Economics of Education Review 94 (2023) 102385

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Impact of online computer assisted learning on education: Experimental evidence from economically vulnerable areas of China

Yu Bai [a], Bin Tang [b,*], Boya Wang [b], Di Mo [c], Linxiu Zhang [d,e], Scott Rozelle [f], Emma Auden [f], Blake Mandell [g]

[a] School of Economics & China Institute for Vitalizing Border Areas and Enriching the People, Minzu University of China, Beijing, China
[b] Center for Experimental Economics in Education (CEEE), Faculty of Education, Shaanxi Normal University, No. 620 West Chang'an Avenue, Chang'an District, Xi'an 710119, China
[c] LinkedIn Corporation, 222 2nd Street, San Francisco, CA 94105, USA
[d] United Nations Environment Programme-International Ecosystem Management Partnership (UNEP-IEMP), Beijing 100101, China
[e] Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[f] Stanford Center on China's Economy and Institutions, Stanford University, Stanford, USA
[g] Yenching Academy of Peking University, Beijing, China

ARTICLE INFO

Keywords:
Education
Computer assisted learning
Randomized controlled trial
Online learning

ABSTRACT

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, these have been mostly offline CAL programs that are difficult and costly to implement. An online CAL (OCAL) may be able to bypass many of offline CAL's implementation problems and enhance the remedial tutoring experience. 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. According to the findings, OCAL improved overall English scores of students in the treatment group relative to the control group by 0.48 standard deviations. This impact is large when compared with offline CAL programs previously evaluated in rural China. We found that OCAL also led to a positive change in the attitudes of students towards English learning and student aspirations for their future education level. We found three possible explanations for OCAL's impact. We believe that online features that enhance the interest-oriented stimulation of the software is the main source of improvement among students. Cost-effectiveness analysis showed that the OCAL program is more cost-effective than traditional offline CAL, which is suitable for policymakers as it indicates high potential for OCAL program expansion.

1. Introduction

Improving education for poor and disadvantaged populations has been a long-standing challenge for both developed and developing countries (Hanushek & Woessmann, 2008, 2012). This is perhaps especially true for developing countries undergoing economic transitions, typically from an economy reliant on labor-intensive manufacturing to one based on skill-based industries. When a significant portion of a country's labor force has relatively low educational attainment, the resulting shortage of human capital impedes further economic advancement (Khor et al., 2016). In addition, disparities between a country's urban and rural education systems may lead to long-term disparities in human capital acquisition and income between urban and rural populations, with serious implications for a country's growth and stability (Li et al., 2010, 2012).

In China, millions of students from rural areas and migrant communities lag behind their urban counterparts in terms of academic achievement. For example, Lai et al. (2013) found that fourth-grade students living in poor, rural areas had significantly lower scores in core subject areas compared to their urban peers. Even though China's government has already passed a number of compulsory education mandates aimed at remedying this issue, such as increasing the number of rural students attending school, or increasing the number of years of schooling attended by rural students, it has placed less emphasis on

* Corresponding author.
E-mail address: tangbin198994@163.com (B. Tang).

https://doi.org/10.1016/j.econedurev.2023.102385
Received 21 March 2019; Received in revised form 24 February 2023; Accepted 28 February 2023
Available online 7 March 2023

improving the overall quality of rural education (Wang et al., 2017a; Wang et al., 2017b; Lai et al., 2013).

One major reason why rural students in China fall behind—and stay behind—urban students academically, is because they are much more likely to lack access to remedial academic support. When urban students fall behind in school, they often have access to affordable remedial tutoring services offered to them by their teachers, commercial sources, or family members (Huang and Du, 2007; Lai et al., 2015). Students in rural areas, however, tend to have far less access to sources offering remedial tutoring (Lai et al., 2013). In addition, studies have found that many rural parents are unable to assist their children academically because they either live and work far away from their children, or they have low educational attainment themselves (Mo et al., 2015).

Internationally, some educators have addressed the problem of unequal access to remedial education for marginalized or disadvantaged populations through an education technology known as computer-assisted learning (CAL). In the current study, we define CAL as any kind of computer-related or computer-based software program created for improving a user's learning outcome in a specific subject area (Rouse & Krueger, 2004). However, we note that there are several different, more narrow definitions of CAL used in the literature. One such definition refers to CAL as the "well-defined" use of software packages designed to develop particular skills, such as improving math computation (Escueta et al., 2017). A second definition refers to CAL as an intelligent tutoring system that transmits knowledge to users using artificial intelligence to meet the specific needs of each learner (Dalgarno, 2001). Finally, a third definition refers to CAL as any learning method that may be run on a computer platform, such as online courses and online learning, and is delivered one-to-one or one-to-many (Schittek et al., 2001). This paper adopts the first and second definitions of CAL. While some forms of CAL are limited to purely giving students exercises in a drill-and-practice format, others offer students remedial learning materials in the form of gamified interfaces with the aim of improving educational outcomes via cultivating a long-term interest in learning (Inal & Cagiltay, 2007; Schaefer & Warren, 2004).

Past studies on CAL's effectiveness have been mostly promising, yet there have been somewhat mixed results. On the one hand, Escueta et al. (2017) identified 29 randomized-controlled trials (RCTs) of CAL experimental studies run in developed countries, among which 20 of the studies showed positive effects on student learning outcomes, eight found no effects, and one found a negative effect . On the other hand, some studies have concluded that the effects of CAL differ according to the target subject, and that the choice of subject could matter when analyzing the mechanisms behind program effects (Almekhlafi, 2006; Bianchi et al., 2022; Zhang & Zou, 2022; Sariman & Cetin, 2018; Seo & Bryant, 2009; Bayturan & Keşan, 2012; Nikou & Economides, 2018; Shannon et al., 2015).

Recently, CAL has become increasingly popular for use in developing countries, and experimental studies on the impact of CAL have been conducted in countries such as India and China. For example, Muralidharan et al. (2019) studied the impact of a personalized technology-aided after-school instruction program and found that math scores and Hindi scores improved by 0.37 SDs and 0.23 SDs, respectively. CAL has also been found to be effective in the context of rural China. Bai et al. (2016), Lai et al. (2013, 2015, 2016), Mo et al. (2014, 2015), and Yang et al. (2013) all offer evidence from randomized experiments showing that CAL can improve learning outcomes among disadvantaged rural students in China. Moreover, aside from improving academic performance, CAL programs have been found to exert beneficial effects on non-academic outcomes, such as a student's self-confidence and their interest in schooling (Lai et al., 2013). In sum, the international literature generally shows that CAL programs have been effective in raising both academic and non-academic outcomes in developing settings.

Yet while CAL programs have been found to be effective in rural China, there exist implementation drawbacks in regards to the type of CAL technology used in previous interventions. To the best of our knowledge, CAL interventions in China have nearly all used offline CAL programs providing tutoring courses powered by a set of gamified remedial exercises and learning materials. Interviews with the Principal Investigators of earlier China-based offline CAL studies revealed that at least 50 h per school per semester had been spent on the software installation, testing, monitoring, troubleshooting, and maintenance required to implement and manage the offline CAL interventions and evaluations, as offline CAL programs must be preloaded onto computers and run without Internet access. Additionally, in-person training workshops were needed for both teachers and students, and virus threats to computers were a constant source of program interruption. These factors thus make offline CAL programs expensive and time-consuming, which in turn may dissuade policymakers from upscaling CAL technology to a large number of schools.

With online computer-assisted learning (OCAL hereafter), many of these problems are remedied, which may be advantageous for improving CAL's effectiveness in the context of rural China. First, OCAL eliminates the need to manually install and maintain software on each computer hardware and lowers the costs and time demands placed on program teachers who are often already overworked. Second, the ability to log in anywhere and at anytime may be able to increase student access, as students can open the program software on any computer they have access to that has Internet connectivity, instead of only through those on which the CAL software is installed. Third, the OCAL software system allows for the integration of features into the program offering interaction between different users (e.g., students can compete with peers on quizzes and earn virtual prizes; the software system can also provide an automated ranking system that may further motivate student learning) (Borgatti & Cross, 2003; Holmes & O'loughlin, 2014). Since offline CAL has already been demonstrated to improve student academic performance in rural China, we hypothesize that OCAL will improve student academic performance by an even larger amount at a lower cost.

To our knowledge, few if any OCAL programs using online remedial tutoring have been rigorously evaluated in rural China. Though there exist some studies on online courses or internet-based CAL in the context of other countries, they provide limited evidence relevant to rural China. In addition, the results of these studies have been mixed. Of studies that have produced positive effects, Deault et al. (2009) used a web-based literacy program to help students with reading; results showed that significant positive effects of the intervention were evident for about half of the reading and related measures, and the program might have helped kids with attention problems. Kelly et al. (2013), another study that found positive effects, used a CAL program that provided online homework support to students, and achieved an effect size of 0.56 SDs on math. There are studies that have found no significant impacts, however, proposing that CAL has no advantage over traditional methods of learning. For example, Rouse and Krueger (2004) evaluated the Fast ForWord computer-based language training program on students and detected null results. Van Klaveren et al. (2017) compared an adaptive CAL program against a static (non-adaptive) one, showing no statistically significant improvement from the adaptive CAL program relative to a non-adaptive CAL program. Therefore, due to the mixed nature of the existing literature, it is difficult to propose an estimate of the effectiveness of OCAL based on previous research; additionally, context-specific studies remain necessary in rural China.

