Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai, China

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,* Qinghe Qu and Xiao Hu
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 Acer Computers and ADOC 2.0 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, 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

Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Rural Minority Areas in Qinghai

Abstract

The education of disadvantaged populations has been a long-standing challenge to the education system in both developed and developing countries. This paper uses a cluster randomized field experiment in 57 schools (26 schools were part of the CAL program; 31 control schools were not) to explore the effects of the CAL program on student academic and non-cognitive outcomes for students in public schools in minority rural areas in China. Our results show that a remedial, game-based CAL program that focused on teaching Standard Chinese (Mandarin) held out of regular school hours improved the standardized Mandarin scores of the students in the treatment schools by 0.14-0.20 standard deviations more than those in the control schools. Moreover, CAL also had significant spillover effects on student standardized math test scores. Still further, our results also show insignificant positive effects of CAL intervention on student non-academic outcomes of interest in studying and metacognition, and significant positive effect on student self-efficacy of Mandarin studying. In general, low-performing students benefited more from the program.

Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Rural Minority Areas in Qinghai

The education of poor and disadvantaged populations, including those from minority subgroups, 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, 2001). In China, although children in both cities and rural areas have nearly universal rates of participation between grades one to nine, there is still a large achievement gap between students from cities/richer coastal regions and ethnic minority students from poor rural areas (Hannum, 1999b; Cao and Nee, 2000). The literature clearly demonstrates differences in the academic performance between Han students and students from different ethnic groups—although the size of the gap differs from ethnic group to ethnic group (Yang et al., 2013).

One possible reason for the underperformance of rural minority students is poor Mandarin language skills. Mandarin (or standard Chinese) is an important skill for individuals when they take competitive exams and for participation in labor markets. In addition, most school textbooks for all subjects are in Mandarin.

Unfortunately, the Mandarin language skills of most rural minority students are substandard. Rural minority students often speak their own languages or dialects at home, and, at most, only use Mandarin at school. Most teachers in rural minority schools are also minorities themselves and often have only limited competency in Mandarin.

Considering the language deficiencies of both students and teachers, a computer-assisted learning (or CAL) program teaching Mandarin language might be useful in helping poor minority students to begin to meet one of their most urgent academic needs: improving their Mandarin language skills. Given the fact that almost all text books are in Mandarin, such a program might also be able to raise overall academic performance. Many existing studies have shown that CAL can be particularly effective in meeting the needs of students that live in environments in which schools and teaching are poor in quality and the home learning environment is inadequate (e.g., Barrow et al., 2008; Benerjee et al. 2007; He et al., 2008; Lai et al., 2012). For these students CAL may be able to act as a substitute for teachers. CAL also might be able to provide the help that parents who are illiterate (or in our case, parents whose Mandarin language skills are too poor to help their children with Mandarin language school work) or those who are too busy. Finally, CAL may also be a way to provide remedial tutoring services when commercialized services are either not available or not affordable.

The overall goal of this paper is to explore the effects of a Mandarin language CAL program on student academic outcomes for minority students in poor rural areas in China. To achieve this goal, we present the results from a randomized field experiment of a CAL Mandarin remedial tutoring program in 57 schools in poor minority rural areas in Qinghai Province, one of the largest and poorest provinces that is densely populated by ethnic minority groups in northwestern China. The program was implemented for one semester (or 13 weeks) in the spring of 2011 and involved all 1889 third-grade students in the 57 sample schools, mostly aged nine to eleven. Around 86% of the students are ethnic minorities, mostly from impoverished rural families.

In launching a study on the effect of a CAL program to address this problem, we contribute to the existing literature in several ways. First, by implementing a randomized controlled trial, our study provides an estimate of the impact of the CAL program on the educational performance of one of China's most disadvantaged group—rural ethnic minority children. Second, since a share of the students are non-Mandarin speakers, and others have poor language backgrounds (and little support from their families), we also will examine the spillovers from learning Mandarin onto the performance in other subjects (in our case—math) in an educational context that teaches in nonnative languages (which in this case in Mandarin). Our study is one of the first studies to look at not only the effect of CAL on language learning but also its spillovers on other subjects due to improved language competency. To our knowledge, the only study in the literature exploring this issue is He et al. (2008). However, in their study, improved math performance via the CAL program in English was due to the fact that teachers implemented the interventions more efficiently, sparing more time for other subjects. This is unlikely to be the case in our study, as in our case, the supervisor teachers did not have the authority to substitute the time of CAL sessions for other subjects. As a result, we believe the improvement in math performance was more likely to be the result of improved Mandarin competency from the CAL program. Third, according to the data collected during the study, we are able to examine the CAL impacts not only on student test scores, but also nonacademic outcomes (e.g., metagcognition in Mandarin and self-efficacy in Mandarin studying). Such noncognitive subjects are also important to the development of students but are seldom measured. Finally, by evaluating the effect of CAL program at the end of one semester (instead of evaluating CAL after an entire school year—as is the case in most existing CAL studies), our study shows that CAL provides significant beneficial effects within months after the implementation. Overall, our work demonstrates that CAL has the potential to provide a relatively quick and effective remedy for the severe underperformance of rural minority children in China (and possibly disadvantaged children in other parts of the world) in language and other subjects.

