Di Mo
Yu Bai
Matthew Boswell
Scott Rozelle

Evaluating the effectiveness of computers as tutors in China

July 2016

Impact Evaluation Report 41

Education

International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding, producing and synthesising high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie impact evaluations

3ie-supported impact evaluations assess the difference a development intervention has made to social and economic outcomes. 3ie is committed to funding rigorous evaluations that include a theory-based design, use the most appropriate mix of methods to capture outcomes and are useful in complex development contexts.

About this report

3ie accepted the final version of this report, *Evaluating the effectiveness of computers as tutors in China,* as partial fulfilment of requirements under grant PW2.06 issued under Policy Window 2. The content has been copy-edited and formatted for publication by 3ie. Due to unavoidable constraints at the time of publication, a few of the tables or figures may be less than optimal. All of the content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are also the sole responsibility of the authors. Comments or queries should be directed to the corresponding author, Matthew Boswell at boswell@stanford.edu

Funding for this impact evaluation was provided by 3ie's donors, which include UK aid, the Bill & Melinda Gates Foundation and the Hewlett Foundation. A complete listing of all of 3ie's donors is available on the 3ie website.

Suggested citation: Mo, D, Bai, Y, Boswell, M and Rozelle, S, 2016, *Evaluating the effectiveness of computers as tutors in China, 3ie Impact Evaluation Report 41*. New Delhi: International Initiative for Impact Evaluation (3ie)

3ie Impact Evaluation Report Series executive editors: Jyotsna Puri and Beryl Leach
Managing editor: Deepthy Menon
Assistant managing editor: Radhika Menon and Bidisha Barooah
Production manager: Pradeep Singh
Copy editor: Shreya Ray
Proofreader: James Middleton
Cover design: John F McGill
Printer: Via Interactive
Cover photo: Stanford University/Rural Education Action Program

Evaluating the effectiveness of computers as tutors in China

Di Mo
Freeman Spogli Institute for International Studies, Stanford University

Yu Bai
Center for Chinese Agricultural Policy, Chinese Academy of Sciences

Matthew Boswell
Rural Education Action Program, Stanford University

Scott Rozelle
Rural Education Action Program, Stanford University

Impact Evaluation Report 41

July 2016

International Initiative for Impact Evaluation

Summary

It has been well documented that links exist between school inputs and academic performance (Banerjee et al., 2003; Glewwe, 2002; Hanushek, 1997, 2002). Integrating Information and Communication Technology (ICT) is one such educational input that is considered to be a promising approach to help disadvantaged students across the world (Ebner and Holzinger, 2007; Bakar et al., 2006; Pal, 2006; Banerjee et al., 2007; Linden, 2008; Lai et al., 2011; Mo et al., 2014a). However, despite the popularity of the use of ICTs in education, researchers have found considerable heterogeneity in the impact of ICT programs on student academic achievement. Two important facets of ICT programs that may affect program effectiveness are which organization implements the program, as well as whether the program is integrated into teaching (computer-assisted instruction, CAI) or not (computer-assisted learning, CAL).

The overall goal of this paper is twofold: (1) to evaluate the impact of a government-implemented CAL program; and (2) to understand whether a CAI program is effective in raising learning outcomes and to compare the relative effectiveness of a CAI program against a CAL program. To meet the first goal, we will compare the government's CAL program implementation to a simultaneous CAL program implemented by a team of researchers. To meet the second goal, we have three specific objectives. First, we measure the impact of a CAI program on student academic performance. Second, we compare the CAI program to a CAL program to investigate the relative effectiveness of CAI. Third, we try to explore the mechanisms that allow CAI programs to be more or less effective than CAL programs.

In examining our first research question, we found that students whose CAL treatment was implemented by the research group improved significantly in their English test scores than students in the government-implemented control group. We investigated potential causes for these differential impacts and our analysis suggests that at the end of the teacher training, teachers who attended the training organized by the research group did not differ in knowledge in software and protocol from the teachers whose training was organized by the government. Nor was there evidence showing that the frequency of the CAL classes differed between the two experimental groups. However, our results showed that when the government implemented the CAL program CAL classes were more likely to replace regular English classes than when the research team implemented the program.

The analysis conducted on our second research question found that although the average impact of any ICT program on student English test scores was insignificant, when the two treatment groups (CAI and CAL) were separated, the CAI program proved to be more effective than the CAL program at raising students' English test scores. Our investigations into whether there were heterogeneous effects depending on students' initial achievement levels and teachers' workload found differential impacts among CAI and CAL programming. First, we found that the lower a student's initial English performance is, the less he or she benefited from the CAL program. In contrast, when the program was integrated into English instruction, it benefited students who performed well as well as poorly. The program supervisor's regular workload was also found to not matter for the CAI program. However, workload mattered for those CAL supervisors who were non-English teachers. When the workload of a non-English teacher was higher than the mean, the program impact was found to be dramatically reduced.

Contents

List of figures and tables

Abbreviations and acronyms

CAI	computer-assisted instruction
CAL	computer-assisted learning
CEEE	Center for Experimental Economics in Education
GDP	Gross Domestic Product
ICT	information and communication technology
OLS	ordinary least squares
RCT	randomized controlled trial
TESOL	Teaching English to Speakers of Other Languages

1. Introduction

There are many examples of studies that establish links between school inputs and academic performance, including experimental studies (Banerjee *et al*., 2003; Glewwe, 2002) and non-experimental ones (Hanushek, 1997, 2002). Among these are a significant number of studies in developing countries that seek to determine whether inputs can reach their intended aim in contexts where resources are limited (Banerjee *et al*., 2003). Integrating Information and Communication Technology (ICT) into education promises to be one such approach to help disadvantaged students across the world (Ebner and Holzinger, 2007; Bakar *et al*., 2006; Pal, 2006; Banerjee *et al*., 2007; Linden, 2008; Lai *et al*., 2011; Mo *et al*., 2014a).

Policymakers in developed and developing countries have tried various ways to use ICT for innovation and improvement in education (Eurydice, 2001; Papanastasiou and Angeli, 2008). In many parts of the world, large investments have been made to integrate ICT into the education system (Trucano, 2005; World Bank, 2015). For instance, Turkey spends 11.7 per cent of its education budget on ICT (Hismanoglu, 2012). China's government also plans to increase investment in ICT in education, with special attention to remote rural areas (Ministry of Education, 2012a, 2012b, 2014). The Ministry of Education established an expert committee to implement the plan. According to the plan, a computer classroom is to be set up in every rural primary school during the 12th Five-Year Plan (Ministry of Education, 2012c).

Despite the popularity of the use of ICTs in education, researchers have found considerable heterogeneity in the impact of ICT programs on student academic achievement. Previous studies in developed and developing countries have shown inconsistent evidence of the educational benefits of integrating ICT into education (e.g. Banerjee *et al*., 2007; Barrow *et al*., 2008; Almekhlafi, 2006; Olibie, 2010; Lai *et al*., 2012; Lai *et al*., 2013; Mo *et al*., 2014b; Angrist and Lavy, 2002; Rouse and Krueger, 2004; Hlas and Vuksanovich, 2007; Dynarski *et al*., 2007; Chien, 2011). For instance, Banerjee *et al*. (2007) have shown that a computer-assisted learning (CAL) program significantly helped students increase math scores. The gains from the program persisted over time, as well, when they were measured one year after the program concluded. A number of other studies of CAL programs in developing countries have found positive impacts (He *et al*., 2008; Mo *et al*., 2014b). In contrast, Angrist and Lavy (2002) found that integrating ICT into Israeli elementary schools led to a mostly negative but insignificant impact on eighth grade students. Rouse and Krueger (2004) have shown that an ICT program had no significant effect on student reading in US schools.

Two important facets of ICT programs and other forms of educational inputs that may affect program effectiveness are: which organization implements the program; and whether programs involve stand-alone learning software (CAL) or the active involvement of teachers (computer-assisted instruction, CAI). Despite the importance of these questions, there is little rigorous empirical evidence to better our understanding of how to best implement ICT-based educational inputs (Duflo, Glennerster and Kremer, 2007). The only known study to have compared a government-implemented program with an NGO-implemented program is an experiment that involved contract teachers in Kenya (Bold *et al*., 2013).

Many studies that look at ICT programs in education do not distinguish between programs that integrate the technology into teaching and those that do not (Angrist, 2002; Linden, 2008). Among the studies that explicitly test the impact of CAI, many are small in scale. For instance, a study in Nigeria only involved a total of 160 students in four classes (Olibie, 2010). In addition, these studies lacked a valid control group or an appropriate identification strategy to identify the impact of CAI or CAL programs (e.g. Rahimi and Yadollahi, 2011; Almekhlafi, 2006; Hughes, 2005). Importantly, to our knowledge, almost no study has compared the relative effectiveness of CAI and CAL programs. Few studies have sought to understand the mechanism for why integration of ICT into teaching either works or does not.

Whether ICT should be integrated into teaching practices has been debated in the literature at some length (Mumtaz, 2000; Granger *et al.*, 2002; Stockwell, 2007; Tondeur *et al.*, 2008; Bingimlas, 2009). On the one hand, researchers have argued that ICT's value in education is to provide a means to make teaching more efficient and interesting (Hismanoglu, 2012; Dina and Ciornei, 2013; Hughes, 2005). Dina and Ciornei (2013) and Hismanoglu (2012) suggest that ICT cannot replace the role of teachers in providing timely feedback to their students. Therefore, it is not only desirable but also necessary for ICT to be integrated into the teaching process. On the other hand, Rahimi and Yadollahi (2011) have suggested that having teachers adapt to the use of ICT may just induce the replacement of traditional teaching with ICT materials. The problem is, that in order to avoid having to spend the effort to fully incorporate ICT into teaching, teachers may simply choose to display ICT materials without adequate adjustment and selection to match student learning levels and pace. In this case, student performance may even be harmed.

Another important point the literature makes is that the effectiveness of different ICT programs may depend in part on the learning level of students at baseline. For example, Huang *et al*. (2014), show that the effectiveness of ICT programs can depend on the initial learning level of students. According to Krashen (1982), calibrating the additional amount of new learning material relative to a student's current level of knowledge is a crucial step to better learning. In other words, the provision of new learning materials is most effective when it is provided at a level just beyond the student's current level of competence. As a result, ICT-based instruction needs to be readily comprehensible to students for there to be effective learning (Hatch, 1978; Long, 1996). Such a theory implies that students should possess a minimum level of knowledge of the subject before ICT can benefit them. In other words, ICT is not a technology that can be effectively used by a student to learn a subject that is totally new to him or her.

In addition, the other factor determining the program's effectiveness is the teacher implementing it. Burston (1996), Jones (2001) and Yang and Huang (2008) show that implementing an ICT program may increase teachers' workload, since they need to make efforts to prepare for and organize the new program. Therefore, a teacher's workload at baseline could affect the effectiveness of an ICT program. If a teacher has a heavy workload, he or she may not be willing to carefully or fully implement an ICT program, whereas a teacher with a lighter workload would be more inclined to implement the program. However, researchers have also suggested that if an ICT program is effective in helping students learn and complements regular teaching, it may require less marginal effort from the teachers to improve student performance (Lam, 2000; Nomass, 2013). In other words, overall teacher effort may decrease if the ICT program can help improve student performance, thus improving the overall efficiency of teaching. As a result, determining which type of ICT programs works in a particular context may depend on the nature of the teacher and how well the ICT program complements teaching practices.

Drawing on the relevant literature, there are two channels that may make the impact of a CAI program different from a CAL program. First, while a CAI program may work for everyone, CAL may not work if students' knowledge level is too low. If the learning level of an individual student is much lower than the level of learning materials that the ICT program provides, he or she may not be able to benefit from the CAL program. Because CAI is rolled out as part of a teaching program, the English teacher could help poorly performing students catch up with the subject so that they may make better use of the ICT program's materials. In contrast, because CAL requires less engagement on the part of instructors (beyond the management of the ICT program), poorly performing students might not be able to keep up with the material being presented by the ICT program. Second, CAI may be more effective if the CAI program and regular teaching complement each other. For example, the English teaching and learning in an ICT program can complement the regular teaching in the English curriculum. Therefore, the required marginal effort to improve students' English level is lower for the English teacher who is running a CAI program than for a non-English teacher who is running a CAL program in English. As a result, we may observe that non-English teachers with higher workloads may be less likely to exert effort to implement the CAL program than the English teachers with higher workloads who implement a CAI program.

A further question exists with regard to who (or which entity) implements ICT programs. How the program implementer affects the effectiveness of a program may be closely related to the incentives of the implementer (Heckman 1991). If staff in the educational system receive little performance incentive, they may not be motivated to exert effort (Dizon-Ross Dupas Robinson, 2016). There may also be a lack of accountability of teachers, principals and government officials within the educational system. In contrast, NGOs and research teams may be more motivated and engaged in program implementation and follow up progress more closely.

Such differences in effort between government and non-government entities may cause the difference in the arrangement of school inputs when a new program is implemented (Allcott and Mullainathan, 2012). Schools or teachers may re-optimize the investment of school inputs in response to a new program (Angrist and Lavy, 2002). Under weak incentives, it may be more likely that government staff or schools do not invest in enough school inputs to implement a new program or cut back school inputs for other educational activities in response to the new program, as compared with the case of an NGO or a research team is the implementer. For example, schools or teachers may react to a new educational program by substituting school inputs that were used in other educational activities (Rouse and Krueger, 2004). The substitution may attenuate the impact of the program if it undermines other productive educational activities.

The overall goal of this paper is twofold: (1) to evaluate the impact of a government-implemented CAL program by comparing it to a CAL program implemented by a research team; and (2) to understand whether a CAI program is effective in raising learning outcomes and to compare the relative effectiveness of a CAI program against a CAL program. To meet the first goal, we compare the government's CAL program implementation to a simultaneous CAL program implemented by a team of researchers. To meet the second goal, we have three specific objectives. First, we measure the impact of a CAI program on student academic performance. Second, we compare the CAI program to a CAL program to investigate the relative effectiveness of CAI. Third, we try to explore the mechanisms that allow CAI programs to be more or less effective than CAL programs.

Both the CAI and CAL interventions discussed in this report use English language learning as the subject of focus. There are several reasons for this choice. First of all, English is an important second language to learn because income levels are correlated with English levels in developed and developing countries (Munshi and Rosenzweig, 2006; He *et al*., 2008). Second, English is one of the main subjects that is tested in high school and college entrance exams in China (McKay, 2002; Bolton, 2012) and represents one-third of the total points on each exam. Third, English teaching and learning is particularly weak in poor rural China (Li, 2002; Zhao, 2003; Hu, 2005). Studies have shown that in China, poor English skills is one of the biggest factors preventing rural students from attending high school (Loyalka, 2014). However, no rigorous studies have tested whether integrating ICT into English education is effective in improving English learning in rural China.

