June 2018

Impact of Online Computer Assisted Learning on Education: Evidence from a Randomized Controlled Trial in China

Yu Bai, Bin Tang*, Boya Wang, Di Mo, Linxiu Zhang, Scott Rozelle, Emma Auden and Blake Mandell

Abstract

Education of poor and disadvantaged populations has been a long-standing challenge for education systems in both developed and developing countries. In China, millions of students in rural areas and migrant communities lag far behind their urban counterparts in terms of academic achievement. When they fall behind, they often have no way to catch up. Many of their parents have neither the skills nor the money to provide remedial tutoring; rural teachers often do not have time to give students the individual attention they need. Given this, there is growing interest by both educators and policymakers in helping underperforming students catch up using computer assisted learning (CAL). While CAL interventions have been shown to be effective internationally and elsewhere in China, traditional software-based CAL programs are difficult and costly to implement. An online version of CAL (OCAL), however, may be able to bypass many of offline CAL's implementation problems and enhance the remedial tutoring experience. Unfortunately, there is little empirical evidence on whether OCAL programs can be effective in improving the quality of rural primary school education in developing countries. The objective of this paper is to examine the impact of an OCAL intervention on the academic and non-academic performance of students and to explore the mechanism behind OCAL's impact. Importantly, we also aim to assess the cost effectiveness of the new OCAL program versus traditional CAL interventions. To achieve these objectives, we carried out a randomized controlled trial (RCT) involving over 1650 fifth grade students in 44 schools in rural areas and migrant communities across China. Students in the 22 treatment schools attended two 40-minute OCAL sessions during their computer class each week for one semester; the students in the other 22 schools were in the control group and did not receive any intervention. According to our findings, OCAL improved overall English scores of students in the treatment group relative to the control group by 0.56 standard deviations. This impact is large when compared with offline CAL programs. We found that OCAL also led to a positive change in the attitudes of students towards English learning and towards student aspirations for their future education level. We found three possible explanations for OCAL's impact. After rejecting the possibility of the Hawthorne Effect or self-efficacy-induced changes, we believe interest-oriented stimulation is the main source of improvement among students. The chance for comparison and competition with peers, as well as customized remedial question banks tailored to each student's individual needs, likely contributed to the measured increases in academic performance among students in our sample. Cost-effectiveness analysis showed that the OCAL program is more cost-effective than traditional offline CAL, a comparison which is significant for policymakers as it indicates high potential for OCAL program expansion.

Keywords: Education, computer assisted learning, randomized controlled trial, online learning

---

Working Paper 329

June 2018

reap.fsi.stanford.edu

Impact of Online Computer Assisted Learning on Education: Evidence from a Randomized Controlled Trial in China

Introduction

Education among poor and disadvantaged populations has been a long-standing challenge for education systems in both developed and developing countries (Hanushek & Woessmann, 2008, 2012; Pal, Pawar, Brewer, & Toyama, 2006). From a national perspective, when the share of the labor force with relatively high levels of education is low, it can slow down economic growth (Khor et al., 2016). Furthermore, an urban-rural education gap has been shown to give rise to long-term disparities in human capital acquisition and income between rural and urban populations (Li, Liu, Luo, Zhang, and Rozelle 2010; Li, Li, Wu, and Xiong, 2012), which can have implications for both growth and social stability.

In China millions of students from rural areas and migrant communities lag far behind their urban counterparts in terms of academic achievement. For example, Lai et al. (2013) found that fourth-grade students in poor, rural areas have significantly lower scores in core subject areas than do urban students. Although compulsory education mandates aim to increase the number of rural students in school and the number of years they attend, less emphasis has been placed on improving quality of education for these students (Wang, Bai, Zhang, and Rozelle, 2017; Lai et al., 2013).

One reason that rural students fall and stay behind academically is that they lack support to catch up when they begin to fall behind. When urban students fall behind, they have access to affordable remedial tutoring services from their teachers, commercial sources, or their families (Huang and Du, 2007; Lai, Luo, Zhang, Huang, and Rozelle, 2015). Students from rural areas, however, frequently have less access to these sources of remedial tutoring. Teachers often live far from school and leave immediately at the end of the school day (Lai et al., 2013). Most rural families cannot afford commercial remedial tutoring services (Lai et al., 2013). Many parents are unable to assist their children with academics because they either live and work far away from their children or they have low levels of formal education (Mo et al., 2015).

One way that educators have been able to address this problem internationally is through computer-assisted learning (CAL). CAL is a program that uses computers to offer students remedial learning materials in the form of interesting interfaces and games with the aim of improving educational outcomes, long-term interest in learning, and optimism about learning (Inal & Cagiltay, 2007; Schaefer & Warren, 2004). The CAL approach has been demonstrated to benefit the academic performance of disadvantaged students around the world, including in the US, India, and China (Goolsbee & Guryan, 2006; Barrow et al., 2009; He et al, 2008). For example, Barrow et al. (2009) found that CAL programs in Chicago improved student math scores by 0.17 standard deviations. A study in India found that CAL had a positive effect on the academic performance of students (Banerjee et al., 2007), an effect which was still observable even a full year after the project ended.

CAL also has been proven effective in China. Bai et al. (2016), Lai et al. (2013, 2015, 2016), Mo et al. (2014, 2015), and Yang et al. (2013), offer evidence from six randomized experiments showing that offline CAL can improve learning outcomes among disadvantaged rural students in mainland China. Aside from improving academic performance, CAL also has beneficial effects on non-academic outcomes. For example, CAL has been shown to improve both student interest in learning and self-efficacy in studying. Evidence from Lai et al. (2013, 2014) indicates that an after-school CAL remedial tutoring program not only improved academic performance among rural students in a short period of time, but also significantly improved student interest in schooling and levels of self-confidence. In sum, the literature clearly shows that offline CAL programs have been overwhelmingly effective in raising both academic performance and non-cognitive outcomes.

Until now, however, CAL interventions have almost all been *offline CAL* programs. In offline CAL, the main tool is software that powers a set of game-based remedial exercises and learning materials. Because it is software-based, these programs must be preloaded onto computers and are run without Internet access. In-person training workshops are also needed for both teachers and students. Being software-based and lacking interface with the internet, monitoring and evaluation all must either be done during visits by the study team to each school or through long-distance interactions between the study team and teachers (involving phone calls and/or email exchanges).

Hence, despite the effectiveness of CAL programs, there remain fundamental drawbacks to CAL: simply put, CAL is logistically complex and relatively expensive to implement. Interviews with the Principle Investigators (PIs) of earlier China-based traditional offline CAL studies revealed that at least 50 hours per school per semester were spent on the software installation, testing, monitoring, troubleshooting and maintenance required to implement and manage CAL interventions and evaluations. Furthermore, virus threats to computers were a constant source of program interruption. Finally, PIs needed to use paper-based evaluations, which required an additional substantial investment of physical and human resources. These factors make offline CAL programs expensive and time-consuming, which in turn may have dissuaded policymakers from upscaling CAL to a large number of schools.

The logistical complexity of software-based CAL has stimulated the search for alternative interventions. One strategy that should be able to retain the effective elements of offline CAL while eliminating the associated implementation difficulties is to modify traditional offline CAL into an online interactive version. In fact, online CAL (OCAL) would seem to offer several potential benefits. First, eliminating the need to manually install and maintain the software on each computer hardware would certainly lower the costs and time demands placed on the CAL team by software-related activities. Second, the ability to log in anywhere and anytime might be able to increase student access.

Third, the OCAL software system might allow us to integrate social functions into the program that could allow interaction between users (e.g., students could compete with peers on quizzes and earn virtual prizes; the system could also provide an automated ranking system that could further motivate student learning) (Borgatti and Cross, 2003; Holmes and O'loughlin, 2014).

