Each column reports the results of one regression of the student academic outcome on the corresponding variables in rows (1)–(15). Schools without substitution refer to schools, which did not arrange any of the CAL sessions during their in-school hours.

*** $p < 0.01$, ** $p < 0.05$.

[a] The baseline Mandarin score is the score on the standardised Mandarin test that was given to all students in the sample before the CAL programme.

[b] The baseline math score is the score on the standardised math test that was given to all students in the sample before the CAL programme.

[c] 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.

To assess whether there was a confounding influence from selection bias, we first compared the school and student characteristics between these two types of schools. The set of student characteristics were the same as those we used for testing the randomisation of the programme treatment. The school-level characteristics included the number of students, the teacher–student ratio, the proportion of female teachers, an index of the availability of school facilities and the proportion of students receiving poverty subsidies. When comparing the differences between the two types of schools, we found no significant differences between the two types of schools student or school-level characteristics (Tables 8 and 9).

We also added a placebo test to gauge the sample selection problem of this approach. To be more specific, we re-estimated the impacts of our CAL phase I programme on the student by comparing schools that had arranged at least some of the CAL sessions during regular school hours during phase II to the control schools. If there existed serious selection bias due to differences between treatment schools that used some regular school hours for CAL sessions and those that

Table 8. Differences in student characteristics between schools that took up in-school hours for CAL sessions and those that did not.

Variables	A dummy variable of whether the student attended schools that took up regular school hours for CAL sessions or not
(1) Baseline Mandarin score (units of standard deviation)[a]	0.02
	(0.03)
(2) Baseline math score (units of standard deviation)[b]	−0.06
	(0.04)
(3) Female (1 = yes; 0 = no)	−0.00
	(0.02)
(4) Tibetan	−0.11
	(0.17)
(5) Tu	0.00
	(0.18)
(6) Hui	−0.02
	(0.18)
(7) Salar	−0.13
	(0.15)
(8) Only child (1 = yes; 0 = no)	−0.05
	(0.05)
(9) Father illiterate (1 = yes; 0 = no)	0.04
	(0.06)
(10) Mother illiterate (1 = yes; 0 = no)	0.05
	(0.04)
(11) Ever used a computer (1 = yes; 0 = no)	−0.05
	(0.06)
(12) Access to other modern technologies[c]	0.11
	(0.09)
(13) Family off-farm	−0.04
	(0.05)
(14) Live with both parents	−0.01
	(0.04)
(15) Observations	1441

Robust standard errors clustered at the school level are shown in parentheses.
*** $p < 0.01$, ** $p < 0.05$.
[a] The baseline Mandarin score is the score on the standardised Mandarin test that was given to all students in the sample before the CAL programme.
[b] The baseline math score is the score on the standardised math test that was given to all students in the sample before the CAL programme.
[c] 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. The coefficients in the column report the results of one regression of the a dummy variable of whether the student attended schools that took up regular school hours for CAL Mandarin or math sessions or not on the student characteristics in the row variables.

Table 9. Differences in school characteristics between schools that took up in-school hours for CAL sessions and those that did not.

	(1) Number of students	(2) Teacher–student ratio	(3) The proportion of female teachers	(4) School facility index[a]	(5) The proportion of students receiving poverty subsidies
Substitution (yes = 1; no = 0)	20.09 (30.03)	−0.003 (0.008)	−0.0193 (0.076)	−0.046 (0.053)	−0.12 (0.123)
Observations	26	26	26	26	26

Robust standard errors are shown in parentheses. Each column reports the results of one regression of the corresponding school characteristics on a dummy variable of whether the school took up regular in-school hours for CAL Mandarin or math sessions or not.