The rest of the paper is organized as follows. The next section presents the sampling, data and methods of the two studies. The third section discusses the results of each study. We draw conclusions in the final section.

2. Policy context

In its 12th Five-Year Plan, China's central government earmarked the funds to place a computer room in every rural school. This measure appears to have forestalled the consequences of a sustained gap in information and computing technologies across urban and rural areas, and to help raise academic outcomes among underperforming rural students. However, it provides no guidance on how new computing facilities should be used. This, in part, is a function of China's decentralized approach to education where regional and local governments are commonly encouraged to experiment and identify ways to leverage central investments that target policy goals.

Nevertheless, provincial authorities charged with expending the central funds have questions about how best to leverage such an investment. Success in China's schools depends on excelling in a rigorous test-based curriculum. Relative to their urban counterparts, rural students are at a disadvantage in this system, a truth borne out by the under-representation of rural youth in China's upper secondary and tertiary schooling cohorts. Local authorities have introduced computer rooms sporadically in rural schools, but evidence suggests that such additions have not narrowed this disparity in access. In fact, casual observation suggests much of the investment is being wasted on ineffective projects.

The Center for Experimental Economics in Education (CEEE) conducted three large-scale randomized control trials (RCTs) to evaluate a game-based computer remedial tutoring program designed to remediate learning in two core subjects of the national curriculum—mathematics and Chinese. The program, a CAL, is also designed to be managed by teachers with no ICT experience. The results of the efficacy trials indicate large impacts on educational performance and non-cognitive outcomes among intervention students (Lai *et al*, 2013).

During the qualitative phase of the previous evaluations, educators—from teachers to principals and district superintendents—commonly reported that CAL made their schools more exciting places to be. In particular, educators from schools that already had computers were enthusiastic about CAL because it provided a simple means to integrate their machines into the school day. Although happy to have computer rooms, educators admitted to not always knowing what to do with them because their machines typically came without educational software or a curriculum, and few staff knew how to use them. It is unclear to what extent students even used the computers. On account of this enthusiasm, a consortium of prefectural leaders has succeeded in convincing provincial authorities to consider CAL as a component of the computing facilities roll-out.

The Shaanxi Policy Implementation and Analysis Office in the Shaanxi Province of western China has implemented the CAL program as an RCT, with the assistance of the evaluation team from CEEE. The evaluation is designed to determine whether a computer room with a CAL program can improve student academics when compared to computer rooms without CAL. The results of the evaluation will inform the government's decision to integrate CAL into the mandated expansion of rural computer classrooms. Shaanxi is a province of more than 60 million people and is considered an innovator among provinces in western China.

3. Sampling, data and methods

3.1 Sampling and the process of randomization

We conducted a clustered RCT of CAL and CAI in rural schools during the 2013/14 academic year. A total of 120 primary schools in poor minority areas in China's Qinghai Province are included in our study. A total of 5,574 fourth-grade students participated in the experiment.

We adopted several steps to choose our sample. First, to focus our study on students from poor rural areas, we restricted our sample frame to Haidong Prefecture, a poor minority area in Qinghai Province in northwest China. Among 31 provinces in mainland China, Qinghai ranked 30th in terms of total GDP in 2013 (National Bureau of Statistics of China, 2014). The annual per capita net income of the selected prefecture was only around RMB6,150 (approximately USD990) (National Bureau of Statistics of China, 2014). Second, all six counties in the prefecture were selected to be included in our sampling frame. Five of the counties are nationally designated poor counties (ibid.).

Third, after choosing the counties, we identified 127 schools with sufficient numbers of computers. To do this, we obtained a comprehensive list of schools in the six counties from the local education bureau. In order to run the CAI or CAL programs in the schools, we restricted our sample to schools that met the minimum requirement of computer facilities. The minimum requirement is a ratio of computers to average class size of at least 0.25. This ratio ensures that, if two students share one computer, the school has enough computers to allow a class of students to break up into two groups and take turns to participate in the program. We found a total of 130 schools that met the requirement. We dropped three schools that did not have English teachers. At the end, we had a total of 127 schools in our sample.

Our sample for the CAI intervention was composed of fifth-grade students. According to the national curriculum, primary students start learning English in third grade (Ministry of

Education, 2001). However, due to a lack of English teachers, many schools in minority areas do not start teaching English until the fourth or fifth grades (Zheng, 2009). Therefore, the fifth grade was chosen to make sure that all students had started English learning and the schools had English teachers that could participate in the CAI treatment.

We randomly divided the 127 sample schools into two treatment groups and one control group. We calculated that we required 22 schools each for two intervention arms and 80 schools for the control arm to detect a standardized effect size for the outcome variable of 0.20 with 0.80 power at the five per cent significance level. We assumed an intra-cluster correlation (ICC) of 0.1, a pre- and post-intervention correlation of 0.5. We also assumed that there were 50 observations in each cluster on average. Since we were interested in the contrast between each of the treatment groups and the control group, the sample size of the control group should be larger than the sample size of the each of the treatment groups (Duflo *et al*., 2007). Therefore, we included all the rest of the 83 schools from the sampling frame to be our control group.

Based on power calculations, we randomly chose 22 schools to receive the CAI intervention (i.e. the program sessions were instructed by an English teacher). We randomly chose another 22 schools to receive the CAL treatment (i.e. the program sessions were supervised by a non-English teacher). The final 83 schools were assigned to the control group. There were a total of 1,234 (16.9 per cent) students in the 22 CAI treatment schools, 1,066 (16.93 per cent) students in the 22 CAL treatment schools, and 3,996 (63.47 per cent) students in the control group.

We used a set of student characteristics to check the validity of the random assignment (Tables 1 and 8). When doing so, we regressed each variable from the set of student characteristics (see below for more details on the definition of the variables) on a dummy for treatment group 1 and a dummy for treatment group 2. We found that the pair-wise differences among the three groups (i.e. treatment group 1 and the control group, treatment group 2 and the control group, and treatment group 1 and treatment group 2) were all statistically insignificant for all the student characteristics (rows 1 to 11, columns 1 and 3).

Although at the time of the baseline survey the sample included 5,574 students, there was 6 per cent attrition by the end of the study. For various reasons (mainly school transfers and absences due to illness or injury) at the time of the evaluation survey, we were able to follow up with 5,253 students (Figure 1). Of the 1,818 treatment group 1 students, 1,694 remained in our sample at final evaluation. Of the 1,933 treatment group 2 students, 1,839 remained in our sample at final evaluation. Of 1,823 control group students, 1,720 remained in our sample at final evaluation.

To assess whether the attrition affected the validity of the randomization, we conducted two tests. We first regressed a dummy for attrition on a dummy for treatment group 1 and a dummy for treatment group 2 (Table 2 and Table 9). This was to test whether the attrition rates differed by the treatment and control groups. We found that the differences in attrition rates between treatment group 1 and the control group, and between treatment group 2 and the control group were small and statistically insignificant (rows 1 and 2, column 1). The F-test also showed that treatment groups 1 and 2 did not differ in attrition rates (p-value=0.32). In other words, the intervention did not seem to affect the likelihood of attrition. In the second test, we included only the sample students after attrition and regressed each variable from a set of student characteristics on a dummy for treatment group 1 and a dummy for treatment group 2 (Tables 3 and 10). The results showed that none of the coefficients of the treatment group dummies were statistically significant (rows 1 to 11, columns 1 and 3), indicating there was no evidence that the attrition affected the validity of our randomization.

Finally, because we are interested in the impact of the English CAL software on the earliest English learners at school, we included grade four students in the experiment. During the canvass survey, we asked English teachers about the progress of the English classes. The national curriculum for English starts in the third grade. However, due to a lack of English teachers in rural schools, schools vary substantially in the pace of English teaching in third grade. Therefore, we decided to include fourth-grade students whose English classes follow the national curriculum and therefore are more similarly paced. In total, there were 5,574 fourth-grade students in the initial sample (see Figure 1).

After identifying the 127 sample schools, we randomly chose 40 schools to receive the CAL intervention implemented by our research team (treatment group 1), 40 schools to receive the CAL intervention implemented by the government (treatment group 2), and 40 schools to have no intervention (the control group). A total of 1,818 fourth-grade students were included in treatment group 1 (Figure 1). Treatment group 2 included 1,933 fourth-grade students. The 1,823 fourth-grade students in the rest of the 40 schools served as the control group while the CAL intervention group, included treatment groups 1 (research team) and 2 (government).

The core part of the intervention in treatment group 1 and treatment group 2 was a computer-assisted English remedial tutoring program. To standardize the program implementation, we designed a detailed CAL curriculum and implementation protocol. By following the detailed instructions in the protocol, the assigned teacher-supervisors could easily organize and manage the CAL classes at school.

The difference between the two treatment groups was the implementer. The implementer of the CAL program in treatment group 1 was our research team, while the implementer in treatment group 2 was the county education bureau. Although all the training materials and handouts for teachers, including the software and protocol, were the same for the two groups, the research team and the government implemented the CAL program independently.

In order to introduce the different parts of the intervention, we divide the rest of the subsection into three parts. The first part introduces the core part of the intervention, the CAL program. The second part describes how the CAL program was implemented among the two treatment groups. The third part describes the control group.

3.1.1 The CAL program

The set of software for our program was designed for improving students' basic competences in the uniform national English curriculum. The first piece of the software was obtained from a commercial IT company. This software provided animated reviews and game-based remedial exercises in English for fifth-grade students. The content covered all skills and knowledge in the English national curriculum. Our program team developed another piece of the software package. It provided a large number of additional exercise questions. We worked with teachers and experts from the organization Teaching English to Speakers of Other Languages (TESOL) to choose the questions. We integrated the questions into an animation-based game interface to make it interesting for the students.

Under the supervision of a local teacher-supervisor, trained by either our research group or the government, the students in the treatment groups were supposed to have two 40-minute CAL sessions per week during the computer classes. The sessions were mandatory. The content of each session emphasized core competencies in the uniform national English-language curriculum. During each session, two students shared one computer and played educational games matched to the material covered that week in regular English classes. In a typical session the students first watched an animated video that reviewed the material and then played games to practice the relevant skills, such as vocabulary, grammar and reading comprehension. According to our protocol, the teacher-supervisors needed only to help students with scheduling, computer hardware issues and software operations.

One of the most important jobs of the teacher-supervisor, was to make sure the CAL sessions were proceeding at a pace that matched the regular English classes. Each teacher-supervisor was given a manual that contained detailed instructions for implementation protocol. The manual contained four main sections: (a) the detailed CAL curriculum; (b) CAL classroom rules for students and teacher-supervisors; (c) the responsibilities of the teacher-supervisors when supervising the CAL sessions; and (d) tutorials (in words and graphic illustration) on basic computer operations, CAL software use and troubleshooting. We took care in presenting the protocol in a way that was accessible to teachers without high levels of education or experience with computer use. Any of them would be equipped to manage the CAL program and conduct the classes.

The teacher-supervisor did not need to teach during the CAL classes. Because organizing the CAL classes required no knowledge of the English language, the teacher-supervisor did not need to be an English teacher. The protocol actually required a non-English teacher to be the teacher-supervisor. We also suggested that since the CAL classes were implemented during computer class time,[1] the computer teacher should conduct the CAL sessions. The protocol also set the rules that during the CAL sessions, if a student had an English-related question, he or she should be encouraged to discuss it with their teammate rather than the teacher-supervisor. We compensated each teacher-supervisor with an allowance of RMB500 (USD82) every semester.

Our protocol encouraged schools to open the computer room during lunch breaks or other free class slots for interested students to use the CAL software under the supervision of teachers (henceforth, these sessions are called open voluntary CAL sessions). It was not mandatory for schools to conduct open voluntary CAL sessions.

---

[1] There is in general a shortage of curricula to use during computer class, particularly in poor rural areas (Yang *et al*. 2013). Although the education bureau requires that in these rural public school computer classes need to be scheduled weekly, we found no school in the baseline that had employed computers and/or educational software for instructional purposes in core academic subjects. In the computer classes, teachers typically teach students basic computer operations (sometimes even without practicing on the computer), such as using the mouse and keyboard, typing, web browsing, file and folder operations, etc.) At the endline of the experiment, no control group school was found to have started using any educational software.

3.1.2 Implementation of the CAL program

In order to implement the program in the treatment group 1 schools, the research team organized a half-day training about the CAL software and protocol for teacher-supervisors. During the training, we showed video tutorials that introduced the CAL program, the content and operation of the software and the implementation protocol. These videos were produced in a way that allowed for standardized training in the two treatment groups. To do that, we carefully covered everything that teachers needed to know to organize the CAL sessions. For example, we recorded in a step-by-step fashion how each type of game should be played, what problems students might have and what the solutions were. After the video demonstration, the teachers attended a hands-on session to practice with the software.[2]

We followed up on the progress of the program implementation throughout the academic year. We sent volunteers to help with software installation soon after the training was done. We made regular phone calls every two months to the teachers to make sure that they followed the protocol. We also had a hotline for teachers to contact us with problems.

Before the county education bureau organized the teacher training among the treatment group 2 schools, the higher level of government, the prefecture education bureau (PEB) organized a government training ("training of the trainers"). In this, they sent official documents to county-level education bureaus explaining the program and stating the time and location of the training. Second, each county education bureau assigned one official to be the director of the program and participate in the government training. During the training, the director of the PEB introduced the main goals of the program. The research team showed the video tutorials and demonstrated how to use them for teacher training. We also carefully explained the implementation protocol and the materials that were needed for the training and the CAL program. We required the officials to organize hands-on sessions for the teachers to practice and ask questions. All the materials were distributed to the officials at the end of the training. The PEB gave a closing speech about the next steps in implementing the program, including organizing teacher training, software installation and monitoring progress.

After the government training, each county education bureau organized a teacher training within the county. Only the schools in treatment group 2 were asked to send teachers to the training. During the training the officials demonstrated how to organize the CAL classes and had teachers practice using the software.

3.1.3 CAI intervention group

In order to facilitate the incorporation of the ICT program into English teaching practices during CAI classes, we carefully designed and compiled a CAI implementation protocol. The protocol included a curriculum, a lesson-by-lesson English Teaching Plan (henceforth, the Plan) and teachers' responsibilities. During a one-day intensive training before the program,

[2] One teacher per school was invited to the training in the treatment groups and all the teachers who were invited attended the training.

we trained fifth-grade English teachers[3] on the implementation protocol. To make sure that help was provided whenever there was a problem with the hardware, the software package or the protocol, we offered "24/7 consultation hotlines" to teachers.

