Computer Assisted Learning as Extracurricular Tutor? Evidence from a Randomized Experiment in Rural Boarding Schools in Shaanxi

Fang Lai
LICOS, Katholic University Leuven
and
Center for Chinese Agricultural Policy, Institute for Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences

Linxiu Zhang,* Xiao Hu and Qinghe Qu
Center for Chinese Agricultural Policy, Institute for Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences

Yaojiang Shi
School of Economic Management, Northwest University of Xi'an

Matthew Boswell and Scott Rozelle
Rural Education Action Project, Freeman Spogli Institute, Stanford University

We would like to acknowledge Dell Inc. for their generous support for REAP's Technology and Human Capital theme area. The hard work of dozens of volunteers from the Chinese Academy of Sciences, Northwest University of Xi'an, and Qinghai Minorities University made this paper possible.

*Corresponding Author:
Center for Chinese Agricultural Policy
Institute for Geographical Sciences and Natural Resource Research
Chinese Academy of Sciences
No. 11-A Datun Road, Chaoyang District,. Beijing 100101, P. R. China

Working Paper 235

April 2012

reapchina.org/reap.stanford.edu

Computer Assisted Learning as Extracurricular Tutor? Evidence from a Randomized Experiment in Rural Boarding Schools in Shaanxi

Abstract

The education of disadvantaged populations has been a long-standing challenge to the education system in both developed and developing countries. Although computer-assisted learning (CAL) has been considered one alternative to improve learning outcomes in a cost-effective way, the empirical evidence of its impacts on improving learning outcomes is mixed. This paper uses a clustered randomized field experiment in 72 schools (36 schools were part of the CAL program; 36 control schools were not) to explore the effects of a CAL program on student academic and non-academic outcomes for students in rural public schools in China. Our results show that a remedial, game-based CAL program in math held outside of regular school hours with boarding students in poor rural public schools improved the standardized math scores of the boarding students in the treatment schools by 0.12 standard deviations more than those in the control schools. Students from disadvantaged family backgrounds benefited more from the program. However, CAL did not have any significant impact on either Chinese language standardized test scores or non-academic outcomes. Our results did not find that the CAL program had any spillovers—either positive or negative—on the non-boarding students who were in the same treatment schools.

Key Words: Education; Development; Computer Assisted Learning; Random Assignment; Test Scores; China; Rural schools

JEL Codes: I20; I21; I28; O15

Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Rural Schools in Shaanxi

The education of poor and disadvantaged populations has been a long-standing challenge to education systems in both developed and developing countries (e.g. Glewwe and Kremer, 2006; Planty et al., 2008; World Bank, 2004). In China, although children in both cities and rural areas have nearly universal rates of participation between grades one to nine, there is still an achievement gap between urban and rural students—especially students from poor rural areas. In 2005 over 80 percent of urban students graduated from academic or vocational high schools (Wang et al., 2009; Ministry of Education [MOE], 2006). However, less than 40 percent of rural students from poor counties graduated from high school. In China's large municipalities almost 50 percent of students graduated from college or some other tertiary educational institution. In contrast, less than 5 percent of students from poor rural areas who started grade one in the mid-1990s matriculated into a college in the 2000s (Liu et al. 2008). The high rates of return to higher education in China (e.g., Wang et al., 2007; Li et al., 2005) and the fact that access to higher education facilitates access to formal jobs with benefits and other rights mean that the poor performance of rural students is likely to reinforce the dangerously high and rising inequality trends that have been documented by Li et al. (2011).

In fact, China's rural-urban academic achievement gap starts as early as elementary school. A recent paper (Lai et al., 2012) using standardized tests given to students in urban schools and schools in poor rural areas shows that, on average, the academic progress of a fourth-grade student in China's urban areas is significantly higher than that of an average fourth-grade student in poor rural areas This indicates that in terms of academic progress, elementary students in China's rural areas are far behind their urban counterparts.

Why are rural students—especially those from poor rural areas—scoring so much lower than urban students on standardized tests? There are many possible reasons. School facilities and teachers are systematically better in urban areas (World Bank, 2001; Wang et al., 2009). There is greater investment per capita in urban students compared to rural students (Ministry of Education and National Bureau of Statistics [MOE/NBS], 2004). Parents of urban students also have more resources for the education of their children because income per capita of households in urban areas is, on average, three times higher than income per capita in rural areas (CNBS, 2011).

Another important factor is access to remedial tutoring resources. Many studies have shown that effective remedial tutoring can significantly improve the test scores of low-performing students (e.g., Banerjee et al., 2007). When urban students fall behind in their studies, remedial tutoring services either from their teachers during or after school or from commercial sources are readily available and affordable. On average, parents of urban students also have much higher levels of education and more time to help their children with their studies at least at the elementary school level (Huang and Du, 2007). However, for students in poor rural areas, remedial tutoring is nearly nonexistent. High teaching burdens and logistical difficulties prevent rural teachers from offering after school teaching sessions. Teachers often live far away from school. Even if teachers were willing, a large share of students have to walk long distances between home and school each day and schools are required to send them home immediately at the end of the last period of class. Commercial remedial tutoring services are largely unavailable in the countryside, and even if they were, they are too expensive for impoverished rural families. Finally, parents of students in poor rural areas are often too busy to help their children. Indeed many rural parents do not live at home because they have migrated to distant cities for work. Often those that do live at home are so poorly educated that they are unable to assist their children with their coursework if they fall behind.

To bridge the rural-urban gap in educational inputs (including resources for after-school learning activities, such as remedial tutoring), efforts have been made in other countries to provide adequate educational inputs such as textbooks and school facilities for rural or disadvantaged populations in both developed and developing countries. Unfortunately these initiatives seem to have been unsuccessful in promoting learning outcomes. For example, researchers have examined the effect of interventions focusing on providing traditional educational inputs, either in the form of textbooks and flipcharts (Glewwe et al., 2002, 2004), teacher training (Jacob and Lefgren, 2004) and/or other monetary or in-kind educational inputs in both developing and developed countries (e.g. Hanushek et al., 1986, 1995). Most of the research has found that spending on educational inputs alone does not seem to be effective in raising educational performance.

As a consequence, researchers are actively exploring other ways of delivering educational inputs in order to better improve learning outcomes. Computer-assisted learning (CAL) is one such alternative (e.g. Banerjee et al., 2007; Barrow, 2008; Linden, 2008). Computer-assisted learning entails the use of computers and modern computing technologies, embodied in both software and hardware devices, to enhance learning via computerized instruction, drills and exercises (Kirkpatrick and Cuban, 1998; President's Committee of Advisors on Science and Technology, 1997). By integrating regular class materials into interesting and interactive interfaces and games, computers (as well as other devices) are thought to hold promise for making the learning process a more engaging experience for students (Inal and Cagiltay, 2007; Schaefer and Warren, 2004).

Computer Assisted Learning also can meet several needs of students that live in environments in which schools are poor in quality and the home learning environment is inadequate. For these students, CAL may be able to act as a substitute for teachers (or tutors) when the teachers are not available or have too little training and/or motivation to provide adequate instruction to the students either during or after school hours. CAL may also be a way to provide remedial tutoring services when commercialized services are either not available or not affordable. Finally, CAL might be able to provide the help that parents who are illiterate or too busy cannot provide. In these senses, CAL may be effective in poor rural areas in developing countries, where schools are plagued with poor facilities and unqualified teachers and computer technologies are relatively new and frequently out of reach of the purchase options for most families.

Despite its promise, the empirical evidence on the effectiveness of CAL in promoting learning is at most mixed. Early studies in Israel and other developed countries, such as the United States, found little consistent evidence of the beneficial effects of the application of computer technologies in school instruction on student academic achievement (e.g. Angrist and Lavy, 2002; Fuchs and Woosmann, 2004; Goolsbee and Guryan, 2006). In particular, Angrist and Lavy (2002) found that integrating computer technologies into school instruction in Israeli elementary schools led to slightly lower math test scores of eighth-grade students. While such findings would be a concern for the extension of similar projects elsewhere, an important limitation of the early studies is that they often evaluated the provision of hardware and/or software with little attention to how computers were actually used in the classroom.

Later research efforts (especially in the U.S.) went beyond the early studies by evaluating well-defined individual CAL programs using randomized experiments and found mixed evidence of the effectiveness of CAL. For example, both Dynarski et al. (2007) and Krueger and Rouse (2004) found no significant gain in test scores in either math or reading from CAL programs for U.S. students. In contrast, Barrow et al. (2008) found a computer-assisted learning program improved student math test scores on state-administered standardized tests by 0.17 standard deviations in Chicago schools.

The existing literature has several limitations that have contributed to the ambiguity in the assessment of the potential effectiveness of the use of CAL programs—especially for possible extension in developing countries. First, the majority of CAL evaluations have been done in the context of developed countries, where educational resources are abundant and computers are not novel to the students. Thus, it may not be surprising that many studies have found no significant beneficial effects of CAL on learning outcomes. However, in developing countries (or in underserved populations in developed countries), where educational resources (including school facilities, teachers, and parents) are often highly constrained, and access to technologies such as computers are limited, CAL might be expected to address the urgent needs of remedial education and engage students with technologies that are fresh and new to them. In fact, evaluations of CAL in the context of developing countries, although relatively few in number, mostly show positive effects on student test scores (Banerjee et al., 2007; He et al., 2008; Lai et al., 2012; Linden, 2008).

Second, instead of being supplementary to regular school time, many of the CAL programs in the existing literature often interfere with the regular school curriculum (as students are pulled out of class for CAL sessions). As a consequence, part of the full impact of CAL may be being offset by the negative effects of missing classes (Angrist and Lavy, 2002; Krueger and Rouse, 2004). These substitution effects might have created a downward bias in the estimation of the genuine impacts of CAL interventions. Linden (2008) found that when CAL was implemented as an in-school program (i.e., as a substitute to the regular school inputs), student test scores improved less than they would have improved if students were participating in after-school CAL programs. Hence, an after-school CAL program that is supplementary to regular school time/inputs might be a better intervention on which to measure the genuine effect of CAL on learning outcomes.

Finally, besides academic performance, CAL might also have beneficial effects on non-academic outcomes. For example, CAL might improve the interest that students have in learning or the student self-efficacy of studying.[1] These non-cognitive outcomes, to our knowledge, have seldom been examined in the literature. An exception is Lai et al. (2012), which found that an after-school CAL remedial tutoring program not only improved the academic performance of the students in migrant schools in Beijing in a short period of time, but also significantly improved the student’s interest in schooling and levels of self-confidence.

[1] Self-efficacy of studying is a psychological concept that measures one's belief in one's ability to succeed in learning and problem-solving in a certain subject. One's sense of efficacy of studying can play a major role in how one approaches goals, tasks, and challenges related in the study of a subject. Individuals with higher levels of self-efficacy in studying typically take control over their own learning experience and are more likely to participate in class and prefer hands-on learning experiences.

