Teaching the Language of Wider Communication, Minority Students, and Overall Educational Performance: Evidence from a Randomized Experiment in Qinghai Province, China

FANG LAI
Katholieke Universiteit Leuven and Center for Chinese Agricultural Policy, Institute of Geographic Sciences and Natural Resource Research, Chinese Academy of Sciences

LINXIU ZHANG, QINGHE QU, and XIAO HU
Center for Chinese Agricultural Policy, Institute of Geographic Sciences and Natural Resource Research, Chinese Academy of Sciences

YAOJIANG SHI
Shaanxi Normal University

MATTHEW BOSWELL and SCOTT ROZELLE
Stanford University

The education of poor and disadvantaged populations, particularly those from minority subgroups, has been a long-standing challenge to education systems in both developed and developing countries (e.g., World Bank 2001, 2004; Glewwe and Kremer 2006; Planty et al. 2008). For example, over the past decade in the United States the high school dropout rate of Hispanic students has remained at least twice as high as that of white students (Aud et al. 2011). Using data from the German Microzensus and the German Socio-Economic Panel, Alba, Handl, and Müller (1994) found that, relative to young Germans with identical sociodemographic characteristics, Italian, Turkish, and Yugoslav children are overrepresented in the lowest academic track of the German school system.

One reason for the underperformance of ethnic minority students might be low competency in the national language or the language of wider communi-

We would like to acknowledge Acer Computers and the Asia-Pacific Economic Cooperation (APEC) Digital Opportunity Center (ADOC) 2.0, and the Natural Science Foundation of China (grant 71333012), for their generous support for the Rural Education Action Program'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 article possible.

Electronically published May 4, 2015

cation (LWC). Studies worldwide have documented that language difficulties negatively affect the academic performance of students whose mother tongue is not the LWC. For example, August and Shanahan (2006) and Genesee et al. (2006) show that language barriers were among the major factors for the relative low performance of non-English-speaking minority groups in the United States. As a result, improving national language skills has become one of the most urgent academic needs for ethnic minority groups or immigrants/refugees whose mother tongue is not the LWC.

Usually, there are two ways to deal with the language barriers faced by minority students in their academic progress. One way is to implement *bilingual education* and teach in the mother tongue of the minority group in the early years of their education to secure adequate academic progress. In the bilingual education regime the instruction language is gradually switched to the LWC in later years of schooling. The other approach is keeping the LWC as the language of instruction while making efforts (e.g., through specially targeted courses) to improve the LWC skills of the students who are nonnative LWC speakers (henceforth, *keeping instruction in the LWC*).

While there have been studies on the effectiveness of these two approaches (bilingual education and keeping instruction in the LWC), the findings are inconclusive. Although the overall findings from these studies and an 8-year-old meta-analysis of these studies are in favor of bilingual education (Bamgbose 1984; Yates 1995; IDRC 1997; Cheung and Slavin 2005), most recent large-scale and long-term evaluations do not find significant benefits of bilingual education over keeping instruction in the LWC programs (Slavin et al. 2009, 2011). Recent studies agree that high-quality program design is much more important than the language of instruction (August and Shanahan 2006; Slavin et al. 2011). In other words, it does not really matter whether a school system offers bilingual education or keeps instruction in the LWC; it is the quality of the language program that matters.

While the existing literature is informative, there are several limitations. First, most of the recent methodologically rigorous studies directly compare the relative effectiveness of bilingual education versus LWC programs. There are few recent studies that compare individual programs (either bilingual or LWC programs) with a pure control group, particularly within large-scale, randomized controlled trials. Second, there is generally a dearth of methodologically adequate studies outside of the United States. Most studies in the literature have studied bilingual or English-only (or LWC-only) programs for English language learners, predominantly for Spanish speakers in the United States.

Apart from these critical deficiencies in the existing research, there has also been almost no rigorous investigation of the potential spillover/crossover effects of language-focused interventions on minority student performance in subjects other than language. Among the limited number of studies that are (even remotely) related to this topic, none have investigated the spillover effects of language interventions on minority students directly. Master et al. (2013) used multiyear data from New York City and an instrumental variable (IV) approach (the IV was the 2-year lagged values of the endogenous variable—teacher quality measured by "teacher value-added") to demonstrate that "high-quality English teachers" have long-term positive impacts not only on language performance but also on math performance. Employing a randomized controlled trial approach in India, He, Linden, and MacLeod (2008) also found positive spillovers onto student math performance from a computer-based program that teaches students English. However, according to the authors' analyses, the mechanism of the improved math performance was not better English language skills of students but rather the result of more effective program time management by the teacher supervisors. To our knowledge, there has not been an in-the-field experiment designed explicitly to identify the impact of a program that was created to improve LWC/national language skills and also examine the subsequent effects on overall academic performance.

The overall objective of this study is to examine the impact of an LWC learning program on the overall academic performance (and other outcomes) of students who have limited competency in the LWC. To achieve this research objective, we present the results from a randomized field experiment of a computer assisted learning (CAL) program that teaches rural ethnic minority students LWC language skills and examines their performance in both LWC skills and math. Specifically, the CAL program teaches Mandarin language skills to non-Mandarin-speaking students in 57 schools in Qinghai Province in northwestern China. In this part of Qinghai, around 86% of the students are ethnic minorities whose mother tongues are not Mandarin. The academic performance of non-Han (minority) students in this area of China lags substantially behind Han (nonminority) students (Yang, Wang et al. 2013).