*** $p < 0.01$, ** $p < 0.05$.

[a] The school facility index is the mean value of a set of 0/1 dummy variables indicating the availability (0 = unavailable; 1 = available) of some school facilities, including enclosure, playground, reading room, library, running water, hot water and clinic.

did not, we should expect differential CAL treatment effect estimates between the results of this placebo test and those using the whole sample. To be more specific, we should expect to find insignificant phase I treatment effects from the placebo tests. However, the results of the placebo were consistent with the findings using the whole sample (Table 10, row 1, columns 1–4). Based on this, we believe that the lack of improvement in CAL programme effects during the phase II programme is unlikely to have been caused by sample selection (or systematic differences between schools that used some regular school hours for CAL sessions and those that did not).

The benefits of CAL come with a cost in several aspects of the programme. 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, maintenance costs and compensation for CAL teacher–supervisors. Another major source of the programme cost is compensation for the teacher–supervisors. On average, one teacher–supervisor was in charge of 30 students, and we paid each teacher 120 yuan ($18) per month on average. Therefore, the cost in teacher compensation was $4.8 per student ($18 divided by 30 and then times 8 (months)). In addition to the cost related teacher–supervisor compensation, the maintenance cost amounts to $1 per student; the teacher training cost is around $1.8 per student; the software is negligible when averaged by the number of students. Therefore, the total per-student cost of our CAL programme is $7.6. Given that the average programme effect is 0.15 standard deviations, the CAL programme costs $5.1 per tenth of a standard deviation improvement in the student test score.

Effects of the CAL intervention on student non-academic outcomes

Our findings about the effects of CAL intervention on the self-efficacy of students who were studying Mandarin also support our hypothesis that language barrier might have offset the potential beneficial CAL treatment effects on the student academic outcomes. Our analysis shows that the phase I CAL intervention had significantly positive effects on the self-efficacy of Mandarin studying, whereas the phase II CAL intervention did not have such a benefit (Table 11, row 1, columns 2 and 3). One conclusion we can draw from this finding is that the language barriers or difficulties the minority students had with the CAL math programme might also be one important reason for absence of improvement in the CAL treatment schools on the math performance during the phase II CAL programme over the phase I CAL programme. For example, students may become frustrated if they perform poorly in a math game due to their limited Mandarin comprehension. The difficulties in the math content itself might also exacerbate their confusion and frustration from the language difficulties. This language barrier, in turn, may

Table 10. Placebo test: the estimates of the phase I CAL treatment effects for treatment schools that used at least some regular school hours for CAL sessions during the phase II programme.

Dependent variable: student test scores on the final evaluation of the phase I programme	Sample: All control and treatment schools (for comparison) (1) Mandarin	(2) Math	Sample: control schools and treatment schools that used regular school hours for CAL sessions during phase II (3) Mandarin	(4) Math
(1) Treatment (1 = the treatment group; 0 = the control group)	0.19***	0.23***	0.16**	0.25***
	(0.06)	(0.07)	(0.07)	(0.09)
(2) Baseline Mandarin score (units of standard deviation)[a]	0.53***	0.16***	0.53***	0.16***
	(0.03)	(0.03)	(0.04)	(0.04)
(3) Baseline math score (units of standard deviation)[b]	0.21***	0.50***	0.21***	0.51***
	(0.03)	(0.03)	(0.03)	(0.04)
(4) Female (1 = yes; 0 = no)	0.12***	0.08	0.12***	0.10**
	(0.03)	(0.04)	(0.03)	(0.05)
(5) Tibetan	−0.04	0.01	−0.05	0.09
	(0.07)	(0.07)	(0.09)	(0.07)
(6) Tu	−0.09	−0.11	−0.07	−0.13
	(0.07)	(0.06)	(0.09)	(0.07)
(7) Hui	0.01	−0.19	0.00	−0.16
	(0.12)	(0.10)	(0.13)	(0.11)
(8) Salar	−0.20	0.07	−0.26	0.10
	(0.22)	(0.13)	(0.27)	(0.17)
(9) Only child (1 = yes; 0 = no)	−0.05	0.03	−0.09	0.01
	(0.06)	(0.06)	(0.07)	(0.07)
(10) Father illiterate (1 = yes; 0 = no)	−0.03	−0.05	−0.03	−0.08
	(0.05)	(0.04)	(0.05)	(0.05)
(11) Mother illiterate (1 = yes; 0 = no)	−0.07**	0.02	−0.08**	0.03
	(0.03)	(0.04)	(0.04)	(0.04)
(12) Ever used a computer (1 = yes; 0 = no)	0.05	0.02	0.04	0.05
	(0.05)	(0.05)	(0.05)	(0.05)
(13) Access to other modern technologies[c]	0.28**	0.35***	0.26**	0.33***
	(0.11)	(0.09)	(0.11)	(0.10)
(14) Family off-farm	−0.04	−0.11	−0.06	−0.11
	(0.06)	(0.07)	(0.07)	(0.07)
(15) Live with both parents	0.03	0.05	0.03	0.07
	(0.04)	(0.05)	(0.04)	(0.05)
(16) Observations	1583	1583	1471	1471