The overall goal of this paper is to explore the nature of the effects of CAL on student academic and non-academic outcomes for underserved student populations in a developing country. To reach this goal, we specifically pursue three objectives. First, we examine the immediate impacts of an after-school CAL math program on student academic performance in math (as measured by standardized test scores). Second, we examine the spillovers of math-focused CAL program on student academic performance in other subjects (in our case, the subject of Chinese language). Finally, we investigate the impacts of CAL on non-academic student outcomes.

To meet our goal, in this paper we present the results from a randomized field experiment of a CAL remedial tutoring program in 72 boarding schools in poor rural areas in Shaanxi Province, one of the largest (by population) and poorest provinces in northwest China. The program lasted for one semester in Spring 2011 and involved 2726 third-grade and fifth-grade boarding students, mostly aged nine to twelve. Because we only provided the CAL program to boarding school students, the participants were mostly from very poor rural families (Luo et al., 2009).

We chose boarding schools and boarding students as the main subjects of our study for two reasons. First, the trend in China is to move more towards larger, centralized schools with boarding facilities. Hence, this will be the type of schools that will be most common in rural China in the coming years. Second, according to Luo et al. (2009), the most vulnerable students in China's schooling system tend to be those that live as boarders. As boarding students, they live in the school dormitories five days a week. However, in such schools the teachers are often busy and live far away from the school. Many of the non-boarding school students also live far away from school (though not far away enough to have to board) and long commuting times (usually by walking) means that students are not allowed to stay after school. As a consequence, schools in poor rural areas almost never offer after school tutoring. In addition, when boarding school students return home during the weekend, because of the poverty of the families, their parents typically cannot afford commercial tutoring. The parents of many students are also working away from home and so the children sometimes live with their grandparents or other relatives. Even if the parents were at home, they frequently have too little education to be able to effectively tutor their children during the weekend.

The rest of the paper is organized as follows. The first section briefly lays out the context of the study—rural public boarding schools in China. The next section reviews the study's approach, including the research design and sampling, an explanation of the intervention, a description of the data and an explanation of the statistical approach. The subsequent sections then present the results, discuss the findings and conclude.

Context: Rural Public boarding Schools after China's School Merger Program

Demographic change, increased fiscal capacity and the government's resolve to try to provide higher quality education to rural students have triggered a fundamental change in China's rural education policy. Between 1951 and 2000 one of China's main educational goals was to put a school in every village (MOE, 1992). In the late 1980s and early 1990s China reached a point where there were almost 700,000 schools in the nation's 800,000 villages. By the late 1990s, however, fast income growth, demographic transition and the One Child Policy had greatly reduced the number of children in each age cohort in China's villages. Enrollment in primary schools in China's rural areas dropped (MOE, 1999). As a consequence, class sizes in many rural schools fell sharply.

In the late 1990s and the early 2000s, at a time when the central government decided it had the fiscal resources to increase the quality of rural education, China's educational leadership changed policy direction (Liu et al., 2010). In 1999 the Ministry of Education launched an aggressive School Merger Policy. According to the policy, education officials closed down small, remote schools and focused their attention on improving the teaching and facilities in larger, centralized schools.

In theory, and as was demonstrated by empirical evidence (e.g. Zhuo, 2006; Liu et al., 2010), rural students do benefit from improved educational quality by having access to larger, more centrally located educational facilities which can be built in such a way as to take advantage of scale economies. In principle, in larger centralized schools, better teachers can be hired. Facilities can be built to higher quality standards and equipped better. In larger schools, teachers are able to focus on students in a single grade and, in many cases, on a single subject. In contrast, teaching points (that is, small one-room schools in villages that often are staffed by only one teacher who is responsible for teaching children from Kindergarten to grade two, three, or four) are remotely located and sometimes accommodate fewer than 10 students. In such teaching points, the curriculum is often restricted to math and Chinese language—with little supplementary teaching of English, science, art, music or other types of courses. Central schools are supposed to offer a richer curriculum. In fact, research has shown that the merger policy has improved the quality of education—at least in terms of meeting the policy goals of the government: hiring more qualified teachers and improving the infrastructure of schools (Zhuo, 2006; Liu et al., 2010). The policy also has been widespread. The number of schools fell from around 580,000 in 1999 to 270,000 in 2006 (MOE, 1999, 2006).

While the School Merger Policy was successful in a number of dimensions, there were a number of unanticipated consequences that triggered a series of actions, responses and reactions. One of the most notable problems with the school merger program was that the distance between students' homes and schools increased dramatically (Ma, 2009). Commuting time increased. In many places commuting itself was dangerous and parents worried about the safety of their children (Xie, 2008). In response, the merger program expanded its scope and a new program was launched to build boarding facilities, encouraging or mandating that students (at least those that lived far away from school) live at school during the week away from their family. By the mid-2000s most students that needed a place to board had access to dormitory rooms (albeit in some schools the facilities were still quite rudimentary).

The fast growth of rural boarding schools has generated a lot of concerns. Young children of elementary school age have to leave the comfort and familiarity of their homes and the care of their parents to live in dormitories far away from their friends and family (Pang, 2006). The new living environment also may take a toll on the psychological and physical health of students and thus affect learning (Luo et al., 2009).

Moreover, although the opportunity for after-school learning activities has grown with the increase of boarding school students, few schools have acted to provide such activities. Even though teacher quality and school facilities of centralized rural boarding schools are better than those in small remote village schools or teaching points, they are still inadequate in providing effective after-school tutoring or learning activities. Teachers often live in county seats far away from the schools and have to leave for home right after school. Even though some of the teachers live at schools, they often do not have enough time or energy to organize after-school learning activities with heavy teaching and classroom/dormitory management responsibilities. Many of them are also unable to provide effective after-school tutoring due to their limited education or training. There are also limited facilities that boarding students could use on their own to improve learning or even for entertainment after school. Therefore, after school is dismissed around 4:30 pm, there are few productive after-school activities to involve the boarding students. They often have an early dinner around 4:30 pm or 5 pm and go to bed early. This type of schedule is definitely not an ideal way to efficiently educate the boarding students.

As a result, in part, empirical studies have shown negative educational consequences for rural boarding schools. Shi (2004) and Yue et al. (2012) have shown that when boarding schools are poorly managed, children perform worse in school. Other studies have found that the poor nutrition and health in boarding schools (relative to the home environment) are correlated with poor educational performance (Luo et al., 2009; Luo et al., 2010). Shi et al. (2009) provides evidence that students who transfer from their own village's teaching points into boarding facilities in a centrally located township school have more behavioral and psychological problems.

The education of boarding students in rural areas has become a challenge to China's education system. As discussed in the introduction, there is a large gap in academic achievement between rural and urban students (Lai et al., 2012). As we have seen in this section, boarding students may be thought of as the most vulnerable of the vulnerable. Given China's policy direction for rural education, the number of rural boarding schools will almost certainly continue to increase in the coming years. Therefore, from a policy perspective it is critical to begin to create a productive after-school learning environment for rural boarding students as a way to both improve the efficiency and equity of China's education.

Sampling, Data and Methods

Sampling and the Process of Randomization

We conducted a clustered (at the school level) RCT of Computer-Assisted Learning (CAL) in Shaanxi rural schools in the Spring semester of 2011. A total of 5943 students in 72 schools of Shaanxi rural schools are involved in our study. Among these students, 2726 students are boarding students. These boarding students constitute the main sample for our study. The other 3074 students, who are non-boarders in the same schools, serve as additional controls to check for spillover effects. The non-boarders did not receive the CAL program.

Choosing the sample consisted of several steps. First, to focus our study on poor rural students, we restricted our sample frame to four counties randomly selected out of the ten counties in Ankang Prefecture, the prefecture that covers one of the poorest areas in the southern part of Shaanxi Province. Shaanxi Province is a large (a population of nearly 40 million), rural (more than 60 percent of the population live in rural areas) and poor province in northwestern China. The average per capita income of these four counties is only around 4000RMB (around $600) per year in 2011, which is far below rural China's average per capita income of 6977RMB in the same year (CNBS, 2011). Three out of the four sample counties are nationally-designated poverty counties in China.[2]

After choosing the counties, we obtained a comprehensive list of all *wanxiao* (or all elementary schools with six full grades, grade one through grade six) in each of the four counties

[2] There are 592 national designated poverty counties among the more than 2000 county-level jurisdictions in China. The Leading Group of the Alleviation of Poverty gave counties the designation in the 1990s based on the severity of the level of poverty in the county.

from the Department of Education of Ankang Prefecture. We used two criteria to choose our sample schools. We called each school to confirm whether the school was a boarding school for both third and fifth-grade students and excluded schools with too few boarders (i.e, less than 16 boarding students in either grade). We excluded all schools if they did not use text books in their math classes that were based on China's "uniform national math curriculum." This exclusion criterion was used because these schools would not meet the requirements of our CAL program (which provided remedial tutoring material that was centered on the uniform national math curriculum—an issue we elaborate more on below). Eventually, we included all 72 schools that met the above two criteria in our sample.

Within the sample schools, we included only third-grade and fifth-grade students in the 72 schools in our sample. We chose third-grade and fifth-grade students for several reasons. First, for aforementioned reasons, we designed the program to target boarding students. For safety and management concerns, many schools only provide boarding to students that are in grade 3 and above. For this reason, we did not choose students from grade 1 or 2. Second, given the limited number of computers in each school's computer room and the schedule of boarding students, the CAL program could only accommodate students from two grade levels (instead of four).[3] We

[3] We only had 240 computers for the 36 treatment schools, which was not enough to accommodate students of the same grade (or even the same class) simultaneously. Therefore, the students of the same grade needed to break into several groups and took turns in using the computer for CAL sessions. Moreover, the CAL protocol requested each student should have had two 40-minute sessions per week, yet in most boarding schools, school was over around 4:30 pm and students went to bed around 8 pm. Students also ate dinner during that period of time. In

first excluded the sixth-grade students because they were fully occupied with taking the elementary school graduation test and many principals did not want to give them time to participate in the CAL program. Moreover, the program started in the Spring 2011. Because we were thinking of extending the program into the next academic year, the sixth-grade students would have already graduated and exited our sample. Finally, we chose the third and fifth graders instead of third and fourth graders or fourth and fifth graders because third and fifth graders offer a sharper comparison of the intervention effects by age group.

So who was included in the sample? In fact, all of the third-grade and fifth-grade students in the 72 sample schools were included in the sample, though the boarding students in treatment schools received the CAL intervention. In total, there were 5943 students in the sample, among whom 2726 were boarding students. Among the boarding students, 1155 were third-grade students and 1571 were fifth-grade students (Figure 1).

Although at the time of the baseline survey, the main sample included a total of 72 schools and 2726 boarding students, there was some attrition by the end of the study and a few students were not included in our analysis. For various reasons (mainly because of school transfers and extended absences due to illness or injuries) by the time of the evaluation survey we were only be able to follow up with 2613 boarding students in the 72 sample schools (Figure 1, final row). In other words, 2613 out of the initial 2726 students were included in our evaluation survey and were part of the subsequent statistical analysis. There were 31 attrited

---

addition, each week, students leave for home early on Friday afternoon, and thus the schools could not arrange any CAL sessions on Friday. As a result, it was infeasible for the CAL program to accommodate boarding students of all of the four grades (third to sixth).

students from the third grade and 82 attrited students from the fifth grade. Older students, students who were the only child of their family (only weakly significant at the 10% level) and those who had lower Chinese test scores (for third-graders only) were more likely to leave the sample (Table 1, columns 1 to 3).