To fill in these gaps, 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 have four sub-objectives. For our first sub-objective, we examine OCAL's impact on student academic performance measured by their English score, and use math scores to test for potential spillover effects. We chose English as our OCAL program subject for several reasons. First, English is one of the main subjects used to test students as part of the competitive examination system in China that allows students to compete for positions in high schools and colleges (Bolton & Graddol, 2012; McKay, 2002;

Table 1
Pre-balance check: comparison of outcome variables, student and family characteristics, teacher and school characteristics between treatment and control groups in the total samples.

	Total		Treatment		Control		Difference (treatment-control)	
	Mean	SD	Mean	SD	Mean	SD	Mean	P-value
Outcome variables								
(1) Standardized baseline English test score (standard deviation)	0.00	1.00	0.07	0.97	-0.05	1.02	0.12	0.60
(2) Standardized baseline Math test score (standard deviation)	0.00	1.00	-0.04	1.07	0.03	0.94	-0.07	0.74
(3) Self-efficacy in baseline (10-40 points, standard deviation)	0.00	1.00	-0.04	1.04	0.03	0.97	-0.08	0.47
(4) Like English classes (0-100 points, standard deviation)	0.00	1.00	0.16	0.89	-0.12	1.06	0.28*	0.05
(5) Like English teacher (0-100 points, standard deviation)	0.00	1.00	0.13	0.90	-0.10	1.06	0.23*	0.09
(6) Like school (0-100 points, standard deviation)	0.00	1.00	0.01	0.99	-0.01	1.01	0.01	0.88
(7) Education level students want to achieve (1=college or above; 0=below college)	0.70	0.46	0.70	0.46	0.70	0.46	0.00	0.93
Student characteristics								
(8) Female (1=yes; 0=no)	0.52	0.50	0.54	0.50	0.52	0.50	0.02	0.45
(9) Student's age (years)	10.90	0.80	10.93	0.79	10.87	0.80	0.07	0.53
(10) Ethnic (1=Han; 0=minority)	0.96	0.20	0.95	0.22	0.97	0.18	-0.02	0.55
(11) Computer use (1=ever used computer; 0=never used computer)	0.90	0.31	0.87	0.33	0.91	0.28	-0.04	0.33
Family characteristics								
(12) Family Asset index	0.01	1.00	0.04	1.00	-0.02	0.99	0.06	0.68
(13) Only child (1=yes; 0=no)	0.31	0.46	0.26	0.44	0.35	0.48	-0.09	0.22
(14) father working on farm (1=yes; 0=no)	0.41	0.49	0.44	0.50	0.39	0.48	0.05	0.51
(15) mother working on farm (1=yes; 0=no)	0.39	0.49	0.44	0.50	0.35	0.48	0.09	0.26
(16) father has at least a high school degree(1=yes;0=no)	0.68	0.47	0.68	0.47	0.68	0.47	-0.00	1.00
(17) mother has at least a high school degree(1=yes;0=no)	0.63	0.48	0.63	0.48	0.63	0.48	0.00	0.97
(18) Family size bigger than 5 (1=yes; 0=no)	0.21	0.41	0.21	0.41	0.21	0.41	-0.00	0.85
(19) At least one parent lived out of home (1=yes; 0=no)	0.31	0.46	0.32	0.47	0.30	0.46	0.02	0.66
English teacher characteristics								
(20) Teacher is female (1=yes; 0=no)	0.91	0.29	0.89	0.31	0.92	0.27	-0.03	0.76
(21) English teacher's age (years)	32.55	6.46	32.37	7.20	32.69	5.84	-0.32	0.88
(22) Civil service teacher (1=yes; 0=no)	0.90	0.31	0.80	0.40	0.97	0.18	-0.17	0.16
(23) English teaching hours per week	9.19	4.60	9.82	4.49	8.71	4.62	1.11	0.43
(24) working year (year)	10.12	7.22	10.34	7.73	9.95	6.80	0.39	0.87
(25) Education level of English teacher (1=College and above; 0=below college)	0.35	0.48	0.25	0.43	0.44	0.50	-0.19	0.25
(26) Teachers attitudes towards computer (1-5 points)	4.26	0.31	4.29	0.35	4.23	0.27	0.07	0.54
School characteristics								
(27) Area (sq.km)	0.01	0.02	0.01	0.01	0.01	0.02	-0.00	0.59
(28) Computer-student ratio	0.11	0.07	0.11	0.08	0.11	0.07	0.01	0.75
(29) Rural school (1=yes; 0=no)	0.41	0.49	0.48	0.50	0.36	0.48	0.12	1.49
Observations	1650		714		936			

Notes: *significant at 10%; ** significant at 5%; *** significant at 1%.
As for the p-value between treatment and control groups, we did a regression of above variables (including outcome variables, student and family characteristics, teacher and school characteristics) respectively on treatment (1=treatment group, 0=control group). We listed p-value of each regression's coeffecient here to show whether the treatment group differed from control group on these variables.

Davey et al., 2007). In fact, the subject of English takes up about one-third of the test for both the high school and college entrance exams. Second, English teaching and English learning is particularly weak in rural China (Ma et al., 2021; Wang et al., 2017b). Studies have shown that a low English score is one of the most important factors keeping rural students from attending high school in China (Loyalka, 2014; Li et al., 2021), and rural English teachers are also of notoriously poor quality (Hu, 2003, 2005, 2009; Mo et al., 2020). Given the importance of English for advancing in school in China, and due to the current status of English education in rural China's schools, we decided it would be most appropriate to test an OCAL program focused on improving English skills among rural students. For our second sub-objective, we measure OCAL's impact on non-academic outcomes, in particular on student self-efficacy and student attitudes towards teachers, courses, and schools. For our third sub-objective, we attempt to explain the mechanism behind OCAL's effectiveness by examining a number of possible reasons behind its impact. Finally, we conduct a cost-effectiveness analysis to determine OCAL's feasibility in relation to traditional offline CAL interventions.

To meet these objectives, we will 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. Results show that the OCAL intervention improved the English performance of students in the treatment group (relative to those of the control group) by 0.48 standard deviations, which is larger than the impact (0.15 SD) found in previous offline CAL studies focused on English learning in rural China (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.

After rejecting the possibility that the results were due to either a Hawthorne Effect or self-efficacy-induced changes, we find there are three possible explanations for OCAL's large impact in our experiment: (1) interest-oriented stimulation from OCAL's interactive, gamified nature and the chance for comparison and competition with peers; (2) customized remedial question banks tailored to each student's individual needs; and (3) compared to math, there is room for easier gains in English proficiency. Finally, our analysis indicates that our OCAL intervention has a higher cost-effectiveness ratio than traditional offline CAL programs. Therefore, we conclude that OCAL appears to offer a more efficient solution for improving education in rural China, as it has both larger impacts and lower costs.

Our study adds to the literature in three ways. First, this paper is the first to measure the effect of OCAL on learning outcomes among an underserved population in China using experimental design to rule out potential bias. Second, this paper contributes to our understanding of the overall mechanism of the impacts. Third, because this paper focuses on the impact of OCAL on English learning in rural China, it makes a contribution to the scant literature on CAL in the context of second-language acquisition.

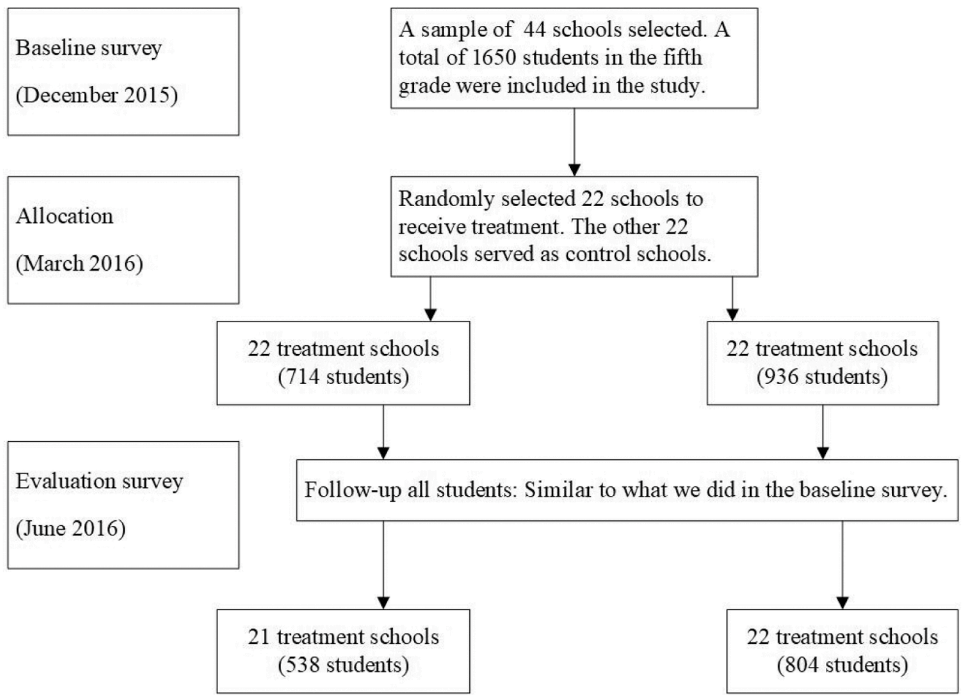

Fig. 1. Experiment profile.