Sampling, Data and Methods

Sampling and the Process of Randomization

We conducted a clustered RCT of CAL in rural schools during the spring semester of 2011. A total of 1889 students in 57 elementary schools in poor minority areas in China's Qinghai Province participated in our study. We adopted a three-step process to arrive at this sample.

First, we chose our sample counties in a way that allows us to focus our study on the impact of the CAL program on improving the Mandarin language skills of students in rural minority areas. In order to do so, we first chose Qinghai Province, a poor, rural province. More than 60 percent of the population lives in rural areas. Qinghai also has a large minority population. Ethnic minorities make up around 46 percent of the province's population. The share of the rural population that is minority is even higher. We then chose three counties in Haidong Prefecture, Huzhu, Hualong and Xunhua. Each of the three counties in our sample has the characteristics that fit our research purpose: Each county is poor, rural and is a minority autonomous county. Second, after choosing the counties, we sampled 57 schools. To do so, we obtained a comprehensive list of all schools. We called each school to confirm the number of minority students in that school and excluded schools with less than 20% minority students (since minorities were the main focus of our study). For implementation convenience we also excluded schools that were either too big (more than 150 third-grade students) or too small (less than 20 third-grade students). There were 70 schools that met all the two criteria above. Because of our limited computer provision, we then randomly chose 57 schools for our final sample.[1]

Finally, because the Mandarin CAL software we used was only available at the third-grade level, we only examined third-grade students in our sample schools. All of the third-grade students in the 57 sample schools were included in the study. In total, there were 1889 third-grade students included in the sample (Figure 1), among whom 86% belonged to a minority ethnic group.

After choosing the 57 schools for our sample, we randomly chose (without stratification) 26 schools from these 57 schools to receive the CAL intervention. The 764 third-grade students in the 26 treatment schools constitute the treatment group (Figure 1). The 1125 third-grade students in the other 31 schools served as the control group.

Although at the time of the baseline survey the main sample included a total of 57 schools and 1889 third-grade students, there was some attrition by the end of the study. For various reasons (mainly because of school transfers and absences due to illness or injuries) by

---

[1] Our sample selection proceeded as follows. According to our original design, we needed 30 treatment and 30 control schools. This was done on the basis of the power calculations. We then proceeded to randomly choose 60 schools from a list that were available in the three sample counties. The list contained the school name; location; and the number of students in grade 3. From this information, we randomized the group into 30 treatment and 30 control schools. Before starting, however, it was discovered that 3 of the schools had been closed. Of the three schools, two were supposed to be assigned to the treatment school group; one was supposed to be assigned to the control school group. This means that at this time, the number of schools available for the experiment was 57 (28 in the treatment group and 29 in the control group). It was at this time that we decided how many computers were needed for the treatment schools. This number was roughly calculated as half of the number of grade 3 students. When the computers were delivered to the project site, it was discovered that there were insufficient number of computers. We thus randomly excluded two of the treatment schools from the treatment group in order to equip the remaining treatment schools with enough number of computers. We used these two schools as control schools in the analysis. This action served to reduce the number of treatment schools to 26 and increased the number of control schools to 31. We calculate the power with our sample of 57 schools. The power to detect a treatment effect of size around 0.2 and significant at the five percent level is around 0.8 with intraclass correlation coefficient (ICC) equal to 0.09 (which is the actual ICC of our study), R-square equal to 0.5 and total number of clusters equal to 57 (with 30 observations in each cluster on average).

the time of the evaluation survey we were only be able to follow up with 1717 third-grade students in the 57 sample schools (Figure 1, final row). Among the 764 treatment students, 680 remained in our sample at final evaluation. Among the 1125 control students, 1037 remained in our sample of final evaluation.

To understand the nature of those that attrited and assess if the attrition affected the nature of the sample (or the validity of the randomization), we regressed each variable from a set of student characteristics on a constant, a dummy for attrition, a dummy for treatment and an interaction of the attrition and treatment dummies. According to our findings (Table 1, row 2), we see that those that attrited have lower Mandarin scores (column 1), a higher share of mothers that were illiterate (column 6), were less likely to live with both parents (column 8) and were less likely to be Tibetan and more likely to be Hui (columns 12 and 13). However, as seen from the fact that none of the coefficients of the interaction variables were statistically significant (row 4), there is no evidence that the attrition affected the validity of our randomization.