We do not consider the attrition to be a serious problem for our study for two reasons. First, the attrition rate was as low as 4%, and thus is unlike to have any substantial influence on our subsequent analysis. Second, when comparing the attrited students in the treatment group to those in the control group, we found they had similar characteristics (Table 1, column). This suggests that, in general, the factors leading to attrition were largely the same for both groups.

After choosing the 72 schools for our sample, we randomly chose 36 schools from these 72 schools to receive the CAL intervention. As the CAL intervention only engaged third- and fifth-grade boarding students, the 1275 boarding students of the third and fifth grades in the 36 treatment schools constitute the treatment group (Figure 1). Among these students, 553 are third-grade students and 722 are fifth-grade students. The 1451 boarding students of the same grade (602 from the third grade and 849 from the fifth grade) in the other 36 schools served as the control group. Due to attrition, there were 2613 students left in our final analytic sample, among whom, 1205 were from the 36 treatment schools, and 1408 were from the control schools. We used a set of student characteristics to check the validity of the random assignment, and found that with the exception of the dummy variable of *both parents at home* (significant at the 5% level for the whole sample, and 10% level for the sample of fifth-grade students), the differences between the treatment and control groups were not only statistically insignificant for all student characteristics but also small in magnitude in most cases (Table 2, columns 1 to 3).

Experiment Arms/Interventions

Excluding the non-boarding students in the 72 sample schools, who would go home after school and had no access to our CAL program, our experiment focused fully on one treatment group (the boarding students of the 36 treatment schools) and one control group (the boarding students of the 36 control schools).

CAL Intervention Group (the boarding students in the 36 treatment schools)

The main intervention involved computer-assisted math remedial tutoring sessions which were designed to complement the regular in-class math curriculum for the spring 2011 semester. Under the supervision of two teacher-supervisors trained by our research group, the students in the treatment group had two 40-minute CAL sessions per week after school at a time after the end of the school day when the non-boarding students had left for home. The sessions were mandatory and attendance was taken by the teacher-supervisors. The content (instructional videos and games) of each session was exactly the same for all students in each of the treatment groups and emphasized basic competencies in the uniform national math curriculum.

During each session, two students shared one computer and played math games designed to help students review and practice the basic math material that was being taught in their regular school math classes. In a typical session, the students first watched an animated video that reviewed the material that they were receiving instruction on during that particular week during their regular math class sessions. The students then played math games to practice the skills introduced in the video lecture. The math games also used animated characters. When playing the games, the students first worked out the solutions with pencils/pens on scratch paper and then submitted the answers using the keyboards and the mice of their computers. If a student had a math-related question, he/she was encouraged to discuss with his/her teammate (the one with whom he/she shared the computer). In other words, students were encouraged to try their best to work out the solutions to all math-related questions together as a team of two. The students were not supposed to consult the other teams or the teacher-supervisor. According to our protocol, the teachers were only allowed to help students with scheduling, computer hardware issues and software operations. In fact, in our observations, the sessions were so intense that the attention of the students were fully on the computer and, while there was a lot of interaction between the members of the two-person teams, there was little communications among the groups or between any of the groups and the teacher-supervisor.

Our research team took great care in preparing the necessary hardware, software, CAL curriculum and program implementation protocol in a way that would both facilitate smooth implementation of the CAL program and prevent confounding influences that might bias our results. As the first step of these efforts, to meet the hardware requirements of the CAL program, we acquired (by way of donation from Dell, Inc.) 240 brand new identical desktop computers. Our CAL software package was installed on these desktops. We then removed all pre-installed software that would not be used during the CAL intervention (such as Windows built-in games and Microsoft Excel). We also disabled the Internet and USB functions on all of the computers. By doing so, we were able to prevent school teachers or other students from using the program computers for other purposes that might affect the operation of the regular CAL program. It was also impossible to upload/install or download software or other material. This was done in part to help avoid the interruptions that might otherwise be caused by accidental deletion of the CAL software or the introduction of viruses. It also was done so that our evaluation of the program effects would not be capturing any other confounding influences (spillovers) if students were able to learn from (or be distracted by) other sources of information that might be accessed, for example, through the Internet. This also avoided the situation that might occur if teachers/students from the control classes were able to copy our CAL software onto other computers.

Both the third and the fifth grade CAL software packages were composed of two individual pieces of software. The first piece of software was a commercial, game-based math-learning software program that was obtained via donation. This package was adopted because it did exactly what we designed the CAL program for. The software provided remedial tutoring material (both animated reviews and remedial questions) in math for the third- and fifth-grade students following the national uniform math curriculum. The designers of the program also set up their software so it could be used in conjunction with the material that students were learning in their math class on a week by week basis.

We developed the second piece of software by ourselves. Our software package (henceforth, the CAL software) was developed to provide the students with a large number of practice questions. The questions were all asked and answered (by the students) in game-based exercises. In choosing the math questions to include in the CAL software, we consulted experienced elementary school math teachers in both public schools in the cities and in the rural areas, as well as experts who were key committee members of the Center for Examination of Beijing, an institute that designs city-wide uniform tests for elementary schools in Beijing. Under their direction and assistance, we chose questions for the CAL software from several commercially available books of practice questions. In order to make the games attractive to students, we recruited volunteers from the Tsinghua University's Department of Computer Science and Graphics Design, one of the top computer science departments in China, to design the animation/picture-based game interface. By combining the commercial software and the

CAL software, we had enough content and exercise games to cover the math course materials for the entire spring 2011 semester and the material was sufficient to provide 80 minutes of remedial tutoring per week (two sessions times 40 minutes per session).

We also produced and included in the CAL software package an audio-enhanced PowerPoint tutorial to demonstrate to the students in a step-by-step fashion how to use each software program. The tutorial also taught students a number of basic computer operations. We exerted a great deal of effort to draft the tutorial in a way that third-grade students with low levels of literacy could understand. The words were simple. We made extensive use of graphic illustrations. An audio file of the same content was also inserted into each PowerPoint slide so that the students who had low levels of reading comprehension could still understand the material being taught in the tutorial.

With both software and hardware ready, we then designed a detailed CAL curriculum and implementation protocol. The protocol was targeted mainly at the teacher-supervisors that were charged with implementing the CAL program in each school. The CAL curriculum was designed so that the progress of the CAL program would match the progress of school instruction on a week by week basis. This was done so that our CAL sessions provided timely review and practice of the knowledge and skills that were introduced and covered as part of their regular math class. One of the most important jobs of the teacher-supervisor was to make sure the weekly CAL sessions were proceeding on a pace that matched the pace in the students' regular math classes. To avoid confounding the effect of the CAL intervention itself with any influence of additional teaching inputs to the students by the teacher supervisors, we requested none of the teacher-supervisors should be math teachers or homeroom teachers of the third- and fifth-grade students. Moreover, teacher-supervisors were not allowed to help students solve the math problems during the CAL sessions.

The implementation protocol was presented in a manual for each grade. Each manual, which was given to the teacher-supervisor as a bound, printed-out booklet and contained detailed instructions. The manual contained four main sections: a.) the detailed CAL curriculum; b.) CAL classroom rules for both students and teacher-supervisors; c.) the responsibilities of the teacher-supervisors when supervising the CAL sessions (what to do and what not do to); and d.) tutorials (in both words and graphic illustration) on basic computer operations, CAL software use and troubleshooting. As in the case of the tutorials (described above), we took care when drafting the protocol so that it was presented in a way that teachers/principals with neither high levels of education nor deep experience with computer use would be able to easily understand the CAL program and the instructions covering computer and software use.

To ensure that the protocol would be properly implemented, we requested that each school assign two teachers to supervise all of the CAL sessions according to the protocol. The teacher-supervisors' five main responsibilities included: a.) taking attendance; b.) making sure that the CAL curriculum in each session was matched to the curriculum being taught in the students' math class; c.) managing the CAL classrooms so that order was maintained; d.) providing immediate assistance when students experienced difficulty in computer and/or math game software operations (but they were not supposed to instruct the students in math); and e.) taking care of the CAL desktops and keeping close contact with our research group/volunteers regarding technical support or CAL management questions. Because this work was clearly beyond the scope of their normal classroom duties, we compensated the teacher-supervisors with a monthly stipend of 100 yuan (approximately 15 USD). This amount was roughly equal to 15 percent of the wage of a typical rural teacher. To prepare teacher-supervisors for their duties, before the spring 2011 semester started, all teacher-supervisors of the 36 treatment schools were required to attend a two-day mandatory training that was held at a central site. The project budget covered room and board and transportation costs for the teachers during the training period.

To further ensure that the teacher-supervisors (and the students under their supervision) strictly followed the protocol, we recruited volunteers from universities in Ankang Prefecture and directed them to pay visits to the treatment schools during the implementation of CAL. During the visits, the volunteers were instructed to attend the CAL sessions and observe whether the protocol was being strictly implemented. The volunteers did not announce their visits to the schools in advance. They also were instructed to avoid all unnecessary interactions with students and teachers so that they would not interrupt the sessions or provide additional assistance to CAL session management, which might confound the program effect. When irregularities were found, the volunteers took notes and informed the CAL management team after the school visit. If the irregularities were so serious as to hinder the normal progress of the CAL intervention, the project manager either called or visited the teacher-supervisor.

Finally, we also provided technical support and free computer repairs and maintenance for the entire semester. We offered what we called a “24/7 consultation hotline” to answer all CAL-related questions, ranging from computer and CAL software operations to classroom management. In addition to monitoring the CAL sessions, our program volunteers also conducted basic on-site computer maintenance during their school visits (twice a semester). They also picked up defective desktops and accessories for repair and reinstalled replacement or repaired laptops and accessories to replace the defective ones.

CAL Control Group (the boarding students in the 36 control schools)

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

Additional Control Group (the non-boarding students in all the 72 sample schools)

To examine the spillover effects of the CAL intervention, we used the 3074 non-boarding students in the third and fifth grade in the 72 sample schools as an additional control group. The CAL intervention might not only affect the performance of the boarding students in the treatment schools, but also that of the non-boarding students in the same school via interactions between these two groups of students. This indirect effect of CAL could be in either direction. On the one hand, non-boarding students could indirectly benefit from the CAL intervention by learning from the boarding students who receive the CAL intervention. On the other hand, the CAL intervention could also hurt the non-boarding students if they felt they were being neglected by the school and this might lower their motivation and the level of efforts in learning during the regular school year. By comparing the academic performance of the non-boarding students in the treatment and control schools, we were able to examine whether the CAL intervention generated any spillovers to the non-boarding students of the treatment schools.

Data Collection

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

In each round of the survey, the enumeration team visited each school and conducted a two-block survey. In the first block students were given a standardized math test and a standardized Chinese test. The math test included 29-31 questions (tests in different grades and rounds included slightly different numbers of questions). The Chinese test included 27-42 questions. Students were required to finish tests in each subject in 25 minutes. All students took the math test first and then they took the Chinese test. Our enumeration team monitored the test and strictly enforced the time limits and tried to make sure there was no cheating. We use the scores of the students on the math and Chinese tests as our measures of student academic performance.