Can OCAL programs effectively improve student learning? To our knowledge, an OCAL program focused on online remedial tutoring has never been evaluated. Even the literature on long-distance learning, which is in some ways comparable to OCAL, provides little help as its conclusions on the effectiveness of long-distance learning are mixed. Some papers have shown significant impacts from long-distance learning, citing qualities such as interactivity (Millbank, 1994), active learning (Willis, 1992; Cavanaugh, 1999), visual imagery (Ravitch, 1987), and effective communication (Horton, 1994) as reasons for its efficacy. Others have found no significant impacts and propose that long-distance learning has no advantage over traditional methods of learning (Phipps and Merisotis, 1999; Levy and Beaulieu, 2003; DiBiase, 2000). Therefore, it is difficult to propose even an estimate of the effectiveness of OCAL.

The overall goal of this paper is to examine the impact of OCAL on the academic and non-academic performance of rural and migrant students in China. To achieve this goal, we identify four particular objectives. First, we examine OCAL's impact on student academic performance on standardized English exams, using math scores to test for potential spillover effects. Second, we measure its impact on non-academic measures, in particular on student self-efficacy and student attitudes towards teachers, courses and schools. Third, we try to explain the mechanism behind OCAL's effectiveness. Finally, we conduct a cost-effectiveness analysis to determine OCAL's feasibility in relation to traditional, offline CAL interventions.

To meet these objectives, in this paper we present the results of a randomized controlled trial (RCT) involving an OCAL intervention with over 1650 fifth-grade students in 44 schools located in rural areas and migrant communities across China. The online CAL intervention took the form of a game-based software designed to complement the regular English curriculum by providing remedial tutoring. Half of the schools were designated as treatment schools, and the other half were designated as control schools. Students in treatment schools partook in the OCAL intervention by attending 40-minute OCAL sessions during computer class twice per week for one academic semester. Control students did not have any access to OCAL.

Our results show that the OCAL intervention for remedial English improved overall English scores of students in the treatment group (relative to those of the control group) by 0.56 standard deviations, which is more than triple the impact (0.15 standard deviations) found in previous offline CAL studies focused on English learning (Bai et al., 2016). We also find that OCAL improved the attitudes of students towards English class but had no significant effect on their attitudes towards their English teachers. We found three possible explanations for OCAL's impact. After rejecting the possibility that the results were due to either a Hawthorne Effect or self-efficacy-induced changes, we believe interest-oriented stimulation is the main source of improvement among students. Specially, both the chance for comparison and competition with peers, as well as customized remedial question banks tailored to each student's individual needs, may have contributedlargely to the measured increases in academic performance among students. Finally, our analysis indicates that our OCAL intervention has a higher cost-effectiveness ratio than do traditional offline CAL programs. Therefore, we conclude that the OCAL approach (at least according to our initial trial) appears to offer a more efficient solution with both larger impacts and lower costs.

The rest of this paper is organized as follows. The next section reviews the study's approach, including research design and sampling, a more in-depth explanation of the intervention, descriptions of the data collected, and the statistical approach. The remaining sections present the results, discuss the findings, and conclude.

Sampling, data and methods

Sampling and the randomization process

We conducted a clustered (at the school level) randomized controlled trial (RCT) of an online computer-assisted learning (OCAL) intervention with 1650 fifth-grade rural and migrant students in the spring semester of 2016. The RCT was carried out in 12 provinces in various regions of China, from which we randomly chose 44 elementary schools. All students in these schools were rural students; that is, they had agricultural *hukou* status. From these 44 schools, a total of 1650 students from 56 classes were involved in our study. We included only fifth-grade students in this experiment. Third-grade and fourth-grade students were excluded since they had just begun learning English, making it difficult to design effective remedial study materials and measure their impact. Sixth-grade students were excluded because they were in the process of preparing for middle school entrance exams (Davey, De Lian and Higgins, 2007).

After the baseline survey, each of the 44 schools in our sample was randomly assigned to either the treatment or control group. This assignment carefully followed a predetermined protocol. During the baseline survey, both the enumerators and the participants were unaware of the groups to which students from each school would be assigned. To ensure that the treatment and control groups were comparable in terms of key characteristics at baseline, we used Bruhn and McKenzie's method (2009) to pre-balance along several key variables for randomization. These key variables were student characteristics, including age, gender, and whether students had ever used a computer; family characteristics, including parent educational attainment and migrant status; and teacher characteristics and attitudes. For a full list of control variables balanced during the randomization process, see Appendix 1. Following Bruhn and McKenzie (2009), we re-randomized several times until the key baseline variables listed in the Appendix were balanced between the treatment and control groups. In the section below, we show the balance between the treatment and control group samples.

After randomization, 22 schools were assigned to receive the OCAL intervention. In total, 714 fifth-grade students from the 22 treatment schools were assigned to the treatment group. The rest 936 students from the other 22 schools served as the control group. Due to attrition, there were 1342 students left in our final analytic sample, with 538 from the treatment schools and 804 from the control schools. In the section below, we describe how we examined for the presence of bias among these dropouts (that is, was there a bias between treatment and control groups in terms of what types or subgroups of students dropped out).

Experimental Interventions

The intervention involved computer-assisted English remedial tutoring sessions designed to complement the entire 2016 spring semester's regular in-class English curriculum. The primary school English curriculum follows the National Standard Curriculum. To be clear, our OCAL tutoring sessions did not introduce new material separate from this curriculum, but rather provided remedial materials and questions that matched up week by week with the content of the standard English curriculum that the students were studying in class. Questions were chosen from official textbooks and exercise books with the help of primary school English teachers. Under the supervision of two teacher-supervisors trained by the research group, the students in each treatment group school had two 40-minute OCAL sessions per week.

Our protocol instructed that OCAL sessions be given during the 'computer class' period because this period is reserved for teaching non-academic material.[1] Based on our

[1] While OCAL increased time spent on English learning, it may have also taken away from time spent acquiring computer knowledge. However, principals and experts on rural education in China informed us that students generally make very minimal gains in computer-related knowledge during class, so we assume

surveys, students in the computer classes offered in most of China's rural schools are taught only basic computer operations (e.g. how to use a mouse, type in Chinese, and navigate Microsoft Office's software suite). When schools lack computer teachers, computer class time is frequently used as an open study hall where students can work on assignments from other teachers.

The OCAL program's instructional videos and games were designed to improve basic competency among students in the uniform national English curriculum. The software's content and exercises, identical for all students in the treatment group, covered the English course materials for 80 minutes of remedial tutoring (two 40-minute sessions) per week across the entire 2016 spring semester. During each session, students used computers to play English games intended to help them review and practice basic material taught in their English classes. In a typical session, the students would watch an animated video reviewing that week's in-class material before playing English games with animated characters to practice the skills covered in the video. Teachers were only allowed to assist students with scheduling, computer hardware issues, and software operations, and could not answer questions regarding the online material. According to our observations, there was little communication between students and the teacher-supervisors.

We chose English as our program subject for several reasons. First, English is one of the main subjects that is used to test students as part of the competitive exam system in

---

the loss in this tradeoff is probably small.

China that allows students to compete for positions in high school and college (Bolton & Graddol, 2012; McKay, 2002). In fact, English represents around one third of the total points in the high school entrance exam and the college entrance exam. Second, English teaching and English learning is particularly weak in poor rural China (Hu, 2005; Li, 2002; Zhao, 2003). Studies have shown that low English scores are one of the most important reasons keeping rural students from attending high school in China (Loyalka, 2014). English teachers are also of notoriously poor quality in rural China (Hu, 2005, 2009). Given the importance of English for advancing in school and the current status of English education in rural schools, we decided it would be appropriate to test an OCAL program focused on improving English skills among students.