After excluding those that attrited, we used a set of student and school characteristics to check the validity of the random assignment for our final analytic sample by regressing each variable on the treatment dummy variable. Specifically, the individual characteristics included measures of whether the students were *male*, their ethnic group (*Han, Hui, Salar, Tu and Tibetan*), county, whether they were the *only child* in the family, whether their father was illiterate (*father illiterate*), whether their mother was illiterate (*mother illiterate*), and whether their parents were still farmers and worked on a farm (*family off-farm*). The school-level characteristics included the number of students, the teacher-student ratio, the proportion of female teachers, an index of the availability of school facilities, and the proportion of students receiving poverty subsidies. We found the differences between the treatment and control groups were statistically insignificant (Table 2, columns 5 to 6). The F-tests of joint significance of the individual characteristics and school characteristics were also insignificant with p-values bigger than 0.1.[2]

Experiment Arms/Interventions

Our experiment focused fully on one treatment group (the third-grade students of the 26 treatment schools) and one control group (the third-grade students of the 31 control schools). The CAL Intervention was the only intervention going on in these schools at the time.

CAL Intervention Group (the grade 3 students in the 26 treatment schools)

The intervention involved computer-assisted Mandarin remedial tutoring sessions, which were designed to complement the regular in-class Mandarin language 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 during lunch break or after school. The sessions were mandatory. The content of each session emphasized basic competencies in the uniform national Mandarin language curriculum.

During each session, two students formed a team to share one computer and played Mandarin games designed to help students review and practice the basic Mandarin material that was being taught in their regular school Mandarin classes during the same week. In a typical session, the students first watched an animated video that reviewed the material that they learned during that particular week during their regular Mandarin class sessions. The students were able to play Mandarin games to practice the knowledge and skills (including vocabulary, reading

[2] We also used the same set of student and school characteristics to check the validity of the random assignment for the whole sample (i.e., the sample that included both the students in our final analytic sample and those who attrited), and found that the differences between the treatment and control groups were also statistically insignificant for all student and school characteristics (results not reported for brevity).

comprehension, and grammar) introduced in class. If a student had a Mandarin-related question, he/she was encouraged to discuss with his/her teammate. 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. According to our in-class observations, the sessions, in fact, were so intense that the attention of the students was fully on the computer and there was little communications among the groups or between any of the groups and the teacher-supervisor.

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

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. 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 26 treatment schools were required to attend a two-day mandatory training that was held at a central site.

To further ensure that the teacher-supervisors (and the students under their supervision) strictly followed the protocol, we recruited volunteers from universities in Haidong Prefecture and directed them to visit the treatment schools during the implementation of CAL.[3] None of the visits were announced to the schools in advance. During the visits, the volunteers were instructed to attend the CAL sessions and observe whether the protocol was being strictly implemented. 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, as such interactions might confound the program effect. The research team also provided technical support and free computer repairs and maintenance for the entire semester.

CAL Control Group (the third-grade students in the 26 control schools)

Third-grade students in the 31 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 group took their regular Mandarin classes at school as usual.

---

[3] It is true that we (the outside research team) visited the schools and gave baseline surveys and tests. However, these were done in both treatment and control schools. Besides this, there was almost no interaction between the research team and the students. In one or two sessions students from local colleges visited the schools. However, they were there for only short periods of time and had nearly zero interaction with the students and/or teachers. In these visits it was unclear that they were really concerned fully with the CAL program. Hence, it is unlikely that any significant part of the observed treatment effect was from a Hawthorne effect. We took many precautions to ensure that any scope for Hawthorne effects was minimal.

Data Collection

The research group conducted two rounds of surveys in the 57 control and treatment schools. The first-round survey was a baseline survey conducted with all third graders in the 57 schools in late December 2010 right before the closing of the fall 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 survey, the enumeration team visited each school and conducted a two-part survey. In the first part students were given a standardized Mandarin test and a standardized math test.[4] The Mandarin test included 40-45 questions (tests in different rounds included slightly different numbers of questions). The math test included 25 questions. Students were required to finish tests in each subject in 25 minutes. All students took the Mandarin test before taking the math test. Our enumeration team strictly enforced time limits and proctored the examinations. We use the scores of the students on the Mandarin and math tests as our measures of student academic performance.

In the second part of the survey, enumerators collected data on the characteristics of students and their families. From this part of the survey we are able to create demographic and socioeconomic variables. The dataset includes measures of whether the students were *male*, their ethnic group (*Han, Hui, Salar, Tu and Tibetan*), county, whether they were the *only child* in the family, whether their father was illiterate (*father illiterate*), whether their mother was illiterate (*mother illiterate*), whether their parents were still farmers and worked on the family farm or if

[4] Again, some experts from the Center for Examination of Beijing helped us pick questions for the tests from official examination books and exercise books.

their parents worked off-farm (*family off-farm*) and whether students had *ever used a computer* or had ever had *access to other modern technologies*. To create indicators of parental care, during the survey the students were also asked whether their parents had migrated to some other location outside of his/her home town or whether their parents stayed at home for most of the time during the semester (*live with both parents*).