In the second block enumerators collected data on the characteristics of students and their families. From this part of the survey we are able to create demographic and socioeconomic variables. The dataset includes measures of each student's *age* (measured in years), whether the student is *female*, grade, county, whether one is the *only child* of his or her family, father's education level (*father has at least high school degree*), mother's education level (*mother has at least high school degree*), whether their parents are still farmers and work only on the farm (*family off-farm*) and poverty status (whether one receives a *poverty subsidy* at school). To create indicators of parental care, during the survey the students were also asked whether their parents had migrated to some other location outside of his/her home town or whether their parents stayed at home for most of the time during the semester (*both parents at home*).

Importantly, in the second block students were also asked to answer questions that could help us measure their noncognitive traits, such as their attitudes toward schooling and the level of metacognition and the self-efficacy of studying math (i.e. one's belief in their ability to excel in solving math problems). To create the indicator depicting the student's attitude toward schooling (*like school*), the students were asked to rate their attitude toward school on a 0-100 scale, where "0" indicates "extremely hate school" and "100" indicates "extremely enjoy school." The indicators of metacognition and the self-efficacy of studying math were created from the responses of students to a seventeen-item psychological scale measuring metacognition,[4] and a seven-item psychological scale for the self-efficacy of studying math.[5]

---

[4] Metacognition is defined as "cognition about cognition," or "knowing about knowing." It refers to a level of thinking that involves active control over the process of thinking that is used in learning situations. Planning the way to approach a learning task, monitoring comprehension and evaluating the progress towards the completion of a task: these are skills that are metacognitive in their nature. Similarly, maintaining motivation to see a task to completion is also a metacognitive skill. The ability to become aware of distracting stimuli – both internal and external – and sustain effort over time also involves metacognitive or executive functions. Metacognition helps people to perform many cognitive tasks more effectively (Metcalfe and Shimamura, 1994).

[5] To measure the self-efficacy of math studying, a professor in psychometrics and measurement in Beijing Normal University helped us choose among the 12 indicators of math attitudes used in TIMSS 2003 and developed a seven-item scale of self-efficacy of math studying that is appropriate to use under the context of elementary schools in China.

For the baseline survey only, information about the access of students to computers and use of any educational software were collected. The information collected in this subblock of the survey allowed us to create variables that include whether the students had *ever used a computer* and whether they had ever had *access to other modern technologies*.

Statistical Methods

We used both unadjusted and adjusted ordinary least squares (OLS) regression analysis to estimate how the academic and non-academic outcomes changed in the treatment group relative to the control group. Our unadjusted analysis regressed changes in the outcome variables (i.e. post-program outcome value minus pre-program outcome value) on a dummy variable of the treatment (CAL intervention) status. We used adjusted analyses as well to control for some systematic differences between the treatment and control groups, improve precision and test for heterogeneous treatment effects (we will describe these approaches in detail in the models below). In all regressions, we accounted for the clustered nature of our sample by constructing Huber-White standard errors corrected for class-level clustering (relaxing the assumption that disturbance terms are independent and identically distributed within classes). The models are presented in order of increasing comprehensiveness.

First, the unadjusted model is:

$$\Delta y_{isgc} = \alpha + \beta \cdot treatment_s + G_g + \varepsilon_{isgc} \tag{1}$$

where Δy_{isgc} is the change in the outcome variable during the program period for child *i* in school *s* , grade *g* and class *c*, $treatment_s$ is a dummy variable for a boarding student attending a treatment school (equal to one for students in the treatment group and zero otherwise),

G_g captures the grade fixed effects (the grade dummies are not included in models using sample of each grade), and ε_{isgc} is a random disturbance term clustered at the class level.

We used several variables to measure the student academic and non-academic outcomes (y_{isgc}). The primary outcome variable of our analysis is the student academic outcome, measured by the student standardized math test score. We also included the student standardized Chinese test score as an additional academic outcome measure. By doing so, we are able to examine if there are any positive or negative spillovers of the CAL intervention to student academic performance in Chinese, the other major subject in China's elementary schools besides math.[6] Importantly, besides variables measuring academic outcomes, we also included three non-academic outcome variables, namely, *like school*, metacognition and self-efficacy of studying math.

By construction, the coefficient of the dummy variable $treatment_s$, β, is equal to the unconditional difference in the change in the outcome (Δy_{isgc}) between the treatment and control groups over the program period. In other words, β measures how the treatment group changed in the outcome levels during the program period relative to the control group.

[6] For example, the CAL program might have improved the student's general learning ability and thus the student Chinese test score might also increase. The CAL program might have also taken up so much of the student's time and energy in learning math that the student had less time and energy to spend on Chinese. In this case, the CAL program in math might negatively affect the student academic performance in Chinese.

To control for any unbalance in the student characteristics (as discussed above: *both parents at home*) and in order to improve the efficiency of the estimation, we built on the adjusted model in equation (1) by including a set of control variables:

$$\Delta y_{isgc} = \alpha + \beta \cdot treatment_s + G_g + \theta \cdot y_{0isgc} + X_{isgc}\gamma + \varepsilon_{isgc} \qquad (2)$$

where all the variables and parameters are the same as those in equation (2), except that we added a set of control variables. Specifically, we controlled for y_{0isgc}, the pre-program outcome value for student i in school s, grade g and class c, and X_{isgc}, a vector of additional control variables. The variables in X_{isgc} are student *and* family characteristics (*female, age, grade, county, only child, father has no high school degree, mother has no high school degree, family off-farm, both parents at home, poverty subsidy, ever used a computer, and access to other modern technology*). By including y_{0isgc} and X_{isgc} as control variables, β in equation (2) provides an unbiased, efficient estimate of the CAL treatment effect.

Results

The data show that boarding students in the treatment group improved significantly more in their math performance than did students in the control group, especially in the case of the third-grade boarding students (Figure 2). The pre-test standardized test scores are lower in the treatment than in the control groups (Panel A, bars labeled with "Before").[7] After the CAL intervention, the treatment group improved significantly more in math than did the control group

[7] The test scores are normalized to standardized scores with mean equal to zero and standard deviation equal to one.

(Panels A and B). The difference in change in standardized math test scores between the two groups is 0.14 standard deviations for the third graders (Panel D). Considering that the program only ran for one semester, the size of the CAL program effect is comparable to the findings in other CAL evaluations that observed beneficial effects of CAL on student performance (e.g., Barrow, 2008; Banerjee et al., 2007; Linden, 2008). From the graph, compared to the fifth-grade boarding students in the control group, fifth-grade boarding students in the treatment group did not seem to have significantly improved their math test scores during the CAL semester (Panels E and F). Part of the reason is that all fifth-grade students had high test scores on the final evaluation math test so that the test score distribution was slightly skewed to the left (Figure 3). In other words, the standardized math test might have been "too easy" for the fifth-grade students. If this were the case the exam might have limited ability to detect difference in the changes in math competency of the students for those scoring in the very uppermost part of the test score distribution.

The multivariate regression analyses (adjusted and unadjusted) are mostly consistent with our graphical descriptive analysis. Using the full sample, including only boarding students from both the third and fifth grade classes, the estimated CAL treatment effect on math test scores is equal to 0.12 standard deviations and is significant at the 5% level using either the unadjusted model (equation (1)) or the adjusted model (equation (2)—Table 3, row 1, columns 1 and 5). When running the multivariate regressions using the grade 3 and grade 5 samples separately, we find that the CAL treatment effect is particularly significant for grade 3 boarding students. The estimated CAL treatment effect using the unadjusted model on math test scores of the grade 3 boarding students is equal to 0.14 standard deviations and is significant at the 10% level using the unadjusted model (row 1, column 2). When we add the additional control variables (using the adjusted model), the estimated treatment effect for grade 3 boarding students increases to 0.18 standard deviations (row 1, column 6) and is significant at the 5% level.

For the fifth-grade boarding students (when we use the entire sample), even though the estimated CAL treatment effect on math test scores is not significant using either the unadjusted model or the adjusted model (Table 3, row 1, columns 3 and 7), there is still evidence of the CAL treatment effects. As stated above, the problem seems to be that the distribution of the standardized math test scores was skewed to the left so that the test might be insensitive to differences in math competency among students who have similarly high scores. When we restrict our sample to students that scored lower than the 70th percentile in the post-CAL math test score (i.e., if we exclude the top 30 percent students in the post-CAL math test score distribution), the estimated CAL treatment effect becomes 0.11-0.12 standard deviations and is significant at the 10% level (row 1, columns 4 and 8). This result is consistent with what we have observed from the graphic evidence (i.e., CAL still has a certain level of positive impacts on the academic performance of the grade 5 boarding students—for students that were not at the top of the left-skewed test score distribution).

Heterogeneous Effects of the CAL Intervention on Student Academic Performance

The estimation results using Equation (2), and including a number of additional interaction terms between the treatment variable and student characteristics, show that, in general, students from disadvantaged family backgrounds benefited more from the CAL intervention (Table 4, rows 2 and 4). For the third-grade students, compared to the students in the control group, students in the treatment group who were not the only child of their parents (and those from families that received poverty subsidies from the government) improved their standardized math test scores by 0.28 standard deviations (0.18 standard deviations) than those from only child homes (those students from families that did not receive poverty subsidies—row 2, columns 1 and 4). For the grade 5 treatment students, when we compare them to the students in the control group, students whose mother had no high school degree improved 0.36 standard deviations more in their math scores than those whose mother had at least high school degree (row 5, column 2), and those whose father had no high school degree improved 0.24 standard deviations more in their math scores than those whose father had at least high school degree (row 5, column 3). Fifth-grade students with lower Chinese test scores on the baseline test also benefited significantly more from our CAL intervention (row 5, columns 5). In other words in these many cases we are finding that students from families that were less able to provide them with tutoring and other support are improving more. One explanation of this is that the CAL is doing exactly what it is designed for: provide remedial tutoring to poor children.[8]

[8] The only exception is that compared to the control group, grade 3 students whose mother had a high school degree benefited more from the CAL intervention than those whose mother had no high school degrees (row 2, column 2). This result is difficult to explain. However, one explanation is that since this heterogeneous treatment effect is only weakly significant at the 10% level it maybe appearing by chance. In addition, we find no significant evidence of CAL intervention heterogeneous program effects for other student demographic and family characteristics (i.e., *female, age and family off-farm*), or for the student baseline math test scores or for the students' access to computers before the program started (results not included in the table for simplicity).

Spillovers and the Impact of CAL on Non-Academic Outcomes

Even though the CAL intervention has positive and significant effects on the academic performance of boarding students among both the grade 3 and grade 5 students, this benefit does not seem to have "spilled over" to non-boarding students in the treatment schools or the student performance in other subjects. Specifically, there is no significant difference in the change of math test scores over the program period between non-boarding students in treatment and control schools (tables of results not included for simplicity). In addition, compared to their counterparts in the control schools, grade 3 and grade 5 boarding students in the treatment schools did not show any significant improvement in their standard Chinese test scores over the program period (tables of results not included for simplicity).