Robust standard errors are shown in parentheses. Each column reports the results of one regression of the student academic outcome on the corresponding variables in rows (1)–(15). Columns (1) and (2) are the same as columns (1) and (2) of Table 5, and are reported here for comparison purpose.
*** $p < 0.01$, ** $p < 0.05$.
[a] The baseline Mandarin score is the score on the standardised Mandarin test that was given to all students in the sample before the CAL programme.
[b] The baseline math score is the score on the standardised math test that was given to all students in the sample before the CAL programme.
[c] 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.

contribute to the lower levels of self-efficacy in Mandarin studying despite improvement in their Mandarin test scores (among students that did not suffer from the substitution effects). With lower self-efficacy in Mandarin studying, as well as the substitution effects, it is not surprising to find no significant improvement in CAL treatment effects during the phase II programme compared to the phase I programme.

We also examined the effects of CAL intervention on student interest in schooling. For the phase II programme, controlling for their initial attitudes at the beginning of the programme, students in the treatment schools 'liked school' significantly more than those in the control schools; this beneficial effect was not observed during the phase I programme (Table 11, columns 5 and 6). Hence, it seems that with the phase II programme, students in the second

Table 11. Ordinary least squares estimators of the impacts of CAL programme on student non-academic outcomes.

	Self-efficacy of Mandarin study			Like School		
	Both cohorts (1)	First cohort (2)	Second cohort (3)	Both cohorts (4)	First cohort (5)	Second cohort (6)
(1) Treatment (1 = the treatment group; 0 = the control group)	0.08**	0.09**	0.05	2.37	0.63	4.19**
	(0.03)	(0.04)	(0.04)	(1.29)	(1.45)	(1.87)
(2) Baseline Mandarin score (units of standard deviation)[a]	0.11***	0.09***	0.15***	2.24***	2.67***	1.70
	(0.01)	(0.02)	(0.02)	(0.74)	(0.86)	(1.05)
(3) Baseline math score (units of standard deviation)[b]	0.04**	0.03**	0.03	0.45	−0.01	0.92
	(0.01)	(0.02)	(0.02)	(0.53)	(0.56)	(0.92)
(4) Female (1 = yes; 0 = no)	0.05**	0.02	0.09***	2.27***	1.85**	2.76**
	(0.02)	(0.02)	(0.03)	(0.76)	(0.87)	(1.19)
(5) Tibetan	−0.01	−0.00	−0.15	1.82	1.16	−1.92
	(0.04)	(0.06)	(0.09)	(2.02)	(1.85)	(3.95)
(6) Tu	0.04	0.02	−0.06	2.80	−0.61	2.77
	(0.04)	(0.04)	(0.11)	(1.78)	(2.07)	(4.32)
(7) Hui	−0.08	−0.09	−0.17**	−3.90	−5.97**	−6.89**
	(0.05)	(0.07)	(0.09)	(2.15)	(2.51)	(3.17)
(8) Salar	−0.00	−0.03	−0.11	4.38**	−0.31	0.42
	(0.04)	(0.10)	(0.10)	(1.98)	(4.20)	(4.35)
(9) Only child (1 = yes; 0 = no)	0.03	0.04	0.03	−0.44	−0.05	−1.01
	(0.03)	(0.03)	(0.04)	(1.07)	(1.33)	(1.77)
(10) Father illiterate (1 = yes; 0 = no)	−0.04	−0.05	−0.02	1.11	0.38	2.30
	(0.02)	(0.02)	(0.03)	(0.77)	(0.99)	(1.39)
(11) Mother illiterate (1 = yes; 0 = no)	0.01	0.01	−0.00	1.05	1.56	0.21
	(0.02)	(0.02)	(0.02)	(0.78)	(1.03)	(1.04)
(12) Ever used a computer (1 = yes; 0 = no)	0.02	0.00	0.08	2.19**	1.69	2.88**
	(0.03)	(0.03)	(0.04)	(0.97)	(1.20)	(1.30)
(13) Access to other modern technologies[c]	0.10**	0.13**	0.04	3.21	5.86**	−0.76
	(0.05)	(0.06)	(0.07)	(2.00)	(2.62)	(3.13)
(14) Family off-farm	0.00	−0.00	0.02	−0.29	−1.49	1.71
	(0.03)	(0.05)	(0.05)	(1.50)	(1.78)	(2.13)