The curriculum was designed to match the pace of English instruction in regular classes. It had detailed information about the link between each unit in the standard curriculum and the unit in software package. It also provided suggestions on how to match specific modules to the learning progress of students. For instance, if a lesson introduced new vocabulary, in the CAI classes the curriculum directed the English teachers to cover the specific modules of listening and spelling exercises for that vocabulary. If a new dialogue was taught in the regular classes, the curriculum took the teachers to the module that had additional dialogues with a similar structure or context. This way, teachers had easy access to the most relevant and suitable modules to enhance student learning as they progressed.

The goal of the Plan was to provide English teachers with various interactive activities that they could organize during the CAI classes. The activities not only provided timely feedback to students, they also aimed to increase interactions between students, and between students and teachers. In the Plan, each lesson was divided into several teaching sessions. Each session had a learning goal that matched the national curriculum. Following the learning goals, we listed example activities that teachers could organize using the materials or exercises provided by the software package. Activities included role-playing games, word puzzle games and various competitions that encouraged student engagement in learning English. These activities were designed in such a way that the materials in the software could be incorporated into the interactions between the students and teachers. Although the Plan had detailed information on how each class could be organized, it was a reference book rather than a compulsory teaching plan. During the training, we emphasized that teachers should adapt their teaching plan to the learning pace and interest of their students.

In the protocol, we also laid out the responsibilities that the English teachers needed to fulfill in implementing the CAI classes. These responsibilities included: (a) taking attendance roll call of each student at the beginning of each class; (b) making sure that the curriculum was followed in such a way that each CAI class matched the curriculum of regular English classes; (c) organizing activities to enhance interaction and learning; (d) providing timely feedback to students, including providing immediate assistance when students experienced difficulty in learning or in computer operations; and (e) keeping close contact with our research group and volunteers regarding technical support or questions regarding the protocol.

To facilitate the implementation of our CAI protocol, we compensated the English teachers who instructed the program sessions with a stipend of RMB500 (approximately USD80) per semester. We gave the stipend to them only if they faithfully implemented the CAI program. To monitor how the English teachers followed the protocol, we recruited enumerators from universities in Haidong Prefecture and sent them to visit the CAI schools during the program

[3] After choosing sample classes, the English teacher teaching the fifth-grade classes in the CAI treatment group was assigned as the teacher-supervisor for the program. The teacher assignment was simple because in the 22 CAI treatment schools, there was only one fifth-grade English teacher in each school.

period. The enumerators randomly selected students from each treatment class and surveyed them about how the CAI classes were conducted.

3.1.4 Control group

A total of 3,996 fifth-grade students in 83 control schools constituted the control group.[4] During the program, students in the control group did not receive any intervention. To avoid any form of spillover effects and Hawthorne effect (Landsberger, 1958), our program team did not visit or contact any control schools except during the baseline and endline surveys. To our knowledge, principals, teachers and students were not informed about our program. The students in control schools took their regular classes at school as before.

3.1.5 Data collection

The research team conducted two rounds of surveys in the 127 sample schools. Phase I of the program was a baseline survey conducted with all fourth-grade students in the 127 schools during the beginning of the fall semester in 2012. The baseline was done before any implementation of the CAL program had begun (and before the randomization assignments were made). Phase II was a final evaluation survey conducted at the conclusion of the program at the end of the spring semester in 2013. The length of the program was one academic year, or a little over nine months.

In each round of the survey, the enumeration team visited each school and conducted a three-part survey. In the first part students were given a standardized English test. The English test included 90 questions. Students were required to finish tests in each subject in 30 minutes. Our enumeration team strictly enforced time limits and proctored the examinations. We used the English test scores as the measure of English academic performance.[5]

In the second part of the survey, enumerators collected data on the students and their families, including demographic and socioeconomic characteristics. The dataset included measures of student gender, student age, whether the student belonged to an ethnic minority, whether the student had used a computer before, whether the student was the only child in the family, whether the mother had finished junior high or above education, whether the father finished junior high or above education, whether the mother had a migrant job, and whether the father had a migrant job.

In the second part of the survey, students were also asked to answer questions about their English teachers, the English classes and other questions about non-cognitive traits. To create indicators for student attitudes toward their English teacher and English classes (whether they liked the English teacher and liked the English class), the students were asked to rate their attitude on a 0–100 scale, where “0” indicated “extremely hate the English teacher/class” and “100” indicated “extremely like the English teacher/class.” We also

[4]The computer:student ratio is uniform across the treatment and control groups (0.13 computers per student).”

[5] All the questions were multiple choice and the tests were graded using STATA. Therefore, the test scores were unlikely to be biased in the grading process. The test questions were based on the standard national curriculum for English language. These test items went through several rounds of selection, validation and piloting by the research team and teacher experts from primary schools. In Appendix 1, we present a few example questions we included in the test.

generated an index to measure the feedback students received in English classes. The index was based on five questions on the interaction between the student and the English teacher, such as "whether the English teacher lets me know my study progress" and "whether the English teacher asks us questions in order to make sure that we understand what he or she teaches". In the second part of the survey we also asked questions to measure the non-cognitive traits, such as self-efficacy (1–4 points),[6] the education level that students want to achieve (1=college or above; 0=below college), selfish and altruistic indicators.[7]

In the third part of the survey, we evaluated the program's implementation. First, we measured the training quality by giving a 15-minute test at the end of the teacher trainings. The test was designed to examine teachers' knowledge of the software and program implementation protocol. All teachers who participated in the trainings completed it. When the county education bureaus conducted teacher trainings, we sent an enumerator to enter the training room after the training was finished and set a test when the teachers were about to leave. Second, during the evaluation survey we asked students and teachers about the frequency of CAL classes during the 2012/13 academic year (whether the CAL classes were organized twice per week). We also asked teachers during which time slots the CAL classes and the open voluntary CAL sessions were arranged. If the CAL sessions replaced a regular class, the teachers were asked to indicate which class.

3.1.6 Data collection

The research group conducted two waves of surveys in the 127 sample schools. The first round of surveys was a baseline survey conducted with all students in 127 schools in September 2013 at the beginning of the autumn semester. It was before any implementation of our program had begun. The second round survey was an evaluation survey conducted at the end of the ICT program in June 2014, a time that coincided with the end of the 2014 spring semester.

In each round of surveys, the enumeration team visited all 127 schools and conducted a three-part survey. In the first part students were given a 30-minute standardized English test and we used the scores of the students as our measure of student academic performance.

---

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

[7] We followed Jiang (2014) by conducting an experiment to elicit student other-regarding preferences. The experiment asks the students to divide a prize either between themselves and other types of people or take the whole prize oneself. There are four different choices that the students can make in an increasing degree of altruism. We generated an altruistic indicator that equals 1 if the student chooses to divide the prize between herself/himself and strangers (the most altruistic choice) and 0 if otherwise. We also generated a selfish indicator if the students chooses to take the whole prize herself/himself.

All of the questions on the English test in the endline survey were different from the questions in the baseline survey. We only chose the questions that did not overlap with the exercises in our software package. Our enumeration team strictly proctored the test and enforced the time limit.

In the second part we collected data on the characteristics of students and their families. From this part of the survey we were able to create demographic and socioeconomic variables. The dataset includes measures of each student's *gender* (1=female; 0=male), *age* (years), *ethnicity* (1=Han; 0=minority), *computer use* (1=has used computer; 0=never used computer), and family characteristics, such as assets, family size and siblings, and whether the father or mother farms full time.

In the third part of the survey, we gathered information about English teachers and schools. For example, we had questions about *English teacher's gender* (1=female; 0=male), *English teacher's ethnicity* (1=Han; 0=minority), *English teacher's age* (years), *whether the English teacher is hired by the education bureau or privately by the school* (1=education bureau; 0=school).[8] We also collected the basic characteristics of the schools, such as *school area* (sq. m.) and *computer:student ratio*.

3.1.7 Statistical methods

We used ordinary least squares (OLS) regression analysis (both with and without control variables) to estimate how the academic outcome changed in the treatment groups relative to the control group and how the program implementation differed between the treatment groups. Our basic OLS analysis regressed the evaluation outcome variables (i.e. post-program outcome value) on the value of outcome variables at baseline and a dummy variable measuring each student's treatment (CAL intervention) status. The basic OLS model is:

$$y_{is} = \alpha + \beta_1 treat_{1s} + \beta_2 treat_{2s} + \theta y_{0is} + \varepsilon_{is} \qquad (1)$$

where $\mathrm{y_{is}}$ is the evaluation outcome variable for child i in school s, $\mathrm{y_{0is}}$ measures the outcome variable of the same child at the baseline, $\mathrm{treat_{1s}}$ is a dummy variable for treatment group 1 (equal to one for students in the treatment group 1 and zero otherwise; treatment group 1 is where the research team implements CAL), $treat_{2s}$ is a dummy variable for treatment group 2 (equal to one for students in the treatment group 2 and zero otherwise; treatment group 2 is where the government implements CAL), and ε_{is} is a random disturbance term clustered at the school level.

We used several variables to measure the student outcomes and program implementation (y_{is}). The primary outcome variable of our analysis is the student academic outcome, measured by the student standardized test scores in English. In addition to measuring academic outcomes, we also measured non-cognitive traits, such as "likes English teacher", "likes English class", and "feedback in English class".

---

[8] If a teacher is hired by the education bureau, he or she is a registered public teacher. The rural schools only have the right to hire contract teachers who cannot be registered by the education bureau. A registered teacher hired by the education bureau gets a higher basic salary and subsidies than the privately hired contract teachers.

By construction, the coefficients of the dummy variables $treat_{1s}$ and $treat_{2s}$,β_1 and β_2, measure the difference in the value of the outcome variables between the treatment groups and the control group. In other words, β_1 and β_2 measure how the treatment groups changed in the outcome levels during the program period relative to the control group.

In order to improve the efficiency of the estimation, we built on the model in equation (1) by including a set of control variables:

$$y_{is} = \alpha + \beta_1 treat_{1s} + \beta_2 treat_{2s} + \theta y_{0is} + X_{is}\gamma + \varepsilon_{is} \quad (2)$$

where all the variables and parameters are the same as those in equation (1), except that we added a set of control variables. Specifically, besides y_{0is}, the pre-program outcome value for student i in school s, we controlled for X_{is}, a vector of additional control variables. The control variables include student individual and family characteristics (rows 2 to 11 in Table 1). In all regressions, we accounted for the clustered nature of our sample by constructing Huber-White standard errors corrected for school-level clustering.

In order to compare the program implementation between the two treatment groups, we adjusted the equation (2) by taking out the variable $treat_{2s}$ and limit the samples to include only the students of the two treatment groups. By doing this, β_1 now measures the difference in the values of the outcome variables between the treatment group 1 and treatment group 2. The program implementation outcomes include teacher training evaluation scores, CAL sessions twice per week, whether regular CAL classes replaced English classes, whether open voluntary CAL sessions replaced English classes, and whether an English teacher was appointed as the CAL teacher-supervisor. In estimation, we also used Huber-White standard errors corrected for school-level clustering.

To evaluate how academic performance changed among students who participated in the CAI or CAL program relative to those who did not, we used unadjusted and adjusted OLS regression analysis. The regression models are presented in order of increasing comprehensiveness.

To test the impact of our program, we estimated average treatment effect by regressing the endline outcome variable (i.e. *standardized endline English test score*) on dummy variables of the treatment status. The model we estimated is:

$$y_{is} = \alpha + \beta_{CAL} \cdot treat_{1s} + \beta_{CAI} \cdot treat_{2s} + \theta \cdot y_{0is} + \varepsilon_{is} \quad (1)$$

where y_{is} is the standardized endline English test score for child *i* in school *s*, $treat_{1s}$ is the dummy variable for the CAL program and $treat_{2s}$ is the dummy variable for the CAI program. Because the treatment was randomly assigned, β_{CAL} and β_{CAI} in equation (1) provide unbiased estimates of the average treatment effect of the CAL program and the CAI program respectively. We also included y_{0is}, which is the English test score of the same child at the baseline. Therefore, the coefficients, β_{CAL} and β_{CAI}, measure the changes in English test score before and after the program that is due to the treatment.

To enhance the efficiency of estimation, we estimated an adjusted model by controlling a set of variables collected at the baseline survey. The model we estimated is:

$$y_{is} = \alpha + \beta_{CAL} \cdot treat_{1s} + \beta_{CAI} \cdot treat_{2s} + \theta \cdot y_{0is} + X_{is}\gamma + \varepsilon_{is} \quad (2)$$

where X_{is} includes a vector of covariates of student, family and school characteristics (such as *gender*, *age, ethnic*, *computer use*, *education level students want to achieve*, *family asset*, *family size and siblings*, *characteristics of English teacher*, *school area*, *computer:student ratio*, etc.)

In order to test for the heterogeneous program effects, we also included interaction terms between the treatment dummy variables and some key covariates in the regression model specified in equation (2). For example, we tested whether the change in English test scores before and after the program differed for students who were better performing at the baseline relative to students who were worse performing. This was done by including in the regression an interaction term between the treatment dummy variables and the variable of baseline English test score. The model we estimated is:

$$y_{is} = \alpha + \beta_{CAL1} \cdot treat_{1s} + \beta_{CAI1} \cdot treat_{2s} + \beta_{CAL2} \cdot treat_{1s} \cdot y_{0is} + \beta_{CAI2} \cdot treat_{2s} \cdot y_{0is} \quad + \theta \cdot y_{0is} + X_{is}\gamma + \varepsilon_{is} \qquad (3)$$

the coefficient on the interaction term β_{CAL2} and β_{CAI2} indicate the heterogeneous treatment effects by student baseline English test score. In all regressions, we accounted for the clustered design by constructing Huber-White standard errors clustered on the school level (relaxing the assumption that disturbance terms are independent and identically distributed within schools).

4. Results

4.1. Study 1: scholars versus bureaucrats implementing school programs: evidence from a randomized controlled trial of computer-assisted learning in rural China

4.1.1 The impacts of the CAL intervention on student English test scores

The data showed that during the 2013/14 academic year, fourth-grade students whose CAL treatment was implemented by the research group improved significantly more in their English test scores than students in the control group (Table 4). Using equation (1), the estimated CAL treatment effect on English test scores among the treatment group 1 students was equal to 0.16 standard deviations and significant at the five per cent level (row 1, column 1). Using the model controlling for the student characteristics and county dummies in equation (2) and county dummies, the estimated CAL treatment effect of treatment group 1 went up slightly to 0.18 standard deviations (row 1, column 3).

