
Journal of Development Effectiveness

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjde20

# Improving learning by improving vision: evidence from two randomized controlled trials of providing vision care in China

**Xiaochen Ma, Huan Wang, Yaojiang Shi, Sean Sylvia, Lei Wang, Yiwei Qian & Scott Rozelle**

**To cite this article:** Xiaochen Ma, Huan Wang, Yaojiang Shi, Sean Sylvia, Lei Wang, Yiwei Qian & Scott Rozelle (2021): Improving learning by improving vision: evidence from two randomized controlled trials of providing vision care in China, Journal of Development Effectiveness, DOI: 10.1080/19439342.2021.1876139

**To link to this article:** https://doi.org/10.1080/19439342.2021.1876139

Published online: 26 Feb 2021.

JOURNAL OF DEVELOPMENT EFFECTIVENESS
https://doi.org/10.1080/19439342.2021.1876139

Routledge
Taylor & Francis Group

ARTICLE



# Improving learning by improving vision: evidence from two randomized controlled trials of providing vision care in China

Xiaochen Ma[a], Huan Wang[b], Yaojiang Shi[c], Sean Sylvia[d], Lei Wang[c], Yiwei Qian[e] and Scott Rozelle[b]

[a]China Center for Health Development Studies, Peking University, Beijing, China; [b]Stanford Center on China's Economy and Institutions, Stanford University, Palo Alto, CA, USA; [c]Center for Experimental Economics in Education, Shaanxi Normal University, Xi'an, Shaanxi, China; [d]Department of Health Policy and Management, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; [e]Department of Economics, University of Southern California, Los Angeles, CA, USA

**ABSTRACT**

This paper examines the external validity of health intervention by comparing the impacts of providing free eyeglasses on the educational performance of nearsighted children in two settings: rural public schools in Western China and urban private migrant schools in Eastern China. The intervention significantly improves educational outcomes by 0.14 standard deviations in math in rural public schools but not in private migrant schools. The difference in measured impacts is due in part to lower quality schooling in migrant schools in Eastern China. Our findings show that only when school is providing a quality education, health interventions might increase student learnings.

**ARTICLE HISTORY**
Received 17 August 2017
Accepted 10 January 2021

**KEYWORDS**
Field Experiment; health Intervention; educational Economics

**JEL CODES:**
C93; I12; I21; O15

## Introduction

Researchers have been increasingly interested in understanding demand-side constraints to student learning (e.g., Ganimian and Murnane 2016). In this regard, economists have begun to evaluate, as one of the most high-profile approaches to alleviate demand-side constraints, the impacts of certain health and nutrition interventions, such as deworming programmes (Miguel and Kremer 2004), the provision of micronutrient supplements (Maluccio et al. 2009), and nutritious school meals (Vermeersch and Kremer 2005). Despite the frequently cited positive impacts of the interventions on children's health and nutrition, the evidence is mixed when examining the impacts of these interventions on the learning outcomes of students (Evans and Popova 2016; Krishnaratne et al. 2013; McEwan 2015).

Mixed evidence from impact evaluations of similar interventions in differing contexts raises a challenge for policymakers. When studies are run in a single venue or subset of schools, a common criticism is that these studies lack external validity and that the sample is not representative of a nation, region, or sector of an economy (Barrett and Carter 2010; Deaton 2010). As a result, many authors argue for the specificity of the study context. Some argue that the same type of intervention can elicit different results in different contexts due to the strong role of complementarities in specific education and health systems (Culter and Lleras-Muney 2014; Glewwe 2013) and the need for entire systems to improve. Interventions that had large effects in one place may end up producing no or, at the most, small improvements in others if the overall system in the setting of the current intervention is broken (Acemoglu 2010).