2. Method

2.1. 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. When choosing schools, schools needed to meet two criteria: (1) the school needed to have at least one computer room which could meet the maximum capacity of users in any grade-five class; and (2) the school needed to have a computer network with appropriate network speed. Additionally, all sample students in the schools were from rural areas; that is, they had agricultural *hukou* status.

Ultimately, from these 44 schools, a total of 1650 fifth-grade students from 56 classes were involved in our study. We selected 44 schools based on power calculations. Before selecting our sample schools, we employed power calculation to determine how many schools and students were needed for statistically identifying the impact of the OCAL program. We summed a standardized effect size for the outcome variable of 0.20, 0.80 power, a five percent significance level, and an intra-cluster correlation (ICC) of 0.06. We also assumed that there were 40 observations in each cluster on average. Base on these assumptions, we needed 17 schools for each study arm. Considering attrition, we randomly assigned 22 schools to the treatment arm and 22 schools to the control arm.

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 pre-balanced along several key variables for randomization. These key variables were student characteristics, including standardized baseline English test scores, standardized baseline math test scores, age, gender, ethnicity, whether students had ever used a computer; family characteristics, including family assets, parent educational attainment, family size, and migrant status; and teacher and school characteristics. For a full list of control variables balanced during the randomization process, see Table 1. By showing the balance between the treatment and control group samples, it may be seen that main outcome variables such as baseline English and math test scores, self-efficacy, and education expectations were balanced in the pre-balance check. There were no significant differences among student characteristics, family characteristics, teacher, and school characteristics.

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 remaining 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 (Fig. 1). We show that attrition would not affect the validity of this study in the appendix.

2.2. Experimental intervention

The intervention involved computer-assisted English remedial tutoring sessions designed to complement the entire 2016 spring semester's regular in-class English curriculum, which follows the National Standard Curriculum of China. 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 min OCAL sessions per week.

Our protocol required that OCAL sessions be given during the "computer class" period because this period is reserved for teaching non-academic material. Based on our surveys, students taking the computer classes offered in most of China's rural schools were 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 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 min of remedial tutoring (two 40 min 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.

The OCAL program that we implemented had the following features. First, due to OCAL's online nature, students could log into a website to access the 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.

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. Additionally, the nature of offline software-based CAL makes it much easier to ensure that students remain focused on the 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.

2.3. Control group

The fifth-grade students in the 22 control schools that constituted the 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.

2.4. 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 by questionnaire. To be noted, all tests and questionnaires were digital, online surveys. We organized the students to enter the computer room and completed tests and questionnaires on the computer.

To measure academic proficiency, the first block of the survey contained two standardized tests: first a 100-question English test and then a 33-question math test (see Lai et al. 2015).[1] Each of these tests had a 30 min 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 performance and math 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.[2]

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 (1999), 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 with high self-efficacy believes they 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 et al., 1982), self-efficacy refers to a subjective rating given by an individual regarding the control he has over his actions. Self-efficacy enhances cognitive processes and correlates with various other outcomes, including decision-making quality (Bandura et al., 1999). 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 et al., 2001). 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 from 1 to 4 to indicate their agreement with questions or statements (numbers 1 to 4 represent the options of "completely incorrect," "slightly correct," "mostly correct," or "completely correct"). The maximum score for the GSES is 40 points and the minimum score is 10 points.

Student attitudes towards 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 between 0 and 100, with 0 meaning strongly dislike and 100 meaning strongly like.

[1] 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.

[2] 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.

Table 2
Ordinary Least Squares estimators of the impact of OCAL program on standardized English test scores.

	Dependent Variable: Standardized endline English test scores (standard deviation)		
	Variables	Unadjusted model (1)	Adjusted model (2)
(1)	OCAL treatment (1=yes; 0=no)	0.39** (0.15)	0.48*** (0.12)
(2)	Standardized baseline English test score (standard deviation)	0.56*** (0.12)	0.50*** (0.07)
(3)	Controls	N	Y
(4)	Observations	1342	1342
(5)	R-squared	0.36	0.50

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.
See Appendix Table A4 for more details of the regressions.

In the second block of the survey, enumerators collected data on the characteristics of students and their families which were 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*, 1=female and 0=male), 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 (family assets). 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 gave principals and teachers digital questionnaires asking them about background information and school information. Through questionnaires, 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 service teacher (1=civil service 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 1 if a student took the baseline survey but not the endline survey and equaled 0 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.

2.5. 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:

$$y_{isc} = \alpha + \beta \cdot treatment + \theta * y_{0isc} + \varepsilon_{isc} \tag{1}$$

where y_{isc} is the outcome variable after the OCAL program for student 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, Eq. (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 Eq. (1) to build an adjusted model:

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

where all the variables and parameters are the same as those in Eq. (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, Eq. (2) can more effectively estimate the OCAL treatment effect.

3. Results

3.1. 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 2.

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 0.39 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 0.48 SDs (significant at the 1% level).

Our results suggest that this particular OCAL intervention had an effect size larger than that of previous, offline CAL initiatives on student academic performance (Fig. 2). For example, an evaluation of an offline CAL intervention conducted in Qinghai province, which offered

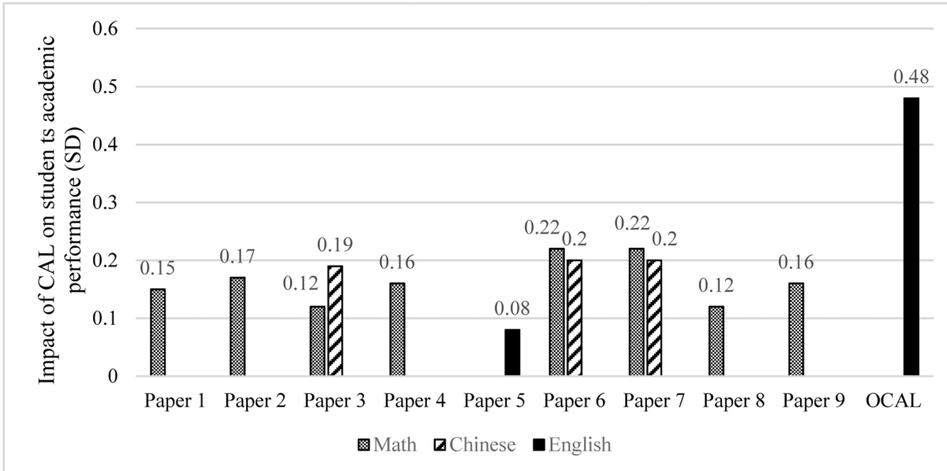

Fig. 2. Impact of CAL on students' academic performance based on literature: Offline CAL compared with OCAL. (see Appendix Table A5 for more details).
Sources: Overview of Existing Literature on Experiments of Chinese CAL Programs

Table 3
Lee bounds of treatment effect.

Number of obs.=1650
Number of selected obs. =1342
Trimming proportion =0.1228

Standardized English scores in endline	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
treatment						
lower	0.28	0.06	4.38	0.00	0.16	0.41
upper	0.69	0.06	10.86	0.00	0.57	0.82

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

	Dependent Variable: Standardized endline Math test scores (standard deviation)		
	Variables	Unadjusted model (1)	Adjusted model (2)
(1)	OCAL treatment (1=yes; 0=no)	-0.12	-0.05
		(0.19)	(0.14)
(2)	Standardized baseline Math test score (standard deviation)	0.49***	0.46***
		(0.05)	(0.05)
(3)	Controls	N	Y
(4)	Observations	1106	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 Math test scores on both treatment variables and a set of control variables for adjusted model.
The standardized baseline Math test score is the score on the standardized math test that is given to all samples students before the OCAL program.
Source: Authors' survey.

remedial tutoring for the same textbook using the same user-interface as this OCAL intervention, led to a 0.15 SD improvement in student English scores (Bai et al., 2016).[3] In order to verify the sensitivity of effect size and consider the attrition, we also adopted method of Lee (2009) and achieve the bounds of the effect size, including the upper bound 0.69 SD and the lower bound 0.28 SD, as shown in Table 3.

3.2. *Explaining the impact of OCAL on student English academic performance*

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

3.3. *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, the test scores of the treatment-group students could have improved because of their awareness of being observed, which led them to try harder in school. In this scenario, the observed impact would not have been due to OCAL.

To address the possibility of a Hawthorne effect, we conducted an analysis of standardized math test scores and found that the OCAL English tutoring had no statistically significant impact, either positive or negative, on math performance (Table 4). The fact that English scores improved while there were no effects on math scores provides evidence that treatment-group students were not 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.

3.4. *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

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

Table 5
Ordinary Least Squares estimators of the impact of OCAL program on student's self-efficacy.

Dependent Variable: student's self-efficacy in endline (10-40 points, standardized deviation)		Unadjusted model (1)	Adjusted model (2)
[1]	OCAL treatment (1=yes; 0=no)	-0.02	-0.08
		(0.06)	(0.07)
[2]	Baseline self-efficacy (10-40 points)	0.32***	0.31***
		(0.03)	(0.04)
[3]	Control variables	N	Y
[4]	Observations	1215	1166
[5]	R-squared	0.10	0.12

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

improvement (Pintrich & De Groot, 1990; Lent et al., 1986; Chemers et al., 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, 2015). Although it is unknown whether student self-efficacy causes improved academic achievement or vice versa, previous studies have demonstrated a link between them.