The CAL treatment implemented by the government did not seem to change student test scores in English relative to the control students (Table 4). Using equation (1), the estimated CAL treatment effect on English test scores among the treatment group 2 students was equal to -0.07 standard deviations and insignificant when using equation (1) (row 2, column 1). Using the model controlling for the various student characteristics and county dummies in equation (2), the estimated CAL treatment effect of treatment group was 2 zero standard deviations and insignificant (row 2, column 3). The F-test showed that the difference between the estimated program effect on the treatment group 1 and the treatment group 2 was significant at the one per cent level. In other words, the CAL program that the research team implemented was much more effective than the program the government implemented.

4.1.2 Causal chain analysis

There are several possible reasons for the difference in treatment effects between the CAL programs the research team and government implemented. First, the teacher training that the government organized may have been less effective than the one the research team organized. The teachers in treatment group 2 may have learned less about the software and the protocol during the training than the teachers in treatment group 1. As a result, due to more problems in software operation or classroom management, CAL classes may have been less effective for treatment group 2. Second, the schools in the treatment group 2 may have not made enough effort to pull together the school inputs to organize the CAL classes twice a week as required by the protocol. Third, another aspect of protocol compliance is that, instead of using the computer classes to organize CAL sessions, schools or teachers may have cut back teacher time in other educational activities, such as English classes.

In order to understand the causal chain of the implementer effect, we tested three hypotheses. First, we tested whether the teacher trainings that the research team organized and the trainings that the government organized differed in teachers' knowledge of the software and protocol. Second, we tested whether the CAL program was conducted twice per week by the schools in the treatment group 1 and the treatment group 2. Third, we tested whether the CAL sessions replaced any regular classes, such as the English classes.

4.1.3 Teacher training quality

The analysis suggests that at the end of the teacher training, teachers who attended the training the research group organized did not differ in knowledge in software and protocol from those whose training the government organized (Table 5). Using equation (2) with only the treatment dummy for treatment group 2 and including only the students of the two treatment groups, the estimated difference in teacher training evaluation scores between the treatment group 2 and the treatment group 1, was -0.07 standard deviations and insignificant (row 1, column 1). In other words, we did not find evidence suggesting that the quality of teacher training the county education bureau organized was significantly worse than the one the research team organized.

4.1.4 Frequency of the CAL classes

There was some evidence that the CAL classes were organized more frequently when the government implemented the treatment than when the research team implemented the treatment in the first semester (Table 5). By design, both treatment group 1 and treatment group 2 were more likely to attend the CAL sessions than the control group (rows 1 and 2, columns 2–5). Based on the teacher survey data, the likelihood of holding CAL sessions twice per week in the first semester and in the second semester of the 2012/13 academic year was not significantly different between the two treatment groups (rows 1 and 2, columns 2 and 3). However, according to student survey data, the classes were more frequently organized in the first semester among treatment group 2 (rows 1 and 2, column 4). According to the student data, 19 per cent more of the treatment group 2 schools held the CAL sessions twice a week than the treatment group 1 in the first semester. The difference was not significant in the second semester (rows 1 and 2, column 5).

4.1.5 Re-arrangement of school inputs

Our results showed that when the government implemented the CAL program, it was more likely to have CAL classes replace regular English classes than when the research team implemented the program (Table 6). Using the control group as the base group (no control school replaced regular English classes with CAL classes), we showed that 39 per cent of the treatment group 2 schools replaced regular English classes with CAL sessions (row 2, column 1). The treatment group 1 had a much smaller percentage of schools that replaced regular English classes with the CAL sessions (14 per cent, row 1, column 1). The difference between the treatment group 1 and treatment group 2 is significant at the one per cent level.

In order to further investigate the school input of teacher time in the program, we looked at whether an English teacher was assigned to be the teacher-supervisor of CAL. Using the control group as the base group (no control school replaced regular English classes with CAL classes), we showed that 32 per cent of the treatment group 2 schools assigned an English teacher to supervise CAL classes (Table 6, row 2, column 3). In contrast, treatment group 1 did not seem to have assigned an English teacher to supervise CAL classes relative to the control group (the coefficient is not significant at the 10 per cent level). The difference between treatment group 1 and 2 is significant at the one per cent level. In other words, when the government implemented the program, schools were more likely to have English teachers supervise the CAL classes and more likely to run the classes during the English class time. In doing so, the workload of the English teachers may have been the same or even reduced, since they did not teach as many English classes as before. They also received an additional allowance of RMB500 per semester.

To further explore the changes in the English classes after the English teachers substituted CAL sessions, we conducted a test to investigate whether treatment group 2 made any difference in student attitudes toward the English classes and teachers (Table 7). We found that students liked their English teachers better in the treatment group 2 relative to the control group. Student ratings of their English teachers were 4.11 points higher when the English teachers were more likely to replace the English classes with CAL sessions (row 2, column 2). The estimated difference in the liking of the English classes between the treatment group 2 and the control group was 2.93 but insignificant (row 2, column 1). In contrast, the treatment group 1 did not seem to differ from the control group in the liking of English classes or teachers (row 1, columns 1 and 2). It seems to suggest that students may have increased their preference toward the teachers and classes when they played the game-based CAL software during English class.

Although students liked their English teachers more, they received less feedback from the teachers in class (Table 7). Treatment group 2 students scored 0.12 points lower in getting feedback in English classes relative to the control group (row 2, column 3). The estimate was significant at the five per cent level. It was not the case for treatment group 1 students, since the English classes were not changed. Treatment group 1 students did not seem to differ from the control group in getting feedback during regular English classes (row 1, column 3).

4.1.6 Heterogeneous effects

We also tested heterogeneous effects to investigate whether better- or worse-performing students benefited differently from the CAL program that the government or the research team implemented (Appendix 3). We included the interaction terms between the treatment variables and student baseline English test scores in equation (2) when running the regression. Our results showed that there was no heterogeneous effect among students in either treatment group 1 or treatment group 2.

4.2. Study 2: which is more effective, computer-assisted learning or teaching? Evidence from a randomized controlled trial in rural schools in China

After combining the two treatment groups and testing the average impact of any ICT program on student English test scores, we found there was no significant impact on students' English performance (Table 11). Both the unadjusted and adjusted models showed that although the program impact was estimated to be 0.05-0.06 standard deviations, it was not significant at the 10 per cent level (column 1, rows 1 and 2). It suggests that there was no detectable average impact of the CAL and CAI programs combined.

By separating the two treatment groups (CAI and CAL), we found that the CAI program was significantly more effective than the CAL program (Table 12). The results show that the program impact differed between the CAI treatment and the CAL treatment. Integration into teaching practices was more effective (Table 12). The CAI program was estimated to have an impact of 0.07 standard deviations and was significant at the 10 per cent level using the unadjusted model (column 1, row 1). By controlling for the covariates at baseline, the impact slightly increased to 0.08 standard deviations. The estimate was significant at the five per cent level (column 2, row 1). In contrast, the coefficient on the CAL treatment was positive but not significant at the 10 per cent level (columns 1 and 2, row 2). The difference between the coefficients on the CAI treatment and the CAL treatment was significant at the one per cent level, suggesting that the two treatments had different impacts on learning.

4.3. Heterogeneous treatment effects

In order to investigate the potential mechanisms that explain why the CAI treatment was more effective than the CAL treatment, we tested the heterogeneous effects of the two programs by student initial performance (Table 13). When we interacted the CAI treatment variable with student English academic performance at the baseline, the coefficient on the interaction term was small in magnitude and insignificant (Table 6, column 1, row 2). It suggests that the better- and worse-performing students benefited similarly from the CAI treatment. However, when we interacted the CAL treatment variable with student baseline English score, we found the coefficient on the interaction team was 0.08 standard deviations and was significant at the 10 per cent level (column 1, row 4). This suggests that the CAL program worked better for the better-performing students than for the worse-performing students. The results were consistent with the literature suggesting that the effectiveness of ICT programs depends on the initial level of knowledge that a student has of that subject (Hatch, 1978; Long, 1996; Gass and Mackey, 2006; Sauro, 2009). For better-performing students, their level of English knowledge was high enough for them to benefit from the CAL program. However, for worse-performing students, they may not have been able to benefit without any instruction from an English teacher. When teachers could provide guidance and corrective feedback in the CAI programs, worse-performing students also benefited.

We also tested the heterogeneous effects of the two programs based on the workload program supervisors faced.[9] First, as CAI required the English teachers to integrate the program into teaching, the time they needed to prepare for and teach the CAI classes may have resulted in an increased workload (Burston, 1996; Jones, 2001; Yang and Huang, 2008). However, since the program also provided convenient additional teaching materials, it may have required less marginal effort from the English teachers to improve student performance (Lam, 2000; Nomass, 2013). In other words, English teachers may have been more willing to make the effort if the CAI program complemented the regular English teaching.

When testing the heterogeneous effects by supervisor's workload, we found that the CAI program worked similarly whether the workload of the supervisor was high or low (Table 14). The coefficient on the interaction term between the CAI treatment variable and the supervisor's workload was insignificant at the 10 per cent level (column 1, row 3). However, for the CAL treatment group, the supervisor's workload mattered tremendously. If the supervisor's teaching workload was lower than the mean, the CAL program increased student English performance by 0.16 standard deviations, and it was significant at the 10 per cent level (column 1, row 4). The CAL program had no impact if the supervisor already had a high teaching workload (column 1, row 6). The results seemed to be consistent with the hypothesis that the CAI program was complementary to regular English teaching. On the other hand, for CAL supervisors, the program added two classes to their regular workload and the program was not complementary to their regular work. Therefore, regular workload mattered more for the CAL supervisors than the CAI supervisors.

5. Conclusion

Given China's current effort to put computer rooms in all schools in poor rural areas of western China, the question whether computers can be effectively used to increase student learning, becomes an important policy-relevant one. As more computers are installed in rural schools, policymakers and school officials will need to explore various options to make full use of these machines. Potentially, one of the most cost-effective ways is to implement a CAL program to provide remedial tutoring. But, it is an open question whether the government can implement the program in a way that achieves its goals.

We explored the question by conducting a randomized experiment of implementation of a CAL program by randomly dividing schools into three groups: the research team implemented group 1's CAL program; the government implemented group 2's CAL program; and the third group served as the control. The key contribution of the study was not only in estimating and comparing how the program impacted on two different implementation teams, but also in estimating how the program affected relevant school inputs.

[9] To create the indicator of workload, we required teachers to report their working schedule of each week during the baseline survey. We calculated a teacher's workload based on the number of lessons she/he taught per week. We generated a dummy variable that equals 1 if the number of lessons taught is larger than the mean and the variable equals 0 if otherwise.

Our first result is that the CAL program that provided remedial English tutoring was effective in improving student performance in English. We proved that by following the protocol of the program implementation for an academic year the students' test scores improved by 0.18 standard deviations. This was also the first experimental evidence showing that using computers and appropriate software can effectively improve student learning in a second language among disadvantaged children in China.

Second, we found that the program lost its impact when it was implemented by the government. Reasons such as the government not adequately organizing teacher training or launching the program were ruled out. Our analysis showed that teacher training and the frequency of CAL classes were comparable between the programs implemented by researchers and implemented by the government.

Instead, we found that schools in the government-implemented program were more likely to cut back school inputs in other regular classes by having English teachers serve as the teacher-supervisor of the CAL programs and by replacing English classes with the CAL sessions. One possible reason for this may have been to minimize the workload of teachers. By replacing the English classes with CAL classes, the workload of English teachers and computer teachers remained unchanged (or was even reduced). We also found that the substitution effect took place, as students received less feedback from teachers during English classes, although they used computer-assisted learning software to learn English. This also suggests that the regular English teaching may have not been more effective than the CAL program in improving student English learning.

By speculating about the differences between researcher-implemented programs and government-implemented ones, we think it is likely that effective monitoring may have made the difference in program impact. We think it is unlikely that schools or teachers in the program implemented by the research team were more motivated than those in the program organized by the government. If there was any difference in incentive, the government may have been more likely to motivate schools and teachers because principals or teachers would like to be promoted to administrative positions or higher teacher ranks. However, through regular interviews with the officials during and after implementation we found that the officials made almost no follow-up checks on the schools or the teachers. Even when they did, they rarely asked about whether the details of the protocol had been respected. This may have been a key shortcoming in the government-implementation group.

We show that under the right conditions CAL can play an important role in boosting achievement in second language learning. It is, however, immaterial that local authorities cannot implement the program on their own. This finding has important implications for many other inputs that researchers endorse on the basis of successes achieved when they are implemented under experimental conditions alone. For a more precise understanding of such program impacts, it is imperative for local authorities to manage implementation.

In this paper we also present findings of the relative effectiveness of CAI and CAL ICT programs in rural Chinese schools with varying demographics and characteristics. The main difference between the CAI treatment and the CAL treatment was whether the ICT program was integrated into teaching. The CAL treatment did not involve any English teaching.

Students only used the software for additional practice by themselves. The CAI treatment required teachers to provide instructions and organize learning activities using the software package. For the CAI treatment group, we provided supervisors with an English Teaching Plan containing material on how each module in the package could be matched to the learning progress of students and what kind of activities can be organized using each module. The activities contained in the Plan provided timely feedback to students and aimed to increase interactions between students and between students and teachers. For the CAL treatment group, the supervisors had fewer responsibilities than the CAI teachers. They only needed to help students with scheduling, computer hardware issues and software operations. CAL supervisors were all non-English teachers and there was no English teaching in CAL classes.

The results show that the program impact on student standardized English scores differed between the CAI treatment group and the CAL treatment group. We found that integrating the program into teaching practices was more effective in improving student English scores. The CAI program increased student test scores by 0.08 standard deviations, while the CAL program had no significant impact.

By testing for the potential mechanisms, we found interesting differential effects of the CAL and CAI programs. First, we found that the lower a student's initial English performance, the less he or she benefitted from the CAL program. In contrast, when the program was integrated into English instruction, it benefited worse- and better-performing students. Second, the program supervisor's regular workload did not matter for the CAI program. One possible reason for it is that the program was complementary with the regular English teaching. The CAI supervisors, who were English teachers, may have been able to substitute their effort for regular English teaching with effort for instructing the CAI program or preparing for regular classes. Therefore, the CAI program impact did not differ for English teachers with high or low workloads. However, workload mattered for the CAL supervisors (who were non-English teachers). When the workload of the non-English teacher was higher than the mean, the program impact was dramatically reduced.