In the second part of the survey students were also asked to answer questions that could help us measure their non-cognitive traits. To create indicators for students' attitudes 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 Mandarin were created from the responses of students to a seventeen-item psychological scale measuring metacognition[5] and a seven-item psychological scale for the self-efficacy of studying Mandarin.[6]

Statistical Methods

We used ordinary least squares (OLS) regression analysis (both with and without control variables) to estimate how the academic and non-cognitive outcomes changed in the treatment group relative to the control group. Our basic OLS analysis regressed the endline outcome

---

[5] 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 & Shimamura, 1994).

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

variables (i.e. post-program outcome value) on the value of outcome variables at baseline and a dummy variable of the treatment (CAL intervention) status. We then included in the basic model a set of control variables to control for some systematic differences between the treatment and control groups and improve precision. In all regressions, we accounted for the clustered nature of our sample by constructing Huber-White standard errors corrected for school-level clustering. The models are presented in order of increasing comprehensiveness.

First, the basic model is:

$$y_{is} = \alpha + \beta \cdot treatment_s + \theta \cdot y_{0is} + \varepsilon_{is} \qquad (1)$$

where y_{is} is the endline outcome variable for child *i* in school *s*, y_{0is} measures the outcome variable of the same child at the baseline, $treatment_s$ is a dummy variable for a third-grade student attending a treatment school (equal to one for students in the treatment group and zero otherwise), and ε_{is} is a random disturbance term clustered at the school level.

We used several variables to measure the student academic and non-cognitive outcomes (y_{is}). The primary outcome variable of our analysis is the student academic outcome, measured by the student standardized Mandarin test score. We also included the student standardized math 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 math, the other major subject in China's elementary schools besides Mandarin. Importantly, besides variables measuring academic outcomes, we also included three non-cognitive outcome variables, namely, *like school*, metacognition, and self-efficacy in Mandarin studying.

By construction, the coefficient of the dummy variable $treatment_s$, β, measures the unconditional difference in the change in outcome ($y_{is} - y_{0is}$) 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.

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

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

where all the variables and parameters are the same as those in equation (1), except that we added a set of control variables. Specifically, besides y_{0is}, the pre-program outcome value for student i in school *s*, we controlled for X_{is}, a vector of additional control variables. The variables in X_{is} are student and family characteristics (*male,* ethnic group (*Han, Tibetan, Tu, Hui and Salar*), *only child*, *father illiterate, mother illiterate, family off-farm, live with both parents, ever used a computer, and access to other modern technology*). By including y_{0is} and X_{is} as control variables, β in equation (2) is an unbiased, efficient estimate of the CAL treatment effect.

Results

The data show that third-grade students in the treatment group improved significantly more in their academic performance in Mandarin than students in the control group (Table 3, row 1, columns 1 and 2). Using the sample including all third-grade students in the 57 sample schools, the estimated CAL treatment effect on Mandarin test scores is equal to 0.14 standard deviations and is significant at the 10% level using the basic model (equation (1); Table 3, row 1, column 1).

It increases to 0.20 and becomes more significant (at the 1% level) using the model controlling for the various student and school characteristics (equation (2)—row 1, column 2).[7]

Spillovers in the Student Math Test Scores

Impressively, the impacts of CAL were not only on the student Mandarin test scores—following the subject matter that was the focus of CAL—but also on the student math test scores (Table 3, row 1, columns 3 and 4). Specifically, compared to their counterparts in the control schools, third-grade students in the treatment schools improved their standard math test scores 0.22-0.23 standard deviations more over the program period. These results suggest that CAL intervention did not improve the performance of students in Mandarin at the expense of their performance in math. Instead, the academic benefits from the CAL intervention spilled over to mathematics as well.

There might be several reasons for this impressive positive spillover effect of the CAL intervention to student math performance. First, the CAL intervention might have increased the general learning ability of the students, which helped improve their performance in other subjects including math. Second, the game-based CAL software made learning a more fun and engaging process and thus stimulated the student interest in learning Mandarin as well as other subjects (including math). More importantly, as discussed before, all school subjects were taught in Mandarin and the textbooks were written in Mandarin. Therefore, improved Mandarin language

[7] In assessing the results, one may ask if part of the CAL effect was an "extra teacher effect." In fact, we initially considered adding an additional treatment arm—as was included in Banerjee et al. (2007). The additional treatment would have provided remedial tutoring in the form of traditional teacher-based after school tutoring sessions. However, when we tried to implement this, we were not allowed to do so by the school district. Many Chinese school districts have strict regulations against the running of traditional teacher-based after school tutoring sessions. In addition, as discussed in the body of the paper above, we did not allow teachers to teach Mandarin or interact with students about Mandarin-related questions during the CAL sessions (except that the teachers could answer questions about computer operations from the students). In fact, the scope for interaction was not great since none of the CAL supervisors were Mandarin teachers. There was little ability and no incentive for them to disregard the rules of the program and teach children additional Mandarin.