Contrary to the findings of past studies, our results indicate that the OCAL intervention had no significant effect on student self-efficacy (Table 5). Such a finding would suggest that, different from earlier studies, there may be many reasons why there may not be a link between self-efficacy and improved English academic achievement. For example, it could be a power issue or a measurement issue. However, we used standard scales and scientific measurement methods, and the sample is large enough.

We propose three possible explanations for the lack of improvement in self-efficacy. The first explanation for the insignificant results on the effect of the intervention on student self-efficacy 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 among them or be emotionally difficult due to the ties between language and identity (Cohen & Norst, 1989). A lack of motivation has also been shown to be a barrier for beginners learning Japanese or Spanish (Clément & 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, 2010; Haggarty & Pepin, 2002). However, learning English is more likely to enhance memorization and recall skills for non-native beginners, and thus improvements in English performance do not necessarily improve decision-making ability, leading to a lack of 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 et al., 1998; Turner et al., 2006). Our intervention frame was likely insufficient to change student self-efficacy, as evidenced by the strong correlation between endline self-efficacy scores and baseline scores (Table 5, row 2).

3.5. Possibility 3: interest-oriented stimulation

In this section, we seek to explain why OCAL might be more effective than traditional, offline CAL. Due to OCAL's 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 offline CAL programs).

First, the OCAL software offers online interactive features that motivates peer engagement. Unlike traditional, offline CAL, OCAL programs offer social features that encourage students to engage their peers in competitive, gamified activities, which increases student engagement as a whole 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 then be used to give gifts to others and to play small online games during breaks from the remedial tutoring exercises. Thus, the social and competitive elements of OCAL may be one of the reasons for the large measured impact (Chen et al., 2010).

Second, OCAL's ability to provide data-informed exercises based on each individual student's mastery 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 difficult or too easy for them. Recent literature, Muralidharan et al. (2019), has indicated that a primary mechanism of CAL lies in its power to tailor learning so that students can learn according to their own pace, thereby increasing effective learning inputs

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

Variables		Like English teacher (0-100 points, standardized deviation) (1)	Like English teacher (0-100 points, standardized deviation) (2)	Like English classes (0-100 points, standardized deviation) (3)	Like English classes (0-100 points, standardized deviation) (4)	Education level students want to achieve (1=college or above; 0=below college, standardized deviation) (5)	Education level students want to achieve (1=college or above; 0=below college, standardized deviation) (6)
[1]	OCAL treatment (1=yes; 0=no)	0.09	0.04	0.18**	0.16**	0.05	0.06**
		(0.09)	(0.07)	(0.08)	(0.06)	(0.03)	(0.03)
[2]	Baseline value of the outcome variable	0.47***	0.44***	0.52***	0.47***	0.32***	0.30***
		(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
[3]	control variables	N	Y	N	Y	N	Y
[4]	Observations	1,166	1,166	1,166	1,166	1,166	1,166
[5]	R-squared	0.21	0.25	0.28	0.32	0.11	0.16

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

Table 7
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)	0.52***
		(0.13)
[2]	Whether English teacher is contract teacher (1=yes; 0=no)	0.11
		(0.15)
[3]	OCAL treatment* contract teacher	-0.92***
		(0.25)
[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.

and producing more beneficial outputs.

Support for the interest-oriented stimulation hypothesis is also provided by examining the effects on student attitudes towards their English learning program. When examining the OCAL intervention's effects on student attitudes toward their courses, teachers, and schools, the results indicate that student attitudes toward their English class significantly improved after the OCAL intervention (Table 6, Columns 3 & 4) while student attitudes toward their English teachers did not significantly change (Table 6, Columns 1 & 2). This is perhaps not surprising, as in the intervention the teachers who organized students to use the OCAL software were all computer class teachers and not English teachers. In sum, we believe that what primarily accounts for the increased student interest in English is the interactive, game-based nature of OCAL software, while due to experimental set-up there was no such attitude change towards English teachers.

Table 8
Heterogeneity among different levels of students' English scores in baseline: Ordinary Least Squares estimators of the impact of OCAL program on standardized English test scores.

Dependent Variable: Standardized endline English test scores (standard deviation)									
		25% lower of Baseline English Performance Samples (Part 1)		25%-50% of Baseline English Performance Samples (Part 2)		50-75% of Baseline English Performance Samples (Part 3)		75% upper of Baseline English Performance Samples (Part 4)	
	Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	OCAL treatment (1=yes; 0=no)	0.26*	0.34**	0.45**	0.42***	0.43**	0.47***	0.45*	0.78***
		(0.15)	(0.14)	(0.18)	(0.17)	(0.17)	(0.13)	(0.25)	(0.20)
(2)	Standardized baseline English test score (standard deviation)	0.35*	0.24**	0.53**	0.45***	0.54**	0.43***	0.67***	0.65***
		(0.18)	(0.11)	(0.20)	(0.10)	(0.21)	(0.12)	(0.15)	(0.14)
(3)	Controls	N	Y	N	Y	N	Y	N	Y
(4)	Observations	379	379	327	327	337	337	299	299
(5)	R-squared	0.11	0.29	0.31	0.49	0.29	0.50	0.33	0.57
(6)	p-value of Chow tests		(8) - (2): 0.0021		(8) - (4): 0.0094		(8) - (6): 0.0215		

Notes: *significant at 10%; ** significant at 5%; *** significant at 1%.
Robust standard errors in parentheses clustered at school level.
p-value of Chow test: We compare the coefficients of treatment in Quantile 1 and Quantile 4, Quantile 2 and Quantile 4, Quantile 3 and Quantile 4 and list the p-values of each test in Row 6.

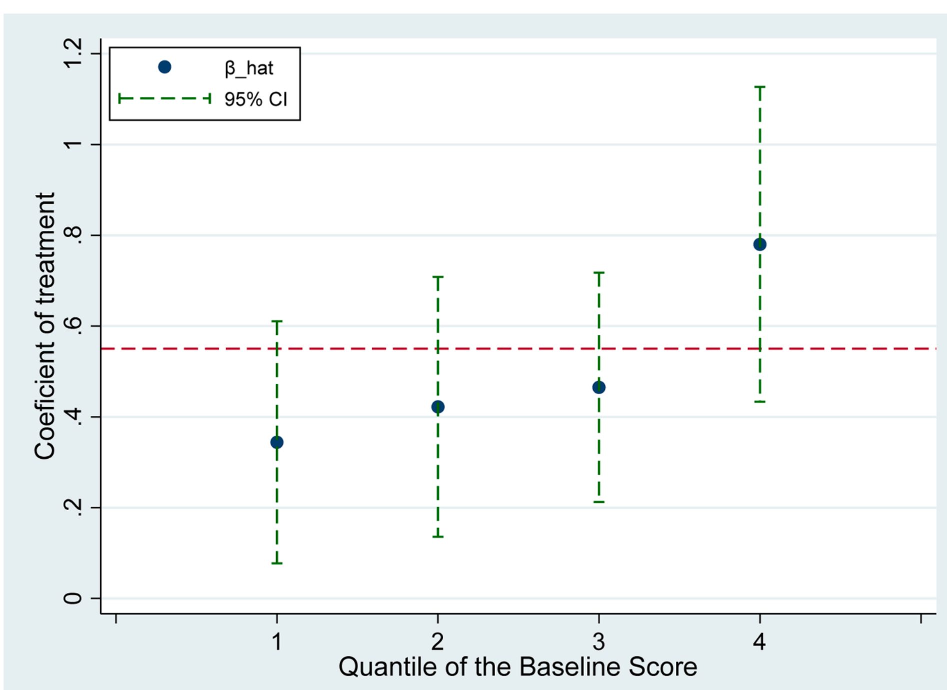

Fig. 3. Heterogeneity among different levels of students' English scores at baseline: Effect of OCAL program on English test scores.

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

	OCAL	Traditional offline CAL
I. Cost		
Training teachers	0.42	0.70
Class Subsidies	6.15	3.32
Software	5.36	5.99
Program Execution Cost (USD/ student)	11.93	10.01
Public Resource Investment (20% of Program Execution Cost) (USD/ student)	2.39	2.00
Social Cost (Program Execution Cost+ Public Resource Investment) (USD/ student)	14.32	12.01
II. Effectiveness		
Program Effect (SD)	0.55	0.18
III. Cost-Effective Ratio		
Program Cost-Effective Ratio (USD/student/SD)	21.69	55.61
Social Cost-Effective Ratio (USD/student/SD)	26.04	66.72
IV. Cost (USD) Per 0.2 SD		
Program Cost	4.33	11.12
Social Cost	5.21	13.34

Notes: We set Cost (USD) Per 0.2 SD in order to compared with results in McEwan(2015).

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, 85(3), 353-394.

3.6. Other factors: low initial level of english

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 easier gains in English proficiency. Although Mandarin, math, and English are all core subjects taught in China's primary schools, disparities in the quality of teachers and school facilities along with varying academic prioritization between urban and rural schools contribute to students having different baseline scores in each of those subjects. Loyalka et al. (2017) investigated the rural-urban educational gap in proficiency levels on core subjects and found 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

Table A2
Test if attrition is correlated with treatment.