(15) Live with both parents	−0.00	−0.01	0.01	0.56	−0.66	1.91
	(0.01)	(0.02)	(0.03)	(0.80)	(1.07)	(1.10)
(16) County dummies	Y	Y	Y	Y	Y	Y
(17) Grade dummies	Y			Y		
(18) Self-efficacy of Mandarin study-baseline	0.16***	0.18***	0.12***			
	(0.02)	(0.03)	(0.04)			
(19) Like school-baseline				0.09***	0.06	0.12***
				(0.03)	(0.04)	(0.04)
(20) Observations	3164	1723	1441	3164	1723	1441

Robust standard errors clustered at the school level. Each column reports the results of one regression of the student non-academic outcome on the corresponding variables in rows (1)–(19).

** Significant at 5%; *** significant at 1%.

[a] The baseline Mandarin score is the score on the standardised Mandarin test that was given to all students in the sample before the CAL programme.

[b] The baseline math score is the score on the standardised math test that was given to all students in the sample before the CAL programme.

[c] 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.

cohort in the treatment schools had more access to CAL than those of the first cohort (with the phase I programme). Whether they gained improvement in test scores or not, they got to play with the animated game-based CAL programme more often than their counterparts in the first cohort. While this may have made their learning experience more interesting, and thus they liked school better than students of the first cohort, a programme that improved learning interest appears to have not necessarily improved the learning efficacy or test scores of the students. Moreover, the increased interest in learning during the phase II programme also indicated that student burn-out is unlikely to be the reason for the absence of improvement of the phase II programme over the phase I programme.

Conclusions

In this paper we present the results from a randomised field experiment of CAL programmes involving two cohorts of grade 3 students (involving a total of 3701 students) in 57 public schools in three poor rural minority counties in Qinghai province, China. To evaluate and compare the effectiveness of the two programmes we randomly chose 26 schools from the sample to serve as the treatment schools. Grade 3 students attending these schools received the CAL intervention. The first cohort of students in the treatment group received the phase I CAL programme, which was an after-school remedial CAL programme in Mandarin. These students were offered 40 min of shared computer time after school, twice a week. During these sessions students played computer-based games that required them to practice using their Mandarin skills. The CAL programme was tailored to the regular school Mandarin curriculum and was remedial in nature, providing the students with drills and exercises that were related to the material that they were learning in class. There was also an animation-based tutoring session that reviewed the lesson of the week. The second cohort of students received the phase II programme, which, besides the Mandarin CAL sessions also included CAL remedial sessions in math. Therefore, the number of CAL sessions doubled during the phase II programme, and many schools used in-school hours instead of after-school hours for at least some of their CAL sessions. The remaining 31 schools served as control schools, with grade 3 students attending these schools composing the control group.