skills via CAL might also have helped the students better understand course materials in math. In particular, according to the national uniform math curriculum, Chinese primary schools begin aggressively introducing arithmetic word problems in the third grade, and thus the questions in the math test of the final evaluation were mostly arithmetic word problems, which demand much higher level of reading comprehension than the questions of the baseline test, which were mainly arithmetic calculation problems. The word count of the math test of the final evaluation is 1355 words, compared to the 836 words for the math test of the baseline evaluation. Therefore, as a result of improved reading comprehension ability via CAL intervention, students in the treatment group might have outperformed those in the control group in the evaluation test as the test mainly consisted of arithmetic word problems.

He et al. (2008) also found that an Indian English Education Curriculum implemented via teacher training improved student math and English score rather than just their English scores. They claimed that this might be due to the fact that teachers implemented the interventions more efficiently and, thus, had more time available for other subjects. This is unlikely to be the case in our study, since the teacher-supervisors did not have the autonomy to substitute the time of CAL sessions for other subjects. The teacher-supervisors in our CAL experiment also did not teach the major subjects to the students in our sample (and so they really would have been unable to provide extra tutoring). Survey data and interviews with students at the end of the program also confirmed that the CAL sessions were not used for other purposes (including classes in other subjects). Moreover, for the schools in our sample, there was also almost no remedial tutoring session offered after school.

Effects of the CAL Intervention on Student Non-academic Outcomes

As for the non-academic outcomes, our results show that the CAL intervention had positive effects on the student non-academic outcomes, though not statistically significant in most of the cases (Table 4, row 1). Compared to students in the control group, the students in the treatment group "like school" more and had higher levels of self-efficacy of Mandarin studying, self-confidence and metacognition. In particular, students in the treatment group had significantly higher levels of self-efficacy of Chinese studying than those in the control group (row 1, column 1).

Heterogeneous Effects of the CAL Intervention on Student Academic and Non-Academic Outcomes for Students with Different Academic Backgrounds

Besides the average impact of the CAL program on student academic and non-academic performance, it is also important to explore whether the CAL program has different impacts for different students. In particular, we are interested in understanding whether the CAL intervention is more effective in the case of students with lower academic backgrounds. To do this, we use Equation (2), but include interaction terms between the treatment variable and student academic background (which is measured using student baseline Chinese and math scores).

The results show that the CAL intervention benefited low-performing students more than it did to high-performing students (Table 5, row 2). Compared to the students in the control group, students in the treatment group with baseline Mandarin test score one standard deviation lower than the mean level improved 0.16 standard deviations more in their Mandarin test scores over the program period (significant at the 1% level; row 2, column 1). Compared to the students in the control group, students in the treatment group with baseline math test score one standard deviation lower than the mean level also improved 0.09 standard deviations more in their

Mandarin test scores over the program period (significant at the 10% level; row 2, column 2). This finding indicates that CAL has been successful in achieving its goal of helping low-performing students catch up. We also tested heterogeneous effects of the CAL intervention on student math performance, but no significant patterns of heterogeneous CAL intervention effects were found (results not included for brevity).[8]

Using the same estimation model we did with the student academic outcome, we found that low-performing students not only benefited more than other students from the CAL intervention in academic performance, but also in some non-academic outcomes. Compared to the students in the control group, students in the treatment group with lower than average baseline test scores in Mandarin and math improved significantly more in their levels of self-efficacy of Mandarin studying than other students in the treatment group over the program period (Table 5, row 2, columns 3 and 4).

Conclusions

In this paper we present the results from a randomized field experiment of a Computer Assisted Learning (CAL) program involving around 1717 third-grade students, aged nine to eleven and from poor rural families, in 57 public schools in poor rural minority areas in Haidong

[8] The results of the heterogeneity analysis looking at the insignificant coefficient of the treatment*baseline test score on math performance should not be surprising. One possible explanation is that this is so because the improvement in math (that we are observing for the full sample) is a product of two things: the CAL program in Mandarin (and the improvement in language from the CAL program); and the beginning level of one's Mandarin language ability (which maybe strongly correlated with the ability to translate improvement in Mandarin competency into improvement in math performance). Hence, although students with lower baseline standardized Mandarin test scores gained more in their Mandarin performance (and this may have contributed to higher math scores), it is also true that they began with lower levels of Mandarin comprehension (and this may have limited the improvement in their math scores). Therefore, it is plausible that even with the relatively higher gains in Mandarin language skills, the lower performing students (at the baseline) were unable to translate those into relatively higher gains in math after the Mandarin-CAL treatment.