So what does this mean? On the one hand, these results suggest that the CAL intervention does not create significant positive spillovers for the non-boarding students who were not covered by the program. The results also suggest that the impacts were only on math test scores—following the subject matter that was the focus of CAL—and not on Chinese test scores. Is this a strike against CAL? In fact, what is perhaps more important is that there is no evidence that the CAL intervention improved the performance of boarding students in math at the expense of the boarding student performance in Chinese (or at the expense of the academic performance of the non-boarding students in the same school).

The CAL intervention also does not appear to have any significant impact on student non-academic outcomes. Compared to students in the control group, the students in the treatment group did not "like school" more. According to our data, boarding school students did not report higher levels of math study efficacy or metacognition after the CAL program. One possible reason for CAL's lack of effect on the student's metacognition and self-reported math study efficacy may be the remedial nature of our CAL program. Due to its remedial nature, our CAL program focused more on repeated exercises rather than creative math learning and problem solving. Consequently, for this reason we perhaps should not be surprised to observe after the program, the students do not believe that they are more capable in math problem solving, in general. This result also helps clarify the mechanism underlying the impact of CAL on test scores. Because of the absence of any non-academic effects of CAL, it does not appear as if the rise of test scores is due to increased interests in schooling or improved metacognition or self-efficacy of studying math brought up by the CAL program. Instead, this result indicates that it is the repetitive remedial drills and exercises of CAL that leads to the improved test scores.

Conclusions

In this paper we present the results from a randomized field experiment of a Computer Assisted Learning (CAL) program involving around 2726 third-grade and fifth-grade boarding students, mostly aged nine to twelve and from poor rural families, in 72 rural public schools in Ankang, Shaanxi. To evaluate the effectiveness of the program we randomly chose 36 schools from the entire sample as treatment schools and the third- and fifth- grade boarding students (and boarding students only) received the CAL intervention. The remaining 36 schools served as control schools and boarding students attending these schools served as the control group. The main intervention was a math CAL program that was held outside of regular school hours. Third and fifth-grade boarding students were offered 40 minutes of shared computer time after school, twice a week. During the sessions students played computer-based games that required them to practice using their knowledge of math and relatively simple problem solving skills. The CAL program was tailored to the regular school math curriculum and was remedial in nature, providing the students with drills and exercises that were related to the material that they were learning in class. There was also an animation-based tutoring session that reviewed the lesson of the week.

Our results indicate that CAL has significant beneficial effects on both student academic and non-academic outcomes, at least in the short term. Two 40-minute CAL math sessions per week increased the student standardized math scores by 0.12 standard deviations. In general, students with disadvantaged family backgrounds benefited more from the program. Moreover, the CAL program did not improve the math performance of the boarding students at the expense of their performance in Chinese or the academic performance of non-boarding students. In fact, CAL did not have any positive or negative spillovers on Chinese test scores of the boarding students or on the academic performance of non-boarding students in the treatment schools.

This paper contributes to the understanding of the effect of CAL on learning outcomes for underserved populations in developing countries in two respects. First, we took care in preparing software and hardware for the program and designing our CAL program implementation and evaluation protocol in order to prevent some potentially confounding influences. Many previous studies reported various shortcomings in program implementation (e.g., schools in the treatment group used program computers for other purposes, such as in the case reported in Banerjee et al., 2007) that might potentially have biased the evaluation results. Our protocol took various measures to prevent such interferences. By implementing the CAL program as a fully supplementary program (that is, it was held outside of regular class), we also eliminated any substitution effects that might have diminished the program effects.

This paper also contributes to the understanding of the effects of CAL by its relatively broad research dimension. Besides providing evidence that the CAL intervention significantly improved student outcomes in underserved populations in developing countries, this paper also examined how this impact changed across different student groups. We also show, unlike that case of Lai et al. (2012), that CAL did not affect non-academic outcomes that may be important to student intellectual development.

Given the significant impacts of CAL on student academic outcomes found in this paper, educational policy makers in China (and in other developing countries) should consider upscaling CAL programs, especially in public schools serving disadvantaged students (e.g., rural public schools in China). Of course, rural public schools in Shaanxi might not be representative of all poor public schools in China or in other developing countries and rural boarding students might not be representative of all rural students. Nonetheless, all public schools in poor rural areas or that serve disadvantaged students do share some common problems: low teacher quality, poor school resources, lack of remedial tutoring and the resulting persistent underperformance of the students. Rural boarding students are also the most vulnerable among all rural students and suffer from difficult living conditions and learning environments as most disadvantaged students do.

Importantly, in order to narrow both the academic achievement gap and digital divide between the rural and urban areas, China's government, and increasingly more governments in developing countries, have committed to making large investments in the computing facilities in rural public schools. However, in many rural schools, after the investment in computing facilities, the computer rooms are locked and the computers are frequently unused because the schools do not know how to properly use them to facilitate student learning. A CAL program, as a complementary input to existing computing resources, has the potential to promote learning outcomes for underserved students by productively using these technologies. Therefore, we believe that the government might want to consider extending CAL programs on a larger scale in China (and in other developing countries) and then rigorously evaluate these new initiatives to inform policies that intend to provide better educational service to the poor. It is hoped that all of these efforts aimed at decreasing the rural-urban achievement gap and digital divide will eventually contribute to the economic equity and sustainable development in China as well as in other developing countries.

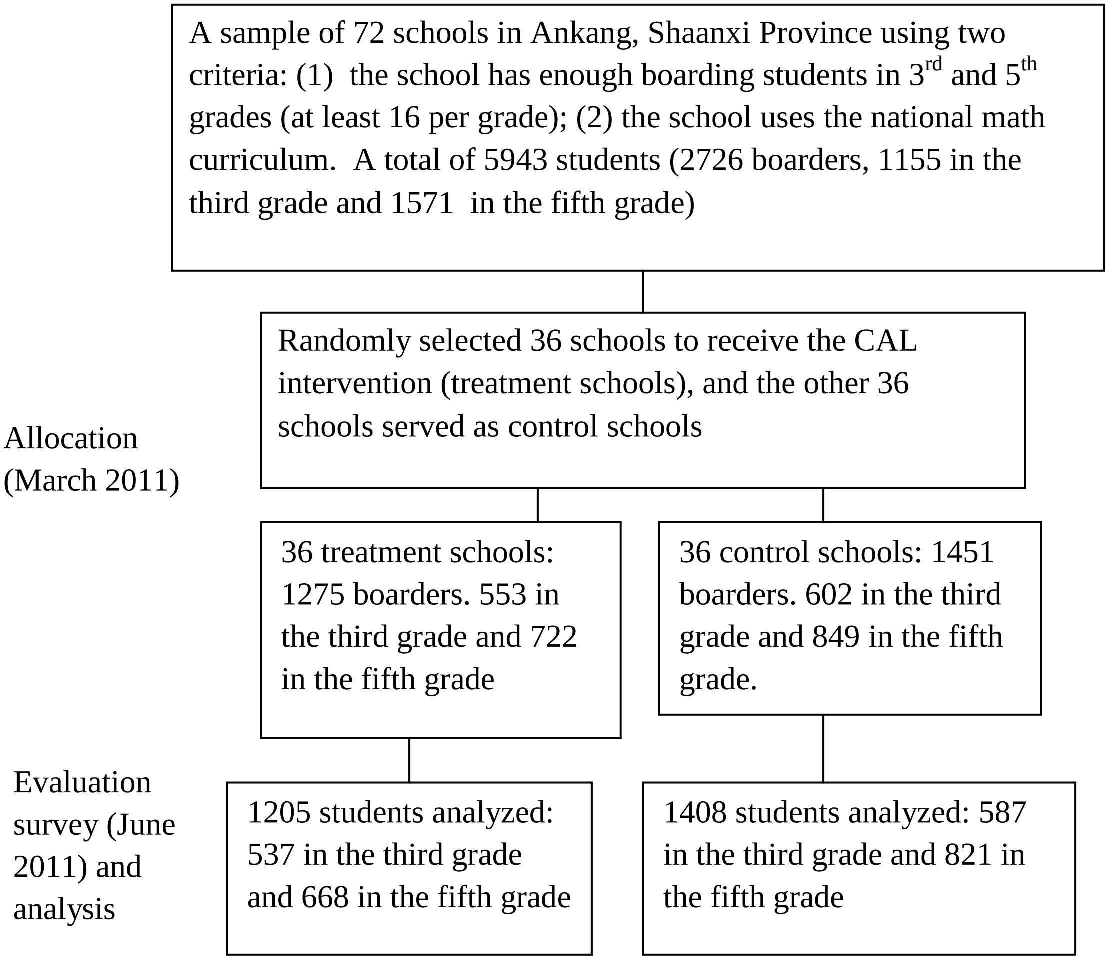
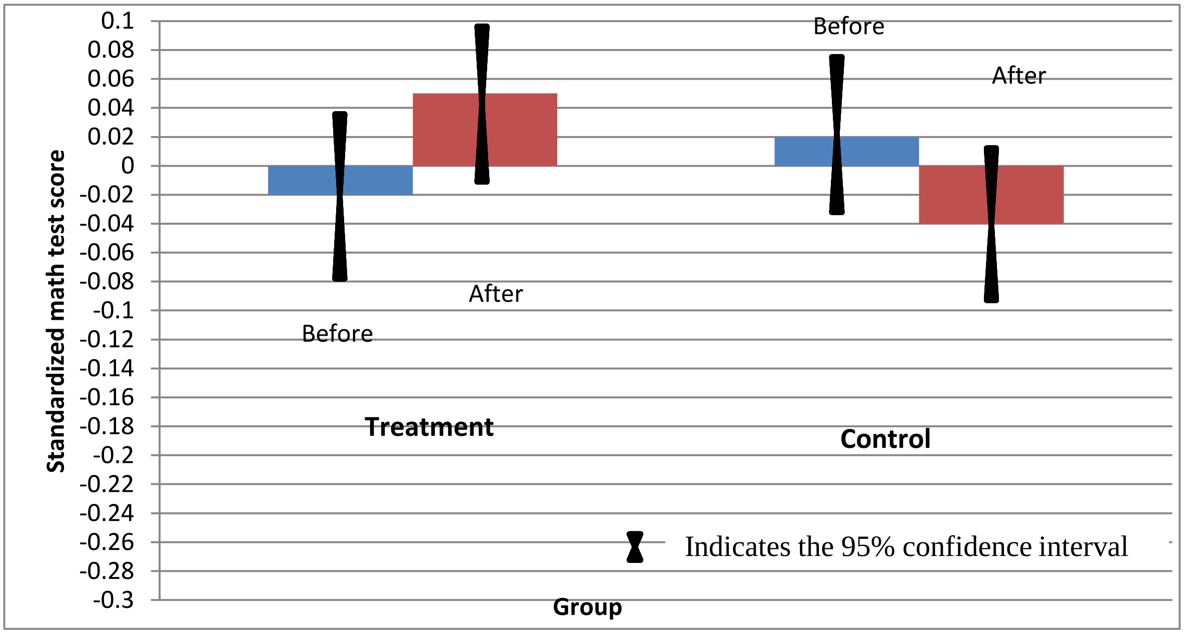
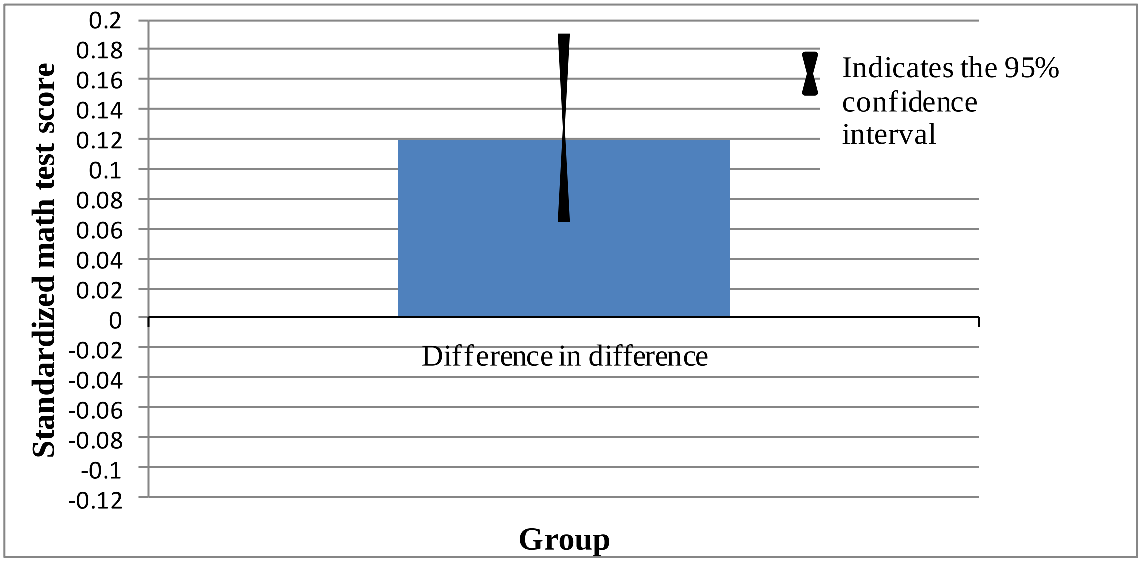