The OCAL program that we implemented had a number of potential advantages over traditional offline CAL software. First, students could log into a website to access the OCAL platform, eliminating the necessity to install software on each individual computer. Second, the software allowed students to adjust the difficulty level of tutoring exercises, a feature which has rarely been used in traditional offline CAL programs evaluated in the past. Third, it featured social networking and gamification components designed to incentivize students to use the software more often and to provide them with interest-oriented stimulation. Using “play coins” earned by completing exercises, students were able to buy virtual gear and outfits and give virtual gifts to friends. They could also play games and engage in friendly competition with classmates to earn additional play coins.

Although not incorporated in our program, OCAL holds the possibility for several other additional advantages. For example, each student could have had a private account that allowed them to use the system anytime and anywhere. Furthermore, the OCAL software could have recorded student progress and given students and teachers frequent and convenient feedback on their performance.

Despite these advantages, there are a few characteristics of OCAL that could cause issues with implementation. Although we tested the networks of all schools in our study and found them to have sufficient internet speed to support our software, internet speeds inevitably vary over time, and a few schools reported trouble connecting to the software. Outside of our study group, there are also a number of rural schools that have computer rooms but do not have internet connections, complicating setup and access to the OCAL system. Additionally, the nature of software-based CAL makes it much easier for us to ensure that students remain focused on the CAL program content throughout the allotted time period. By contrast, OCAL requires students to access the Internet, and in doing so opens up the option for students to surf the web or play other games instead of using OCAL. In our experiment, it was one of the responsibilities of the supervisor to ensure that this did not happen.

OCAL control group

The fifth-grade students in the 22 control schools that constituted the OCAL control group were not given any intervention and attended their English, math, and computer classes as usual. To avoid any spillover effects related to the OCAL intervention, the principals, teachers and students (and their parents) at the control schools were not informed of the OCAL project. The research team visited the control schools only during the baseline and final evaluation surveys.

Data collection

The research group conducted two rounds of surveys of all fifth-grade students at both the treatment and control schools. The first-round survey was a baseline survey conducted before the OCAL program's implementation began in March 2016. The second-round survey was a final evaluation survey conducted in June 2016 at the end of the OCAL program, which coincided with the end of the semester.

In each round of surveys, the enumeration team visited each school and conducted a two-block survey. The first block measured student academic performance and collected data that allowed the assessment of non-cognitive skills. The second block collected data on the characteristics of students and their families.

To measure academic proficiency, the first block of the survey contained two standardized tests—first a 104-question English test and then a 33-question math test (see Lai et al., 2015).[2] Each of these tests had a 30-minute time limit. Our enumeration team proctored the tests, strictly enforcing the time limits and ensuring that there was no cheating. We use English scores as the primary measure of student academic

[2] We invited experts and local teachers to help us pick questions from official examination books and exercise books. We tested the reliability and validity of the test questions in rural schools outside of our sample. Students were required to finish tests in each subject in 30 minutes.

performance and math test scores as a proxy for spillover effects (positive, negative or zero) caused by the English OCAL program. Spillover effects were tested for two reasons. First, the math test could be thought of as a placebo. If there were impacts due to the Hawthorne effect (discussed more below) one might expect math scores to rise as much as English scores. Second, it might be that the use of the computer class time (which at times in some schools was used as a study hall for math and other subjects) for OCAL could have a dampening effect on math scores.[3]

The two non-cognitive outcome variables collected during the first block were self-efficacy and student attitudes towards teachers, courses, and schools. Self-efficacy, the core concept in Bandura's social cognitive theory (1997), indicates the degree to which people are confident in their ability to use the skills they have to perform particular tasks. We measured self-efficacy because we thought it might be one of the channels through which OCAL might affect academic performance. An individual high in self-efficacy believes he can handle various challenges by taking appropriate action. Self-efficacy thus reflects one's sense of control over the environment. In contrast with outcome expectation, which refers to an individual's belief about the likelihood that an action will lead to a specific outcome (Maddux, Sherer, and Rogers, 1982), self-efficacy refers to a subjective rating given by an individual regarding the control he has over his

[3] As mentioned in a previous footnote, we find that we probably intend to over-estimate the impact of OCAL on English. It should be noted that if our results reveal a negative impact on math scores, then OCAL-related gains in English proficiency may be due to a learning tradeoff in which more time is spent learning English at the expense of other subjects. By contrast, if there is no impact on math, it could be because math scores are pulled up by the Hawthorne effect and back down by the tradeoff in class time. Of course, we are not able to identify which combination of these factors are at play in our study.

actions. Self-efficacy enhances cognitive processes and correlates with various other outcomes, including decision-making quality (Bandura, 1997). Because self-efficacy is widely used in research on school environments, mood disorders, mental and physical health, and career choices, it has become a major variable of study in various psychological fields (e.g., clinical, personality, educational, social, and health).

In this paper, we adopt the General Self-Efficacy Scale (GSES), developed in 1981 by clinical psychologist Ralf Schwarzer, to measure self-efficacy (Schwarzer, 1997). The Chinese version of GSES has been shown to have good reliability and validity (Wang and Liu, 2000). The GSES has 10 questions relating to self-confidence levels of individuals when encountering setbacks or difficulties. It adopts a 4-point Likert Scale, wherein students score each item 1-4 to indicate their agreement with questions or statements (numbers 1-4 represent "completely incorrect", "slightly correct", "mostly correct", or "completely correct").

Student attitudes toward teachers, courses, and schools were measured via subjective ratings by students of how much they liked their teachers, courses, and schools. Students rated their preferences on a Likert-style scale between 0 and 100, with 0 meaning strongly dislike and 100 meaning strongly like.

In the second block of the survey, enumerators collected data on the characteristics of students and their families which we used to create demographic and socioeconomic control variables. The dataset included measures of each student's *age* (measured in years), gender (defined by a variable *female*, which is equal to one for girls and zero for boys), grade, county, and whether the *student is an only child*. To gather data on the student's family and household, the survey included measures of their father's education level (*father has at least a high school degree*), mother's education level (*mother has at least a high school degree*), whether their parents are still farmers or work off the farm, and poverty status (whether one receives a poverty subsidy at school). To create indicators of parental care, the survey also asked whether one or both of the student's parents migrated to some other place or if they stayed at home for most of the time during the semester. We also collected data on the characteristics of teachers, including teacher gender (1=female, 0=male), teacher age (years), whether the teacher was a civil teacher (1=civil teacher, 0=contract teacher), English teaching hours per week, years of work experience, educational level (1=college and above; 0=below college), and teacher attitude towards computers. We also collected information on the basic characteristics of the school, such as *area* (sq. km) and *computer-student ratio*.

Students who were present during both the baseline and endline surveys became the sample used to assess OCAL's impact on standardized test scores. During the process of creating the data set, we also created a binary variable for each student called *attrition*, which equaled one if a student took the baseline survey but not the endline survey and equaled zero if the student completed both. Students that did not complete the baseline survey were excluded from the analysis. We examine below how *attrition* related to whether a student was in the control or treatment group. We also check whether attrition was biased toward certain subgroups of students in either the treatment or control group.

Statistical methods

The first step in our statistical analysis was to examine the overall impact of OCAL on the average standardized test scores of the students in our sample. We used both unadjusted and adjusted ordinary least squares (OLS) regression analyses to estimate how the standardized test score outcomes changed in the treatment group relative to the control group. Our unadjusted analysis examined the relationship between a dummy variable for the treatment (OCAL intervention) status and changes in test scores pre-program to post-program. Our adjusted analysis controlled for systematic differences between the treatment and control groups and tested for heterogeneous treatment effects (for more details, see the models below). All regressions accounted for the clustered nature of our sample by constructing Huber-White standard errors to correct for school-level clustering (thus relaxing the assumption that disturbance terms are independent and identically distributed within schools).