We are particularly interested in a CAL program because even though existing empirical studies have shown mixed evidence of the impacts of CAL on student academic outcomes, many of the recent randomized experiments have shown that CAL can be particularly effective in meeting the academic needs of students who live in environments in which schools and teaching are poor in quality and the home learning environment is inadequate (e.g., Banerjee et al. 2007; Barrow, Markman, and Rouse 2008; He et al. 2008; Lai et al. 2013, 2014; Yang, Zhang et al. 2013). For these students CAL may be able to act as a substitute for tutors or higher-quality in-school teaching.

Although there have been other CAL studies in China and elsewhere in recent years, the CAL program that forms the basis for this article is different from these. To our knowledge, the study in this article is the first CAL study conducted in China to focus on implementing a computer assisted Mandarin language learning program for minorities and examining, in particular, its impact on both student Mandarin language skills and performance in other subjects. In the studies in Lai et al. (2013, 2014) and Mo et al. (2014), the CAL programs were designed to deliver remedial tutoring for math, not the Mandarin language. These other CAL studies were also undertaken in regions of China that catered to Han students whose mother tongue was Mandarin—the officially designated national language/LWC in China—and their samples included virtually no non-Han students (Lai et al. 2013, 2014; Mo et al. 2014).

In one paper (Yang, Zhang et al. 2013)—which might be considered a small meta-analysis of the CAL programs that have been run in China over the past several years—a set of coefficients from Qinghai that showed the impact of a Mandarin CAL program on Mandarin language performance was presented simultaneously with two sets of coefficients from the Beijing and Shaanxi CAL program sites. The focus of this article, however, was on the gender difference of the impacts of CAL programs. There was no discussion of the spillover effects from the Mandarin CAL program onto student overall academic performance—the focus of the current study. The implications of these spillovers are important to understand given chronic academic underperformance among nonnative speakers of Mandarin in comparison to native-speaking peers. The only randomized controlled trial that has explored the spillover effects of a CAL language program on student performance in other subjects is He et al. (2008) in India. However, as mentioned above, they explained improved math skills not through improved language skills but rather through the more efficient delivery of the language curriculum. Hence, the results from the current study are the first to demonstrate the effect of a CAL language program on minority students' academic performance in a school system that is committed to keeping instruction in the LWC.

In launching a study on the effect of a CAL program, we contribute to the existing literature in several ways. First, by implementing a randomized controlled trial, our study not only provides an unbiased estimate of the impacts of a CAL language program on student language competency but also provides evidence of the causal impacts of improved language competency on student academic performance in other subjects (in this case, math). Furthermore, this study focuses on a CAL program teaching Mandarin to rural ethnic minority students in China and, hence, is outside the usual environment (teaching Spanish-speaking Latino children English in the United States). Overall, our work demonstrates that a Mandarin CAL language program has the potential to provide a relatively quick and effective remedy for (at least part of) the severe underperformance of rural minority children in China in both Mandarin and other subjects.

The rest of the article is organized as follows. In the next section, we introduce the sample, data, and method of this article. The following two sections report the results and conclude.

Sampling, Data, and Method

Sampling and the Process of Randomization

We conducted a clustered randomized controlled trial of CAL in rural schools during the spring semester of 2011. A total of 1,889 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% of the population lives in rural areas. Qinghai also has a large minority population. Ethnic minorities make up around 46% 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 minority autonomous. 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 its number of minority students 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 the two criteria above. Because of our limited computer provision, we then randomly chose 57 schools for our final sample.[1]

[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 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 three 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 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

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 1,889 third-grade students included in the sample (fig. 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 constituted the treatment group (fig. 1). The 1,125 third-grade students in the other 31 schools served as the control group.

Using data from the treatment and control schools, we can show that the sample is well balanced. To show this, we used a set of student and school characteristics to check the validity of the random assignment for our sample by regressing each variable on the treatment dummy variable (table 1). 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 or not (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 that the differences between the treatment and control groups were statistically insignificant (table 1, cols. 5 and 6). The *F*-tests of joint significance of the individual characteristics and school characteristics also showed that there were no significant differences in the individual and school characteristics between the treatment and control groups.

Although at the time of the baseline survey the main sample included a total of 57 schools and 1,889 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), at the time of the evaluation survey we followed

that we decided how many computers were needed for the treatment schools. This number was calculated as roughly half of the number of grade 3 students. When the computers were delivered to the project site, it was discovered that there were an 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 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 calculated the power with our sample of 57 schools. The power to detect a treatment effect of a size around 0.2 and significant at the 5% level is around 0.8, with an intraclass correlation coefficient (ICC) equal to 0.09 (which is the actual ICC of our study), R^2 equal to 0.5, and total number of clusters equal to 57 (with 30 observations in each cluster on average).