Our results indicate that the CAL interventions had significant beneficial effects on both student academic and non-academic outcomes. Overall, the CAL programme improved student Mandarin and Math test scores by 0.15 standard deviations. Specifically, for the first cohort, two 40-min phase I CAL Mandarin sessions per week increased student standardised Mandarin scores by 0.19 standard deviations. Moreover, the CAL intervention also had significant impact on the math test scores of the students, with the magnitude of the effect equal to 0.23 standard deviations. The CAL intervention also appeared to have some positive impacts on the non-academic outcomes of the students.

However, the findings from the phase II programme illustrates that policymakers/educators need to pay close attention to the classroom context and the interactions between the CAL intervention and existing classroom resources. When the number of sessions doubled (when we added the math subject into the CAL package), schools were forced to shift some of the CAL sessions to periods of time in which they used regular school hours. When this was done, we found no significant improvement in either student Mandarin or math test scores from the phase II CAL programme over the phase I programme in spite of the fact students seemed to like schooling significantly better during the phase II programme period. One interpretation of our results is that this substitute effect (CAL substituting for regular classroom material) may have at least in part offset the positive effects of the CAL intervention. In addition, it is possible that since the math programme was conveyed in Mandarin, it may have created learning difficulties for many students in our sample for whom Mandarin is a second language. This, in turn, may have further decreased the effectiveness of the programme.

In other words, it seems that the augmented phase II programme was beyond the implementation capacity of the schools, and, as a result, it may be that due to these implementation problems, that the CAL programme did not generate the expected benefits on student academic performance.

Given the significant impact of CAL on student academic outcomes found in this paper, educational policy makers in China (and other countries) should consider scaling CAL programmes, especially in public schools serving disadvantaged students (for example, rural schools in China, particularly those in rural minority areas). However, our findings also highlight the importance of the interactions between the CAL programme and the existing context of the classrooms. The phase I programme shows that the CAL programme in Mandarin had significant effects on student Mandarin performance and these effects even spilled over to student math performance. However, findings from the phase II programme remind us that more CAL is not necessarily better. The effectiveness of CAL depends critically on how well the programme fits into the classroom context and the interaction between the programme and existing classroom resources. CAL programmes that exceed the implementation capacity of the schools might not be beneficial.

Notes

1. There are 592 national designated poverty counties among the more than 2000 county-level jurisdictions in China. The Leading Group of the Alleviation of Poverty gave counties the designation in the 1990s based on the severity of the level of poverty in the county.
2. The number of 57 was determined by our budget for field survey and data collection.
3. Please refer to Lai et al. (2015) for detailed description of the implementation of the phase I programme. The main purpose of this paper is to examine the interactions of classroom inputs and the CAL programme by comparing the phase I and phase II programmes, and thus descriptions of the phase I programme implementation is much simplified.
4. In assessing the results, one may ask if part of the CAL effect was an 'extra teacher effect.' In fact, we initially considered adding an additional treatment arm – as was included in Banerjee et al. (2007). The additional treatment would have provided remedial tutoring in the form of traditional teacher-based after school tutoring sessions. However, when we tried to implement this, we were not allowed to do so by the school district. Many Chinese school districts have strict regulations against the running of traditional teacher-based after school tutoring sessions. In addition, as discussed in the body of the paper, we did not allow teachers to teach Mandarin or interact with students about Mandarin-related questions during the CAL sessions (except that the teachers could answer questions about computer operations from the students). In fact, the scope for interaction was not great since none of the CAL supervisors were Mandarin teachers. There was little ability and no incentive for them to disregard the rules of the programme and teach children additional Mandarin.
5. He, Linden, and MacLeod (2008) also found that an Indian English Education Curriculum implemented through teacher training improved student math and English scores rather than just their English scores. They conjectured that this was due to the fact that when teachers implemented the English language CAL sessions, they were able to teach English more efficiently so that during the normal class period, teachers could devote more time to teaching other topics, such as math. While interesting for the study sites in South Asia, this is unlikely to be the case in our study. In our case, teachers were unlikely to spend less time on Mandarin during school and more time on math as the in-school hours for each subject are already delineated according to a strictly followed class schedule. In addition, it is unlikely that this schedule would change as in almost all Chinese schools Mandarin teachers teach Mandarin and math teachers teach math. Therefore, even if the CAL programme was helping improve the student performance in Mandarin, there would be no incentive for Mandarin teachers to allow math teachers to tutor children during their class period. Because of this, the mechanism that is at work in the He et al. case, almost certainly does not apply.
6. We will further explore the possibility of this explanation in the next section.
7. In our sample 86 per cent of the students are rural minority students whose native language is not Mandarin. Therefore, in theory, all (of most) of these minority students might be more or less affected by the language barrier. In particular, as the Salar, Tu and Tibetan minorities have their own spoken languages, it might be more difficult for students of these ethnic groups to understand the content of math programme in Mandarin (52% of our sample are students of these three ethnic minority groups).
8. We will further explore the possibility of this explanation in the next section.