Although we do not know (from our data) the exact reason, interviews with non-English teachers who had high teaching loads revealed that in many cases the CAL program sessions were not offered, some sessions were abbreviated, and some sessions had no supervision (so students either played games or were not directed to the correct lesson). It is possible that this kind of non-compliance with the CAL protocol for the busiest non-English teachers reduced the effectiveness of the treatment for this subset of students.

This paper contributes to the understanding of the conditions under which ICT interventions are effective in raising student outcomes. To do this, we compared two ways of conducting ICT programs in education: integration and non-integration of the program into teaching practices. As we discuss above, integration of the ICT program into teaching appears to be more effective than non-integration in English learning, which happens to be one of the weakest subjects for rural students. Education experts and policymakers should take these results into consideration when extending the ICT programs on a large scale to rural schools in China. The intial knowledge level of students and the workload of teachers may largely influence how effective an ICT program is in improving student learning.

We realize that the project has limitations regarding external validity. The project was implemented in schools in poor rural China where there is a high concentration of ethnic minorities. This suggests that our results are mainly representative of schools with poor resources in China and to an important extent other developing countries. The study may say little about how such projects would work in schools that are more competitive in richer, better-resourced communities that are dominated by one ethnic group.

Figure 1: Study #1 experiment profile

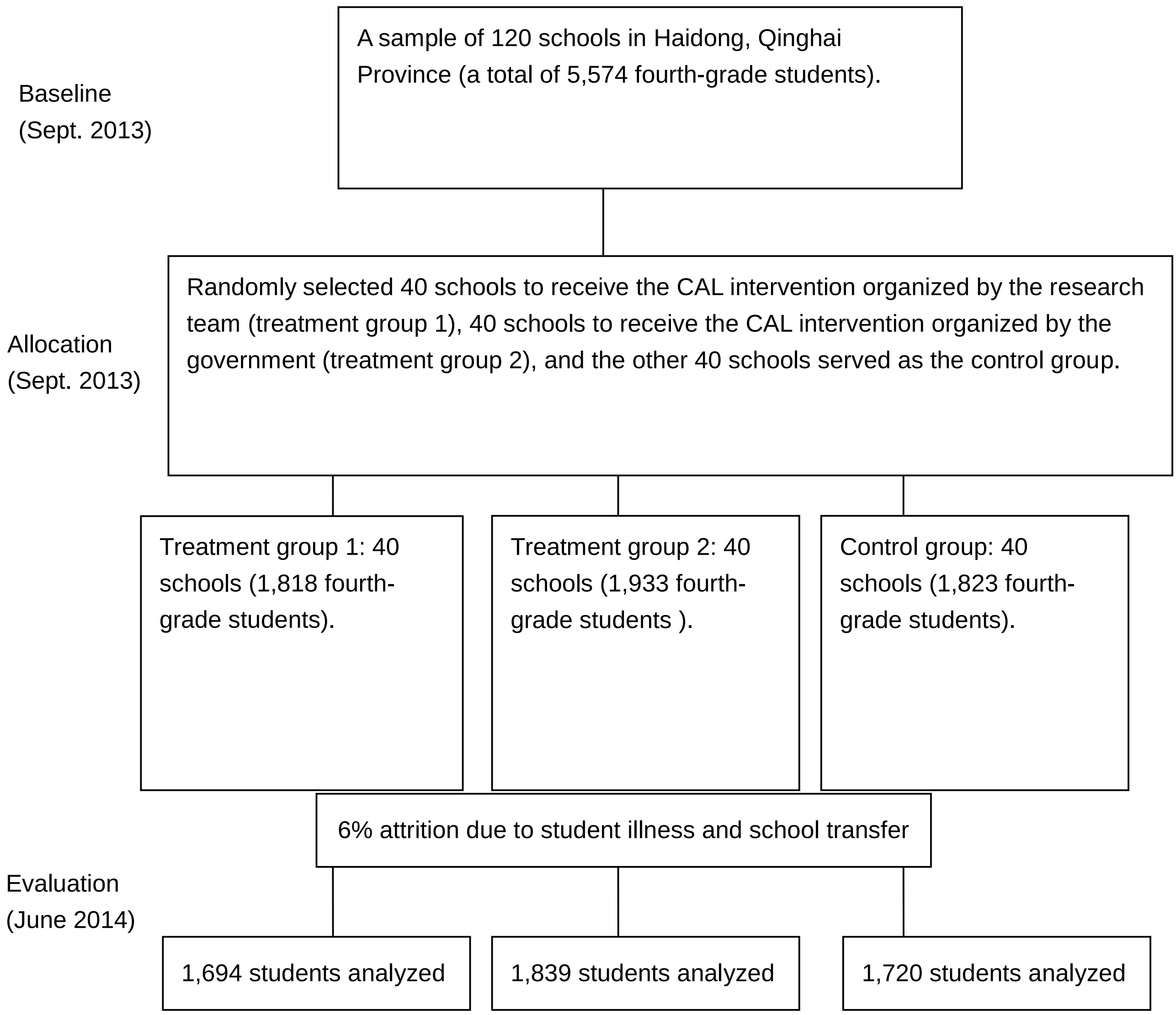


Figure 2: Study #2 experiment profile

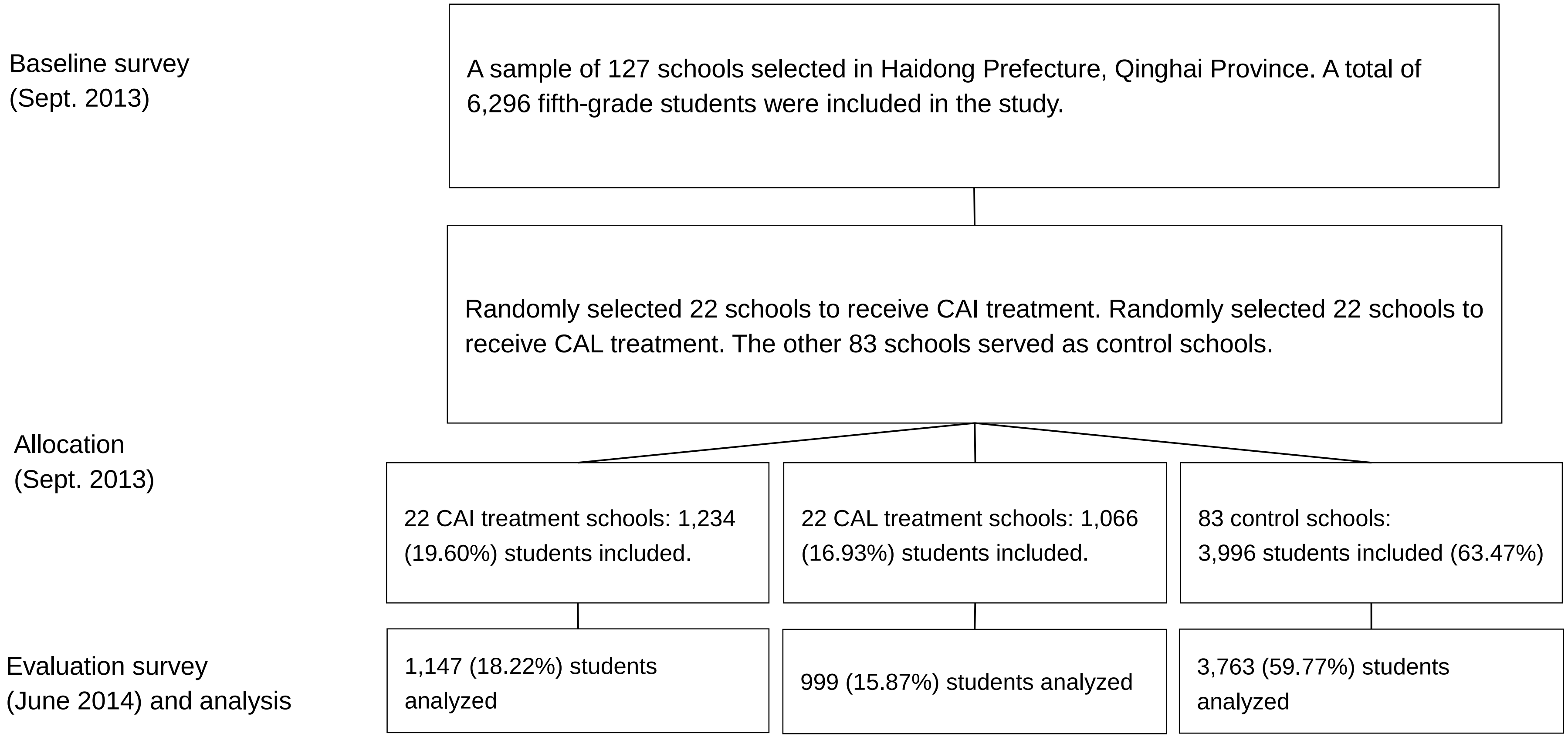


Table 1: Comparison of the characteristics between the treatment groups and the control group of the students before attrition

		Treatment group 1 Research group led intervention		Treatment group 2 Government led intervention		Control group
		Coefficient	S.E.	Coefficient	S.E.	Mean
		(1)	(2)	(3)	(4)	(5)
[1]	Student baseline English score (SD)	-0.02	(0.22)	0.03	(0.18)	0.00
[2]	Student gender (1=male; 0=female)	-0.01	(0.02)	-0.00	(0.02)	0.52
[3]	Student age (year)	0.04	(0.10)	-0.07	(0.10)	10.35
[4]	Student belongs to an ethnic minority (1=yes; 0=no)	-0.00	(0.09)	-0.01	(0.08)	0.52
[5]	Student self-efficacy scale (0-4 pts)	0.00	(0.03)	0.01	(0.03)	2.62
[6]	Used computer before (1=yes; 0=no)	-0.05	(0.07)	-0.00	(0.07)	0.72
[7]	Only child (1=yes; 0=no)	0.02	(0.03)	0.02	(0.03)	1.83
[8]	Mother finished junior high or higher education (1=yes; 0=no)	-0.04	(0.06)	0.03	(0.05)	0.65
[9]	Father finished junior high or higher education (1=yes; 0=no)	0.01	(0.04)	0.03	(0.04)	0.77
[10]	Mother has a migrant job (1=yes; 0=no)	-0.03	(0.03)	-0.02	(0.03)	0.46
[11]	Father has a migrant job (1=yes; 0=no)	-0.03	(0.03)	-0.04	(0.03)	0.68

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

Note: CAL = computer-assisted learning; CAI = computer-assisted instruction

Table 2: Comparisons of the attrition between the treatment and control students

Dependent variable: Attrition (1=yes; 0=no)		
		(1)
[1]	Treatment group 1(1=yes; 0=no)	0.01
		(0.02)
[2]	Treatment group 2 (1=yes; 0=no)	-0.01
		(0.01)
[3]	Constant	0.06***
		(0.01)
[4]	Observations	5,574
[5]	R-squared	0.001

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

Table 3: Comparisons of the characteristics between the treatment groups and the control group of the students after attrition

		Treatment group 1 Research group led intervention		Treatment group 2 Government led intervention	
		Coefficient	S.E.	Coefficient	S.E.
		(1)	(2)	(3)	(4)
[1]	Student baseline English score (SD)	0.01	(0.22)	0.03	(0.18)
[2]	Student gender (1=male; 0=female)	-0.01	(0.02)	0.00	(0.02)
[3]	Student age (year)	0.03	(0.10)	-0.06	(0.10)
[4]	Student belongs to an ethnic minority (1=yes; 0=no)	-0.01	(0.09)	-0.01	(0.08)
[5]	Student self-efficacy scale (0-4 pts)	-0.00	(0.03)	0.01	(0.03)
[6]	Used computer before (1=yes; 0=no)	-0.05	(0.07)	-0.00	(0.07)
[7]	Only child (1=yes; 0=no)	0.01	(0.03)	0.02	(0.03)
[8]	Mother finished junior high or higher education (1=yes; 0=no)	-0.03	(0.06)	0.03	(0.05)
[9]	Father finished junior high or higher education (1=yes; 0=no)	0.01	(0.04)	0.03	(0.04)
[10]	Mother has a migrant job (1=yes; 0=no)	-0.02	(0.03)	-0.02	(0.03)
[11]	Father has a migrant job (1=yes; 0=no)	-0.03	(0.03)	-0.05	(0.03)

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

S.E stands for standard error

Table 4: Ordinary least squares estimators of the treatment effect on standardized English test scores

	Dependent variable: Evaluation English score (SD)			
		(1)	(2)	(3)
[1]	Treatment group 1(1=yes; 0=no)	0.16**	0.16**	0.18***
		(0.07)	(0.07)	(0.06)
[2]	Treatment group 2(1=yes; 0=no)	-0.07	-0.07	0.00
		(0.07)	(0.07)	(0.06)
[3]	Control	No	Yes	Yes
[4]	County dummies	No	No	Yes
[5]	Constant	-0.02	0.45**	0.64***
		(0.05)	(0.19)	(0.16)
[6]	Observations	5,253	5,253	5,253
[7]	R-squared	0.642	0.651	0.697
[8]	F-stat row[1]=row[2]	11.75	11.76	11.41

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

Table 5: Ordinary least squares estimators of the implementer effects on the training quality and the frequency of the CAL classes

		Dependent variable				
		Teacher training evaluation scores (0-20 pts)	CAL sessions twice per week in 1st semester—reported by teacher (1=yes; 0=no)	CAL sessions twice per week in 2nd semester—reported by teacher (1=yes; 0=no)	CAL sessions twice per week in 1st semester—reported by student (1=yes; 0=no)	CAL sessions twice per week in 2nd semester—reported by student (1=yes; 0=no)
		(1)	(2)	(3)	(4)	(5)
[1]	Treatment group 1 (1=yes; 0=no)	-[a]	0.46***	0.34***	0.21***	0.32***
			(0.09)	(0.08)	(0.08)	(0.09)
[2]	Treatment group 2 (1=yes; 0=no)	-0.08	0.59***	0.38***	0.40***	0.28***
		(0.47)	(0.09)	(0.09)	(0.07)	(0.06)
[3]	Control	Yes	Yes	Yes	Yes	Yes
[4]	County dummies	Yes	Yes	Yes	Yes	Yes
[5]	Observations	3,622	5,253	5,253	5,253	5,253
[6]	R-squared	0.482	0.369	0.333	0.207	0.221
[7]	F-stat row[1]=row[2]	-	1.18	0.21	3.59	0.15

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

[a] Control group did not take training evaluation tests because they did not participate in the training. Therefore, in column (1), the control group samples were not included and treatment group 1 is the base group.