**CONTACT** Yaojiang Shi shiyaojiang7@gmail.com Center for Experimental Economics in Education, Shaanxi Normal University, Xi'an, Shaanxi, China

In response to the criticism that impact evaluations lack external validity (Pritchett and Sandefur 2015a, 2015b), new studies have begun to focus on addressing the issue of how the results of the interventions that are being tested are applicable to populations in other contexts. As a part of this movement, there is a small and growing empirical literature that seeks to replicate similar interventions in different contexts. Bauchet et al. (2015) study a replication of an anti-poverty program in South India and find no net impact even though significant effects were found in Bangladesh and other sites due to differences in the nature of local labour markets. In contrast, Banerjee et al. (2015) reviewed six randomised controlled trials (RCTs) to examine the impact of a similar anti-poverty program and found that impacts were not sensitive to context.

Nevertheless, studies that reference the external context of the intervention are more the exception than the rule. After reviewing 92 RCTs published between 2009 and 2014 in leading economics journals, Peters et al. (2016) concluded that most papers that used RCT methodologies did not discuss the issue of external validity and did not provide the information that was necessary to assess potential problems in upscaling to other localities.

The goals of this paper are twofold. First, we measure whether there is a causal impact of vision care of nearsighted students on their educational performance in two different primary school settings. Second, we explain why such a program might have different impacts in different school contexts. In other words, we present evidence about whether a vision care program works as well as evidence about where or when such a vision care program might or might not work.

To achieve our goals, the paper has four main sections. In the first section, we describe the two RCT research settings – one setting in which the public primary schools in a rural area of western provinces generally deliver quality education (Western China program) and the other setting in which the private migrant primary schools in migrant communities of China's coastal urban area deliver poor, low-quality education (Eastern China program). In the second section, we describe the data and analytical approach in which we intervene with a single, standardised vision care treatment. Briefly, in 2012, in a set of 84 rural primary public schools, we implemented an RCT in which we gave students in the 42 treatment schools high-quality vision screenings and refraction, produced glasses, and dispensed them without charge. In a randomly assigned set of 42 control schools, we did not give the students free glasses. In 2013, the same treatment that used the same set of protocols was replicated in a second setting, a set of 94 private migrant primary schools (47 treatment and 47 control schools) by the same research team.

In the third section of the paper, we report the results of these two RCTs, which are implemented in the two different contexts. We found that 20% of the students in the Western China program and 17% in the Eastern China program were nearsighted. Among those with poor vision at baseline, however, only 18% in both study regions were using eyeglasses to correct their vision. In the intention-to-treat (ITT) analysis in the Western China program, providing free eyeglasses and training improved maths test scores by 0.14 standard deviations relative to students in control schools. In contrast, there was no significant impact of the treatment in the Eastern China program.

In the fourth section, we examine three hypotheses that concern why the impact of our provision of free eyeglasses differed between the Western China program and the Eastern China program schools. Differences in compliance rates (Hypothesis 1) cannot account for the differences observed in treatment effects. In fact, the share of students who actually wore the glasses they received in the Eastern China program was significantly higher than that in the Western China program. In addition, we used our data to show that there are no systematic differences in the (at least) observable characteristics of individuals and their families who favour educational performance (Hypothesis 2). After examining differences in the schools in the two settings (that is, the differences in public rural schools and private migrant schools), we show that the differences in measured impacts across the two regions are mainly a matter of the difference in the characteristics of the teachers and schools in the two samples (Hypothesis 3). Overall, our results suggest that vision care programs that are implemented in schools in which teachers are in an environment that is committed to providing quality education, the programs can have a significant impact on learning. Even more generally, our study, in which we replicate an RCT in two regions of China with two different schooling environments, clearly demonstrates that context matters.