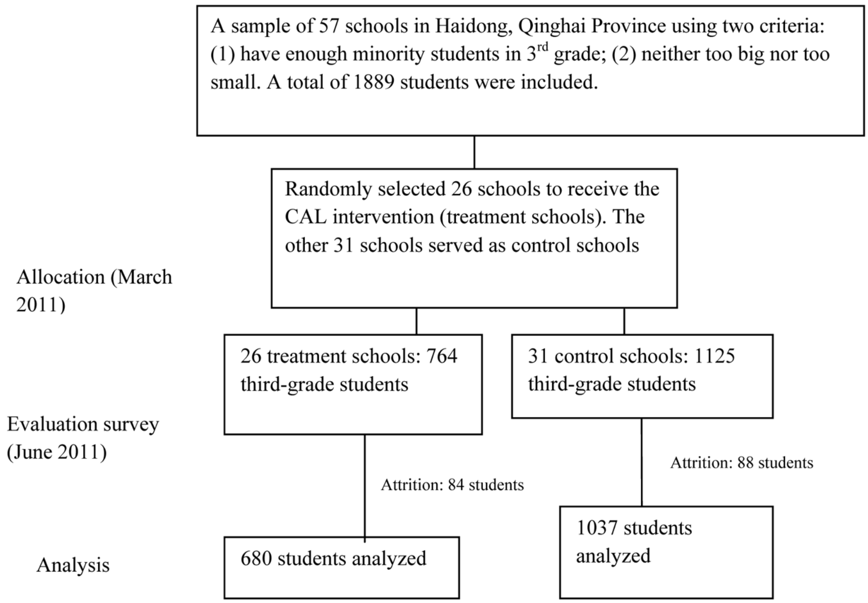

Figure 1. Experiment profile

up with 1,717 third-grade students in the 57 sample schools (fig. 1). Among the 764 treatment students, 680 remained in our sample at final evaluation. Among the 1,125 control students, 1,037 remained in our sample at final evaluation.

To understand the nature of those who attrited and assess whether 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 2), we see that those who attrited (col. 2) have lower Mandarin scores, have a higher share of mothers who were illiterate, were less likely to live with both parents, were less likely to be Tibetan and more likely to be Hui, and were more likely to be from Xunhua. However, as seen from the fact that none of the coefficients of the interaction variables were statistically significant, there is no evidence that the attrition affected the validity of our randomization. Moreover, there is no statistically significant difference in the attrition rate between the treatment and control groups (table 1).

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

TABLE 1
COMPARISON OF STUDENT AND SCHOOL CHARACTERISTICS BETWEEN THE TREATMENT AND CONTROL SCHOOLS (INCLUDING BOTH THE SAMPLE AND ATTRITIONS)

	Treatment (26 Schools, *N* = 764)		Control (31 Schools, *N* = 1,125)		Difference (Treatment − Control)	
	Mean (1)	SD (2)	Mean (3)	SD (4)	Mean (5)	SD (6)
Student-level characteristics:						
Baseline Mandarin score (units of SD)[a]	.01	1.03	.01	.97	.02	.16
Baseline math score (units of SD)[a]	.05	1.02	.03	.98	.08	.14
Male (1 = yes; 0 = no)	.47	.50	.50	.50	.02	.03
Only child (1 = yes; 0 = no)	.13	.33	.14	.34	.01	.02
Father illiterate (1 = yes; 0 = no)	.19	.39	.21	.40	.01	.03
Mother illiterate (1 = yes; 0 = no)	.58	.49	.55	.50	.03	.04
Family off-farm	.07	.26	.07	.26	.00	.02
Live with both parents	.54	.50	.57	.50	.03	.05
Ever used a computer (1 = yes; 0 = no)	.32	.47	.37	.48	.05	.07
Access to other modern technologies[b]	.48	.22	.46	.24	.01	.03
Tibetan	.13	.34	.11	.31	.03	.06
Tu	.22	.41	.21	.41	.01	.09
Hui	.31	.46	.34	.47	.03	.12
Salar	.23	.42	.17	.38	.06	.11
Han	.10	.30	.17	.37	.07	.05
Attrition	.11	.31	.08	.27	.03	.02
School-level characteristics:						
Number of students	198.41	87.84	251.16	111.36	−52.75	31.76
Teacher-student ratio	.06	.02	.05	.01	.00	.00
Proportion of female teachers	.51	.21	.55	.18	.03	.06
School facility index[c]	.57	.14	.57	.19	.00	.05
Proportion of students receiving poverty subsidy	82.87	71.76	86.31	67.48	−3.44	20.90
County Huzhu	.35	.48	.42	.49	.07	.13
County Hualong	.25	.44	.18	.38	.08	.12
County Xunhua	.40	.49	.41	.49	.01	.14

Source. Authors' survey.

[a] Score on the standardized math/Mandarin test that was given to all students in the sample before the computer assisted learning (CAL) program.

[b] 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.

[c] 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.

students of the 31 control schools). The CAL Intervention was the only intervention going on in these schools at the time.