Acknowledgements

We acknowledge financial support from the National Natural Science Foundation of China (grant numbers 71473240 and 71110107028), the 111 Project (Grant number B16031). We would also like to acknowledge Bowei Lee and his family, Acer Computers, DELL, Ford Foundation, Guo Foundation and Narada Foundation for their generous support for REAP's Technology and Human Capital theme area. We are also grateful to the hard work of dozens of volunteers from the Chinese Academy of Sciences, Northwest University of Xi'an, Qinghai Minorities University made this paper possible.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Dr Fang Lai is a research associate at the IMPAQ International. She formerly worked as a research associate at Stanford University and assistant professor at the School of Education in New York University. She obtained her PhD in Agricultural and Resource Economics from University of California at Berkeley. Her work mainly focuses on evaluation of education programs, assessment of teacher and school effectiveness, school choice and educational reform, equity of education, social interaction and peer influence. She has authored and co-authored multiple papers in peer-reviewed academic journal in these fields.

Dr Linxiu Zhang is a professor and deputy director at the Center for Chinese Agricultural Policy (CCAP), Chinese Academy of Sciences (CAS). She is the co-director of a collaborative program called Rural Education Action Program (REAP). She obtained her PhD from Reading University. Her research concentrates on policy relevant studies on rural development in China, particularly on poverty alleviation, labor market development, public investments, and the economics of rural education and healthcare. She has published more than 200 papers in peer-reviewed journals and also received numerous awards. The most recent ones include "Ten Most Outstanding Women in Science" award from CAS (2013), TWAS "Celso-Furtado Prize" in Social Sciences (2013), "Fudan Management Excellence Award" (2014) and a TWAS fellow (2014).

Dr Yu Bai is an assistant professor at the Center for Experimental Economics in Education (CEEE), Shaanxi Normal University in China. He obtained his PhD from the Center for Chinese Agricultural Policy (CCAP), Chinese Academy of Sciences (CAS). His research focuses on rural education in China, particularly on migration, technology and human capital, as well as healthcare.

Dr Chengfang Liu is an associate professor with tenure at the China Center for Agricultural Policy (CCAP), School of Advanced Agricultural Science (SAAS), Peking University (PKU). She obtained her PhD in agricultural and resource economics from the University of California (Davis). For more than 15 years, her research has been in the field of agricultural and rural development, with focus on the provision of rural infrastructure, human capital, and migration. Her work is highly empirical, drawing on a large number of data sets that she and her coauthoring teams have collected in rural China. She has published 80 papers both domestically and internationally, of which 40 in SCI/SSCI indexed journals. Chengfang is currently working on how to carry out rigorous impact assessments in agricultural and rural development policies. She is the principal investigator of Rural Education Action Program (REAP).