Figure 1: Experiment Profile

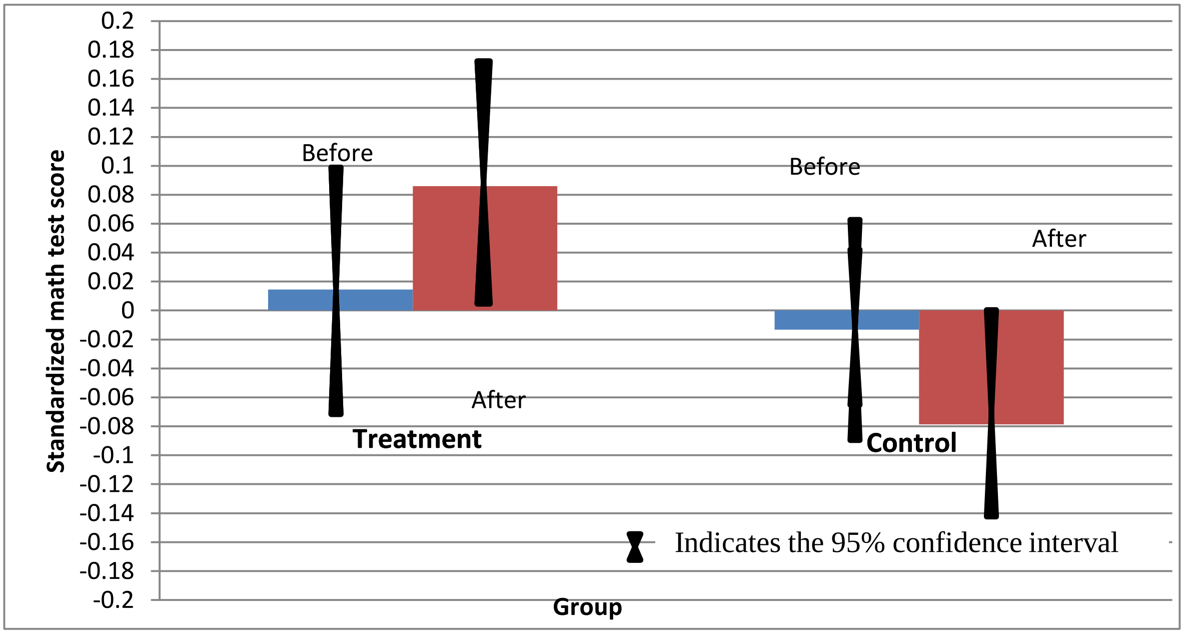


Panel A. Standardized math test scores before and after CAL: the treatment and control groups in both the third and the fifth grades.

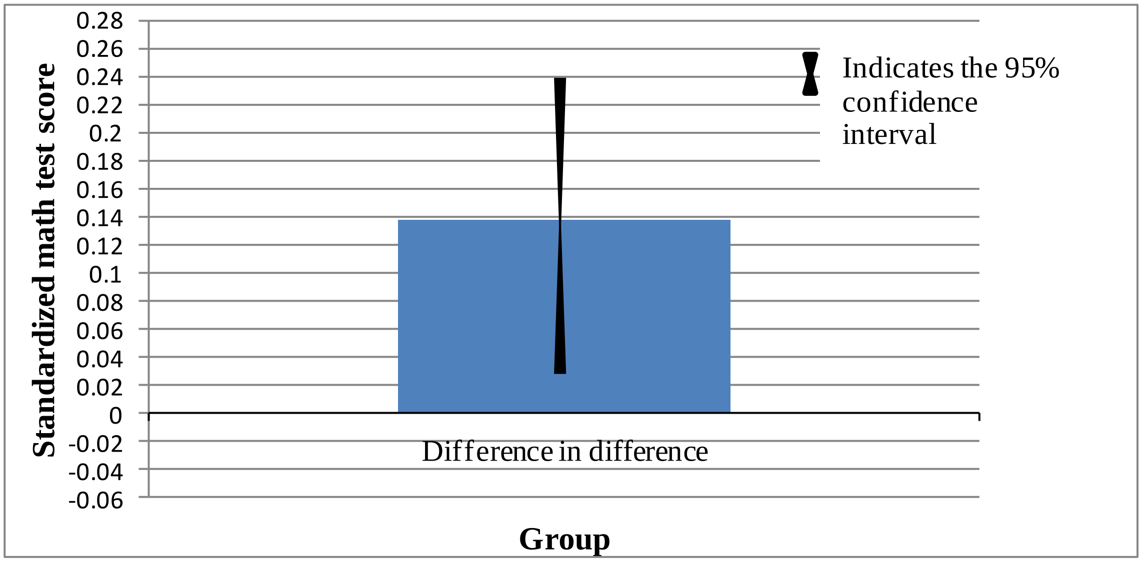


Panel B. Difference in difference in the standardized math test scores before and after the CAL Program between the treatment and control groups in both the third and the fifth grades

Figure 2. Change in the standardized math test scores before and after the CAL program

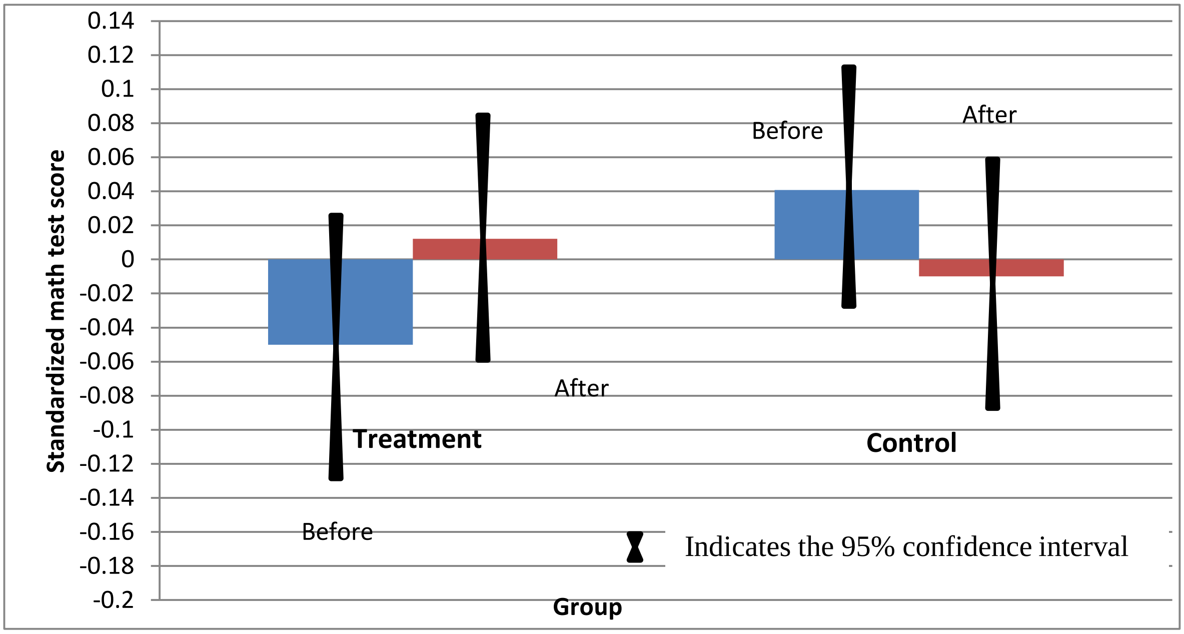


Panel C. Standardized math test scores before and after CAL: the treatment and control groups in the third grade.

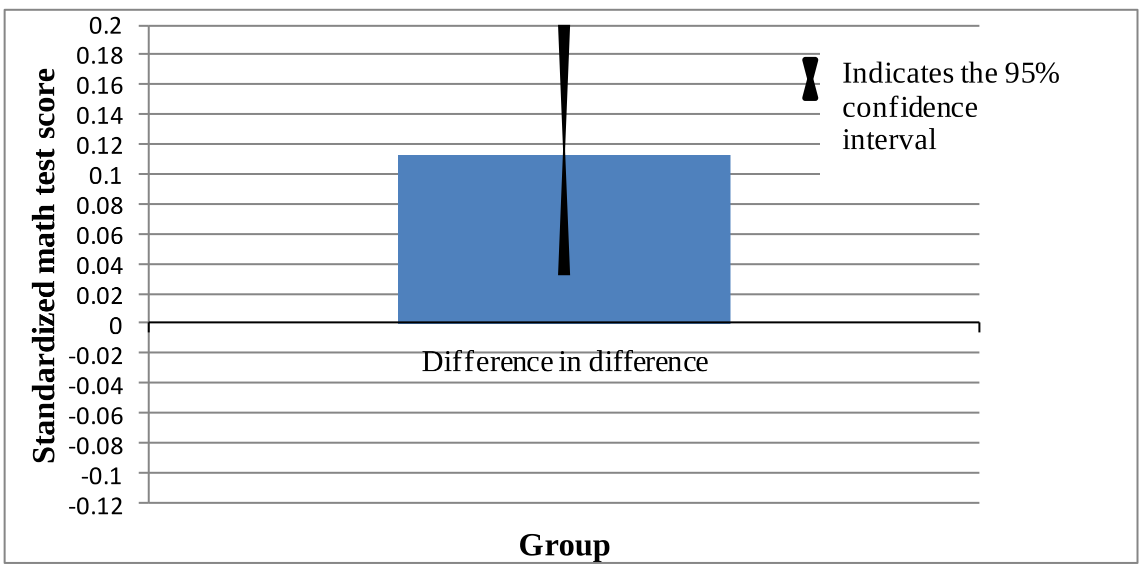


Panel D. Difference in difference in the standardized math test scores before and after the CAL Program between the treatment and control groups in the third grade

Figure 2. Change in the standardized math test scores before and after the CAL program



Panel E. Standardized math test scores before and after CAL: the treatment and control groups in the fifth grade.