CAL Intervention Group (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

TABLE 2

REGRESSIONS OF STUDENT CHARACTERISTICS ON INDICATORS OF ATTRITION AND TREATMENT DUMMIES

	Mean (1)	Attrition (2)	Treatment (3)	Treatment × Attrition (4)
Student-level characteristics:				
Baseline Mandarin score (units of SD)[a]	.00 (1.00)	.33*** (.12)	.00 (.16)	.23 (.18)
Baseline math score (units of SD)[a]	.00 (1.00)	.27 (.15)	.11 (.15)	.26 (.22)
Male (1 = yes; 0 = no)	.49 (.50)	.03 (.06)	.03 (.03)	.05 (.08)
Only child (1 = yes; 0 = no)	.13 (.34)	.01 (.03)	.01 (.02)	.00 (.05)
Father illiterate (1 = yes; 0 = no)	.20 (.40)	.00 (.06)	.02 (.03)	.06 (.07)
Mother illiterate (1 = yes; 0 = no)	.56 (.50)	.12*** (.04)	.03 (.04)	.01 (.07)
Family off-farm	.07 (.26)	.06 (.03)	.00 (.02)	.02 (.05)
Live with both parents	.56 (.50)	.13*** (.05)	.04 (.05)	.12 (.06)
Ever used a computer (1 = yes; 0 = no)	.35 (.48)	.09 (.06)	.06 (.07)	.06 (.09)
Access to other modern technologies[b]	.47 (.24)	.00 (.03)	.01 (.03)	.04 (.04)
Tibetan	.21 (.41)	.14*** (.05)	.02 (.10)	.05 (.07)
Tu	.12 (.32)	.02 (.03)	.02 (.06)	.03 (.05)
Hui	.33 (.47)	.23*** (.05)	.04 (.12)	.01 (.08)
Salar	.20 (.40)	.02 (.05)	.07 (.12)	.07 (.07)
School-level characteristics:				
Number of students	229.82 (105.7)	20.91 (23.97)	−50.48 (31.67)	−26.75 (26.04)
Teacher-student ratio	.06 (.01)	.00 (.00)	.00 (.00)	.00 (.00)
Proportion of female teachers	.53 (.19)	.01 (.03)	.04 (.05)	.09 (.05)
School facility index[c]	.57 (.17)	.01 (.05)	.00 (.05)	.01 (.05)
Proportion of students receiving poverty subsidy	.40 (.27)	.02 (.09)	.10 (.08)	.06 (.10)
County Huzhu	.39 (.49)	.28*** (.07)	.06 (.14)	.00 (.10)
County Hualong	.21 (.41)	.01 (.06)	.08 (.12)	.03 (.08)
County Xunhua	.41 (.49)	.28*** (.08)	.02 (.14)	.03 (.10)

Note. Robust standard errors clustered at the school level in parentheses. Each column is a regression of the row variable on the column variables. *N* = 1,889.

[a] Score on the standardized math/Mandarin test that was given to all students in the sample before the computer assisted learning (CAL) program.

[b] 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.

[c] 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.

*** Significant at 1%.

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 comprehension, and grammar) introduced in class. If a student had a Mandarin-related question, he or she was encouraged to discuss it with his or her teammates. 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 communication 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 who 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 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 (approximately US$15). This amount was equal to roughly 15% 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 attended a 2-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.[2] 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.

According to the notes of the volunteer monitors, almost all schools closely followed the CAL program instructions during implementation. During the end line survey, the interviewers also asked the students how often they had the CAL sessions (more than 99% said they had the sessions regularly), whether any of the CAL sessions were taken up for other classes (none of the students reported that the CAL sessions substituted for other classes), and whether the teacher-supervisors taught them any subject content during the CAL sessions (according to the survey only 0.01% said that the teacher-supervisors occasionally offered them clues on how to solve problems during the CAL sessions when the students asked them for help). Therefore, we found few violations of our rules for CAL implementation according to the student responses. According to records kept by teacher-supervisors, there were almost no cancellations of CAL sessions, except for a very few due to irregular events (e.g., sudden electricity shutdown). It should be noted, however, that whenever a cancellation did occur, all schools arranged make-up sessions so that the CAL sessions were managed according to our schedule.

CAL Control Group (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 type of 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

[2] 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. Apart from 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 almost no interaction with the students 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.

except for during the baseline and final evaluation surveys. The students in the control group took their regular Mandarin classes at school as usual.

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 the 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, 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.[3] 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 data set 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 whether 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 or her hometown 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 noncognitive traits. To create indicators for students' attitudes toward schooling (like school), the students were asked

[3] Personnel from the Center for Examination of Beijing helped us pick questions for the tests from official examination books and exercise books.

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 the self-confidence and the self-efficacy of studying Mandarin were created from the responses of students to a 10-item psychological scale for the self-confidence and a seven-item psychological scale for the self-efficacy of studying Mandarin.[4]

Statistical Method

We used ordinary least squares (OLS) regression analysis (both with and without control variables) to estimate how the academic and noncognitive outcomes changed in the treatment group relative to the control group. Our basic OLS analysis regressed the end line outcome variables (i.e., postprogram 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 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

$$y_{is} = \alpha + \beta \times \text{treatment}_s + \theta \times y_{0is} + \varepsilon_{is}, \tag{1}$$

where y_{is} is the end line 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 noncognitive 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 whether 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 apart from Mandarin. Importantly, in addition to variables

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

measuring academic outcomes, we included three noncognitive outcome variables, namely, like school, self-confidence, and self-efficacy in studying Mandarin.