The unadjusted model that we use in the paper is:

$$y_{isc} = \alpha + \beta \cdot treatment + \theta * y_{0isc} + \varepsilon_{isc}, \quad (1)$$

where y_{isc} is the outcome variable after the OCAL program for child i in class c at school s, *treatment* is a dummy variable for a student attending a treatment school (i.e., equal to one for students in the treatment group and zero otherwise), and ε_{isc} is a random disturbance term clustered at the school level. The outcome variable y_{isc} includes both student academic performance and non-cognitive outcomes. Student academic performance, as measured by scores on standardized English and math tests, is the primary variable on which this study focuses. The non-cognitive outcomes that this study measures are student self-efficacy in learning and student attitudes towards classes, teachers, and schools. Because this study aims to track the learning progress of students, equation (1) also includes y_{0isc}, the baseline academic performance and non-cognitive variables for student *i* in class *c* at school *s*.

To improve the statistical efficiency of the estimation, we add a set of control variables to the unadjusted model in equation (1) to build an adjusted model:

$$y_{isc} = \alpha + \beta \cdot treatment + \theta * y_{0isc} + \gamma * X_{isc} + \delta * T_{isc} + \sigma * S_{isc} + \varepsilon_{isc}, \qquad (2)$$

where all the variables and parameters are the same as those in equation (1), except that we add a vector of additional control variables. The additional control vector contains variables that describe student and family characteristics (X_{isc}: student gender, age, ethnicity, only-child family, family size, family off-farm, at least one parent lives away from home, family assets, ever used a computer); English teacher characteristics (T_{isc}: gender, age, civil teacher, educational level, years of work experience, English teaching hours per week, attitude towards computers); and school characteristics (S_{isc}: area of school, computer-student ratio, boarding or non-boarding school). By including these control variables, equation (2) can more effectively estimate the OCAL treatment effect (*treatment*).*Attrition and validity of randomization*

Although at baseline there was a total of 1650 students, there was an overall attrition rate of 18.7%. For various reasons (mainly school transfers and extended absences due to illness or injuries) during the evaluation survey we only followed up with

1342 students, comprised of 538 from the treatment schools and 804 from the control schools.

To figure out whether attrition affected the composition of the total sample or the validity of randomization, we compared attrition rates between the treatment and control group students as shown in Table 1. The results demonstrate that the attrition pattern does not differ between the treatment and control groups and that attrition rates are not affected by treatment status (Table 1). In conducting the test, we estimated equation (1) with the attrition status as the dependent variable and without control variables. The results show attrition rates are not correlated with treatment status when pooling the students (column 1, row 1).

Our analysis indicates that there were no significant differences in student characteristics between the treatment and control groups before attrition (Table 2). Similarly, after dropping data from the attrited students, we checked the validity of the randomization by running a regression with the OCAL treatment dummy variable. We found similar results among sample students before and after attrition (Table 3). In other words, student characteristics are well-balanced between the treatment and control groups both before (during the baseline) and after attrition (at the endline).

Results

The impact of OCAL on English academic performance

To estimate OCAL's impact, we ran ordinary least squares (OLS) regressions on the unadjusted model (1) and adjusted model (2) using student scores on standardized English and math tests as indicators of student academic performance. The main results are reported in Table 4.

According to the analysis, OCAL had a large, statistically significant impact on English learning among students. Specifically, results from the unadjusted model (1), which does not include control variables, indicate that the treatment improved student English scores by 0.45 SDs (significant at the 5% level). Results from the adjusted model (2), which adds control variables including individual, class and school characteristics, demonstrate that OCAL improved student English scores by 0.56 SDs (significant at the 1% level).

Our results suggest that this particular OCAL intervention had an effect size more than three times larger than that of previous, offline CAL initiatives on student academic performance (Figure 2). This effect size is quite large. Qinghai's offline CAL English program, which offered remedial tutoring for the same textbook using the same user-interface as this OCAL intervention, led to 0.15 SD improvements in student English scores (Bai et al., 2016). [4]

Explaining the impact of OCAL on student English academic performance

[4] It should be noted that the previously-mentioned English CAL program in Qinghai was focused on low-performing minority students, and it is unclear how much of the difference between our results might be attributable to differences in experimental setting.

It is of interest to explore why OCAL has such a relatively larger impact on student academic performance. We propose several possible explanations and evaluate the validity of each one.

Possibility 1: Hawthorne Effect

The Hawthorne effect refers to the tendency of individuals who are aware that they are being observed by or cared for by others to change their behavior and perform better even though they were not being treated with any additional intervention (Parsons, 1974). One possible explanation for the success of this OCAL program is that the project team's intervention led treatment-group students to feel cared for, and thus motivated these students to academically outperform control-group students on various tests. In other words, if the Hawthorne effect was at play in our experiment, it could have meant that the test scores of the treatment-group students could have improved because their awareness of being observed, which led them to try harder in school and the observed impact was not due to OCAL.

To address this possibility, analysis of standardized math test scores show that OCAL English tutoring had no statistically significant impact, either positive or negative, on student math performance (Table 5). The fact that English scores improved while there were no effects on math scores would provide evidence that treatment-group students were impacted by the Hawthorne effect. Also, control-group students and their teachers were unaware of their involvement in a study. This means that the control-group students also were not impacted by the Hawthorne effect. Thus, we believe that our data support the conclusion that it was not the Hawthorne effect that led to OCAL's impacts.

Possibility 2: Self-efficacy

Educators have recognized that student beliefs about their academic capabilities play an essential role in their motivation to achieve, thus indicating self-efficacy may be a mediating factor in academic performance (Zimmerman, 2000). Various studies have supported a positive, statistically significant relationship between self-efficacy and academic improvement (Pintrich and De Groot, 1990; Lent, Brown, Larkin, 1986; Chemers, Hu, and Garcia, 2001; Choi, 2005). In previous evaluations of offline CAL programs, improvements in student self-efficacy and math scores have tended to occur together (Mo et al., 2013; Mo et al., 2015). Although it is unknown whether student self-efficacy causes improved academic achievement or vice versa, previous studies have demonstrated a link between them.

Our results indicate that the OCAL intervention had no significant effect on student self-efficacy (Table 6). Such a finding would suggest that, unlike earlier studies (which focused on the impact of CAL on math scores), there may not be a link between self-efficacy and improved academic achievement (in general). In other words, the rise in test scores from English OCAL may have a different mechanism than offline CAL's mechanism for math learning.

We propose three possible explanations for the lack of improvement in self-efficacy. The first explanation relates to the nature of language study. Because the students in our study are learning English as a second language, its unfamiliarity may provoke anxiety or even hostility amongst them or be emotionally difficult due to the ties between language and identity (Cohen and Norst, 1989). A lack of motivation has also been shown to be a barrier for beginners learning Japanese or Spanish (Clément and Kruidenier, 1985). Thus, the increased exposure to English may, at least in the short term, provoke negative or mixed emotions from students despite improving their proficiency. Such a response would not be expected to be associated with a rise of self-efficacy.

The second explanation lies in the differing nature of language and math study. Learning math requires students to exercise their logical thinking abilities. Improvement in logical thinking abilities through the study of math may also lead to improvement in decision-making abilities, which in turn can increase self-efficacy (Schoenfeld, 1992; Haggarty & Pepin, 2002). However, learning English is more likely to enhance memorization and recall skills, and improvements in English performance does not necessarily improve decision-making ability, so there is no concurrent improvement in self-efficacy.

The third explanation is that self-efficacy is difficult to change within a short time frame, which has been demonstrated by multiple RCTs (Dunn, Andersen and Jakicic, 1998; Turner, Mancl and Aaron, 2006). Our intervention frame that was just given in fifth grade was likely insufficient to change student self-efficacy, as evidenced by the strong correlation between endline self-efficacy scores and baseline scores (Table 6, row 2).