Note: CAL = computer-assisted learning; CAI = computer-assisted instruction

Table 6: Ordinary least squares estimators of the implementer effects on substitution of CAL sessions

		Dependent variable	
		Regular CAL classes replaced English classes (1=yes; 0=no)	Appointed English teacher as the CAL supervisor (1=yes; 0=no)
		(1)	(3)
[1]	Treatment group 1 (1=yes; 0=no)	0.14*	0.12
		(0.08)	(0.07)
[2]	Treatment group 2 (1=yes; 0=no)	0.39***	0.31***
		(0.09)	(0.10)
[3]	Control	Yes	Yes
[4]	County dummies	Yes	Yes
[5]	Observations	3,622	3,622
[6]	R-squared	0.338	0.364
[7]	F-stat row[1]=row[2]	4.54	4.12

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

Table 7: Ordinary least squares estimators of the treatment effects on student esteem for English teacher and English class and the feedback students received during English class

		Dependent variable		
		Liking of the English class (0–100 pts)	Liking of the English teacher (0–100 pts)	Feedback in the English class (0–4 pts)
		(1)	(2)	(3)
[1]	Treatment group 1 (1=yes; 0=no)	-0.21	-3.43	-0.02
		(2.84)	(3.44)	(0.06)
[2]	Treatment group 2 (1=yes; 0=no)	2.93	4.11*	-0.12**
		(2.34)	(2.43)	(0.05)
[3]	Control	Yes	Yes	Yes
[4]	County dummies	Yes	Yes	Yes
[5]	Constant	73.33***	76.15***	2.66***
		(6.19)	(6.41)	(0.16)
[6]	Observations	5253	5253	5253
[7]	R-squared	0.099	0.053	0.054

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

Table 8: Comparison of the student characteristics between the treatment and control groups within baseline students

Variable	Difference between the treatment and control groups within baseline students[b] CAI treatment (1)	CAL treatment (2)
Student characteristics		
[1] Standardized baseline English test score (standard deviation)[a]	0.02	-0.04
	(0.10)	(0.10)
[2] Gender (1=female; 0=male)	0.02	0.00
	(0.02)	(0.02)
[3] Age (years)	0.11	0.03
	(0.10)	(0.14)
[4] Ethnic (1=Han; 0=minority)	0.04	0.04
	(0.06)	(0.07)
[5] Computer use (1=ever used computer; 0=never used computer)	0.06	0.01
	(0.07)	(0.07)
[6] Student self-efficacy (1–4 points)	-0.02	0.05
	(0.03)	(0.04)
[7] Education level students want to achieve (1=college or above; 0=below college)	-0.03	0.01
	(0.04)	(0.05)
[8] Selfish (1=yes; 0=no)	0.03	0.01
	(0.02)	(0.02)
[9] Altruistic (1=yes; 0=no)	-0.05	-0.04
	(0.03)	(0.03)
[10] Like English teacher (0–100 points)	2.95	1.50
	(2.00)	(2.27)
Family characteristics		
[11] Log (asset)	-0.04	0.00
	(0.04)	(0.03)
[12] Family size bigger than 5 (1=yes; 0=no)	0.00	-0.01
	(0.03)	(0.03)
[13] Have elder sisters (1=yes; 0=no)	-0.00	-0.04
	(0.02)	(0.03)
[14] Have younger sisters (1=yes; 0=no)	0.01	0.01
	(0.02)	(0.02)
[15] Father farms everyday (1=yes; 0=no)	0.00	0.01
	(0.02)	(0.03)

[16] Mother farms everyday (1=yes; 0=no)	0.01	-0.01
	(0.03)	(0.03)
Teacher and school characteristics		
[17] English teacher's gender (1=female; 0=male)	-0.09	-0.12
	(0.11)	(0.10)
[18] English teacher's ethnicity (1=Han; 0=minority)	-0.02	0.02
	(0.09)	(0.10)
[19] English teacher's age (years)	-0.20	-0.21
	(1.36)	(1.31)
[20] Whether the English teacher is hired by the education bureau or privately by the school (1=education bureau; 0=school)	-0.13	0.05
	(0.09)	(0.10)
[21] Non-English teacher's gender (1=female; 0=male)	-0.13	0.09
	(0.12)	(0.07)
[22] Non-English teacher's ethnicity (1=Han; 0=minority)	0.02	-0.06
	(0.12)	(0.09)
[23] Non-English teacher's age (years)	-1.45	0.74
	(1.47)	(1.32)
[24] Whether the Non-English teacher is hired by the education bureau or privately by the school (1=education bureau; 0=school)	-0.11	0.06
	(0.09)	(0.06)
[25] School area (sq. m.)	2,773.57	2,952.02
	(2,265.85)	(3,876.97)
[26] Computer–student ratio	0.26	0.14
	(0.29)	(0.20)
[27] Observations	6,296	6,296

Source: Authors' survey

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

[a] The standardized baseline English test score is the score on the standardized English test that is given to all samples' students before the ICT program, and it is standardized using the baseline mean and standard deviations for the control group.

[b] The within-school difference between the treatment and control groups is calculated by regressions of each of row variables on two treatment dummy variables, controlling for school dummy variables, with robust standard errors clustered at the school level. The sample includes both the non-attrition and the attrition observations.

Table 9: Comparisons of attrition between the treatment and control groups

Dependent variable: attrition (1=students attrited; 0=student remained)	
	Attrition
Variables	(1)
[1] CAI treatment (1=yes; 0=no)	-0.01
	(0.02)
[2] CAL treatment (1=yes; 0=no)	-0.01
	(0.02)
[3] Constant	0.02***
	(0.01)
[4] Observations	6,296
[5] R-squared	0.03

Source: Authors' survey

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

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

Table 10: Comparison of the student characteristics between the treatment and control groups within unattrited students

Variable	Difference between the treatment and control groups within unattrited students[b]	
	CAI treatment	CAL treatment
	(1)	(2)
Student characteristics		
[1] Standardized baseline English test score	0.02	-0.03
(standard deviation)[a]	(0.10)	(0.10)
[2] Gender (1=female; 0=male)	0.03	0.01
	(0.02)	(0.02)
[3] Age (years)	0.12	0.01
	(0.10)	(0.13)
[4] Ethnic (1=Han; 0=minority)	0.03	0.04
	(0.06)	(0.07)
[5] Computer use	0.06	0.00
(1=ever used computer; 0=never used computer)	(0.07)	(0.08)
[6] Student self-efficacy (1–4 pts)	-0.03	0.05
	(0.03)	(0.04)
[7] Education level students want to achieve	-0.04	0.00
(1=college or above; 0=below college)	(0.04)	(0.05)
[8] Selfish (1=yes; 0=no)	0.03	0.00
	(0.02)	(0.02)
[9] Altruistic (1=yes; 0=no)	-0.05	-0.04
	(0.03)	(0.03)
[10] Like English teacher (0–100 pts)	2.87	1.52
	(2.11)	(2.35)
Family characteristics		
[11] Log(asset)	-0.05	0.00
	(0.04)	(0.03)
[12] Family size bigger than 5 (1=yes; 0=no)	-0.00	-0.01
	(0.03)	(0.03)
[13] Have elder sisters (1=yes; 0=no)	-0.01	-0.04
	(0.02)	(0.03)
[14] Have younger sisters (1=yes; 0=no)	0.01	0.00
	(0.02)	(0.02)
[15] Father farms everyday (1=yes; 0=no)	0.01	0.01
	(0.02)	(0.03)
[16] Mother farms everyday (1=yes; 0=no)	0.01	-0.01

	(0.03)	(0.03)
Teacher and school characteristics		
[17] English teacher's gender (1=female; 0=male)	-0.09	-0.12
	(0.11)	(0.11)
[18] English teacher' ethnicity (1=Han; 0=minority)	-0.02	0.02
	(0.09)	(0.10)
[19] English teacher's age (years)	-0.24	-0.31
	(1.38)	(1.32)
[20] Whether the English teacher is hired by the education bureau or privately by the school (1=education bureau; 0=school)	-0.12	0.04
	(0.09)	(0.10)
[21] Non-English teacher's gender (1=female; 0=male)	-0.13	0.08
	(0.12)	(0.07)
[22] Non-English teacher's ethnicity (1=Han; 0=minority)	0.02	-0.09
	(0.12)	(0.09)
[23] Non-English teacher's age (years)	-1.74	0.68
	(1.47)	(1.30)
[24] Whether the Non-English teacher is hired by the education bureau or privately by the school (1=education bureau; 0=school)	-0.11	0.06
	(0.09)	(0.05)
[25] School area (sq. m.)	2,807.22	2,679.06
	(2,343.78)	(3,693.87)
[26] Computer:student ratio	0.28	0.13
	(0.30)	(0.20)
[27] Observations	5,909	5,909

Source: Authors' survey

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

[a] The standardized baseline English test score is the score on the standardized English test that is given to all samples' students before the ICT program, and it is standardized using the baseline mean and standard deviations for the control group.

[b] The within-school difference between the treatment and control groups is calculated by regressions of each of the row variables on two treatment dummy variables, controlling for school dummy variables, with robust standard errors clustered at the school level. The sample is limited to the unattrited observations.

Table 11: Ordinary least squares estimators of the impact of ICT program on standardized English test scores

Dependent variable: standardized endline English test score (standard deviation)		
Variables	(1) Unadjusted Model	(2) Adjusted Model
[1] Any treatment (1=yes; 0=no)	0.05	0.06
	(0.04)	(0.04)
[2] Standardized baseline English test score	0.71***	0.68***
(standard deviation)[a]	(0.03)	(0.03)
[3] Controls	NO	YES
[4] Observations	5,909	5,909
[5] R-squared	0.72	0.73

Source: Authors' survey

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

The test aims to show the impact of the treatment on student English test scores, the test regresses standardized endline English test scores on treatment variable for unadjusted model and regresses standardized endline English test scores on both treatment variables and a set of control variables for adjusted model. The base group is the control group that did not receive either CAL or CAI treatment.

[a] The standardized baseline English test score is the score on the standardized English test that is given to all samples' students before the ICT program, and it is standardized using the baseline mean and standard deviations for the control group.

Table 12: Ordinary least squares estimators of the impact of ICT program on standardized English test scores.

Dependent variable: standardized endline English test score (standard deviation)		
Variables	(1) Unadjusted model	(2) Adjusted Model
[1] CAI treatment (1=yes; 0=no)	0.07*	0.08**
	(0.04)	(0.04)
[2] CAL treatment (1=yes; 0=no)	0.03	0.03
	(0.06)	(0.05)
[3] Standardized baseline English test score (standard deviation)[a]	0.71***	0.68***
	(0.03)	(0.03)
[4] Controls	NO	YES
[5] Observations	5,909	5,909
[6] R-squared	0.72	0.73

Source: Authors' survey

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

The test aims to show the impact of the treatment on students' English test scores; the test regresses standardized endline English test scores on the treatment variables and a set of control variables.

[a] The standardized baseline English test score is the score on the standardized English test that is given to all samples' students before the ICT program, and it is standardized using the baseline mean and standard deviations for the control group.

Table 13: Ordinary least squares estimators of the heterogeneous impact of ICT program on standardized English test scores of students' baseline score

Dependent variable: standardized endline English test score (standard deviation)	
	(1)
Variables	
[1] CAI treatment (1=yes; 0=no)	0.08**
	(0.04)
[2] CAI treatment* standardized baseline English test score	0.03
	(0.04)
[3] CAL treatment (1=yes; 0=no)	0.03
	(0.05)
[4] CAL treatment* standardized baseline English test score	0.08*
	(0.04)
[5] Standardized baseline English test score (standard deviation)[a]	0.66***
	(0.03)
[6] Controls	YES
[7] Observations	5,909
[8] R-squared	0.73

Source: Authors' survey

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

The test aims to show the heterogeneous impact of the ICT program on standardized English test scores of students' baseline score.

[a] The standardized baseline English test score is the score on the standardized English test that is given to all samples' students before the ICT program, and it is standardized using the baseline mean and standard deviations for the control group.

Table 14: Ordinary least squares estimators of the heterogeneous impact of ICT program on standardized English test scores of the supervisor's teaching workload

Dependent variable: standardized endline English test score (standard deviation)	
	(1)
Variables	
[1] CAI treatment (1=yes; 0=no)	0.11**
	(0.05)
[2] CAI supervisor's workload (1=yes; 0=no)	0.00
	(0.05)
[3] CAI treatment* CAI supervisor's workload	-0.06
	-0.08
[4] CAL treatment (1=yes; 0=no)	0.16*
	(0.08)
[5] CAL supervisor's workload (1=yes; 0=no)	0.02
	(0.04)
[6] CAL treatment* CAL supervisor's workload	-0.19*
	(0.11)
[7] Standardized baseline English test score (standard deviation)[a]	0.67***
	(0.03)
[8] Controls	YES
[9] Observations	5,909
[10] R-squared	0.73

Source: Authors' survey

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

The test aims to show the heterogeneous impact of the ICT program on standardized English test scores of the supervisor's workload.

[a] The standardized baseline English test score is the score on the standardized English test that is given to all samples' students before the ICT program, and it is standardized using the baseline mean and standard deviations for the control group.