## Research setting

### *Average levels of income in the sample regions*

The first program (Western China) was conducted in rural areas of two provinces in Western China: Gansu and Shaanxi. Gansu's GDP per capita of USD 3,976 was ranked the third poorest among China's 31 provincial administrative regions, while Shaanxi's GDP per capita of USD 6,108 was ranked 13th and was similar to that for the country as a whole (USD 6,969) (China National Bureau of Statistics 2008–2014). Despite the higher ranking, Shaanxi still houses a sizeable number of households that live in poverty. Shaanxi has one of the highest numbers of nationally designated poor counties.[1]

The second program (Eastern China) was conducted in migrant communities in Jiangsu Province and Shanghai Municipality, two of the richest areas in China. In 2013, Shanghai's GDP per capita was USD 15,696, and Jiangsu's GDP per capita was USD 12,155.

Provincial yearbooks, however, provide only average income per capita for an entire province. In addition, most of the information in provincial yearbooks concerns the measurement of attributes of those individuals who are formal residents (and not temporary migrants) of the province. Hence, average income figures at the provincial level can disguise the relative poverty of the families in our sample.

Indeed, most of the areas in our sample are dominated by poor households. In Gansu and Shaanxi, the students in most of our schools are from poor rural families. Of the 18 counties in the Western China program sample, 15 are nationally designated poor counties. All of our sample schools are in rural communities. Thus, the average income of most families in our sample is near the poverty line.

Despite living in one of the wealthiest regions in China, the households in the Eastern China program are in many ways more similar to the surveyed families in our Western China sample areas. All students in the Eastern China program region are from families who live in migrant communities. These migrant families are part of a large movement of people from rural to urban areas that has increased dramatically over the past three decades (China National Development and Reform Commission 2015).

Although labour has flowed relatively freely from agriculture to industry, the process of shifting lives, homes, and families has been more difficult. A core aspect of the challenges that migrants face stems from China's *hukou* household registration system, which classifies China's citizens as either rural or urban residents. Without an urban *hukou*, migrants and their families have limited access to urban public services, including housing, health care, social security, and, above all, education for their children. As a result, as Li et al. (2014) argue, although individuals in these migrant communities live in some of the wealthiest regions of China, the families and the schools that serve them more closely resemble communities, families, and schools in underdeveloped rural areas. As discussed in the subsection below, the data show that children and families who live in migrant communities are similar in many observed characteristics to those who live in rural communities.

### *Schools in the sample regions*

In China today, public schools in both rural and urban areas are required to provide free education to children, according to the compulsory education law. This free education, however, is guaranteed only for children whose *hukou* matches the school's location. In the case of migrants, if there are not enough slots in urban public schools, the children of migrant families have no choice but to attend private migrant schools.

In rural areas of Western China, nearly all children attend the nearest public primary school, which is typically located in their township's urban centre (or a nearby large village). Although rural schools are located in the parts of China that are poorer and more remote, they are supported by public funds and are subject to monitoring by national and provincial government/education officials. All rural public schools also are under the direct management and supervision of county-level bureaus of education.

Due to the combined input of upper-level governments (province, prefecture, and county) and the national government, the basic ingredients of a sound school system are present in most rural public schools. The salaries of teachers have been paid by the national government since 2005 and were raised to levels of the incomes of civil servants in 2009 (An, 2018). The curriculum for rural schools were revised and updated in 2010. The national government and the regional and local counterpart government agencies have invested billions into school facilities (Yiu and Adams 2013). Rural public schools are assessed annually by the national government, and the results of these assessments are used to identify (and often address) problems in the school system (Yiu and Adams 2013). In short, although rural areas are poor, schools deliver a basic and consistent level of education.

In migrant communities of Eastern China, because migrant children in cities still retain their rural *hukous*, most are still unable to attend public schools. Instead, migrant students often attend privately-run, tuition-funded, for-profit migrant schools that have little public funding or regulation (Lai et al. 2014). Migrant schools also have been known to be shut down without notice (Chen et al. 2015). Teachers in migrant schools often are not able to meet the same quality standards as required by the public system, in part because their salaries are low and turnover is high (Kwong 2004). Therefore, education in migrant communities is difficult to provide and often not consistent or stable (Lai et al. 2014; Wang et al. 2017).