Possibility 3: Interest-oriented stimulation

In this sub-section we seek to explain why OCAL might be more effective than traditional, offline CAL.Due to its ability to offer interest-oriented stimulation, we propose two reasons to explain this source of change. Both are related to the online-based features that are only possible with OCAL (and not traditional CAL programs) ..

First, the OCAL software offers both motivating peer interactions and the use of the Internet to acquire and update learning resources in a data-informed manner. Unlike traditional, offline CAL, OCAL's online-based social features engage students in competitive, gamified activities with peers, which increases student engagement with the remedial tutoring software. Two students using the OCAL software can compete to answer a question based on materials from their English textbooks. The first competing student to correctly answer a question will receive 'golden coins' (virtual tokens used to measure student progress in the software), which can be used to give gifts to others and play small online games (e.g., Picture Matching and Bubble Round) during breaks from the remedial tutoring exercises. The social and competitive elements of OCAL may be one of the reasons for the large measured impact (Festinger, 1954; Chen, Harper, Konstan and Xin, 2010).

Second, OCAL's ability to provide data-informed exercises based on each individual student's level may increase student learning. The online software automatically chooses appropriate tutoring questions based on student answer history and scores from previous rounds to provide appropriately difficult material. This practice enables students to review what they have learned, correct mistakes, and gain new knowledge at their own pace. Additionally, students can restart rounds to obtain higher scores. Thus, students will spend more time learning and reviewing relevant material and less time feeling frustrated or bored with material that is too hard or too easy.

Support for the interest-oriented stimulation hypothesis also is provided by examining the effects on student's attitudes towards their English learning program. When examining the OCAL intervention's effects on student attitudes toward their courses, teachers, and schools, theresults indicate that student attitudes toward their English class significantly improved after the OCAL intervention (Table 7, Columns 3 & 4) while student attitudes toward their English teachers did not significantly change (Table 7, Columns 1 & 2).

What account for this is because the interactive, game-based OCAL software appears to have attracted the attention of students and made learning English more enjoyable, it may be that students began to enjoy their courses more. However, because of the logistical implementation of the OCAL intervention—with computer teachers helping students use the software and classroom English teachers being uninvolved in its administration—it perhaps is understandable that the improved attitudes among students toward their English courses are not correlated with a change in attitudes toward their English teachers.

Other factors

One final factor that may help explain OCAL's effectiveness is that the difference between the baseline English scores of rural students and their scores in other subjects mean that there is room for easy-to-access gains in English proficiency. Although Mandarin, math and English are all core subjects taught in Chinese primary schools, disparities in the quality of teachers and facilities along with varying academic prioritization contribute to students having different baseline scores in each subject. Loyalka et al. (2017) investigated the rural-urban educational gap in proficiency levels on core subjects, finding that the average rural-urban gap in math is 0.51 SD whereas the gap in English is 0.86 SD. We assume some of OCAL's unprecedented impact lies in the disproportionately low baseline English proficiency of rural students. A lower starting point may mean more of a possibility that the OCAL program can effect a large improvement. Similar to "economic convergence" (Quah, 1996; Button, 1998) which explains why countries with low indices of development can grow faster than those with high indices, CAL interventions have most effectively improved English scores among rural students (in comparison to other core subject scores, as seen in Figure 2), which have a significantly lower baseline level than student scores in math or Mandarin. We believe this phenomenon may partially explain OCAL's large impact on English performance.

In summary, we believe the difference between OCAL and offline CAL's impact can be explained by the low baseline English level of students in our sample, OCAL's interactive nature, and its ability to offer targeted exercises for individual users. We also have tried to demonstrate how the Hawthorne effect and changes in self-efficacy do not appear to be the reason for OCAL's effectiveness. In short, the effectiveness of the program in this study, at least in part, appears to come from interest-oriented stimulation. Although our study cannot conclusively link OCAL's success to the opportunity for interaction between online users and the system's data-driven responses to varying student proficiency, we believe these factors can explain part of OCAL's effectiveness.

Cost-effectiveness analysis

To provide a reference for the expansion of the OCAL program, we analyze the project's cost-effectiveness to better understand its potential for implementation on a larger scale (Table 9). We used the same cost-effectiveness methodology developed in Bai (2016), which calculates a scaled-up program cost according to the original project's actual cost. The program's main costs are from teacher training and class subsidies, as well as software design, development, updates, and maintenance. Our calculations assume a class size of 30 students (given the study's average class size of 30.5 students).

The cost to train teachers includes communication costs (3 instances ∗ 10 RMB/instance = 30 RMB), training materials (20 RMB), and trainer remuneration (30 RMB). The teacher training subtotal is 80 RMB/teacher, which is equivalent to 2.67 RMB/student. Class subsidies are given to program teachers for implementing the intervention in class, and cost 2 classes/week ∗ 12 weeks ∗ 50 RMB/class = 1200 RMB/teacher. This comes out to 1200 / 30 = 40 RMB/student. The cost to design and develop the software is a one-time expenditure of 100. Assuming that the software will last for 5 years, its per-student unit cost is 100,000 RMB / 5 years / 44 classes / (30 students/class) = 15.15 RMB/student. Software updates and maintenance (which includes renting servers and debugging) costs 3000 RMB / 44 classes / (30 students/class) = 2.27 RMB/student. Adding these together, the total software cost is 15.15 + 2.27 = 17.42 RMB/student. We can approximate the public resource investment as 20% of the program execution cost (Auriol and Warlters, 2012). Social costs include costs of the program's execution and public resource investment.

As shown in the Table 9 figures, we compute the OCAL program cost per standard deviation raised to be 128.77 RMB/student in total, whereas that of offline CAL per standard deviation raised is 213.95 RMB/student. Thus, according to our analysis, the OCAL program is shown to be much more cost-effective than traditional offline CAL. The comparison should be significant for policymakers as it indicates high potential for OCAL program expansion.

Conclusion

In this paper we present the results from a randomized field experiment on an OCAL program involving 1650 students in fifth grade. The main intervention was an English OCAL remedial tutoring program that was held outside of the students' regular English course. Our results indicate that OCAL has significant beneficial effects on academic outcomes. Two 40-min OCAL English sessions per week increased student standardized English scores by 0.56 standard deviations. OCAL also significantly increased student interest in learning. The effect of OCAL on academic performance was less significant for students whose English teachers were contract teachers.

This paper contributes in several ways to our understanding of OCAL's potential impact in developing countries. First, we are the first to measure the effect of OCAL on learning outcomes among an underserved population in a developing economy; we are also the first to explore in depth its advantages over traditional offline CAL. Second, in order to estimate the effect of the OCAL program, we tried to design our OCAL program implementation and evaluation protocol to prevent potentially confounding influences from affecting the impacts. Furthermore, we administered math tests to create a placebo, allowing us to examine for the Hawthorne effect and spillover.

This paper also contributes to our understanding of the mechanism of OCAL. We found three possible explanations for the impact, and after rejecting the possibility of the Hawthorne Effect and self-efficacy-induced changes, we believe interest-oriented stimulation is the main source of improvement among students. The chance for comparison and competition with peers, as well as customized remedial question banks tailored to each student's individual needs, both may be contributing to the measured increases in academic performance among students.

One limitation of the paper is that we have not been able to quantify precisely why the OCAL intervention was so much larger than previous traditional CAL interventions. We believe that the nature of engagement, the interaction among users, and other sources of features of the program that we believe contribute to interest-oriented stimulation as well as the ability of game to adjust to be customized to the student's learning level may behind the impressive measured impact. However, which component accounts for how much of the gain is beyond the scope of this paper and may indeed form the basis of an important research project in the future.