Figure 3: Baseline English test scores by treatment and control groups

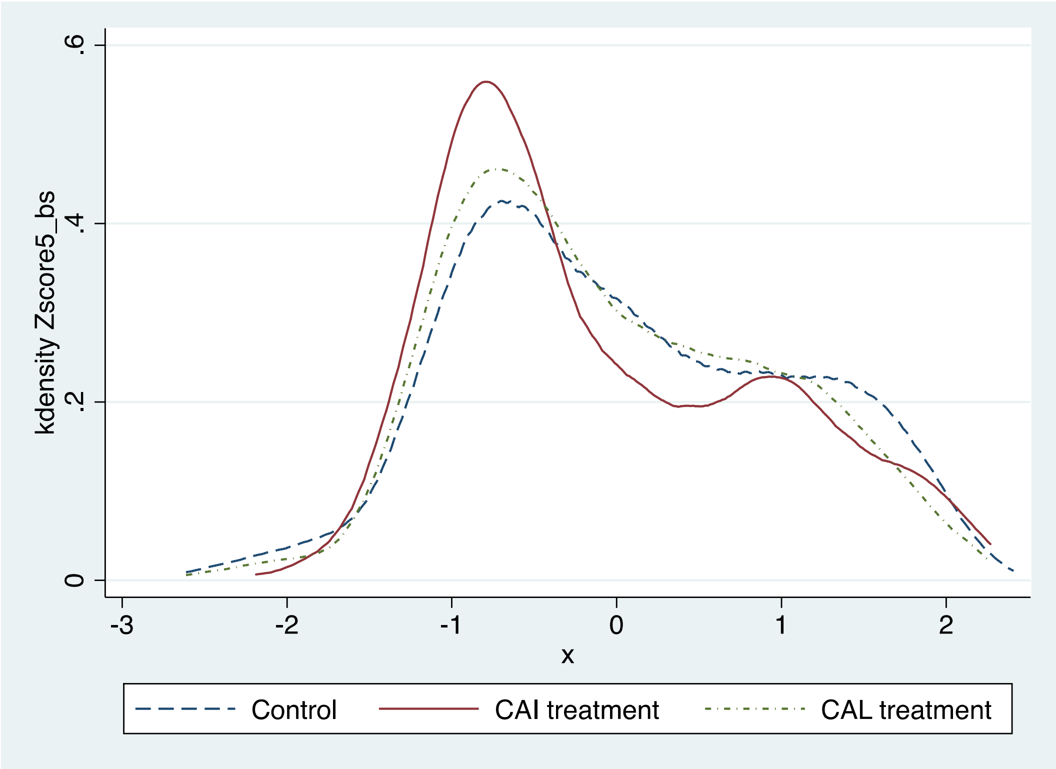
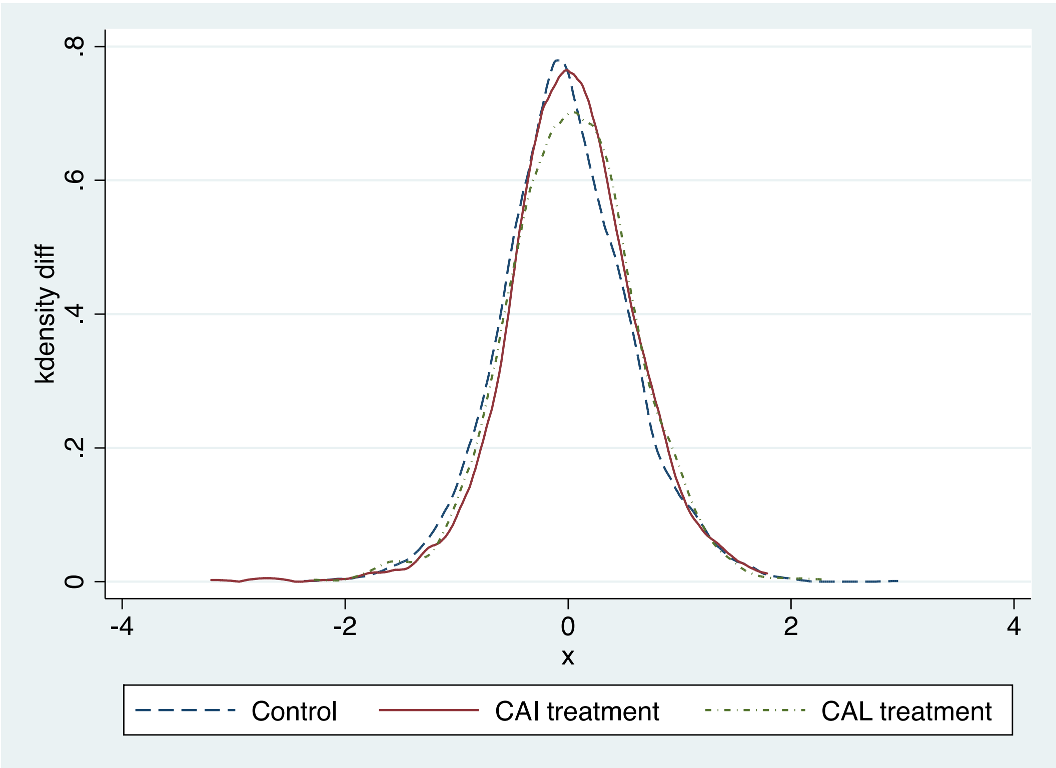


Figure 4: Change in English test scores between baseline and endline by treatment and control groups



Appendix A: Example test items in the English test

( ) 1. How many ______ are there in your school?

A. teachers B. teacher C. teacheres

( ) 2. It is time ______. We are late ______ school.

A. for get up; to B. to get up; for C. to get up; to

( ) 3. --- ________________

--- It's pink.

A. What colour is it?

B. What time is it?

C. Where is it?

( ) 4. ---__________________

--- No, it's rainy.

A. What's the weather like in Cairo?

B. Is it rainy in London?

C. Is it sunny in London?

( ) 5. --- Is this _____ farm?

--- Yes, it is _____.

A. your; mine B. yours; my C. your; my

Appendix B: Cost-effectiveness calculations

	CAL run by English teachers) (1)	CAL run by non-English teachers (2)	CAL run by the research group (3)	CAL run by the government (4)
Panel A. Costs				
Programmatic costs (per student)				
Teacher training	3.40	–	4.53	–
Stipend for teachers	16.21	–	21.58	–
Software	4.81		5.99	
Cost of public funds (per student)				
Cost of taxation	4.88	–	6.42	–
Total costs (per student)				
Programmatic	24.42	–	32.09	–
Social	29.31	–	38.51	–
Panel B. Benefits				
Program effect (English score SD)	0.08	N.S.	0.18	N.S.
Panel C. Cost effectiveness (Cost of proving 0.01 English score SD)				
Programmatic	305.31	N.S.	178.29	N.S.
Social	366.37	N.S.	213.95	N.S.

Notes: N.S. = not significant

All costs in RMB (exchange rate as of Sept. 2013 was RMB6.3/USD1).

a) The cost of the teacher training was approximately 210 yuan per teacher including fixed costs of producing materials (20 yuan), teacher's time costs are calculated based on a daily adult wage of 120 yuan for eight hours of work, and two-way transportation costs to and from county seat (assumed average distance is 35 km and 1 yuan per km traveled).

b) We compensated all the teachers who instructed the program sessions with a stipend of 500 yuan per semester.

c) We assumed the cost of software was around 3,000 yuan. Cost of installing software per school by our research team is calculated as the sum of travel costs to the school seat, time costs, hotel costs and food costs (147 yuan).

d) In the absence of good estimates for China, we use a deadweight loss from taxation of 20 per cent of programmatic costs (Auriol and Warlters, 2012).

f) Effects not significant (N.S.) for the CAL run by the government group.

We also present "programmatic" cost-effectiveness (the direct monetary program costs to the implementing organization) and social cost-effectiveness calculations. We calculate social costs as the sum of: (a) programmatic costs; and (b) the cost of public funds. The key finding is that the CAL program run by the research group is the most cost-effective treatment in terms of programmatic and social costs. In this group, we calculate the programmatic cost of improving English scores by 0.01 SD to be 178.29 yuan per child (or USD28.29). The relative social cost is 213.95 yuan (USD33.96) per child. In the group where CAL was run by the English teachers, we calculate the programmatic cost of improving English score by 0.01 SD to be 305.31 yuan (USD48.46) per child. The relative social cost is 366.37 yuan (USD58.15) per child.

Appendix C: Ordinary least squares estimators of the heterogeneous impact of ICT program on standardized English test scores of students' baseline score

Dependent variable: standardized endline English test score (standard deviation)	
	(1)
[1] Treatment group 1 (1=yes; 0=no)	0.16**
	(0.07)
[2] Treatment group 1 * standardized baseline English test score	-0.03
	(0.05)
[3] Treatment group 2 (1=yes; 0=no)	-0.07
	(0.07)
[4] Treatment group 2 * standardized baseline English test score	0.00
	(0.05)
[5] Standardized baseline English test score (standard deviation)[a]	0.79***
	(0.04)
[6] Control	Yes
[7] County dummies	Yes
[8] Observations	5,253
[9] R-squared	0.642

Source: Authors' survey

* significant at 10 per cent; ** significant at five per cent; *** significant at one per cent. Robust standard errors in parentheses clustered at school level.

[a] The standardized baseline English test score is the score on the standardized English test that is given to all samples' students before the program, and it is standardized using the baseline mean and standard deviations for the control group.

References

Allcott, H, and Mullainathan, S, 2012. External validity and partner selection bias. *NBER* Working Paper No. 18373, September 2012

Almekhlafi, A.G., 2006. The effect of computer assisted language learning (CALL) on United Arab Emirates: English as a foreign language (EFL) school students' achievement and attitude. *Journal of Interactive Learning Research*, 17(2), pp.121–42.

Angrist, J. and Lavy, V., 2002. New Evidence on Classroom Computers and Pupil Learning*. *The Economic Journal*, 112(482), pp.735–65.

Bakar, A., Inal, Y., and Cagiltay, K., 2006. Use of Commercial Games for Educational Purposes: Will Today's Teacher Candidates Use them in the Future? *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Vol. 2006, No. 1, pp.1,757–62.

Banerjee, A., Cole, S., Duflo, E., and Linden, L., 2007. Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics.* 122 (3): pp.1,235–64.

Barrow, L., Markman, L., and Rouse, C.E. (2008). *Technology's edge: The educational benefits of computer-aided instruction* (No. w14240). National Bureau of Economic Research.

Bingimlas, K.A., 2009. Barriers to the successful integration of ICT in teaching and learning environments: A review of the literature. *Eurasia Journal of Mathematics, Science & Technology Education*, 5(3), pp.235–45.

Bold, T., 2013. Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education. *Center for Global Development* Working Paper No. 321

Bolton, K., and Graddol, D., 2012. English in China today. *English Today*, 28(03), pp.3–9.

Burston, J., 1996. CALL at the Crossroads: Myths, Realities, Promises, and Challenges. *Australian Review of Applied Linguistics*, 19(2), pp.27–36.

Chien, Y.C., 2011. *Effects of Computer-assisted Language Learning (CALL) Instruction on the Acquisition of Passive Grammatical Forms by Post-secondary English as a Second Language (ESL) Students*. Doctoral dissertation, University of Central Florida Orlando, Florida.

Dina, A.T. and Ciornei, S.I., 2013. The advantages and disadvantages of computer assisted language learning and teaching for foreign languages. *Procedia-Social and Behavioral Sciences*, 76, pp.248–52.

Dizon-Ross, R., Dupas, P., and Robinson, J., 2016. Governance and the Effectiveness of Public Health Subsidies. *Stanford University* Working paper, March 2, 2016

Duflo, E., Glennerster, R., and Kremer, M., 2007. Using randomization in development economics research: A toolkit. *Handbook of development economics*, 4, pp.3,895–962.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L. and Sussex, W., 2007. Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort. Report to Congress. National Center for Education Evaluation and Regional Assistance.

Ebner, M. and Holzinger, A., 2007. Successful implementation of user-centered game based learning in higher education: An example from civil engineering. *Computers & education*, 49(3), pp.873–90.

Eurydice, 2001. Basic indicators on the incorporation of ICT into European Education Systems: Facts and figures. 2000/01. Annual Report. European Commission Directorate General for Education and Culture: Brussels.

Gass, S.M. and Mackey, A., 2006. Input, interaction and output: An overview. *AILA Review*, 19(1), pp.3–17.

Glewwe, P., 2002. Schools and skills in developing countries: education policies and socioeconomic outcomes. *Journal of Economic Literature*, Volume 40, Number 2, 1 June 2002, pp. 436-482(47)

Glewwe, P.W., Hanushek, E.A., Humpage, S.D. and Ravina, R., 2011. School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010 (No. w17554). National Bureau of Economic Research.

Graddol, D. (2006). *English Next* (Vol. 62). London: British Council.

Granger, C.A., Morbey, M.L., Lotherington, H., Owston, R.D. and Wideman, H.H. (2002). Factors contributing to teachers' successful implementation of IT. *Journal of Computer Assisted Learning*, 18(4), 480–88.

Hanushek, Eric A., 1997. Assessing the effects of school resources on student performance: an update. *Educational evaluation and policy analysis* June 20, 1997 vol. 19 no. 2 141-164

Hanushek, E. A., 2002. New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement. *Journal of Labor Economics* 27, 3 (July 2009).

Hatch, E.M. (1978). *Second language acquisition: A book of readings*. Newbury House Pub.

He, F., Linden, L. and MacLeod, M., 2008. How to teach English in India: Testing the relative productivity of instruction methods within the Pratham English language education program. New York, United States: Columbia University. Mimeographed document.

Heckman, J.J., 1991. Randomization and Social Policy Evaluation. *NBER Technical* Working Paper No. 107

Hismanoglu, M., 2012. Prospective EFL Teachers' Perceptions of ICT Integration: A Study of Distance Higher Education in Turkey. *Educational Technology & Society*, 15(1), pp.185–96.

Hlas, A.C., and Vuksanovich, M. (2007). Computer assisted language learning in Spanish elementary school foreign language classrooms: The role of CD-ROMS. *Hispania*, p/769–83. www.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_711/201006/xxgk_88602.html.

Hu, G., 2005. English language education in China: Policies, progress, and problems. *Language Policy*, 4(1), pp.5–24.

Huang, W., Di, M., Yaojiang, S., Linxiu, Z., Matthew, B., Scott, R., 2014. Computer Technology in Education: Evidence from a Pooled Study of Computer Assisted Learning Programs among Rural Students in China. REAP working paper. Available at: http://reap.fsi.stanford.edu/publication/computer-technology-education-evidence-pooled-study-computer-assisted-learning-programs.

Hughes, J., 2005. The role of teacher knowledge and learning experiences in forming technology-integrated pedagogy. *Journal of technology and teacher education*, 13(2), pp.277–302.

Imbens, G.W. and Wooldridge, J.M., 2007. Cluster and Stratified Sampling. *What's New in Econometrics Lecture Note*, *8*.

Jiang, T., 2014. Other-Regarding Preferences and Other-Regarding Cheating—Experimental Evidence from China, Italy, Japan and the Netherlands. (December 1, 2014).

Jones, J. F., 2001. CALL and the responsibilities of teachers and administrators. *ELT Journal*, 55(4), pp.360–67.

Krashen, S., 1982. Principles and practice in second language acquisition (pp. 1982-1982). Pergamon: Oxford.

Lai, F., Luo, R., Zhang, L., Huang, X. and Rozelle, S., 2011. Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing. REAP working paper.

Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., Boswell, M., Rozelle, S., 2013. Computer assisted learning as extracurricular tutor? Evidence from a randomised experiment in rural boarding schools in Shaanxi. *Journal of Development Effectiveness*, 5(2), pp.208–31.

Lai, F., Zhang, L., Qu, Q., Hu, X., Shi, Y., Boswell, M. and Rozelle S., 2012. Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai, China. REAP working paper.