Variables	attrition[a] (1=students attrited; 0=student remained) (1)
(1) OCAL treatment (1=yes; 0=no)	0.11 (0.11)
(2) Constant	0.14***(0.04)
(3) Observations	1650
(4) R-squared	0.02

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

Table A1
Attrition check: comparison of outcome variables, student and family characteristics, teacher and school characteristics between attrited samples and non-attrited samples.

	Total		Attrited samples		Non-attrited samples		Difference (Attrited samples - Non-attrited samples)	
	Mean	SD	Mean	SD	Mean	SD	Mean	P-value
Outcome variables								
(1) Standardized baseline English test score (standard deviation)	0.00	1.00	-0.14	0.95	0.03	1.01	-0.18	0.32
(2) Standardized baseline Math test score (standard deviation)	0.00	1.00	-0.20	1.05	0.04	0.99	-0.25*	0.07
(3) Self-efficacy in baseline (10-40 points, standard deviation)	0.00	1.00	0.12	1.02	-0.03	0.99	0.15*	0.06
(4) Like English classes (0-100 points, standard deviation)	0.00	1.00	0.00	1.08	-0.00	0.98	0.00	1.00
(5) Like English teacher (0-100 points, standard deviation)	0.00	1.00	-0.06	1.09	0.01	0.98	-0.07	0.62
(6) Like school (0-100 points, standard deviation)	0.00	1.00	-0.09	1.08	0.02	0.98	-0.11	0.32
(7) Education level students want to achieve (1=college or above; 0=below college)	0.70	0.46	0.73	0.44	0.69	0.46	0.04	0.38
Students characteristics								
(8) Female (1=yes; 0=no)	0.52	0.50	0.53	0.50	0.52	0.50	0.01	0.88
(9) Student's age (years)	10.90	0.80	10.89	0.85	10.90	0.78	-0.00	0.97
(10) Ethnic (1=Han; 0=minority)	0.96	0.20	0.96	0.19	0.96	0.20	0.00	0.85
(11) Computer use (1=ever used computer; 0=never used computer)	0.90	0.31	0.86	0.35	0.90	0.29	-0.04	0.24
Family characteristics								
(12) Family Asset index	0.01	1.00	-0.07	1.03	0.03	0.99	-0.10	0.38
(13) Only child (1=yes; 0=no)	0.31	0.46	0.30	0.46	0.32	0.46	-0.02	0.74
(14) father working on farm (1=yes; 0=no)	0.41	0.49	0.39	0.49	0.42	0.49	-0.03	0.58
(15) mother working on farm (1=yes; 0=no)	0.39	0.49	0.39	0.49	0.39	0.49	-0.00	0.92
(16) father has at least a high school degree(1=yes;0=no)	0.68	0.47	0.70	0.46	0.68	0.47	0.03	0.56
(17) mother has at least a high school degree(1=yes;0=no)	0.63	0.48	0.68	0.47	0.62	0.49	0.06	0.20
(18) Family size bigger than 5 (1=yes; 0=no)	0.21	0.41	0.18	0.39	0.22	0.41	-0.04	0.13
(19) At least one parent lived out of home (1=yes; 0=no)	0.31	0.46	0.27	0.44	0.32	0.47	-0.05	0.24
English Teacher characteristics								
(20) Teacher is female (1=yes; 0=no)	0.91	0.29	0.84	0.37	0.93	0.26	-0.09	0.41
(21) English teacher's age (years)	32.55	6.46	31.14	4.98	32.87	6.71	-1.73	0.10
(22) Civil service teacher (1=yes; 0=no)	0.90	0.31	0.88	0.33	0.90	0.30	-0.02	0.76
(23) English teaching hours per week	9.19	4.60	9.00	4.74	9.23	4.57	-0.23	0.88
(24) working year (year)	10.12	7.22	9.52	5.88	10.25	7.48	-0.74	0.62
(25) Education level of English teacher (1=College and above; 0=below college)	0.35	0.48	0.36	0.48	0.35	0.48	0.01	0.97
(26) Teachers attitudes towards computer (1-5 points)	4.26	0.31	4.22	0.26	4.26	0.32	-0.04	0.42
School characteristics								
(27) Area (sq.km)	0.01	0.02	0.01	0.02	0.01	0.02	0.00	0.49
(28) Computer-student ratio	0.11	0.07	0.09	0.07	0.11	0.07	-0.03	0.18
(29) Rural school (1=yes; 0=no)	0.41	0.49	0.26	0.44	0.44	0.50	-0.18	0.12
Observations	1650	308	1342					

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

Table A3
Balance check: comparison of outcome variables, student and family characteristics, teacher and school characteristics between treatment and control groups after attrition.

	Non-attrited sample after attrition		Treatment		Control		Difference (treatment-control)	
	Mean	SD	Mean	SD	Mean	SD	Mean	P-value
Outcome variables								
(1) Standardized baseline English test score (standard deviation)	0.03	1.01	0.09	1.00	-0.00	1.01	0.09	0.69
(2) Standardized baseline Math test score (standard deviation)	0.04	0.99	0.01	1.07	0.06	0.93	-0.05	0.80
(3) Self-efficacy in baseline (10-40 points, standard deviation)	-0.03	0.99	-0.10	1.03	0.02	0.96	-0.12	0.24
(4) Like English classes (0-100 points, standard deviation)	0.00	0.98	0.15	0.87	-0.10	1.04	0.26*	0.08
(5) Like English teacher (0-100 points, standard deviation)	0.01	0.98	0.13	0.88	-0.06	1.03	0.19	0.14
(6) Like school (0-100 points, standard deviation)	0.02	0.98	0.04	0.97	0.01	0.99	0.03	0.76
(7) Education level students want to achieve (1=college or above; 0=below college)	0.69	0.46	0.70	1.00	0.69	0.46	0.01	0.77
Students characteristics								
(8) Female (1=yes; 0=no)	0.52	0.50	0.54	0.50	0.51	0.50	0.03	0.27
(9) Student's age (years)	10.90	0.78	10.93	0.77	10.87	0.79	0.06	0.55
(10) Ethnic (1=Han; 0=minority)	0.96	0.20	0.95	0.22	0.96	0.19	-0.02	0.58
(11) Computer use (1=ever used computer; 0=never used computer)	0.90	0.29	0.89	0.31	0.91	0.28	-0.02	0.58
Family characteristics								
(12) Family Asset index	0.03	0.99	0.07	1.00	0.00	0.98	0.07	0.66
(13) Only child (1=yes; 0=no)	0.32	0.46	0.27	0.44	0.35	0.48	-0.08	0.29
(14) father working on farm (1=yes; 0=no)	0.42	0.49	0.46	0.50	0.39	0.49	0.06	0.45
(15) mother working on farm (1=yes; 0=no)	0.39	0.49	0.46	0.50	0.35	0.48	0.11	0.20
(16) father has at least a high school degree(1=yes;0=no)	0.68	0.47	0.67	0.47	0.68	0.47	-0.01	0.84
(17) mother has at least a high school degree(1=yes;0=no)	0.62	0.49	0.61	0.49	0.63	0.48	-0.02	0.77
(18) Family size bigger than 5 (1=yes; 0=no)	0.22	0.41	0.22	0.42	0.22	0.41	0.01	0.78
(19) At least one parent lived out of home (1=yes; 0=no)	0.32	0.47	0.35	0.48	0.30	0.46	0.05	0.29
English Teacher characteristics								
(20) Teacher is female (1=yes; 0=no)	0.93	0.26	0.88	0.33	0.96	0.20	-0.08	0.30
(21) English teacher's age (years)	32.87	6.71	32.76	7.95	32.95	5.74	-0.19	0.94
(22) Civil service teacher (1=yes; 0=no)	0.90	0.30	0.77	0.42	0.98	0.13	-0.21*	0.09
(23) English teaching hours per week	9.23	4.57	9.65	4.53	8.95	4.57	0.70	0.62
(24) working year (year)	10.25	7.48	10.36	8.46	10.18	6.76	0.18	0.95
(25) of English teacher (1=College and above; 0=below college)	0.35	0.48	0.27	0.45	0.41	0.49	-0.14	0.42
(26) Teachers attitudes towards computer (1-5 points)	4.26	0.32	4.29	0.37	4.24	0.27	0.05	0.68
School characteristics								
(27) Area (sq.km)	0.01	0.02	0.01	0.01	0.01	0.02	-0.01	0.34
(28) Computer-student ratio	0.11	0.07	0.13	0.08	0.10	0.07	0.02	0.29
(29) Rural school (1=yes; 0=no)	0.44	0.50	0.57	0.50	0.36	0.48	0.20	0.28
Observations	1342		538		804			

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

unprecedented impact in this study lies in the disproportionately low baseline English proficiency of rural students, as a lower starting point may mean that the OCAL program can effect a larger 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 Fig. 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 specific.

We thusly believe that the measured differences between OCAL and offline CAL's impact on rural students' English learning may be explained by three factors: (1) the low baseline English level of the students in our sample; (2) OCAL's interactive nature; and (3) OCAL's ability to offer targeted learning exercises for each student. Additionally, we demonstrated how the Hawthorne effect and changes in self-efficacy do not appear to be reasons for OCAL's effectiveness. In short, the effectiveness of the OCAL program in this study, at least in part, appears to come from interest-oriented stimulation. Of course, although our study cannot conclusively link OCAL's large impact on English learning to the above three factors, we believe nonetheless that these factors help to explain OCAL's effectiveness.