By construction, the coefficient of the dummy variable treatment$_s$, β, measures the difference in the outcome variable between the treatment and control groups conditional on the baseline value of the outcome variable. In other words, β measures, holding the baseline value of the outcome variable constant, how well the treatment group does in terms of the outcome variable 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:

$$y_{is} = \alpha + \beta \times \text{treatment}_s + \theta \times y_{0is} + X_{is}\gamma + \varepsilon_{is}, \qquad (2)$$

where all the variables and parameters are the same as those in equation (1), except that we added a set of control variables. Specifically, apart from y_{0is}, the preprogram 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

Table 3 reports the estimated treatment effects of the CAL program on student standardized Mandarin and math test scores, with each column reporting estimated coefficients from one regression of the outcome variable on explanatory variables in the corresponding rows. Column 1 reports the estimated coefficients from regressions of the standardized Mandarin scores on the treatment dummy variable and student baseline Mandarin test scores. Column 2 reports the results of the same regression but includes a set of control variables (as shown in the corresponding rows). Columns 3 and 4 follow the same models as columns 1 and 2, respectively, except the subject is math rather than Mandarin. The top row of this table reports the estimated treatment effects of each model.

The data show that grade 3 students in the treatment group improved significantly more in their academic performance in Mandarin than students in the control group (table 3, cols. 1 and 2). Using the sample including all grade 3 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

TABLE 3

ORDINARY LEAST SQUARES ESTIMATORS OF THE IMPACTS OF CAL PROGRAM ON STUDENT ACADEMIC OUTCOMES

	Mandarin		Math	
	(1)	(2)	(3)	(4)
Treatment (1 = treatment group; 0 = control group)	.14*	.20***	.23**	.22***
	(.07)	(.07)	(.09)	(.07)
Control variable:				
Baseline Mandarin score (units of SD)[a]	.64***	.49***		Yes
	(.03)	(.03)		
Baseline math score (units of SD)[a]		Yes	.61***	.50***
			(.02)	(.03)
Student characteristics[b]		Yes		Yes
School characteristics[c]		Yes		Yes
R^2	.48	.53	.39	.44

Source. Authors' survey.

Note. Dependent variable: standardized post-CAL (computer assisted learning) test score. Robust standard errors clustered at the school level in parentheses. Each column reports the results of one regression of the student final standardized Mandarin or math test scores over the program period on the row variables. N = 1,717.

[a] Score on the standardized math/Mandarin test that was given to all students in the sample before the CAL program.

[b] Includes 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).

[c] Includes number of students, teacher-student ratio, proportion of female teachers, school facility index (as described in table 1), and proportion of students receiving poverty subsidy.

* Significant at 10%.

** Significant at 5%.

*** Significant at 1%.

at the 10% level using the basic model (eq. [1]; table 3, col. 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 (eq. [2]; table 3, col. 2).

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, cols. 3 and 4). Specifically, compared to their counterparts in the control schools, grade 3 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.

Although the measured effect on math scores is larger than the measured effect on language scores, it is important to note that our findings should not be interpreted to mean that the program helped students to learn more math than Mandarin. The main reason for this is that, in general, these two subjects almost certainly have different levels of sensitivity to program intervention. In other words, in general, an increase of 0.1 standard deviations in math scores from some program intervention is typically thought to be easier to achieve than an increase of 0.1 standard deviations in Mandarin scores. The literature in education supports this view (e.g., Klein et al. 2000). Specifically, Rich (2013) has documented that student math performance is often more sensitive to external factors such as teacher effect and educational interventions than student performance in language.

There might be several reasons for this 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 above, it may be that the remedial tutoring that improved the Mandarin skills of children who are not proficient in the LWC had a positive effect on other subjects that were being taught using the LWC. Specifically, all school subjects were taught in Mandarin, and the textbooks were written in Mandarin. Therefore, improved Mandarin language skills via CAL might also have helped the students better understand course materials in math. According to the national uniform math curriculum, Chinese primary schools begin aggressively introducing arithmetic word problems in grade 3. Indeed, the grade 3 math textbooks that were being used by the students in our sample schools contained tens of thousands of words and more than 1,000 unique Chinese characters. Reflecting this, the word count of the math test of the final evaluation is more than 1,300 words. Therefore, it is entirely plausible that as a result of improved reading comprehension ability via CAL intervention, students in the treatment group outperformed those in the control group.

He et al. (2008) also found that an Indian English Education Curriculum implemented via teacher training improved student math and English scores 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). As discussed above, 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 were almost no remedial tutoring sessions offered after school.

Of course, we cannot completely exclude the possibility that the improvement in the math performance (and also in the Mandarin subject) of the students was the result of some other factors instead of due to improved language skills. For example, it is possible that the program might make students more interested in learning all subjects, thus improving their performance in math. Moreover, having an extra teacher (in this case, the teacher-supervisor of the CAL sessions) might also have some effect on student performance. However, as mentioned above, the CAL program was carefully designed and implemented to minimize interactions between the teacher-supervisors and students. On the basis of our data, we believe these efforts were effective. During both the baseline and end line surveys, we asked students about their interest in studying (i.e., the indicator of "like school") and found no significant increase in the treatment group compared to the control group (table 4, col. 3). As for the extra teacher effect, there also is little reason to believe that the schools intentionally assigned high-quality teachers as the CAL session supervisors. Moreover, in the literature, according to our extensive search, there is no conclusive empirical evidence that having an extra teacher (regardless of quality) could systematically improve student performance. For these reasons we argue that while we cannot rule out the possibility of confounding influences mentioned above, given the information from our data and in-the-field observations, we believe such effects, if there are any, only had a narrow influence on the validity of our empirical results.