***Comparison of Individual, Family, Teacher, and School Characteristics between Eastern and Western China School Settings***

In this subsection, we describe differences in individual, family, teacher, and school characteristics in the two RCT research settings. If we implement an identical experiment in the two areas, it might be possible to identify the source of observed differences in the impact of the vision care treatment. This can be accomplished through an examination of differences in the nature (and quality) of public rural schools in the Western China sample and the private migrant schools in the Eastern China sample as well as among the children who attend the schools.

### *Individual and family characteristics*

We find that the students in the two regions are similar in terms of most individual and family baseline characteristics (Table 1). Specifically, the average age in both regions is 11 years, with a standard deviation of 0.9 years, and about half of the students in both regions are boys. The students also are almost identical in terms of the share of wearing eyeglasses at the baseline, at 18% in the Western China program and 18% in the Eastern China program.

Although most of the characteristics are the same, there are four exceptions. First, the students in the migrant schools in the Eastern China program are about 0.575 dioptres (D) in spherical equivalent less myopic than are their counterparts in the Western China program. The difference is statistically significant at the 1% level and should be interpreted as clinically significant as well (Elliott and Howell-Duffy 2015).

Second, students in the Eastern China program had lower baseline maths test scores than did their counterparts in the Western China program. When we pool samples of both programs and standardise the baseline test scores, using the entire distribution, we find that students in the Eastern China program were about 0.40 standard deviations lower than students in the Western China program ($p$-value <0.01). Although one might worry that this difference is a factor that could affect the nature of the comparison, it also could be symptom of the quality of schools, as discussed below.

**Table 1.** Baseline characteristics.

| | | Western China Program | Eastern China Program | Difference | |
|---|---|---|---|---|---|
| | | | | (2)–(1) | *n* |
| Variable | | (1) | (2) | (3) | (4) |
| (1) | Age (years) | 10.884 | 10.943 | 0.059 | 1,435 |
| | | [0.942] | [0.921] | (0.082) | |
| (2) | Gender (1 = male) | 0.515 | 0.503 | −0.013 | 1,437 |
| | | [0.500] | [0.500] | (0.025) | |
| (3) | Severity of myopia (refractive power) | −1.617 | −1.042 | 0.575*** | 1,436 |
| | | [1.813] | [1.937] | (0.117) | |
| (4) | Wear eyeglasses at baseline (1 = yes) | 0.183 | 0.178 | −0.005 | 1,414 |
| | | [0.387] | [0.383] | (0.025) | |
| (5) | Baseline maths score | 0.376 | −0.020 | −0.396*** | 1,436 |
| | | [0.959] | [1.007] | (0.087) | |
| (6) | Parent with high school education (1 = yes) | 0.201 | 0.309 | 0.108*** | 1,397 |
| | | [0.401] | [0.463] | (0.026) | |
| (7) | Household asset value (1,000 RMB) | 29.507 | 36.136 | 6.629** | 1,353 |
| | | [34.406] | [32.792] | (3.000) | |
| (8) | Teacher experience less than 3 yrs (1 = yes) | 0.241 | 0.355 | 0.110 | 176 |
| | | [0.430] | [0.481] | (0.069) | |
| (9) | Teacher without rank (1 = yes) | 0.000 | 0.404 | 0.404*** | 176 |
| | | [0.000] | [0.493] | (0.054) | |
| (10) | Recruitment of contract teachers (1 = yes) | 0.675 | 0.926 | 0.251*** | 176 |
| | | [0.471] | [0.264] | (0.057) | |
| (11) | School history less than 10 years (1 = yes) | 0.108 | 0.484 | 0.375*** | 176 |
| | | 0.313 | 0.502 | (0.064) | |

*Note*. Rows 1–7 present individual-level results. Rows 8–11 present school-level results. Columns 1 and 2 show means with standard deviations reported in brackets. Column 3 shows coefficients estimated by regressing each of the baseline characteristics on the program dummy, with standard errors clustered at the school level in parentheses (except the last four variables, which use school-level data). Severity of myopia is refractive power as defined by spherical equivalent. Baseline maths score are based on the standardised maths score pooling data from two programs at the baseline. Household assets value are calculated by summing the value, as reported in the China Rural Household Survey Yearbook (Department of Rural Surveys, National Bureau of Statistics of China 2013), of items on a list of 13 owned by the family.