Panel F. Difference in difference in the standardized math test scores before and after the CAL Program between the treatment and control groups in the fifth grade

Figure 2. Change in the standardized math test scores before and after the CAL program

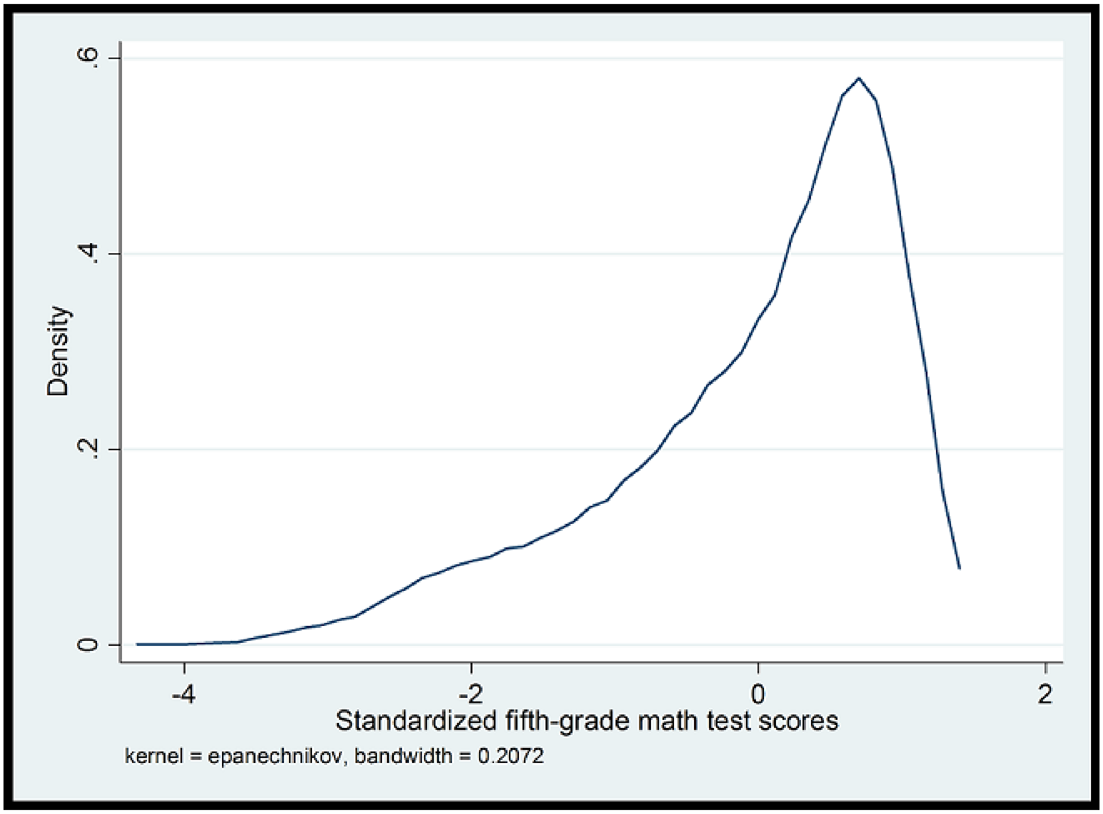


Figure 3. The distribution of the standardized fifth-grade math test scores of the final evaluation test

Table 1. Comparisons of the student characteristics between the attrited students and those remaining in the sample and the characteristics of attrited students between the treatment and control groups

		Difference between attrited and non-attrited students[e]			Difference between the treatment and control groups within attrited students[f]
		All	Third Grade	Fifth Grade	All Attrited students
		(1)	(2)	(3)	(4)
(1)	Baseline Chinese score[a]	0.01	0.02***	0.00	0.01
	(units of standard deviation)	[0.00]	[0.01]	[0.01]	[0.06]
(2)	Baseline math score[b]	0.00	-0.01	0.00	0.01
	(units of standard deviation)	[0.00]	[0.01]	[0.01]	[0.06]
(3)	Female (0=no; 1=yes)	-0.01	-0.02	-0.01	0.06
		[0.01]	[0.01]	[0.01]	[0.09]
(4)	Age (years)	-0.08***	-0.05***	-0.10***	0.02
		[0.01]	[0.01]	[0.02]	[0.04]
(5)	Only child (0=no; 1=yes)	-0.02*	-0.02*	-0.01	-0.02
		[0.01]	[0.01]	[0.01]	[0.09]
(6)	Father has at least high	0.00	0.00	0.00	0.16
	school degree (0=no; 1=yes)	[0.01]	[0.02]	[0.02]	[0.18]
(7)	Mother has at least high	0.00	0.00	0.00	-0.12
	school degree (0=no; 1=yes)	[0.01]	[0.02]	[0.02]	[0.27]
(8)	Family off-farm	0.01	0.00	0.02	0.1
	(0=no; 1=yes)	[0.01]	[0.02]	[0.02]	[0.22]
(9)	Poverty subsidy	-0.01	-0.01	-0.01	0.01
	(0=no; 1=yes)	[0.01]	[0.01]	[0.01]	[0.13]
(10)	Both parents at home	-0.01	-0.01	0.00	-0.04
	(0=no; 1=yes)	[0.01]	[0.01]	[0.01]	[0.10]
(11)	Ever used a computer (1=yes; 0=no)	-0.02	0.00	-0.04	0.02
		[0.01]	[0.01]	[0.02]	[0.16]
(12)	Access to other modern technologies[c]	-0.01	0.00	-0.01	-0.17
		[0.02]	[0.02]	[0.03]	[0.22]
(13)	Grade	Y			Y
(14)	Observations	2726	1155	1571	113

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in brackets clustered at the class level

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

[c] Access to other modern technologies is the mean value of a set of 0/1 dummy variables including whether the student has used cell phone, internet, game console, CAL software, and videos for learning assistance.

[d] The sample includes both the sample observations(non-attrition) and the attrition observations.

[e] The differences between attrited and non-attrited students in columns (1) and (3) are calcuated by regressiong the indicator of attrition on the row characteristics for each grade, or controlling for grade dummies for all students. Colunmn (2) indicates that there was no attritions among the third-grade students.

[f] The sample is limited to the attrited observations.

The differences between the treatment and control students among the attrited students in column (4) are calcuated by regressing the indicator of the treatment dummy on the row characteristics, restricting the sample to attrited students for students in each grade, or controlling for grade dummies for all students.

Table 2. Difference in characterisitcs between the students in the treatment and control groups

Dependent variable: whether the student received CAL treatment (0=no; 1=yes)				
		All	Third Grade	Fifth Grade
		(1)	(2)	(3)
(1)	Baseline Chinese score[a] (units of standard deviation)	-0.01 [0.02]	-0.04 [0.03]	0.01 [0.03]
(2)	Baseline math score[b] (units of standard deviation)	-0.01 [0.02]	0.03 [0.03]	-0.03 [0.03]
(3)	Female (0=no; 1=yes)	0.02 [0.02]	0.03 [0.04]	0.02 [0.03]
(4)	Age (years)	0.03 [0.02]	0.01 [0.03]	0.03 [0.03]
(5)	Only child (0=no; 1=yes)	-0.01 [0.03]	0.04 [0.05]	-0.04 [0.03]
(6)	Father has at least high school degree (0=no; 1=yes)	0.02 [0.04]	0.00 [0.05]	0.05 [0.05]
(7)	Mother has at least high school degree (0=no; 1=yes)	0.07 [0.04]	0.09 [0.05]	0.01 [0.06]
(8)	Family off-farm (0=no; 1=yes)	-0.04 [0.04]	0.00 [0.06]	-0.07 [0.05]
(9)	Poverty subsidy (0=no; 1=yes)	0.02 [0.04]	0.04 [0.06]	0.00 [0.05]
(10)	Both parents at home (0=no; 1=yes)	0.06** [0.02]	0.06 [0.04]	0.05* [0.03]
(11)	Ever used a computer (1=yes; 0=no)	0.07 [0.08]	0.00 [0.10]	0.15 [0.12]
(12)	Access to other modern technologies[c]	0.03 [0.08]	0.13 [0.11]	-0.05 [0.12]
(13)	Grade	Y		
(14)	Observations	2613	1124	1489
(15)	R-squared	0.03	0.03	0.05

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in brackets clustered at the class level

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

[c] Access to other modern technologies is the mean value of a set of 0/1 dummy variables including whether the student has used cell phone, internet, game console, CAL software, and videos for learning assistance.

The differences between the treatment and control students in columns (1)-(3) are calcuated by regressions of the indicator of the treatment (0=control; 1=treatment) on the row characteristics for students in each grade or controlling for grade dummies for all students.

Table 3. Ordinary Least Squares estimators of the impacts of the CAL program on student academic outcomes

		Dependent variable: standardized post-CAL math test score - standardized baseline math test score							
		All	Third Grade	Fifth Grade: all	Fifth grade: up to the 70th percentile in post-CAL math score	All	Third Grade	Fifth Grade_all	Fifth grade: up to the 70th percentile in post-CAL math score
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	Treatment	0.12**	0.14*	0.11	0.12*	0.12**	0.18**	0.07	0.11*
		[0.06]	[0.08]	[0.08]	[0.07]	[0.05]	[0.08]	[0.07]	[0.07]
(2)	Baseline Chinese score[a]					0.22***	0.22***	0.22***	0.20***
	(units of standard deviation)					[0.02]	[0.04]	[0.03]	[0.03]
(3)	Baseline math score[b]				-0.48**	-0.53***	-0.56***	-0.51***	0.65
	(units of standard deviation)				[0.04]	[0.02]	[0.03]	[0.03]	[0.43]
(4)	Female (0=no; 1=yes)					-0.12***	-0.13**	-0.10**	-0.60***
						[0.03]	[0.05]	[0.04]	[0.04]
(5)	Age (years)					-0.06***	-0.02	-0.09***	-0.01
						[0.02]	[0.03]	[0.03]	[0.06]
(6)	Only child (0=no; 1=yes)					-0.03	0.00	-0.05	-0.10***
						[0.04]	[0.05]	[0.05]	[0.03]
(7)	Father has at least high					-0.07	-0.01	-0.12	-0.03
	school degree (0=no; 1=yes)					[0.05]	[0.07]	[0.08]	[0.05]
(8)	Mother has at least high					-0.09	-0.08	-0.14*	-0.1
	school degree (0=no; 1=yes)					[0.06]	[0.08]	[0.08]	[0.10]
(9)	Family off-farm					0.06	-0.06	0.14*	-0.14
	(0=no; 1=yes)					[0.06]	[0.09]	[0.08]	[0.15]
(10)	Poverty subsidy					0.03	0.02	0.04	0.08
	(0=no; 1=yes)					[0.04]	[0.06]	[0.05]	[0.10]
(11)	Both parents at home					0.06**	0.01	0.10***	0.02
	(0=no; 1=yes)					[0.03]	[0.05]	[0.04]	[0.05]
(12)	Ever used a computer (1=yes; 0=no)					-0.08	-0.16*	0.09	0.07
						[0.07]	[0.08]	[0.11]	[0.05]
(13)	Access to other modern technologies[c]					0.06	0.00	0.09	0.06
						[0.08]	[0.12]	[0.09]	[0.11]
(14)	Grade	Y				Y			
(15)	County					Y	Y	Y	Y
(16)	Observations	2613	1124	1489	909	2613	1124	1489	909
(17)	R-squared	0.01	0.02	0.01	0.01	0.26	0.29	0.25	0.31

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in brackets clustered at the class level

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

[c] Access to other modern technologies is the mean value of a set of 0/1 dummy variables including whether the student has used cell phone, internet, game console, CAL software, and videos for learning assistance.