Effects of the CAL Intervention on Student Nonacademic Outcomes

Table 4 reports the estimated treatment effects of the CAL program on student nonacademic outcomes, with each column reporting estimated coefficients from one regression of the outcome variable (as the column title) on explanatory variables in the corresponding rows. Columns 1–3 report the estimated coefficients from regressions of the student scores on self-efficacy in Mandarin study, self-confidence, and how much students like schooling, respectively, on the treatment dummy variable. All regressions controlled for the student baseline values of the nonacademic outcomes as well as a set of control covariates including student and school characteristics. The top row of this table reports the estimated CAL treatment effects for each nonacademic outcome.

TABLE 4
ORDINARY LEAST SQUARES ESTIMATORS OF THE IMPACTS OF CAL PROGRAM ON STUDENT NONACADEMIC OUTCOMES

	Self-Efficacy of Studying Mandarin (1)	Self-Confidence (2)	Like School (3)
Treatment (1 = treatment group; 0 = control group)	.07*	.02	.37
	(.04)	(.04)	(1.46)
Baseline values of the nonacademic outcome indicator and other control variables (student and school characteristics)	Yes	Yes	Yes
R^2	.15	.12	.05

Source. Authors' survey.
Note. Dependent variables: indicators of nonacademic outcomes. Baseline math/Mandarin score is the score on the standardized math/Mandarin test that was given to all students in the sample before the computer assisted learning (CAL) program. Other control variables are 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 1), and proportion of students receiving poverty subsidy. Each column reports the results of one regression of the student nonacademic outcome on the corresponding row variables. Robust standard errors clustered at the school level in parentheses. $N = 1{,}715$.
* Significant at 10%.

Our results show that the CAL intervention had positive effects on the student nonacademic outcomes, although they are not statistically significant in most of the cases (table 4). Compared to students in the control group, the students in the treatment group like school more and had higher levels of self-efficacy of studying Mandarin and self-confidence. In particular, students in the treatment group had significantly higher levels of self-efficacy of studying Mandarin than those in the control group (col. 1), yet the differences between the treatment and control groups in the other two indicators were insignificant (cols. 2 and 3).

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

Apart from the average impact of the CAL program on student academic and nonacademic performance, it is also important to explore whether the CAL program has different impacts for different students. In particular, as this CAL language program is remedial in nature (helping students with difficulties in LWC), 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 Mandarin and math scores).

The main purpose of the heterogeneity analysis is to detect the existence of heterogeneous program effects. For this reason we believe that a linear function form is enough for detecting the heterogeneous effect and describing its general pattern. All the same, we explored alternative specifications for heterogeneity analysis. Specifically, we plot the change in the student Mandarin standardized test scores over the program period for each group (treatment group and control group) versus the baseline Mandarin standardized test scores (fig. 2). According to our findings, the difference in the change in test scores over the program period between the treatment and control groups decreases monotonically as the baseline Mandarin test score goes up. Therefore, we conclude that the linear specification is a reasonable choice in our case.

The results of the heterogeneous CAL treatment effects across student academic background are summarized in table 5, with each column reporting estimated coefficients from one regression of the outcome variable (as the column title) on explanatory variables in the corresponding rows. Column 1 reports the estimated coefficients from regressions of the standardized student Mandarin scores on the treatment dummy variable, student baseline Mandarin test scores, and the interactions of these two variables. Column 2 reports the results of the same regression but uses student baseline math test scores instead of Mandarin test scores on the right-hand side. Columns 3 and 4 follow the same models as columns 1 and 2, respectively, except we change the outcome from Mandarin test scores to student self-efficacy of Mandarin study. Like table 4, all regressions controlled for a set of student and school characteristics. Table 5 reports how the estimated treatment effects of each model vary across the distribution of the test scores. It also reports the *q*-value FDR (false discovery rate, or probability of falsely reject the null hypothesis) of each test of heterogeneous effect.

The results show that the CAL intervention benefited low-performing students more than it did high-performing students (table 5, interactions). Compared to the students in the control group, students in the treatment group with a baseline Mandarin test score 1 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; col. 1). Compared to the students in the control group, students in the treatment group with a baseline math test score 1 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; col. 2). This finding indicates that CAL has been successful in achieving its goal of helping low-performing

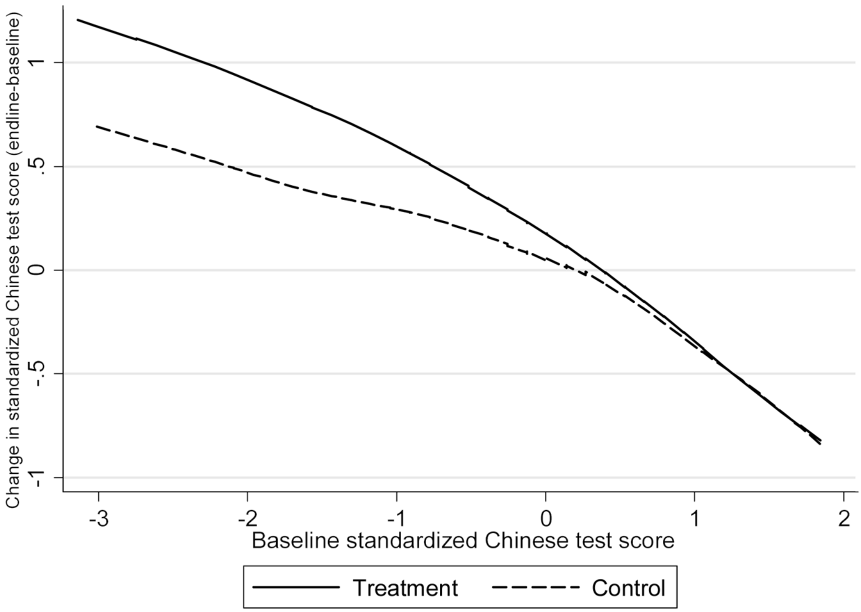

Figure 2. Change in student standardized Mandarin test score during the program period versus student baseline standardized Mandarin test score.