Prefecture, Qinghai. To evaluate the effectiveness of the program we randomly chose 26 schools from the sample as treatment schools, and the third-grade students attending these schools received the CAL Mandarin after-school remedial program. The remaining 31 schools served as control schools, with third-grade students attending these schools as the control group. Third-grade students in the treatment group were offered 40 minutes of shared computer time after school, twice a week. During these sessions students played computer-based games that required them to practice using their Chinese knowledge and skills. The CAL program was tailored to the regular school Chinese 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 outcomes, at least in the short term. Two 40-minute CAL Chinese sessions per week increased student standardized Chinese scores by 0.14-0.20 standard deviations. In general, low-performing students benefited more from the program. Moreover, the CAL intervention also had significant impact on the math test scores of the students, with the magnitudes of the effects varying from 0.22 to 0.23 standard deviations. Still further, the CAL intervention appeared to have some positive impacts on the non-academic outcomes of the students, though not statistically significant in most cases. In particular, the CAL intervention significantly improves the student self-efficacy of Chinese studying, with low-performing students benefited most from the intervention.

Given the significant impact of CAL on student academic outcomes found in this paper, educational policy makers in China (and in other developing countries) should consider scaling CAL programs, especially in public schools serving disadvantaged students (e.g., rural public schools in China, particularly those in remote rural minority areas). Of course, public schools in rural minority areas in Qinghai might not be representative of all poor public schools in China or in other developing countries. Nonetheless, all public schools in poor rural areas 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 students. Rural minority students are also the most vulnerable among all rural students and suffer from disadvantageous living conditions and learning environment.

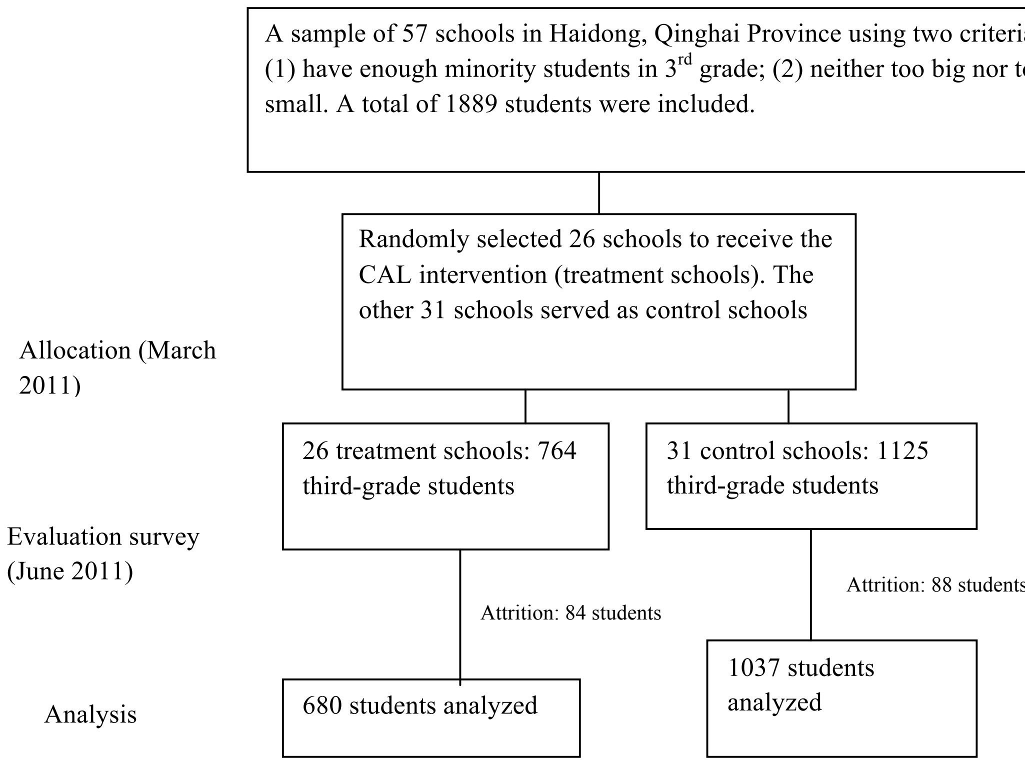

Figure 1: Experiment Profile

Table 1. Regressions of student characteristics on indicators of attrition and treatment dummies.