Each column reports the results of one regression of the change in student standardized math test scores over the program period on the corresponding variables in rows (1) to (15).

Table 4. The Ordinary Least Squares estimators of the heterogeneous program effect on students with different characteristics

Dependent variable: standardized post-CAL math test score - standardized baseline math test score

	(1)	(2)	(3)	(4)	(5)
	Only child (0=no; 1=yes)	Mother has no high school degree (0=no; 1=yes)	Father has no high school degree (0=no; 1=yes)	Poverty subsidy (0=no; 1=yes)	Baseline Chinese score
Third grade					
(1) Treatment	0.25***	0.15*	0.18**	0.11	0.19**
	[0.08]	[0.08]	[0.08]	[0.09]	[0.08]
(2) Treatment interacted with the corresponding column variable	-0.28***	-0.30*	0.04	0.18*	0.1
	[0.10]	[0.16]	[0.13]	[0.10]	[0.06]
(3) # of observations	1124	1124	1124	1124	1124
Fifth grade					
(4) Treatment	0.07	-0.27	-0.15	0.16	0.07
	[0.07]	[0.07]	[0.07]	[0.10]	[0.07]
(5) Treatment interacted with the corresponding column variable	0.00	0.36**	0.24*	-0.06	-0.11*
	[0.10]	[0.15]	[0.13]	[0.09]	[0.06]
(6) # of observations	1489	1489	1489	1489	1489

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in brackets clustered at the class level.

For each grade, each column reports the results of one regression of the change in student standardized math test scores over the program period on the treatment dummy,the interaction of the treatment dummy and the corresponding variable in columns (1) to (5), controlling for the county dummies, the grade dummies, female, age, only child, father has no high school degree, mother has no high school degree, family off-farm, both parents at home, poverty subsidy, ever used a computer and access to other modern technology.

We also examined heterogeneous program effect across the other student characterisitics included the model, and none of them are significant. Therefore we did not include the relevant results in the table for simplicity

Reference

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

Banerjee, A., Cole, S., Duflo, E., and Linden, L. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics*. 122 (3): 1235-1264.

Barrow, L., Markman, L., and Rouse, C. 2008. "Technology's Edge: The Educational Benefits of Computer-Aided Instruction." *National Bureau of Economic Research working paper No. 14240*.

CNBS [China National Bureau of Statistics]. 2011. *China National Statistical Yearbook, 2011*. China State Statistical Press: Beijing, China.

Dynarski, M. 2007. "Effectiveness of Reading and Mathematics Software Products: Findings from the 1st Student Cohort." *Mathematica Report.*

Fuchs, T., and Woosmann, L. 2004. "Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at Schools." *Brussels Economic Review*, 47: 359-585.

Glewwe, P., and Kremer, M. 2006. "Schools, Teachers, and Education Outcomes in Developing Countries." *Handbook on the Economics of Education.* (New York, NY: Elsevier), 2: 945-1017.

Glewwe, P., and Moulin, S. 2002. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." BREAD Working Paper, Cambridge, MA,

Glewwe, P., and Zitzewitz, E. 2004. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics,* 74: 251-268.

Goolsbee, A., and Guryan, J. 2006. “The Impact of Internet Subsidies in Public Schools.” *The Review of Economics and Statistics*, MIT Press, 88 (2): 336-347,

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

———.1995. “Interpreting Recent Research on Schooling in Developing Countries.” *World Bank Research Observer*, 10: 227-246.

He, F., Linden, L., and MacLeod, M. 2008. “How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program.” *Working Paper.* Columbia University Department of Economics.

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

Inal, Y., and K. Cagiltay. 2007. “Flow Experiences of Children in an Interactive Social Game Environment.” *British Journal of Educational Technology*, 38 (3): 455-464.

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

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

Krueger, A., and Rouse, C. 2004. “Putting Computerized Instruction to the Test: A Randomized Evaluation of a 'Scientifically-based' Reading Program.” *Economics of Education Review*, 23: 323-338.

Lai, F., Liu, C., Luo, R., Zhang, L., Ma, X., Bai, Y., Sharbono, B and Rozelle, S. 2012. "Private Migrant Schools or Rural/Urban Public Schools: Where Should China Educate Its Migrant Children?" REAP working paper.

Lai, F., Luo, R., Zhang, L., Huang, X. and Rozelle, S. 2012. "Does Computer-Assisted Learning Improve Learning Outcomes?" Evidence from a Randomized Experiment in Migrant Schools in Beijing. REAP working paper.

Li, H. P., W. Liu, N. Ma, and J. Zhang. 2005. "Does Education Pay in Urban China? Estimating Returns to Education Using Twins." Discussion Papers, Chinese University of Hong Kong, Department of Economics. Retrieved Feburary 8, 2010, from http://www.gsm.pku.edu.cn/store/object/2006824194540611.pdf.

Li, S., Sato, H. and Sicular T., eds. 2011. "Rising Inequality in China: Challenge to a Harmonious Society." University of Western Ontario Department of Economics working paper.

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

Liu, C., Zhang, L., Luo, R., Rozelle, S., and Loyalka, P. 2010. "The Effect of Primary School Mergers on Academic Performance of Students in Rural China." *International Journal of Educational Development* 30 (6), 570-585.

Liu, C., L. Zhang, R. Luo, and S. Rozelle. 2008. "Early Commitment on Financial Aid and College Decision Making of Poor Students: Evidence from a Randomized Evaluation in Rural China." Retrieved September 28, 2009, from http://stanfordreap.win.aplus.net/doc/pdf/207_Liu%20et%20al_ECFA%20and%20college%20decision%20of%20poor%20students%20in%20rural%20China.pdf.

Luo, R., Shi Y., Zhang L., Liu C., Rozelle S., Sharbono B. 2009. "Malnutrition in China's Rural Boarding Schools: the Case of Primary Schools in Shaanxi Province." *Asia Pacific Journal of Education* 29 (4), 481-501.

Luo, R., Zhang, L., Liu, C., Rozelle, S., Sharbono, B., Adams, J. 2010. "Behind Before They Start: The Challenge of Early Childhood Education in Rural China." REAP working paper No. 209, CCAP, Beijing, China.

Ma, Y. 2009. "The Rebound of Dropout Rate after School Merging Program in Rural China." Retrieved January 10, 2010, from: http://news.xinhuanet.com/edu/2009-11/12/content_12440098.htm.

Metcalfe, J., & Shimamura, A. P. 1994. "Metacognition: knowing about knowing." Cambridge, MA: MIT Press

MOE (Ministry of Education). 1992. *Educational Statistics Yearbook of China*. Beijing: People's education press.

———. 1999. *Educational statistics yearbook of China*. Beijing: People's Education Press.

———. 2006. *Educational statistics yearbook of China*. Beijing: People's Education Press.

MOE and NBS (Ministry of Education and National Bureau of Statistics). 2004. *China Educational Finance Statistical Yearbook*. Beijing: China Statistics Press.

Pang, L. 2006. "Primary and High School Consolidation in Rural China: Issues, Reasons, and Solutions." *Education Development Research,* 2, 1-6.

Planty, M., Hussar, W., Snyder, T., Provasnik, S., Kena, G., Dinkes, R., KewalRamani, A., and Kemp, J. 2008. *The Condition of Education 2008* (NCES 2008-031). National Center for

Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

President's Committee of Advisors on Science and Technology. 1997. "Panel on Educational Technology." *Report to the President on the Use of Technology to Strengthen K-12 Education in the United States*, Washington DC: Office of the President.

Shaefer, S. and Warren, J. 2004. "Teaching Computer Game Design and Construction." *Computer-Aided Design*, 36: 1501-1510.

Shi, R. 2004. "Experience Drawn from Research on School Consolidation in Foreign Countries." *Comparative Education Research*, 12.

Shi, Y., Zhang, L., Liu C., Li, H., Rozelle, S., Sharbono, B. 2009. "Community Service, Social Responsibility and Educational Performance in Rural China's Middle Schools: Evidence from a Case Study of Northwest China." REAP Working Paper No 208, CCAP, Beijing, China.

Wang, X., B. Fleisher, H. Li, S. Li. 2007. "Access to Higher Education and Inequality: The Chinese Experiment." IZA Discussion Paper No. 4030. Available at SSRN: http://ssrn.com/abstract=996402.

Wang, X., C. Liu, L. Zhang, R. Luo, Y. Shi, S. Rozelle, and B. Sharbono. 2009. "What is Keeping the Poor out of College? Enrollment Rates, Educational Barriers and College Matriculation in China." Retrieved January 28, 2010, from http://reap.stanford.edu/publications/what_is_keeping_the_poor_out_of_college_enrollment_rates_educational_barriers_and_college_matriculation_in_china/

World Bank. 2004. *World Development Report 2004: Making Services Work for the Poor.* New York, NY: Oxford University Press.

World Bank. 2001. *China: Challenges of Secondary Education*. Second Education Series 22856. Washington D.C., World Bank.

Xie, M. 2008. "The Increasing Burden on Rural Family after School Merging Program in Rural China." Retrieved January 10, 2010, from http://www.gd.xinhuanet.com/newscenter/2008-09/03/content_14298403.htm.

Yue, A., Shi, Y., Chang, F., Yang, C., Wang, H.,Yi, H., Luo, R., Liu, C., Zhang, L., Chu, J. and Rozelle, S. 2012. "Dormitory Management and Boarding Students in China's Rural Elementary Schools." REAP working paper.

Zhuo, M. 2006. "Primary School Consolidation in Rural China: Advantages, Disadvantages and Solutions." Education Forum, The Education Department of Fujian Province. Retrieved January 10, 2010, from: http://218.85.72.34/newsInfo.aspx?pkId=10103.