Our results also show that OCAL is cost-effective and may be a practical option (relative to traditional CAL) for policymakers looking to use computer technology to help solve educational disparities. Based on our analysis, OCAL is significantly more cost-effective than traditional offline CAL. Education policymakers in China (and in other developing countries, as well as underserved communities in developed countries) who are considering implementing large-scale CAL programs might consider OCAL as a feasible, low-cost alternative. Furthermore, the cost-effectiveness of OCAL has significant implications for China, as the Chinese government has committed to make large investments in computing facilities and internet access in rural schools. This paper demonstrates that an OCAL program could be used as a complementary input to existing computer resources and has the potential to narrow the urban-rural achievement gap and help disadvantaged populations.

References

Bai, Y., Mo, D., Zhang, L., Boswell, M., & Rozelle, S. (2016). The impact of integrating ICT with teaching: Evidence from a randomized controlled trial in rural schools in China. Computers & Education, 96, 1-14.

Bandura, A. (1997). Self-efficacy: The exercise of control. Macmillan.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. The Quarterly Journal of Economics, 122(3), 1235-1264.

Barrow, L., Markman, L., & Rouse, C. E. (2009). Technology's edge: The educational benefits of computer-aided instruction. American Economic Journal: Economic Policy, 1(1), 52-74.

Borgatti, S. P., & Cross, R. (2003). A relational view of information seeking and learning in social networks. Management science, 49(4), 432-445.

Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. American economic journal: applied economics, 1(4), 200-232.

Button, K. (1998). Infrastructure investment, endogenous growth and economic convergence. The annals of regional science, 32(1), 145-162.

Cavanaugh, C. S. (1999). The effectiveness of interactive distance education technologies in K-12 learning: A meta-analysis.

Chemers, M. M., Hu, L. T., & Garcia, B. F. (2001). Academic self-efficacy and first year college student performance and adjustment. Journal of Educational psychology, 93(1), 55.

Chen, Y., Harper, F. M., Konstan, J., & Li, S. X. (2010). Social comparisons and contributions to online communities: A field experiment on movielens. American Economic Review, 100(4), 1358-98.

Choi, N. (2005). Self‐efficacy and self‐concept as predictors of college students' academic performance. Psychology in the Schools, 42(2), 197-205.

Clément, R., & Kruidenier, B. G. (1985). Aptitude, attitude and motivation in second language proficiency: A test of Clément's model. Journal of language and Social Psychology, 4(1), 21-37.

Cohen, Y. M. J. N., & Norst, M. J. (1989). Fear, dependence and loss of self-esteem: Affective barriers in second language learning among adults. RELC Journal, 20(2), 61-77.

Davey, G., De Lian, C., & Higgins, L. (2007). The university entrance examination system in China. Journal of further and Higher Education, 31(4), 385-396.

DiBiase, D. 2010. Nature of geographic information: An open geospatial textbook. College of Earth and Mineral Sciences Open Educational Resources Initiative, The Pennsylvania State University. Http://natureofgeoinfo.org.

Dunn, A. L., Andersen, R. E., & Jakicic, J. M. (1998). Lifestyle physical activity interventions: History, short-and long-term effects, and recommendations. American journal of preventive medicine, 15(4), 398-412.

Festinger, L. (1954). A theory of social comparison processes. Human relations, 7(2), 117-140.

Goolsbee, A., & Guryan, J. (2006). The impact of Internet subsidies in public schools. The Review of Economics and Statistics, 88(2), 336-347.

Haggarty, L., & Pepin, B. (2002). An investigation of mathematics textbooks and their use in English, French and German classrooms: Who gets an opportunity to learn what?. British educational research journal, 28(4), 567-590.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. Journal of economic literature, 46(3), 607-68.

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. Journal of economic growth, 17(4), 267-321.

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

Holmes, K. M., & O'loughlin, N. (2014). The experiences of people with learning disabilities on social networking sites. British Journal of Learning Disabilities, 42(1), 1-5.

Horton, W. (1994). How we communicate. Paper presented at the meeting of the Rocky Mountain Chapter of the Society for Technical Communication, Denver, CO, June,1994

Huang, A., and X. Du. 2007. “Comparative Analysis of Urban and Rural Difference in Family Education in China.” Journal of Yibing University (in Chinese) 1: 107–110.

Inal, Y., & Cagiltay, K. (2007). Flow experiences of children in an interactive social game environment. British Journal of Educational Technology, 38(3), 455-464.

Khor, N., Pang, L., Liu, C., Chang, F., Mo, D., Loyalka, P., & Rozelle, S. (2016). China's Looming Human Capital Crisis: Upper Secondary Educational Attainment Rates and the Middle-income Trap. The China Quarterly, 228, 905-926.

Lai, F., Luo, R., Zhang, L., Huang, X., & Rozelle, S. (2015). Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing. Economics of Education Review, 47, 34-48.

Lai, F., Zhang, L., Bai, Y., Liu, C., Shi, Y., Chang, F., & Rozelle, S. (2016). More is not always better: evidence from a randomised experiment of computer-assisted learning in rural minority schools in Qinghai. Journal of Development Effectiveness, 8(4), 449-472.

Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., ... & 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), 208-231.

Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. Journal of counseling psychology, 33(3), 265.

Levy, S., & Beaulieu, R. (2003). Online distance learning among the California community colleges: Looking at the planning and implementation. The American Journal of Distance Education, 17(4), 207-220.

Li, H., Li, L., Wu, B., & Xiong, Y. (2012). The end of cheap Chinese labor. Journal of Economic Perspectives, 26(4), 57-74.

Li, X., Liu, C., Luo, R., Zhang, L., & Rozelle, S. (2010). The challenges facing young workers during rural labor transition. China Agricultural Economic Review, 2(2), 185-199.

Loyalka, P., Chu, J., Wei, J., Johnson, N., & Reniker, J. (2017). Inequalities in the pathway to college in China: When do students from poor areas fall behind?. The China Quarterly, 229, 172-194.

Maddux, J. E., Sherer, M., & Rogers, R. W. (1982). Self-efficacy expectancy and outcome expectancy: Their relationship and their effects on behavioral intentions. Cognitive Therapy and Research, 6(2), 207-211.

Merisotis, J. P., & Phipps, R. A. (1999). What's the difference?: Outcomes of distance vs. traditional classroom-based learning. Change: The Magazine of Higher Learning, 31(3), 12-17.

Millbank, G. (1994). Writing multimedia training with integrated simulation. Paper presented at the Writers' Retreat on Interactive Technology and Equipment.

Vancouver, BC: The University of British Columbia Continuing Studies.

Mo, D., Huang, W., Shi, Y., Zhang, L., Boswell, M., & Rozelle, S. (2015). Computer technology in education: Evidence from a pooled study of computer assisted learning programs among rural students in China. China Economic Review, 36, 131-145.

Mo, D., Swinnen, J., Zhang, L., Yi, H., Qu, Q., Boswell, M., & Rozelle, S. (2013). Can one-to-one computing narrow the digital divide and the educational gap in China? The case of Beijing migrant schools. World development, 46, 14-29.

Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., ... & Rozelle, S. (2014). Integrating computer-assisted learning into a regular curriculum: evidence from a randomised experiment in rural schools in Shaanxi. Journal of development effectiveness, 6(3), 300-323.

Pal, J., Pawar, U. S., Brewer, E. A., & Toyama, K. (2006, May). The case for multi-user design for computer aided learning in developing regions. In Proceedings of the 15th international conference on World Wide Web (pp. 781-789). ACM.

Parsons, H. M. (1974). What Happened at Hawthorne?: New evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies. Science, 183(4128), 922-932.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of educational psychology, 82(1), 33.

Quah, D. T. (1996). Twin peaks: growth and convergence in models of distribution dynamics. The economic journal, 1045-1055.

Ravitch , D. (1987). Technology and the curriculum. In M.A. White (Ed.)

Schaefer, S., & Warren, J. (2004). Teaching computer game design and construction. Computer-Aided Design, 36(14), 1501-1510.