This paper finds that online computer-assisted learning has a significant positive impact on English language performance. Moreover, this finding suggests that the role of computer-assisted learning in helping students to acquire a second language in developing countries may be significant. In fact, recent literature has shown that computer-assisted learning can help students improve their second language acquisition. One such study states that computer-assisted learning can provide a more accurate pronunciation, vocabulary, and language usage scenario than a teacher, facilitating immersion in a well-constructed context (Machado, 1997). Second, computer-assisted language learning can increase learners' confidence as well as improve learners' pronunciation and vocabulary through recorded speech or computerised pronunciation. According to AbuSeileek (2012), a computer-based language learning environment allowed participants to blind their identities and experience reduced levels of anxiety that may be invoked by face-to-face interaction using an unfamiliar language; these features thus proved helpful in developing the participants' communication skills. Third, student academic achievement and learning satisfaction may be effectively enhanced, and students may be encouraged to learn better, through OCAL's adaptive learning features that set tailored learning tasks in a stage-based way (Hui et al., 2008; Dehghanzadeh et al., 2021). We believe that the above theories and findings provide viable reasons for why OCAL programs can produce a large improvement in students' English skills.

3.7. Heterogeneity

Apart from measuring the impacts of the OCAL program on student academic performance and non-academic outcomes, it is also crucial to explore whether OCAL has differential impacts for various types of students, teachers, and schools. The aim of the heterogeneity analysis is

Table A4
Ordinary Least Squares estimators of the impact of OCAL program on standardized English test scores.

	(1) Unadjusted_model	(2) Adjusted_model
VARIABLES	Standardized endline English test score	Standardized endline English test score
OCAL treatment (1=yes; 0=no)	0.39**	0.48***
	(0.15)	(0.12)
Standardized baseline English test score	0.56***	0.50***
	(0.12)	(0.07)
Female (1=yes; 0=no)		-0.21***
		(0.04)
Student's age (years)		-0.04*
		(0.02)
Ethnic (1=han; 0=minority)		0.03
		(0.09)
Computer use (1=ever used computer; 0=never used computer)		0.07
		(0.10)
Family Asset index		0.01
		(0.03)
Only child (1=yes; 0=no)		-0.04
		(0.06)
Father working on farm (1=yes; 0=no)		-0.10
		(0.06)
Mother working on farm (1=yes; 0=no)		0.03
		(0.08)
Father has at least a high school degree(1=yes;0=no)		-0.05
		(0.06)
Mother has at least a high school degree(1=yes;0=no)		-0.01
		(0.05)
Family size bigger than 5 (1=yes; 0=no)		-0.05
		(0.05)
At least one parent lived out of home (1=yes; 0=no)		-0.14***
		(0.04)
Teacher is female(1=yes; 0=no)		0.26**
		(0.12)
English teacher's age (years)		-0.03*
		(0.02)
Civil service teacher (1=yes; 0=no)		0.67***
		(0.17)
English teaching hours per week		-0.01
		(0.01)
Working year (years)		-0.00
		(0.02)
Education level of English teacher (1=College and above; 0=below college)		-0.02
		(0.09)
Teachers attitudes towards computer (1-5 points)		-0.22*
		(0.11)
Area (sq.km)		-0.47
		(2.19)
Computer-student ratio		1.26**
		(0.62)
Rural school (1=yes; 0=no)		0.35**
		(0.13)
Constant	-0.17*	1.30
	(0.09)	(0.85)
Observations	1,342	1,342
R-squared	0.36	0.50

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

therefore to explore the existence of heterogeneous program effects across subgroups. A linear function has been formalized to identify the heterogeneous effects (Lai et al., 2015).

Although we found no significant heterogeneous effect on student gender and school characteristics, compared to students whose English teachers were regular civil service teachers, those with short-term contract teachers had a significant reduction in test scores 0.40 SD at the 5% level (Table 7). In rural China, contract teachers have a lower income and a less stable salary than civil service teachers, so they may be less incentivized to participate in a new intervention. Additionally, it is a possibility that contract English teachers not only do not teach English well, but may even exert a negative impact on students' English learning. Another salient point is that civil service teachers are under the administration of school principals and the local education department. The OCAL program in this paper was implemented by our research group as well as by the local education department. Since the local education department coordinated the implementation of the program, we believe that civil service teachers would attach more importance and consideration to the program than contract teachers.

In order to figure out whether there exists heterogeneity among different levels of student English scores at baseline, we conducted the analysis shown in Table 8. To do this, we divided students in each school into four subgroups organized from low to high according to their baseline English scores. The first subgroup was the lowest 25th percentile of student, the second subgroup was the 25th to 50th percentile of students, the third subgroup was the 50th to 75th percentile of students, and the fourth subgroup was the 75th percentile students. In each subgroup, regressions were made following Model (1) and (2) respectively; columns (1)(3)(5)(7) are the results of not adding control variables and columns (2)(4)(6)(8) are the results of adding control variables. According to Table 8, there is a larger effect for students in the upper quantile, or students belonging to the 75th percentile of baseline English scores. Additionally, Fig. 3 shows the heterogeneity among diverse learning start level: The effect increases steadily with the upward trend by quantile. As may be concluded from Table 8 and Fig. 3, the effect of OCAL on English learning is better for students with higher English scores at baseline.

3.8. *Cost-effectiveness analysis*

We now analyze the OCAL intervention's cost-effectiveness to better understand its potential for implementation on a larger scale (Table 9). To do so, we referred to the cost calculation used in Auriol and Warlters (2012). Using the cost-effectiveness methodology developed in Bai (2016) and obtaining the cost of a traditional offline CAL intervention from that same study, we calculated a scaled-up program cost according to our original project's actual cost. The program's main costs were from teacher training and class subsidies, as well as software design, development, updates, and maintenance. These costs are all compared with those of the control group of this study which is used as the benchmark. Our calculations assume a class size of 30 students (given the study's average class size of 30.5 students).

The detailed cost calculations are as follows. First, the cost to train teachers includes communication costs (3 instances $*$ 10 RMB/instance = 30 RMB = 4.62 USD[4]), training materials (20 RMB = 3.08USD), and trainer remuneration (30 RMB = 4.62 USD). The teacher training subtotal is 80 RMB/teacher, which is equivalent to 2.67 RMB/student (or 0.42 USD). Second, 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 = 184.62 USD/ teacher. This comes out to 184.62 / 30 = 6.15 USD/student. Third, the cost to design and develop the software is a one-time expenditure of 100,000 RMB. Assuming that the software will last for 5 years, its per-student

[4] On conversion, we assume 6.5 RMB= 1 USD

Table A5
Overview of existing literature on experiments of Chinese CAL programs.

PaperNo.	Year	Location	Duration (in months)	Grade	Delivering Method	Sessions per week	Subject and β (in SD)	s.e.	Heterogeneous Effects	Source
1	Fall 2010	Beijing	2	3rd	Offline CAL	Two 40-min sessions	Math 0.15	0.04	Father without high school diploma 0.2SD; Father without college diploma 0.35SD	Lai et al. (2015)
2	2010	Beijing	9	3rd	laptops with learning/ remedial tutoring software	Not mentioned	Math 0.17	0.10	Not mentioned	Mo et al. (2013)
3	2011	Shaanxi, Qinghai, Beijing	6-8	3rd, 5th	Offline CAL	Two 40-min sessions	Beijing Math 0.12 Shaanxi Math 0.10 Qinghai Chinese 0.19	0.05 0.06 0.07	no difference in gender	Yang et al. (2013)
4	2011	Shaanxi	9	3rd, 5th	In-school offline CAL	Two 40-min sessions	Math 0.16	0.06	Boarding student benefit less **	Mo et al. (2014)
5	2013	Qinghai	9	5th	CAI(Computer Assisted Instruction)	Two 40-min sessions	English 0.08	0.04	No difference in academic performce at baseline	Bai et al. (2016)
6	Spring 2011	Qinghai	3-4	3rd	Offline CAL	Two 40-min sessions	Mandarin 0.20 Math 0.22	0.07 0.07	Low-performing students benefit more***	Lai et al. (2015)
7	2011-2012 2012-2013	Qinghai	9	3rd	Offline CAL	Two 40-min sessions	Mandarin 0.20 Math 0.22	0.07 0.08	Not mentioned	Lai et al. (2016)
8	Spring of 2011	Shaanxi	3-4	3rd, 5th	Offline CAL	Two 40-min sessions	Math 0.12	0.05	Students from poorer families benefit more ***	Lai et al. (2013)
9	2013-2014	Qinghai	9	4th	Offline CAL implemented by NGO	Two 40-min sessions	English 0.16	0.07	Not mentioned	Mo et al. (2020)

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

unit cost is 100,000 RMB / 5 years / 22 classes / (30 students/class) = 30.30 RMB/student = 4.66 USD/student. Software updates and maintenance (which includes renting servers and debugging) costs 3000 RMB / 22 classes / (30 students/class) = 4.54 RMB/student = 0.70 USD/student. Adding these together, the total software cost is 4.66 + 0.70 = 5.36 USD/student. Therefore, the Program Execution Cost (USD/ student) is 11.93 (training teachers + class subsidies + software). We can approximate the public resource investment as 20% of the program execution cost (Auriol & Warlters, 2012). Social costs include costs of the program's execution and public resource investment.