VARIABLES	(1) Baseline Mandarin score (units of standard deviation)[a]	(2) Baseline math score (units of standard deviation)[b]	(3) Male (1=yes; 0=no)	(4) Only child (1=yes; 0=no)	(5) Father illiterate (1=yes; 0=no)	(6) Mother illiterate (1=yes; 0=no)	(7) Family off-farm	(8) Live with both parents	(9) Ever used a computer (1=yes; 0=no)	(10) Access to other modern technologies[c]	(11) Tu	(12) Tibetan	(13) Hui	(14) Salar
(1) Mean	0.01	-0.03	0.49	0.13	0.20	0.56	0.07	0.56	0.35	0.47	0.12	0.21	0.33	0.20
(2) Attrition	-0.33***	-0.27	0.03	-0.01	-0.00	0.12***	0.06	-0.13***	0.09	-0.00	-0.02	-0.14***	0.23***	0.02
	(0.12)	(0.15)	(0.06)	(0.03)	(0.06)	(0.04)	(0.03)	(0.05)	(0.06)	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)
(3) Treatment	0.00	-0.11	-0.03	-0.01	-0.02	0.03	-0.00	-0.04	-0.06	0.01	0.02	0.02	-0.04	0.07
	(0.16)	(0.15)	(0.03)	(0.02)	(0.03)	(0.04)	(0.02)	(0.05)	(0.07)	(0.03)	(0.06)	(0.10)	(0.12)	(0.12)
(4) Treatment*attrition	0.23	0.26	0.05	-0.00	0.06	0.01	0.02	0.12	0.06	0.04	0.03	-0.05	-0.01	-0.07
	(0.18)	(0.22)	(0.08)	(0.05)	(0.07)	(0.07)	(0.05)	(0.06)	(0.09)	(0.04)	(0.05)	(0.07)	(0.08)	(0.07)
(5) Observations	1,889	1,889	1,889	1,884	1,874	1,885	1,889	1,889	1,889	1,889	1,889	1,889	1,889	1,889

Robust standard errors in parentheses

** significant at 5%; *** significant at 1%

[a,b] The baseline math/Mandarin score is the score on the standardized math/Mandarin 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.

Table 2. Comparison of student and school characteristics between the treatment and control schools

		Treatment	(26 schools, 680 obs)	Control	(31 schools, 1037 obs)	The raw difference (treatment-control)	
		Mean	SD	Mean	SD	Mean	SD
		(1)	(2)	(3)	(4)	(5)	(6)
Student level characteristics							
(1)	Baseline Mandarin score (units of standard deviation)[a]	0.00	1.08	0.00	1.00	0.00	0.04
(2)	Baseline math score (units of standard deviation)[b]	-0.11	1.05	0.00	1.00	-0.10	0.04
(3)	Male (1=yes; 0=no)	0.46	0.50	0.49	0.50	-0.03	0.03
(4)	Only child (1=yes; 0=no)	0.13	0.33	0.14	0.34	-0.01	0.02
(5)	Father illiterate (1=yes; 0=no)	0.19	0.39	0.21	0.40	-0.02	0.03
(6)	Mother illiterate (1=yes; 0=no)	0.57	0.50	0.54	0.50	0.03	0.04
(7)	Family off-farm	0.06	0.25	0.07	0.25	0.00	0.02
(8)	Live with both parents	0.54	0.50	0.58	0.49	-0.04	0.05
(9)	Ever used a computer (1=yes; 0=no)	0.30	0.46	0.37	0.48	-0.06	0.07
(10)	Access to other modern technologies[c]	0.47	0.22	0.46	0.24	0.01	0.03
(11)	Tibetan	0.13	0.34	0.11	0.31	0.02	0.06
(12)	Tu	0.24	0.43	0.22	0.41	0.02	0.10
(13)	Hui	0.29	0.45	0.32	0.47	-0.04	0.12
(14)	Salar	0.24	0.43	0.17	0.38	0.07	0.12
(15)	Han	0.10	0.31	0.17	0.38	-0.07	0.05
School level characteristics							
(16)	Number of students	177.96	78.66	213.90	102.36	-35.94	24.00
(17)	Teacher-student ratio	0.06	0.02	0.06	0.02	0.00	0.00
(18)	Proportion of female teachers	0.51	0.21	0.54	0.19	-0.03	0.05
(19)	School facility index[d]	0.56	0.14	0.58	0.18	-0.02	0.04
(20)	Proportion of students receiving poverty subsidy	0.45	0.31	0.36	0.23	0.09	0.07

Source: Authors' survey.

** significant at 5%; *** significant at 1%. Robust standard errors.

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

The p-value of the F-test of the joint significance of the student individual variables is 0.83, and p-value of the F-test of the joint significance for school-level characteristics is 0.22 .

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

	Dependent variable: standardized post-CAL test score				
		Mandarin (1)	Mandarin (2)	Math (3)	Math (4)
(1)	Treatment (1=the treatment group; 0=the control group)	0.14* (0.07)	0.20*** (0.07)	0.23** (0.09)	0.22*** (0.07)
Control variables:					
(2)	Baseline Mandarin score (units of standard deviation)[a]	0.64*** (0.03)	0.49*** (0.03)	0.49*** (0.03)	Y
(3)	Baseline math score (units of standard deviation)[b]		Y		0.50*** (0.03)
(4)	Student characteristics[c]		Y		Y
(5)	School characteristics[d]		Y		Y
(6)	Observations	1717	1717	1717	1717
(7)	R-squared	0.48	0.53	0.39	0.44

Source: Authors' survey.