Dr Yaojiang Shi is a professor and the founding director of the Center for Experimental Economics in Education (CEEE), Shaanxi Normal University, China. His research focuses on the emergence and development of economic enterprises in rural China and China's education reforms. He uses empirical research to identify important leverage points for education policy that addressed the needs of the rural poor. He has published more than 80 papers in peer-reviewed journals. He and his team has produced 25 policy reports and sent directly to national and provincial government leaders for review.

Dr Fang Chang is an assistant professor at the Center for Experimental Economics in Education (CEEE), Shaanxi Normal University in China. Her research focuses on rural development as well as rural education. She obtained her PhD in public economics from China's Northwestern University. She has published more than 10 papers in peer-reviewed journals. Recently, she has been working on teaching methods and teacher performance payment at the compulsory level in rural China.

Dr Scott Rozelle is the Helen Farnsworth Senior Fellow in the Food Security and the Environment Program of Stanford University's Freeman Spogli Institute for International Studies. He is the co-director of Rural Education Action Program (REAP). Rozelle's research focuses almost exclusively on China and is concerned with three general themes; a) agricultural policy, including the supply, demand, and trade in agricultural commodities, b) the rural environment; and c) issues of poverty alleviation with a focus on rural education. He is the chair of the Board of Academic Advisors of the Center for Chinese Agricultural Policy. He has won numerous awards including International Collaboration Award of Chinese Academy of Science, China International Science and Technology Cooperation Award, Friendship Award of the People's Republic of China.

References

Angrist, J., and V. Lavy. 2002. "New Evidence on Classroom Computers and Pupil Learning." *The Economic Journal* 112 (482): 735–765.

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 2008. "Technology's Edge: The Educational Benefits of Computer-Aided Instruction." National Bureau of Economic Research working paper No. 14240.

Dynarski, M., R. Agodini, S. Heaviside, T. Novak, N. Carey, L. Campuzano, B. Means, R. Murphy, W. Penuel, H. Javitz, D. Emery, and W. Sussex. 2007. *Effectiveness of Reading and Mathematics Software Products: Findings from the 1st Student Cohort*. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences.

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.

Glewwe, P., and M. Kremer. 2006. "Schools, Teachers, and Education Outcomes in Developing Countries." Chapter 16 in *Handbook of the Economics of Education*, Vol. 2, edited by E. Hanushek and F. Welch, 945–1017. New York, NY: Elsevier.

Goolsbee, A., and J. Guryan. 2006. "The Impact of Internet Subsidies in Public Schools." *The Review of Economics and Statistics*, MIT Press 88 (2): 336–347. doi:10.1162/rest.88.2.336

Hanushek, E. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24: 1141–1177.

Hanushek, E. A. 1995. "Interpreting Recent Research on Schooling in Developing Countries." *The World Bank Research Observer* 10: 227–246. doi:10.1093/wbro/10.2.227.

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*. New York, NY: Columbia University. Mimeographed document.

Inal, Y., and K. Cagiltay. 2007. "Flow Experiences of Children in an Interactive Social Game Environment." *British Journal of Educational Technology* 38 (3): 455–464. doi:10.1111/bjet.2007.38.issue-3.

Jacob, B., and L. Lefgren. 2004. "The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago." *The Journal of Human Resources* 39 (1): 50–79. doi:10.2307/3559005.

Kirkpatrick, H., and L. Cuban. 1998. "Computers Make Kids Smarter – Right?." *Technos Quarterly for Education and Technology* 7 (2): 26–31.

Lai, F., R. Luo, L. Zhang, X. Huang, and S. Rozelle. 2013a. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing." REAP Working paper.

Lai, F., L. Zhang, X. Hu, Q. Qu, Y. Shi, Y. Qiao, M. Boswell, and S. Rozelle. 2013b. "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. 2015. "Teaching the Language of Wider Communication, Minority Students, and Overall Educational Performance: Evidence from a Randomized Experiment in Qinghai Province, China." *Economic Development and Cultural Change* 63 (issue 4): 753–776. doi:10.1086/681233.