Schonfeld, I. S. 1992. "Assessing stress in teachers: Depressive symptoms scales and neutral self-reports of the work environment". In Stress and well-being at work: Assessment and interventions for occupational mental health, Edited by: Quick, J. C., Murphy, L. R. and Hurrell, J. J. Jr. 270–285. Washington, DC: American Psychological Association.

Schwarzer, R., Bäßler, J., Kwiatek, P., Schröder, K., & Zhang, J. X. (1997). The assessment of optimistic self□beliefs: comparison of the German, Spanish, and Chinese versions of the general self□efficacy scale. Applied Psychology, 46(1), 69-88.

Turner, J. A., Mancl, L., & Aaron, L. A. (2006). Short-and long-term efficacy of brief cognitive-behavioral therapy for patients with chronic temporomandibular disorder pain: a randomized, controlled trial. Pain, 121(3), 181-194.

Wang, X., Bai, Y., Zhang, L., & Rozelle, S. (2017). Migration, Schooling Choice, and Student Outcomes in China. Population and Development Review, 43(4), 625-643.

Wang, C., Hu, Z., Liu, Y. (2001). Evidences for Reliability and Validity of the Chinese

Version of General SelfEfficacy Scale. Chinese Journal of Applied Psychology, 1, 006.

Willis, B. (1992). Instructional development for distance education. (ERIC Document Reproduction Service No. ED 351 007).

Yang, Y., Zhang, L., Zeng, J., Pang, X., Lai, F., & Rozelle, S. (2013). Computers and the academic performance of elementary school-aged girls in China's poor communities. Computers & Education, 60(1), 335-346.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. Contemporary educational psychology, 25(1), 82-91.

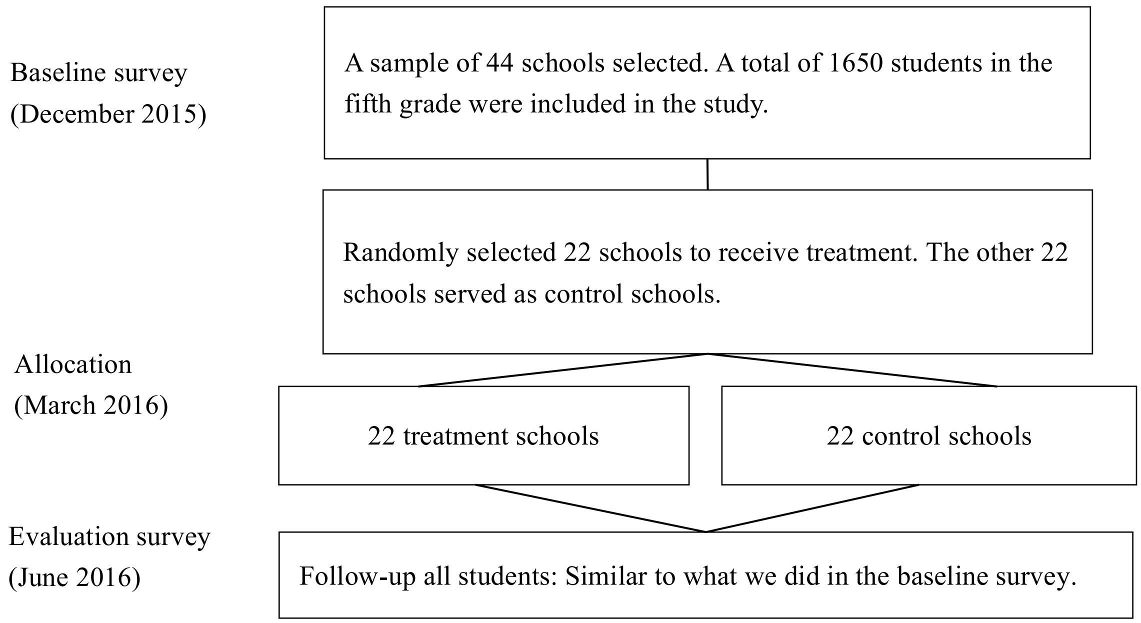
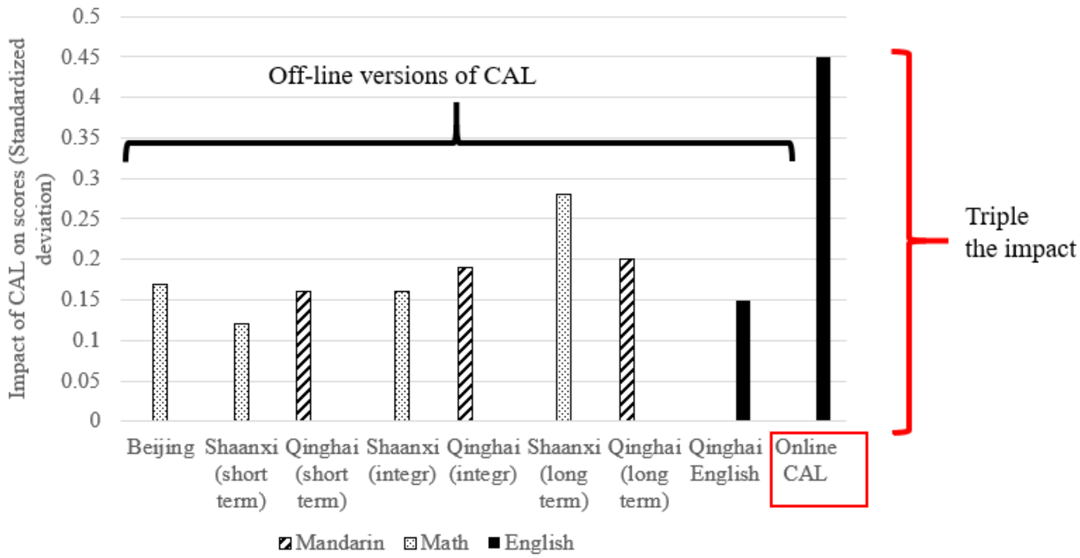


Figure 1. Experiment profile



Sources: Our research team's previous publish paper on CAL

Figure 2. Impact of CAL on scores: off-line compared with online.

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

VARIABLES	attrition[a] (1=students attrited; 0=student remained) (1)
[1] OCAL treatment (1=yes; 0=no)	0.11 (0.11)
[2] Observations	1650
[3] R-squared	0.02

Notes: *significant at 10%; ** significant at 5%; *** significant at 1%.
Robust standard errors in parentheses clustered at school level.
[a] The test aims to show whether attrition rates are different between the treatment and control groups. The test regresses attrition status on the treatment variable.

Table 2. Comparison of the student characteristics between the treatment and control groups within baseline students.

Variable		Difference between treatment and control groups within baseline	
		OCAL treatment (1)	
Students characteristics			
[1]	Standardized baseline English test score (standard deviation)	0.12	(0.22)
[2]	Standardized baseline math test score (standard deviation)	-0.07	(0.20)
[3]	Female (1=yes; 0=no)	0.02	(0.02)
[4]	Student's age (years)	0.07	(0.11)
[5]	Ethnic (1=Han; 0=minority)	-0.02	(0.03)
[6]	Computer use (1=ever used computer; 0=never used computer)	-0.04	(0.04)
[7]	Like English classes (0-100 points)	8.29*	(4.16)
[8]	Like English teacher (0-100 points)	6.39*	(3.72)
[9]	Like school (0-100 points)	0.31	(2.06)
[10]	Education level students want to achieve (1=college or above; 0=below college)	0.00	(0.04)
Family characteristics			
[11]	Asset index	0.06	(0.14)
[12]	Only child (1=yes; 0=no)	-0.09	(0.07)
[13]	father farm's everyday (1=yes; 0=no)	0.05	(0.08)
[14]	mother farm's everyday (1=yes; 0=no)	0.09	(0.08)
[15]	At least one parent has junior high school degrees or above (1=yes; 0=no)	0.01	(0.05)
[16]	Family size bigger than 5 (1=yes; 0=no)	-0.00	(0.03)
[17]	At least one parent lived out of home (1=yes; 0=no)	0.02	(0.05)
Teacher and school characteristics			
[18]	teacher is female (1=yes; 0=no)	-0.03	(0.09)
[19]	English teacher's age (years)	-0.32	(2.15)
[20]	Civil service teacher (1=yes; 0=no)	-0.17	(0.12)
[21]	English teaching hours per week	1.11	(1.40)
[22]	work year (year)	0.39	(2.39)
[23]	Education level of English teacher (1=College and above; 0=below college)	-0.27	(0.36)
[24]	Teachers attitudes towards computer (1-5 points)	0.07	(0.11)
[25]	Area (sq.km)	-0.00	(0.01)
[26]	Computer-student ratio	0.01	(0.02)
[27]	Rural school (1=yes; 0=no)	0.12	(0.18)
[28]	Observations	1650	