** significant at 5%; *** significant at 1%. Robust standard errors clustered at the school level.

[a,b] The baseline math/Mandarin score is the score on the standardized math/Mandarin test that was given to all students in the sample

[c] Student characteristics include male (1=yes; no=0), only child (1=yes; no=0), father illiterate (1=yes; no=0), mother illiterate (1=yes; no=0), family off-farm (1=yes; no=0), live with both parents (1=yes; no=0), ever used a computer (1=yes; no=0), Access to other modern technologies, and ethnic groups (Han, Tibetan, Tu, Hui and Salar).

[d] School characteristics include number of students, teacher-student ratio, proportion of female teachers, School facility index (as described in Table 2), and proportion of students receiving poverty subsidy.

Each column reports the results of one regression of the student final standardized Mandarin or math test scores over the program period on the variables in rows (1) to (5).

Table 4. Ordinary Least Squares estimators of the impacts of CAL program on student non-academic outcomes

	Dependent variable: Indicators of non-academic outcomes	Self-efficacy of Mandarin studying	Self-confidence	Like school	Metacognition in Mandarin
		(1)	(2)	(3)	(4)
(1)	Treatment (1=the treatment group; 0=the control group)	0.07*	0.02	0.37	0.05
		(0.04)	-0.04	(1.46)	(0.04)
(2)	Baseline values of the non-academic outcome indicator and other control variables (student and school characteristics)	Y	Y	Y	
(3)	Observations	1715	1715	1715	1717
(4)	R-squared	0.15	0.12	0.05	0.06

Source: Authors' survey.

** significant at 5%; *** significant at 1%. Robust standard errors clustered at the school level.

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

Other control variables include student characteristics including male (1=yes; no=0), only child (1=yes; no=0), father illiterate (1=yes; no=0), mother illiterate (1=yes; no=0), family off-farm (1=yes; no=0), live with both parents (1=yes; no=0), ever used a computer (1=yes; no=0), Access to other modern technologies, and ethnic groups (Han, Tibetan, Tu, Hui and Salar), and school characteristics including number of students, teacher-student ratio, proportion of female teachers,School facility index (as described in Table 2), and proportion of students receiving poverty subsidy.

Each column reports the results of one regression of the student non-academic outcome on the corresponding variables in rows (1) to (2).

Table 5. The Ordinary Least Squares estimators of the heterogeneous program effect on students with different levels of academic performance

	Dependent variable:	Standardized post-CAL Mandarin test score		Self-efficacy of Mandarin studying	
		(1)	(2)	(3)	(4)
(1)	Treatment (1=the treatment group; 0=the control group)	0.19***	0.19***	0.08*	0.07*
		[0.06]	[0.06]	[0.04]	[0.04]
(2)	Interactions: Treatment*	Baseline Mandarin score (units of standard deviation)[a]	Baseline math score (units of standard deviation)[b]	Baseline Mandarin score (units of standard deviation)	Baseline math score (units of standard deviation)
		-0.16***	-0.09*	-0.06**	-0.09***
		[0.05]	[0.04]	[0.03]	[0.03]
(3)	Control variables	Y	Y	Y	Y
(4)	Observations	1717	1717	1717	1717
(5)	R-squared	0.24	0.23	0.23	0.23

Source: Authors' survey.

** significant at 5%; *** significant at 1%. Robust standard errors clustered at the school level.

Control variables include the baseline math/Mandarin score, student characteristics including male (1=yes; no=0), only child (1=yes; no=0), father illiterate (1=yes; no=0), mother illiterate (1=yes; no=0), family off-farm (1=yes; no=0), live with both parents (1=yes; no=0), ever used a computer (1=yes; no=0), access to other modern technologies, and ethnic groups (Han, Tibetan, Tu, Hui and Salar), and school characteristics including number of students, teacher-student ratio, proportion of female teachers,School facility index (as described in Table 2), and proportion of students receiving poverty subsidy.

Each column reports the coefficients of the treatment dummy variable (in row (1)) and the interaction of the treatment dummy variable and the corresponding student characteristics in row (2) from one regression.

Each regression regresses the dependent variable on the treatment dummy variable and the interaction of the treatment dummy variable and the corresponding student characteristics in row (2) in that column, controlling for all of the control variables.

Reference

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*.

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.

Hannum, E. 1999b. Poverty and Basic-Level Schooling in the People's Republic of China: Equity Issues in the 1990s. *Prospects: The Quarterly Journal of Comparative Education* 29: 561–77.

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

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

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.

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.

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