Lam, Y. (2000). Technophilia vs. technophobia: A preliminary look at why second-language teachers do or do not use technology in their classrooms. *Canadian Modern Language Review*, 56(3), pp.390–420.

Landsberger, H.A., 1958. Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry.

Li, S., 2002. Status and development of English teaching in minority areas (in Chinese). *Foreign Language Teaching & Research in Basic Education,* 11, pp.34–39.

Linden, L.L., 2008. Complement or Substitute? The Effect of Technology on Student Achievement in India. InfoDev.

Long, M., 1996. The role of the linguistic environment in second language acquisition. In: W.C. Ritchie and T.K. Bathia, eds, *Handbook of research on second language.* San Diego, CA: Academic Press, pp.413–68.

Loyalka, P., Liu, L., Chen, G., Zheng, X., 2014. The Cost of Disability in China. *Demography*

February 2014, Volume 51, Issue 1, pp 97-118.

McKay, S.L., 2002. *Teaching English as an international language: Rethinking goals and perspectives*. New York NY*:* Oxford University Press.

Ministry of Education, 2001. Notice of "Compulsory Education Curriculum Pilot Program". the Ministry of Education.

Ministry of Education, 2012a. Notice of “Accelerating the Process of Information Technology Integration into Education.” Ministry of Education.

Ministry of Education, 2012b. Ten-Year Development Plan for Integrating Information Technology into Education. Available at: www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s5892/201203/xxgk_133322.html [Accessed 13 March 2012].

Ministry of Education, 2012c). 12th Five-Year Plan for Integrating Information Technology into Education.

Ministry of Education, 2014. Notice of “Working Priorities for Integrating Information Technology into Education.” Ministry of Education. Available at www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s7062/201403/165870.html [Accessed 12 March 2014].

Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., and Rozelle, S., 2014a. Integrating computer-assisted learning into a regular curriculum: evidence from a randomised experiment in rural schools in Shaanxi. *Journal of Development Effectiveness*, 6(3), pp.300–23.

Mo D., Zhang L., Wang J., Huang W., Shi Y., Boswell M., Rozelle S., 2014b. The persistence of gains in learning from computer assisted learning (CAL): Evidence from a randomized experiment in rural schools in Shaanxi province in China. REAP working paper.

Mumtaz, S., 2000. Factors affecting teachers' use of information and communications technology: a review of the literature. *Journal of information technology for teacher education*, 9(3), pp.319–42.

Munshi, Kaivan, and Mark Rosenzweig, 2006. Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy. *American Economic Review*, 96(4), pp.1,225–52.

National Bureau of Statistic of China, 2010. The 6th Population Census of the People’s Republic of China, 2010.

National Bureau of Statistics of China, 2014. China National Statistical Yearbook, 2014. China State Statistical Press: Beijing, China.

Nomass, B.B. (2013). The impact of using technology in teaching English as a second language. *English Language and Literature Studies*, 3(1), p.111.

Olibie, E.I., 2010. Using computer-assisted language learning to improve students' English language achievement in universal basic education. *International Journal of Educational Research and Technology*, 1(1), pp.66–71.

Pal, J., Pawar, U.S., Brewer, E.A. and Toyama, K., 2006. The case for multi-user design for computer aided learning in developing regions. In: *Proceedings of the 15th international conference on World Wide Web*. ACM, (May), pp.781–89.

Papanastasiou, E.C. and Angeli, C., 2008. Evaluating the use of ICT in education: Psychometric properties of the survey of factors affecting teachers teaching with technology (SFA-T3). *Educational Technology & Society*, 11(1), pp.69–86.

Prashant L., Zhaolei S., James C., Natalie J., Jianguo W. and Scott R., 2014. Is the high school admissions process fair? Explaining inequalities in elite high school enrollments in developing countries. REAP working paper.

Rahimi, M. and Yadollahi, S., 2011. ICT use in EFL classes: A focus on EFL teachers' characteristics. *World Journal of English Language*, 1(2), p.17.

Rong, M. (2007). Bilingual education for China's ethnic minorities. *Chinese Education & Society*, 40(2), pp.9–25.

Rouse, C.E. and Krueger, A.B., 2004. Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program. *Economics of Education Review*, 23(4), pp.323–38.

Sauro, S., 2009. Computer-mediated corrective feedback and the development of L2 grammar, *Language Learning & Technology*, 13(1), pp.96–120.

Schwartzer, R., 1992. Self-efficacy in the adoption and maintenance of health behaviours: Theoretical approaches and a new model. *Self-Efficacy: Thought Control of Action. Washington, DC: Hemisphere*, pp.217–43.

Stockwell, G., 2007. A review of technology choice for teaching language skills and areas in the CALL literature. *ReCALL*, 19(02), pp.105–20.

Tondeur, J., Valcke, M. and Van Braak, J., 2008. A multidimensional approach to determinants of computer use in primary education: Teacher and school characteristics. *Journal of Computer Assisted Learning*, 24(6), pp.494–506.

Trucano, M., 2005. Knowledge Maps: ICTs in Education - What Do We Know about the Effective Uses of Information and Communication Technologies in Education in Developing Countries? *Non-Journal* 77, 2005

Wang, Q., 2007. The national curriculum changes and their effects on English language teaching in the People's Republic of China. In: *International handbook of English language teaching* Springer US, pp.87–105.

World Bank, 2014. World Development Indicators 2014, Washington DC: World Bank, https://openknowledge.worldbank.org/handle/10986/18237.

Yang, J., 2005. English as a third language among China's ethnic minorities. *International Journal of Bilingual Education and Bilingualism*, 8(6), pp.552–67.

Yang, S.C. and Huang, Y.F., 2008. A study of high school English teachers' behavior, concerns and beliefs in integrating information technology into English instruction. *Computers in Human Behavior*, 24(3), pp.1,085–103.

Zhao, H., 2003. Exploring Resources of Local Courses an Effective Way of Developing Teaching Body in Ethnic Minority Regions. *Journal of Research on Education for Ethnic Minorities*, (4), pp.60–64.

Zheng, H., 2009. English teaching and learning in rural China primary school (in Chinese). *Journal of Teaching and Management* (12), pp.82–83.

Publications in the 3ie Impact Evaluation Report Series

The following reports are available from http://www.3ieimpact.org/en/publications/3ie-impact-evaluation-reports/3ie-impact-evaluations/

A triple win? The impact of Tanzania's Joint Forest Management programme on livelihoods, governance and forests, *3ie Impact Evaluation Report 34.* Persha, L and Meshack, C (2016)

The effect of transfers and preschool on children's cognitive development in Uganda, 3ie Impact Evaluation Report 32. Gillian, DO and Roy, S (2016)

The effect of conditional transfers on intimate partner violence: evidence from Northern Ecuador, 3ie Impact Evaluation Report 33. Banerjee, A, Duflo E, and Gallego, F (2016)

Removing barriers to higher education in Chile: evaluation of peer effects and scholarships for test preparation 3ie Impact Evaluation Report 36. Banerjee, A, Duflo E, Gallego, F (2016)

Sustainability of impact: dimensions of decline and persistence in adopting a biofortified crop in Uganda 3ie Impact Evaluation Report 35. McNiven, S, Gilligan, DO and Hotz, C (2016)

Micro entrepreneurship support programme in Chile, 3ie Impact Evaluation Report 40. Martínez, CA, Puentes, EE and Ruiz-Tagle, JV (2016)

Thirty-five years later: evaluating the impacts of a child health and family planning programme in Bangladesh, 3ie Impact Evaluation Report 39. Barham, T, Kuhn, R, Menken, J and Razzaque, A (2016)

Can egovernance reduce capture of public programmes? Experimental evidence from India's employment guarantee, 3ie Impact Evaluation Report 31. Banerjee, A, Duflo, E, Imbert, C, Mathew, S and Pande, R (2015)

Smallholder access to weather securities in India: demand and impact on production decisions, 3ie Impact Evaluation Report 28. Ceballos, F, Manuel, I, Robles, M and Butler, A (2015)

What happens once the intervention ends? The medium-term impacts of a cash transfer programme in Malawi, 3ie Impact Evaluation Report 27. Baird, S, Chirwa, E, McIntosh, C, and Özler, B (2015)

Validation of hearing screening procedures in Ecuadorian schools, 3ie Impact Evaluation Report 26. Muñoz, K, White, K, Callow-Heusser, C and Ortiz, E (2015)

Assessing the impact of farmer field schools on fertilizer use in China, 3ie Impact Evaluation Report 25. Burger, N, Fu, M, Gu, K, Jia, X, Kumar, KB and Mingliang, G (2015)

The SASA! study: a cluster randomised trial to assess the impact of a violence and HIV prevention programme in Kampala, Uganda, 3ie Impact Evaluation Report 24. Watts, C, Devries, K, Kiss, L, Abramsky, T, Kyegombe, N and Michau, L (2014)

Enhancing food production and food security through improved inputs: an evaluation of Tanzania's National Agricultural Input Voucher Scheme with a focus on gender impacts, 3ie Impact Evaluation Report 23. Gine, X, Patel, S, Cuellar-Martinez, C, McCoy, S and Lauren, R (2015)

A wide angle view of learning: evaluation of the CCE and LEP programmes in Haryana, 3ie Impact Evaluation Report 22. Duflo, E, Berry, J, Mukerji, S and Shotland, M (2015)

Shelter from the storm: upgrading housing infrastructure in Latin American slums, *3ie Impact Evaluation Report 21*. Galiani, S, Gertler, P, Cooper, R, Martinez, S, Ross, A and Undurraga, R (2015)

Environmental and socioeconomic impacts of Mexico's payments for ecosystem services programme, *3ie Impact Evaluation Report 20*. Alix-Garcia, J, Aronson, G, Radeloff, V, Ramirez-Reyes, C, Shapiro, E, Sims, K and Yañez-Pagans, P (2015)

A randomised evaluation of the effects of an agricultural insurance programme on rural households' behaviour: evidence from China, 3ie Impact Evaluation Report 19. Cai, J, de Janvry, A and Sadoulet, E (2014)

Impact of malaria control and enhanced literacy instruction on educational outcomes among school children in Kenya: a multi-sectoral, prospective, randomised evaluation, 3ie Impact Evaluation Report 18. Brooker, S and Halliday, K (2015)

Assessing long-term impacts of conditional cash transfers on children and young adults in rural Nicaragua, *3ie Impact Evaluation Report 17*. Barham, T, Macours, K, Maluccio, JA, Regalia, F, Aguilera, V and Moncada, ME (2014)

The impact of mother literacy and participation programmes on child learning: evidence from a randomised evaluation in India, 3ie Impact Evaluation Report 16. Banerji, R, Berry, J and Shortland, M (2014)

A youth wage subsidy experiment for South Africa, 3ie Impact Evaluation Report 15. Levinsohn, J, Rankin, N, Roberts, G and Schöer, V (2014)

Providing collateral and improving product market access for smallholder farmers: a randomised evaluation of inventory credit in Sierra Leone, 3ie Impact Evaluation Report 14. Casaburi, L, Glennerster, R, Suri, T and Kamara, S (2014)

Scaling up male circumcision service provision: results from a randomised evaluation in Malawi, 3ie Impact Evaluation Report 13. Thornton, R, Chinkhumba, J, Godlonton, S and Pierotti, R (2014)

Targeting the poor: evidence from a field experiment in Indonesia, 3ie Impact Evaluation Report 12. Atlas, V, Banerjee, A, Hanna, R, Olken, B, Wai-poi, M and Purnamasari, R (2014)

An impact evaluation of information disclosure on elected representatives' performance: evidence from rural and urban India, 3ie Impact Evaluation Report 11. Banerjee, A, Duflo, E, Imbert, C, Pande, R, Walton, M and Mahapatra, B (2014)

Truth-telling by third-party audits and the response of polluting firms: Experimental evidence from India, 3ie Impact Evaluation Report 10. Duflo, E, Greenstone, M, Pande, R and Ryan, N (2013)

No margin, no mission? Evaluating the role of incentives in the distribution of public goods in Zambia, *3ie Impact Evaluation Report 9*. Ashraf, N, Bandiera, O and Jack, K (2013)

Paying for performance in China's battle against anaemia, *3ie Impact Evaluation Report 8*. Zhang, L, Rozelle, S and Shi, Y (2013)

Social and economic impacts of Tuungane: final report on the effects of a community-driven reconstruction programme in the Democratic Republic of Congo, *3ie Impact Evaluation Report 7*. Humphreys, M, Sanchez de la Sierra, R and van der Windt, P (2013)

The impact of daycare on maternal labour supply and child development in Mexico, *3ie Impact Evaluation Report 6*. Angeles, G, Gadsden, P, Galiani, S, Gertler, P, Herrera, A, Kariger, P and Seira, E (2014)

Impact evaluation of the non-contributory social pension programme 70 y más in Mexico, *3ie Impact Evaluation Report 5.* Rodríguez, A, Espinoza, B, Tamayo, K, Pereda, P, Góngora, V, Tagliaferro, G and Solís, M (2014)

Does marginal cost pricing of electricity affect groundwater pumping behaviour of farmers? Evidence from India, 3ie Impact Evaluation Report 4. Meenakshi, JV, Banerji, A, Mukherji, A and Gupta, A (2013)

The GoBifo project evaluation report: Assessing the impacts of community-driven development in Sierra Leone, *3ie Impact Evaluation Report 3*. Casey, K, Glennerster, R and Miguel, E (2013)

A rapid assessment randomised-controlled trial of improved cookstoves in rural Ghana, *3ie Impact Evaluation Report 2*. Burwen, J and Levine, DI (2012)

The promise of preschool in Africa: A randomised impact evaluation of early childhood development in rural Mozambique, *3ie Impact Evaluation Report 1*. Martinez, S, Naudeau, S and Pereira, V (2012)

The use of information and communication technology (ICT) in schools is considered a promising approach for improving learning outcomes, particularly for disadvantaged students in developing countries. However, researchers have so far found mixed evidence of the impact of ICT on students' academic outcomes. In this study, researchers examined the factors that influence the effectiveness of ICT as an educational tool in China. They found that a researcher-implemented computer-assisted learning programme for providing remedial English tutoring was more effective in improving student performance in English compared to a government-implemented programme. The study also concludes that a computer-assisted instruction programme that required teachers to provide instructions led to an improvement of English test scores, when compared to only providing the software without teaching support.

Impact Evaluation Series

International Initiative for Impact Evaluation
202-203, Rectangle One
D-4, Saket District Centre
New Delhi – 110017
India

3ie@3ieimpact.org
Tel: +91 11 4989 4444

www.3ieimpact.org