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).[5]

Using the same estimation model we did with the student academic outcome, we found that low-performing students benefited more than other students from the CAL intervention not only in academic performance but also in some nonacademic 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

[5] 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 may be 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.

TABLE 5

ORDINARY LEAST SQUARES ESTIMATORS OF THE HETEROGENEOUS PROGRAM EFFECT ON STUDENTS WITH DIFFERENT LEVELS OF ACADEMIC PERFORMANCE

	Standardized Post-CAL Mandarin Test Score		Self-Efficacy of Studying Mandarin	
	(1)	(2)	(3)	(4)
Treatment (1 = treatment group; 0 = control group)	.21***	.19***	.08*	.07*
	(.07)	(.06)	(.04)	(.04)
Interactions:				
Treatment × baseline Mandarin score (units of SD)[a]	.16***		.06**	
	(.05)		(.03)	
Treatment × baseline math score (units of SD)[a]		.09*		.09***
		(.04)		(.03)
q-value (false discovery rate)	.06	.18	.25	.06
Control variables	Yes	Yes	Yes	Yes
R^2	.24	.23	.23	.23

Source. Authors' survey.
Note. 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 1), and proportion of students receiving poverty subsidy. Each column reports the coefficients of the treatment dummy variable and the interaction of the treatment dummy variable and the corresponding student characteristics 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 that column, controlling for all of the control variables. Robust standard errors clustered at the school level in parentheses. $N = 1,717$.
[a] Score on the standardized math/Mandarin test that was given to all students in the sample before the computer assisted learning (CAL) program.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

self-efficacy of studying Mandarin than other students in the treatment group over the program period (table 5, cols. 3 and 4).

It is possible that while testing heterogeneous effects across a large number of characteristics (11 student characteristics), we might by chance find that the academic background characteristics are significant. To ensure integrity of our analysis, we conducted an additional FDR calculation to examine whether the observed heterogeneous treatment effects across student academic backgrounds were genuine. For controlling the FDR, we used the sharpened two-stage *q*-values that were introduced in Benjamini, Krieger, and Yekutieli (2006). We find that at least for the heterogeneous effect of CAL treatment on student Mandarin test scores across student baseline Mandarin test scores and the heterogeneous effect of CAL treatment on student self-efficacy of studying Mandarin across student baseline math test scores, the *q*-values (the smallest FDR, or the probability of falsely rejecting the null hypothesis) was only 0.06 (table 5, cols. 1 and 4). In other words, the heterogeneous effects were significant when controlling FDR = 0.06. This indicates that the observed heterogeneous CAL treatment effects across student academic backgrounds (measured by the student baseline Mandarin test scores and math test scores) are unlikely to have just happened by chance.

Conclusion

On the basis of a randomized experiment using a computer assisted learning (CAL) program, we examined the impact of the language of wider communication (LWC) learning on overall academic performance (as well as other outcomes). The sample of students in our study mostly have limited competency in the LWC. The study involves more than 1,700 third-grade students, age 9–11, in 57 public schools in poor rural minority areas in Haidong Prefecture, Qinghai. To evaluate the effectiveness of the LWC learning program, we randomly chose 26 schools from the sample as treatment schools. As part of the treatment the third-grade students attending these schools received the Mandarin after-school remedial program. The remaining 31 schools served as control schools. The program was tailored to the regular school Mandarin curriculum and was remedial in nature.

Our results indicate that improving the LWC skills of ethnic minority students has significant positive effects on student language and math test scores as well as on nonacademic outcomes. Two 40-minute Mandarin sessions per week increased student standardized Mandarin scores by 0.14–0.20 standard deviations. In general, low-performing students benefited more from the program. Perhaps more importantly, the intervention that was focused fully on teaching the LWC also had a 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. The intervention appeared to have positive impacts on the nonacademic outcomes of the students, although impacts were not statistically significant in most cases. So, in short, the study highlights the importance of efforts in improving student LWC skills to the academic performance of students with limited LWC competency. Specifically, our study provides rigorous evidence that a CAL program is a quick and effective means to improve student LWC skills where educational resources are scarce and that improved LWC competency results in better academic performance in other subjects (in this case, math).

The findings of our study are important for several reasons. First, the results are important because ours is one of the first studies in the context of a developing country to show this relationship (that LWC skills contribute to overall academic performance). Second, our findings can be interpreted as being truly causal—since the study was done in the context of large-scale randomized controlled trial. Third, our results show that the positive relationship between efforts in improving LWC skills and overall academic performance holds when the LWC is not English; indeed this is a study of the effect of teaching Mandarin on math test scores, which is outside the usual environment (teaching Spanish-speaking Latino children English in the United States) in which these studies have been conducted in the past. Overall, our study shows that a CAL-based LWC skills program has the potential to be a relatively quick and effective remedy to the underperformance of disadvantaged students in developing countries with limited LWC competency.