*$p < .10$, **$p < .05$, ***$p < .01$

Third, parents in the Eastern China program are slightly more educated than are their counterparts in the Western China program. In the Eastern China program, the share of parents with at least a high school education (31%) is 11 percentage points higher than the parents of the students in the Western China program (20%; *p*-value <0.01).

Fourth, the families in the Eastern China program, as would be expected, given their status as migrant workers (versus a family who lives in a rural village), are wealthier than are their counterparts in the Western China program. The average value of a household's assets in the Eastern China program (36,000 RMB) is about 7,000 RMB (1,000 USD) higher than that in the Western China program (29,000 RMB).

The above comparisons of the students and families in the two samples (at baseline) provide the first indication of the poor quality of education in the Eastern China program site. It is well known in the education literature that one of the main determinants of student learning outcomes is the education level of the parents (Haveman and Wolfe 1995). If this is true in China (which, indeed, it is; Brown and Park 2002), we would expect the learning outcomes of the students in the Eastern China program to be higher than those of students in the Western China program. At baseline, however, the results are exactly the opposite: Students in the Western China program outperformed students in the Eastern China program. One interpretation of these findings is that the quality of education in the Eastern China program schools is lower than that in the Western China program schools.

### *Teacher and school characteristics*

We find the quality of private migrant schools in the Eastern China program appears to be poorer than that of rural public schools in the Western China program. This difference in quality appears when we examine four variables.

First, schools in the Eastern China program appear to have a higher share of inexperienced teachers, an indication of lower quality of teaching. About 24% of teachers in the Western China program have less than three years of teaching experience; the comparable number in the Eastern China program is 36% (or 50% higher in relative terms). Note that, although the difference in teaching experience is relatively large, it is not statistically significant (in part, perhaps, due to the lower statistical power in distinguishing differences between schools at the school level, as the comparison of teacher and school characteristics uses school-aggregated instead of individual data, as do the student and family characteristics comparisons).

Second, schools in the Eastern China program have a higher share of teachers who do not have any formal 'rank.' In China, teacher rank is a credential based on an annual assessment of the teacher by local school administrators. Teacher rank is often used by policymakers and school administrators to screen for the quality of teachers when hiring and/or making assignment and compensation decisions (Chu et al. 2015). According to this indicator, there are large differences between the Western and Eastern China programs. Every teacher in the Western China program has a teacher rank. In the Eastern China program, however, 40% of teachers do not hold a rank ($p$-value <0.01).

Third, schools in the Eastern China program are more likely to recruit contract teachers, an indication of the shortage of qualified teachers (Wang 2002). In China, it has been shown that contract teachers are less qualified than teachers who are part of the formal teaching system (Wang 2002). Termed *daike*, or temporary teachers, contract teachers often are individuals who have been unable to become part of a permanent teaching staff and, as such, often hold only temporary positions (Robinson and Yi 2008). Wang et al. (2018) also find poor quality among contract teachers. Although 68% of schools in the Western China program recruited contract teachers, the number was even higher (93%) in the Eastern China program (or 37% higher in relative terms; $p$-value <0.01).

Fourth, schools in the Eastern China program have significantly shorter histories than do their Western China program counterparts. Studies within and outside China have documented that school-closing policies target lower-performing migrant schools (Engberg et al. 2012; Zhao and Parolin 2012). When schools are closed, school principals often move their school to another location and reopen. In addition, in rural areas, such as the schooling environment in Western China, a policy of closing low-quality schools was pursued in the early 2000s under the guidelines of the school merger policy (Mo et al. 2012). In the case of rural public schools, however, closed schools are almost never reopened. As such, it is logical to think of using the variable 'years of school history' as an indicator of the provision of quality education. If this is the case, the schools in the Eastern China program are of lower quality. Although around 1 in 10 schools in the Western China program (11%) have school histories that span less than 10 years, nearly half of the schools in the Eastern China program (48%) have school histories of less than 10 years (a difference of almost five times; $p$-value <0.01).