Linden, L. 2008. "Complement or Substitute? The Effect of Technology on Student Achievement in India." Working Paper. Columbia University Department of Economics.

McEwan, P. 2014. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Review of Educational Research*. doi:10.3102100346543l4553127. October 7, 2014.

Mo, D. 2015. "The Impact of Government Upscaling Efforts on the Effectiveness of Computer Assisted Learning Programs in Rural China." Working paper. Stanford, CA: REAP, Stanford University.

Mo, D., J. Swinnen, L. Zhang, H. Yi, Q. Qu, M. Boswell, and S. Rozelle. 2013. "Can One Laptop per Child Reduce the Digital Divide and Educational Gap? Evidence from a Randomized Experiment in Migrant Schools in Beijing." *World Development* 46: 14–29. doi:10.1016/j.worlddev.2012.12.019.

Mo, D., L. Zhang, R. Luo, Q. Qu, W. Huang, J. Wang, Y. Qiao, M. Boswell, and S. Rozelle. 2014. "Integrating Computer-Assisted Learning into a Regular Curriculum: Evidence from a Randomised Experiment in Rural Schools in Shaanxi." *Journal of Development Effectiveness* 6 (3): 300–323. doi:10.1080/19439342.2014.911770.

Planty, M., W. Hussar, T. Snyder, S. Provasnik, G. Kena, R. Dinkes, A. KewalRamani, and J. Kemp. 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., and A. 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.

Schaefer, S., and J. Warren. 2004. "Teaching Computer Game Design and Construction." *Computer Aided Design* 36: 1501–1510. doi:10.1016/j.cad.2003.10.006.

Yang, Y., H. Wang, L. Zhang, S. Sylvia, R. Luo, Y. Shi, W. Wang, and S. Rozelle.2013. 'The Han-Minority Achievement Gap, Language and Returns to Schools in Rural China.' REAP working paper #258.

Appendix. Comparison of student and family characteristics and student access to computer and other modern technologies between the treatment and control groups for all students (including attrited students)

Dependent variable: the treatment dummy variable	(1) First cohort	(2) Second cohort
(1) Baseline Mandarin score (units of standard deviation)[a]	0.00	0.00
	(0.16)	(0.16)
(2) Baseline math score (units of standard deviation)[b]	−0.10	−0.04
	(0.15)	(0.15)
(3) Female (1 = yes; 0 = no)	−0.03	−0.00
	(0.03)	(0.03)
(4) Tibetan	0.03	0.01
	(0.07)	(0.06)
(5) Tu	−0.00	0.04
	(0.09)	(0.09)
(6) Hui	−0.02	−0.03
	(0.12)	(0.13)
(7) Salar	0.05	−0.05
	(0.11)	(0.04)
(8) Only child (1 = yes; 0 = no)	−0.01	−0.01
	(0.02)	(0.02)
(9) Father illiterate (1 = yes; 0 = no)	−0.02	0.02
	(0.03)	(0.04)
(10) Mother illiterate (1 = yes; 0 = no)	0.03	−0.00
	(0.04)	(0.04)
(11) Ever used a computer (1 = yes; 0 = no)	−0.05	−0.02
	(0.07)	(0.04)
(12) Access to other modern technologies[c]	0.01	−0.02
	(0.03)	(0.03)
(13) Family off-farm	−0.00	0.00
	(0.02)	(0.02)
(14) Live with both parents	−0.03	0.00
	(0.04)	(0.05)
(15) Observations	1981	1720

Robust standard errors clustered at the school level are shown in parentheses.
*** $p < 0.01$, ** $p < 0.05$.
[a] The baseline Mandarin score is the score on the standardised Mandarin test that was given to all students in the sample before the CAL programme.
[b] The baseline math score is the score on the standardised math test that was given to all students in the sample before the CAL programme.
[c] 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.
The difference between the treatment and control groups conditional on the other variables is calculated by regressing each of the row variables on the treatment dummy variable.