References

Alba, R. D., J. Handl, and W. Müller. 1994. "Ethnische Ungleichheit im Deutschen Bildungssystem." *KölnerZeitschrift für Soziologie und Sozialpsychologie* 46:209–37.

Aud, S., W. Hussar, G. Kena, K. Bianco, L. Frohlich, J. Kemp, and K. Tahan. 2011. *The Condition of Education 2011*. NCES 2011-033. Washington, DC: National Center for Education Statistics.

August, D., and T. Shanahan, eds. 2006. *Developing Literacy in Second-Language Learners*. Mahwah, NJ: Erlbaum.

Bamgbose, A. 1984. "Mother-Tongue Medium and Scholastic Attainment in Nigeria." *Prospects* 14, no. 1:87–93.

Banerjee, A., S. Cole, E. Duflo, and L. Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122, no. 3:1235–64.

Barrow, L., L. Markman, and C. Rouse. 2008. "Technology's Edge: The Educational Benefits of Computer-Aided Instruction." NBER Working Paper no. 14240, National Bureau of Economic Research, Cambridge, MA.

Benjamini, Y., A. M. Krieger, and D. Yekutieli. 2006. "Adaptive Linear Step-Up Procedures That Control the False Discovery Rate." *Biometrika* 93, no. 3:491–507.

Cheung, A., and E. Slavin. 2005. "Effective Reading Programs for English Language Learners and Other Language-Minority Students." *Bilingual Research Journal* 29, no. 2:241–67.

Genesee, F., K. Lindholm-Leary, K. Saunders, and D. Christian. 2006. *Educating English Language Learners: A Synthesis of Research Evidence*. New York: Cambridge University Press.

Glewwe, P., and M. Kremer. 2006. "Schools, Teachers, and Education Outcomes in Developing Countries." In *Handbook on the Economics of Education*, ed. E. Hanushek and F. Welch, 945–1017. New York: Elsevier.

He, F., L. Linden, and M. MacLeod. 2008. "How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program." Unpublished manuscript, Department of Economics and School of International and Public Affairs, Columbia University.

IDRC (International Development Research Center). 1997. *Languages of Instruction, Policy Implications for Education in Africa*. Ottawa: International Development Research Center.

Klein, S., L. Hamilton, D. McCaffrey, and B. Stecher. 2000. "What Do Test Scores in Texas Tell Us?" RAND Issue Paper no. 202, RAND Corporation, Santa Monica, CA.

Lai, F., R. Luo, L. Zhang, X. Huang, and S. Rozelle. 2014. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing." REAP Working Paper no. 228, Rural Education Action Program, Stanford, CA.

Lai, F., L. Zhang, X. Hu, Q. Qu, Y. Shi, Y. Qiao, M. Boswell, and S. Rozelle. 2013. "Computer Assisted Learning as Extracurricular Tutor? Evidence from a Randomised Experiment in Rural Boarding Schools in Shaanxi." *Journal of Development Effectiveness* 5, no. 2:208–31.

Master, B., S. Loeb, and J. Wyckoff. 2013. "Learning That Lasts: Unpacking Variation in Teachers' Effects on Students' Long-Term Knowledge." Unpublished manuscript, Center for Education Policy Analysis, Stanford University.

Mo, D., L. Zhang, R. Luo, Q. Qu, W. Huang, J. Wang, Y. Qiao, M. Boswell, and S. Rozelle. 2014. "Integrating Computer Assisted Learning into a Regular Curriculum: Evidence from a Randomized Experiment in Rural Schools in Shaanxi." *Journal of Development Effectiveness* 6, no. 3:300–323.

Planty, M., W. Hussar, T. Snyder, S. Provasnik, G. Kena, R. Dinkes, A. Kewal Ramani, and J. Kemp. 2008. *The Condition of Education 2008*. Washington, DC: National Center for Education Statistics.

Rich, M. 2013. "In Raising Scores, 123 Is Easier than ABC." *New York Times*, May 29.

Slavin, R., C. Lake, B. Chambers, A. Cheung, and S. Davis. 2009. "Effective Reading Programs for the Elementary Grades: A Best-Evidence Synthesis." *Review of Educational Research.* 79, no. 4:1391–1466.

Slavin, R., N. Madden, M. Calderón, A. Chamberlain, and M. Hennessy. 2011. "Reading and Language Outcomes of a Multiyear Randomized Evaluation of Transitional Bilingual Education." *Educational Evaluation and Policy Analysis* 33, no. 1:47–58.

World Bank. 2001. *China: Challenges of Secondary Education*. 2nd ed. Washington, DC: World Bank.

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

Yang, Y., H. Wang, L. Zhang, S. Sylvia, R. Luo, Y. Shi, W. Wang, and S. Rozelle. 2013. "The Han-Minority Achievement Gap, Language and Returns to Schools in Rural China." REAP Working Paper no. 258, Rural Education Action Program, Stanford, CA.

Yang, Y., L. Zhang, J. Zeng, X. Pang, F. Lai, and S. Rozelle. 2013. "Computers and the Academic Performance of Elementary School-Aged Girls in China's Poor Communities." *Computers and Education* 60:335–46.

Yates, R. 1995. "Functional Literacy and the Language Question." *International Journal of Educational Development* 15, no. 4:437–47.