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

*significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3. 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	
		OCAL treatment (1)	
Students characteristics			
[1]	Standardized baseline English test score (standard deviation)	0.09	(0.22)
[2]	Standardized baseline math test score (standard deviation)	-0.05	(0.22)
[3]	Female (1=yes; 0=no)	0.03	(0.03)
[4]	Student's age (years)	0.06	(0.10)
[5]	Ethnic (1=Han; 0=minority)	-0.02	(0.03)
[6]	Computer use (1=ever used computer; 0=never used computer)	-0.02	(0.04)
[7]	Like English classes (0-100 points)	7.39*	(4.18)
[8]	Like English teacher (0-100 points)	5.33	(3.56)
[9]	Like school (0-100 points)	0.55	(1.80)
[10]	Education level students want to achieve (1=college or above; 0=below college)	0.01	(0.04)
Family characteristics			
[11]	Asset index	0.07	(0.16)
[12]	Only child (1=yes; 0=no)	-0.08	(0.07)
[13]	father farm's everyday (1=yes; 0=no)	0.06	(0.08)
[14]	mother farm's everyday (1=yes; 0=no)	0.11	(0.09)
[15]	At least one parent has junior high school degrees or above (1=yes; 0=no)	-0.01	(0.05)
[16]	Family size bigger than 5 (1=yes; 0=no)	0.01	(0.03)
[17]	At least one parent lived out of home (1=yes; 0=no)	0.05	(0.05)
Teacher and school characteristics			
[18]	teacher is female (1=yes; 0=no)	-0.08	(0.08)
[19]	English teacher's age (years)	-0.19	(2.48)
[20]	Civil service teacher (1=yes; 0=no)	-0.21*	(0.12)
[21]	English teaching hours per week	0.70	(1.41)
[22]	work year (year)	0.18	(2.71)
[23]	Education level of English teacher (1=College and above; 0=below college)	-0.14	(0.39)
[24]	Teachers attitudes towards computer (1-5 points)	0.05	(0.12)
[25]	Area (sq.km)	-0.01	(0.01)
[26]	Computer-student ratio	0.02	(0.02)
[27]	Rural school (1=yes; 0=no)	0.20	(0.19)
[28]	Observations	1342	

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

*significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4. Ordinary Least Squares estimators of the impact of OCAL program on standardized English test scores.

Dependent Variable: Standardized endline English test scores (standard deviation)			
		Unadjusted model	Adjusted model
		(1)	(2)
[1]	OCAL treatment (1=yes; 0=no)	0.45**	0.56***
		(0.18)	(0.14)
[2]	Standardized baseline English test score (standard deviation)	0.65***	0.58***
		(0.14)	(0.08)
[3]	Controls	N	Y
[4]	Observations	1342	1342
[5]	R-squared	0.36	0.5

Notes: *significant at 10%; ** significant at 5%; *** significant at 1%.

Robust standard errors in parentheses clustered at school level.

The test aims to show the impact of the 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 standardized baseline English test score is the score on the standardized English test that is given to all samples students before the OCAL program, and it is standardized using the baseline mean and standard deviations for the control group.

Source: Authors' survey.

Table 5. Ordinary Least Squares estimators of the impact of OCAL program on standardized Math test scores.

Dependent Variable: Standardized endline Math test score (standard deviation)			
		Unadjusted model	Adjusted model
		(1)	(2)
[1]	OCAL treatment (1=yes; 0=no)	-0.12 (0.19)	-0.06 (0.14)
[2]	Standardized baseline math test score (standard deviation)	0.47*** (0.05)	0.44*** (0.04)
[3]	Controls	N	Y
[4]	Observations	1141	1106
[5]	R-squared	0.23	0.34

Notes: *significant at 10%; ** significant at 5%; *** significant at 1%.

Robust standard errors in parentheses clustered at school level.

The test aims to show the impact of the treatment on student math test scores, the test regresses standardized endline math 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 standardized baseline English test score is the score on the standardized math test that is given to all samples students before the OCAL program, and it is standardized using the baseline mean and standard deviations for the control group.

Source: Authors' survey.

Table 6. Ordinary Least Squares estimators of the impact of OCAL program on student's self-efficacy.

Dependent Variable: student's self-efficacy (1-4 points)		Unadjusted model (1)	Adjusted model (2)
[1]	OCAL treatment (1=yes; 0=no)	-0.01 (0.04)	-0.05 (0.04)
[2]	Baseline value of the outcome variable	0.32*** (0.03)	0.30*** (0.04)
[3]	School dummy variables and other control variables	(0.03)	(0.04)
[4]	Observations	1215	1166
[5]	R-squared	0.1	0.12

Notes: *significant at 10%; ** significant at 5%; *** significant at 1%.

Robust standard errors in parentheses clustered at school level.

ble 7. Ordinary Least Squares estimators of the impact of OCAL program on student's attitudes.

riables	Like English teacher (0-100 points)	Like English teacher (0-100 points)	Like English classes (0-100 points)	Like English classes (0-100 points)	Education level students want to achieve (1=college or above; 0=below college)	Education lev students want achieve (1=college or above; 0=bel college)
	(1)	(2)	(3)	(4)	(5)	(6)
OCAL treatment (1=yes; 0=no)	2.05	0.38	5.46**	3.95**	0.05	0.06*
	(3.01)	(2.07)	(2.47)	(1.91)	(0.03)	(0.03)
Baseline value of the outcome variable	Y	Y	Y	Y	Y	Y
School dummy variables and other control variables	N	Y	N	Y	N	Y
Observations	1,189	1,189	1,189	1,189	1,189	1,189
R-squared	0.18	0.22	0.28	0.32	0.11	0.15

tes: *significant at 10%; ** significant at 5%; *** significant at 1%.

bust standard errors in parentheses clustered at school level.

Table 8. Ordinary Least Squares estimators of the heterogeneous impact of OCAL on standardized English test scores.

Dependent Variable: Standardized endline English test score (standard deviation)

		(1)
[1]	OCAL treatment (1=yes; 0=no)	0.60***
		(0.14)
[2]	English teacher is contract teacher (1=yes; 0=no)	0.10
		(0.16)
[3]	OCAL treatment* contract teacher	-1.02***
		(0.28)
[4]	Observations	1342
[5]	R-squared	0.50

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

Table 9. Cost-effectiveness analysis on OCAL compared with traditional offline CAL.

	OCAL	Traditional offline CAL
□. Cost		
Training teachers	2.67	4.53
Class Subsidies	40	21.58
Software	17.42	5.99
Program Execution Cost (RMB/ student)	60.09	32.09
Public resource investment (RMB/ student)	12.02	6.42
Social Cost (RMB/ student)	72. 11	38.51
□. Effectiveness		
Program Effect	0.56	0.18
□. Cost-Effective Ratio		
Program Cost-Effective Ratio	107.30	178.28
Social Cost-Effective Ratio	128.77	213.95