As shown in the Table 9, we compute the OCAL program cost per standard deviation raised to be 21.69 USD/student in total, whereas that of offline CAL per standard deviation raised is 55.61 USD/student. As to Social Cost-Effective Ratio, to achieve an impact of per 1 SD, the social cost of OCAL is 26.04 USD/student and offline CAL is 66.72 USD/student. In addition, with reference to McEwan (2015), we also calculated the USD cost per 0.2SD. OCAL takes 4.33-5.21 USD/0.2 SD and offline CAL takes 11.12-13.34 USD. In the summary study of McEwan (2015), the cost-effectiveness ratios of 26 intervention projects are reported. By comparing the results of this paper with them, we can see that the cost of both online and offline CAL belongs to the intervention with less cost, among which OCAL ranks 5th out of all lowest cost interventions and offline CAL ranks 11th. It thus may be seen that interventions using information technology to promote educational output is low in cost and good in effect, which is consistent with the conclusion of Kremer et al. (2013) and proves that the cost-effectiveness analysis in our paper is reasonable and reliable.

In addition, it is worth noting that OCAL's software development cost is higher than the software development cost for offline CAL software, because OCAL requires web-based development-however, this is a one-time development cost. Once OCAL reaches more students, its marginal cost will be further reduced.

From these calculations, it may be stated that OCAL has the cost-benefit advantage over offline CAL if scaled-up. Thus, the OCAL program is much more cost-effective than traditional offline CAL. This comparison between the cost-effectiveness of OCAL versus offline CAL should be considered significant for policymakers to take into consideration, as it indicates that there is a high potential for OCAL program expansion.

4. Conclusion

In this paper, we present results from a randomized field experiment for an OCAL program located in the rural areas and migrant communities of China that involved 1650 rural students in the fifth grade. The main intervention was an English OCAL remedial tutoring program that was given during the school day but outside of the students' regular English class time. Our results indicate that OCAL has significant beneficial effects on English academic outcomes. Two 40-min OCAL English sessions per week for 4 months increased student standardized English scores by 0.48 standard deviations. OCAL also significantly increased student interest in learning. Finally, the effect of OCAL on academic performance was less significant for students whose English teachers were contract teachers and not civil service teachers.

This paper contributes to our understanding of OCAL's potential impact in developing countries in several ways. First, it is the first to measure the effect of OCAL on learning outcomes among an underserved population in China using experimental design to rule out potential bias; we also explore OCAL's advantages over traditional offline CAL in depth, since previous studies on CAL in China generally all used offline CAL. Second, this paper contributes to our understanding of the mechanisms driving OCAL's impacts on academic outcomes. We explored three possible explanations for this, and administered math tests to create a placebo that allowed us to examine the Hawthorne effect. After rejecting the possibility of the Hawthorne effect and self-efficacy-induced changes, we believe that interest-oriented stimulation is the main source of improvement among the sample's students. Specifically, online features that allow for interaction and competition between peers, as well as customized remedial question banks tailored to each student's individual needs, may both contribute to the measured increases in academic performance among sampled students. Third, this paper focuses on the impact of OCAL on English learning in rural China, thereby contributing to the scant literature on CAL in the context of second-language acquisition, as most studies to date have focused on math and Mandarin reading skills (Lai et al., 2015; Mo et al., 2014; Lai et al., 2016).

One limitation of the paper is that we were not able to quantify precisely which feature(s) of the OCAL intervention contributed the largest impact when compared to previous offline CAL interventions conducted in rural China, which should be the subject of future studies. Second, due to data limitations, we were also unable to provide sound explanations for why we produced surprising results such as the lack of an effect on student self-efficacy. Third, in this study's treatment group, students received more overall instructional time (80 min of additional English instruction delivered via OCAL per week) compared to the control group. Therefore, we cannot rule out the possibility that increased schooling time contributed to better student outcomes. In another study, to isolate the technology effects of CAL from other mixed effects, Ma et al. (2020) accounted for this limitation by proposing a theoretical model illustrating the mechanisms by which CAL can affect educational outcomes.

In sum, our results show that online CAL (relative to traditional CAL) is cost-effective and may be a practical option 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 China's government has committed to making large investments in computing facilities and internet access for 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.

Declaration of Competing Interest

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

The corresponding author is responsible for ensuring that the descriptions are accurate and agreed by all authors.

CRediT authorship contribution statement

Yu Bai: Conceptualization, Methodology, Formal analysis, Writing – review & editing, Visualization, Resources. **Bin Tang:** Software, Data curation, Formal analysis, Writing – original draft, Visualization, Writing – review & editing. **Boya Wang:** Data curation, Formal analysis. **Di Mo:** Investigation, Resources, Software. **Linxiu Zhang:** Supervision, Project administration, Funding acquisition. **Scott Rozelle:** Supervision, Funding acquisition. **Emma Auden:** Writing – review & editing. **Blake Mandell:** Writing – review & editing.

Data availability

Data will be made available on request.

Acknowledgements

The authors are supported by Shaanxi Province Philosophy and Social Science Major Theoretical and Practical issues Research Project (2022ND0377), the Fundamental Research Funds for the Central Universities (2022ZYYB23), Natural Science Foundation of Shaanxi Province (2023-JC-QN-0784), National Natural Science Foundation of China (71933003), the National Social Science Foundation of China (22CJL032) and the 111 Project (B16031). The authors would also like to thank Dell's funds for their collaboration. We thank Qinghe Qu, Zhenfang Liang, Xichen Zhang, and all the other interns involved in this study for their tireless efforts.

Appendix to

Impact of Online Computer Assisted Learning on Education: Experimental Evidence from economically vulnerable areas of China

Appendix 1. Attrition test

In the section below, we describe how we tested for the presence of bias among these dropouts (that is, whether there was a bias between treatment and control groups in terms of what types or subgroups of students dropped out).

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 attendance absence 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 as shown in Table A1.

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 (Table A2). 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. In conducting the test, we estimated Eq. (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.

Our analysis indicates that there were no significant differences in student characteristics between the treatment and control groups before attrition (Table 1). 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 A3). From Table A3, we can clearly see that after attrition, the treatment group and control group are balanced in terms of the main outcome variables and a variety of characteristics. 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).

Reference for Appendix Table A5:

[Paper No. 1] 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.

[Paper No. 2] 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.

[Paper No. 3] 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.

[Paper No. 4] 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.

[Paper No. 5] 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.

[Paper No. 6] Lai, F., Zhang, L., Qu, Q., Hu, X., Shi, Y., Boswell, M., & Rozelle, S. (2015). Teaching the language of wider communication, minority students, and overall educational performance: Evidence from a randomized experiment in Qinghai Province, China. *Economic Development and Cultural Change*, 63(4), 753-776.

[Paper No. 7] 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.

[Paper No. 8] 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.

[Paper No. 9] Mo, D., Bai, Y., Shi, Y., Abbey, C., Zhang, L., Rozelle, S., & Loyalka, P. (2020). Institutions, implementation, and program effectiveness: Evidence from a randomized evaluation of computer-assisted learning in rural China. *Journal of Development Economics, 146*, 102487.

Appendix 2. Other Tables.

References

AbuSeileek, A. F. (2012). The effect of computer-assisted cooperative learning methods and group size on the EFL learners' achievement in communication skills. *Computers & Education, 58*(1), 231–239. https://doi.org/10.1016/j.compedu.2011.07.011

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), 121–142.

Auriol, E., & Warlters, M. (2012). The marginal cost of public funds and tax reform in Africa. *Journal of Development Economics, 97*(1), 58–72. https://doi.org/10.1016/j.jdeveco.2011.01.003

Bai, Y. (2016). The impact of computer assisted learning on academic performance of students in Rural China: Evidence from Qinghai. Ph. D. Dissertation. The University of Chinese Academy of Sciences.

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. https://doi.org/10.1016/j.compedu.2016.02.005

Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-efficacy: The exercise of control.

Bayturan, S., & Keşan, C. (2012). The effect of computer-assisted instruction on the achievement and attitudes towards mathematics of students in mathematics education. *International Journal of Global Education (IJGE), 1*(2). ISSN: 2146-9296.

Bianchi, N., Lu, Y., & Song, H. (2022). The effect of computer-assisted learning on students' long-term development. *Journal of Development Economics, 158*, Article 102919. https://doi.org/10.1016/j.jdeveco.2022.102919

Bolton, K., & Graddol, D. (2012). English in China today: The current popularity of English in China is unprecedented, and has been fuelled by the recent political and social development of Chinese society. *English Today, 28*(3), 3–9. https://doi.org/10.1017/S0266078412000223

Borgatti, S. P., & Cross, R. (2003). A relational view of information seeking and learning in social networks. *Management Science, 49*(4), 432–445. https://doi.org/10.1287/mnsc.49.4.432.14428

Button, K. (1998). Infrastructure investment, endogenous growth and economic convergence. *The Annals of Regional Science, 32*(1), 145–162. https://doi.org/10.1007/s001680050067

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. https://doi.org/10.1037/0022-0663.93.1.55

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), 1358–1398. https://doi.org/10.1257/aer.100.4.1358

Choi, N. (2005). Self-efficacy and self-concept as predictors of college students' academic performance. *Psychology in the Schools, 42*(2), 197–205. https://doi.org/10.1002/pits.20048

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. https://doi.org/10.1177/0261927X8500400102

Cohen, Y., & 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. https://doi.org/10.1177/003368828902000206

Dalgarno, B. (2001). Interpretations of constructivism and consequences for computer assisted learning. *British Journal of Educational Technology, 32*(2), 183–194. https://doi.org/10.1111/1467-8535.00189

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. https://doi.org/10.1080/03098770701625761

Deault, L., Savage, R., & Abrami, P. (2009). Inattention and response to the ABRACADABRA web-based literacy intervention. *Journal of Research on Educational Effectiveness, 2*(3), 250–286. https://doi.org/10.1080/19345740902979371

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning, 34*(7), 934–957. https://doi.org/10.1080/09588221.2019.1648298

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. https://doi.org/10.1016/S0749-3797(98)00084-1

Escueta, M., Quan, V., Nickow, A. J., & Oreopoulos, P. (2017). Education technology: An evidence-based review (No. w23744). *National Bureau of Economic Research.* https://doi.org/10.3386/w23744

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. https://doi.org/10.1080/0141192022000005832

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46*(3), 607–668. https://doi.org/10.1257/jel.46.3.607

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. https://doi.org/10.1007/s10887-012-9081-x

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. https://doi.org/10.1111/bld.12001

Hu, G. (2003). English language teaching in China: Regional differences and contributing factors. *Journal of Multilingual and Multicultural Development, 24*(4), 290–318. https://doi.org/10.1080/01434630308666503

Hu, G. (2005). English language education in China: Policies, progress, and problems. *Language Policy, 4*(1), 5–24. https://doi.org/10.1007/s10993-004-6561-7

Hu, G. (2009). The craze for English-medium education in China: Driving forces and looming consequences. *English Today, 25*(4), 47–54. https://doi.org/10.1017/S0266078409990472

Huang, A., & Du, X. (2007). Comparative analysis of urban and rural difference in family education in China. *Journal of Yibing University, 1*, 107–110 (in Chinese).

Hui, W., Hu, P. H., Clark, T. H., Tam, K. Y., & Milton, J. (2008). Technology-assisted learning: A longitudinal field study of knowledge category, learning effectiveness and satisfaction in language learning. *Journal of Computer Assisted Learning, 24*(3), 245–259. https://doi.org/10.1111/j.1365-2729.2007.00257.x

Inal, Y., & Cagiltay, K. (2007). Flow experiences of children in an interactive social game environment. *British Journal of Educational Technology, 38*(3), 455–464. https://doi.org/10.1111/j.1467-8535.2007.00709.x

Kelly, K., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, J., & Goldstein, D. S. (2013). Estimating the effect of web-based homework. In *Proceedings of the international conference on artificial intelligence in education* (pp. 824–827). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39112-5_122.

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. https://doi.org/10.1017/S0305741016001119

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science, 340*(6130), 297–300. https://doi.org/10.1126/science.1235350

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. https://doi.org/10.1016/j.econedurev.2015.03.005

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. https://doi.org/10.1080/19439342.2016.1220412

Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., Bowell, 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), 208–231. https://doi.org/10.1080/19439342.2013.780089

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies, 76*(3), 1071–1102. https://doi.org/10.1111/j.1467-937X.2009.00536.x

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. https://doi.org/10.1037/0022-0167.33.3.265

Li, H., Li, L., Wu, B., & Xiong, Y. (2012). The end of cheap Chinese labor. *Journal of Economic Perspectives, 26*(4), 57–74. https://doi.org/10.1257/jep.26.4.57

Li, H., Meng, L., Mu, K., & Wang, S. (2021). English Language Requirement and Educational Inequality: Evidence of 16 Million College. Stanford Kingcenter Working paper No. 2006. Retrieved from, https://kingcenter.stanford.edu/sites/g/files/sbiybj16611/files/media/file/wp2006_0.pdf.

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. https://doi.org/10.1108/17561371011044298

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. https://doi.org/10.1017/S0305741016001594

Loyalka, P., Shi, Z., Chu, J., Johnson, N., Wei, J., & Rozelle, S. (2014). Is the high school admissions process fair? Explaining inequalities in elite high school enrollments in developing countries (No. 276). Working Paper. Retrieved from: https://kingcenter.stanford.edu/sites/g/files/sbiybj16611/files/media/file/wp2006_0.pdf.

Ma, L., Xiao, L., & Liu, J. (2021). Motivational beliefs of urban and rural students in English as a foreign language learning: The case of China. *Journal of Multilingual and Multicultural Development*, 1–14. https://doi.org/10.1080/01434632.2021.1991933

Ma, Y., Fairlie, R. W., Loyalka, P., & Rozelle, S. (2020). Isolating the "Tech" from Edtech: Experimental evidence on computer assisted learning in China (No. w26953). *National Bureau of Economic Research*. https://doi.org/10.3386/w26953

Machado, P. B. (1997). *The effects of computer-assisted technology on the language acquisition rates of second language acquisition students*. United States International University.

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. https://doi.org/10.1007/BF01183893

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research, 85*(3), 353–394. https://doi.org/10.3102/0034654314553127

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

Mo, D., Bai, Y., Shi, Y., Abbey, C., Zhang, L., Rozelle, S., & Loyalka, P. (2020). Institutions, implementation, and program effectiveness: Evidence from a randomized evaluation of computer-assisted learning in rural China. *Journal of Development Economics, 146*, Article 102487. https://doi.org/10.1016/j.jdeveco.2020.102487

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. https://doi.org/10.1016/j.chieco.2015.09.001

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. https://doi.org/10.1016/j.worlddev.2012.12.019

Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., Qiao, Y., Boswell, M., & 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. https://doi.org/10.1080/19439342.2014.911770

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review, 109*(4), 1426–1460. https://doi.org/10.1257/aer.20171112

Nikou, S. A., & Economides, A. A. (2018). Mobile-based micro-learning and assessment: Impact on learning performance and motivation of high school students. *Journal of Computer Assisted Learning, 34*(3), 269–278. https://doi.org/10.1111/jcal.12240

Parsons, H. M. (1974). What Happened at Hawthorne? New evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies. *Science, 183* (4128), 922–932. https://doi.org/10.1126/science.183.4128.922

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. https://doi.org/10.1037/0022-0663.82.1.33

Quah, D. T. (1996). Twin peaks: Growth and convergence in models of distribution dynamics. *The Economic Journal*, 1045–1055. https://doi.org/10.2307/2235377

Rouse, C. E., & 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), 323–338. https://doi.org/10.1016/j.econedurev.2003.10.005

Sariman, G., & Çetin, D. (2018). The effect of computer aided education on the skill of word stressing in teaching Turkish as a foreign language. *Universal Journal of Educational Research, 6*(8), 1701–1709.

Schaefer, S., & Warren, J. (2004). Teaching computer game design and construction. *Computer-Aided Design, 36*(14), 1501–1510. https://doi.org/10.1016/j.cad.2003.10.006

Schittek, M., Mattheos, N., Lyon, H. C., & Attström, R. (2001). Computer assisted learning. A review. *European Journal of Dental Education: Review Article, 5*(3), 93–100. https://doi.org/10.1034/j.1600-0579.2001.050301.x

Schoenfeld, A. H. (2010). *How we think: A theory of goal-oriented decision making and its educational applications*. Routledge. https://doi.org/10.4324/9780203843000

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. https://doi.org/10.1111/j.1464-0597.1997.tb01096.x

Seo, Y. J., & Bryant, D. P. (2009). Analysis of studies of the effects of computer-assisted instruction on the mathematics performance of students with learning disabilities. *Computers & Education, 53*(3), 913–928. https://doi.org/10.1016/j.compedu.2009.05.002

Shannon, L. C., Styers, M. K., Wilkerson, S. B., & Peery, E. (2015). Computer-assisted learning in elementary reading: A randomized control trial. *Computers in the Schools, 32*(1), 20–34. https://doi.org/10.1080/07380569.2014.969159

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. https://doi.org/10.1016/j.pain.2005.11.017

Van Klaveren, C., Vonk, S., & Cornelisz, I. (2017). The effect of adaptive versus static practicing on student learning-evidence from a randomized field experiment. *Economics of Education Review, 58*, 175–187. https://doi.org/10.1016/j.econedurev.2017.04.003

Wang, C., Hu, Z., & Liu, Y. (2001). Evidences for reliability and validity of the Chinese version of general self efficacy scale. *Chinese Journal of Applied Psychology, 1*, 006.

Wang, H., Smyth, R., & Cheng, Z. (2017a). The economic returns to proficiency in English in China. *China Economic Review, 43*, 91–104. https://doi.org/10.1016/j.chieco.2017.01.004

Wang, X., Bai, Y., Zhang, L., & Rozelle, S. (2017b). Migration, schooling choice, and student outcomes in China. *Population and Development Review, 43*(4), 625–643.

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. https://doi.org/10.1016/j.compedu.2012.08.011

Zhang, R., & Zou, D. (2022). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning, 35*(4), 696–742. https://doi.org/10.1080/09588221.2020.1744666

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82–91. https://doi.org/10.1006/ceps.1999.1016