In summary, our data paint a picture of schools in migrant communities that is consistent with that of the literature (Lai et al. 2014; Wang et al. 2017). Hence, it is possible that the poor quality of education in the Eastern China program schools might be (one of) the reason(s) that we find different impacts when we implement the same experiment in the two research settings. If the compliance rate in the Eastern China program is at least as high as that of Western China program, and if the two experiments are implemented according to a standard protocol by the same research team (as the individuals in the two program areas were similar and, in fact, favoured the children in the Eastern China program schools), we believe that there is a possibility that the differences in measured impact stem from the differences in quality of schooling. In other words, if students in school are not in a setting that provides quality teaching/schooling, we should not expect an intervention to have the same impact as in a school where students are actually learning.

## Data and analytical approach

### Sampling

#### Western China program

Sampling for the Western China program was carried out immediately prior to the implementation of our study during the 2012–13 school year. One prefecture in each province was selected: Tianshui prefecture in Gansu province and Yulin prefecture in Shaanxi province. After choosing the prefectures, we obtained a list of all rural primary schools in the two prefectures from each of the county education bureaus in the prefecture. To eliminate potential spillovers, we randomly selected one school from each township in the sample frame. In total, 84 rural primary schools and 3,645 students were selected (Figure 1). We randomly selected one class of fifth graders (likely age range: 10–12 years).

#### Eastern China program

To ensure the comparison of the two programs, we implemented the sampling in the Eastern China program in a way that was nearly identical to that of the Western China program. Although the program was run one year later (during the 2013–14 academic school year), the Eastern China program also spanned one complete school year. We followed the same protocol of sampling as in the Western China program. All primary schools in migrant communities of one prefecture of Jiangsu province (Suzhou) and all migrant schools in Shanghai Municipality were included in the sampling frame. In total, 94 schools and 4,409 students were selected for our study (Figure 2). We randomly selected one class of fifth-graders from each school (likely age range: 10–12 years).

### Experimental design

#### Western China program

To ensure a balanced sample and to improve the power of the experimental design, we stratified the intervention assignment by location (county), school size, and eye examination results (proportion of myopic students) collected at the baseline. This yielded a total of 42 strata. Our analysis takes this randomisation procedure into account (Bruhn and McKenzie 2009). Within each stratum, one school was randomly assigned to the treatment group and the other, to the control group.

Once the randomised assignment was completed after the baseline survey and eye examination (as described below), the implementation team launched the intervention. In the treatment group, every student was screened, and nearsighted students (or students with poor vision) were identified through a two-step eye examination protocol (described in the Data Collection subsection). A prescription for glasses was produced for each student with poor vision in the treatment group. A letter that contained a description of this program and the student's prescription were sent to parents. In addition, information about vision care and the importance of wearing eyeglasses was given to students in the form of an in-class training program. Based on the prescription, a pair of free eyeglasses was produced for each myopic student. About four weeks after the baseline (October 2012), program optometrists dispensed free eyeglasses in the treatment schools.

In the control group, only the letter was given to the parents, and a prescription was given to the students. No additional training was provided to the students. The students and their families were blind to the cluster RCT.

#### Eastern China program

The intervention in the Eastern China program was designed in a way that was identical to the intervention in the Western China program. Similar to the Western China program, intervention assignment was stratified by location, school size, and eye examination results (proportion of myopic students). This yielded a total of 47 strata. Half of the schools were randomly assigned to the treatment group and the other half, to the control group. In treatment schools, a prescription and

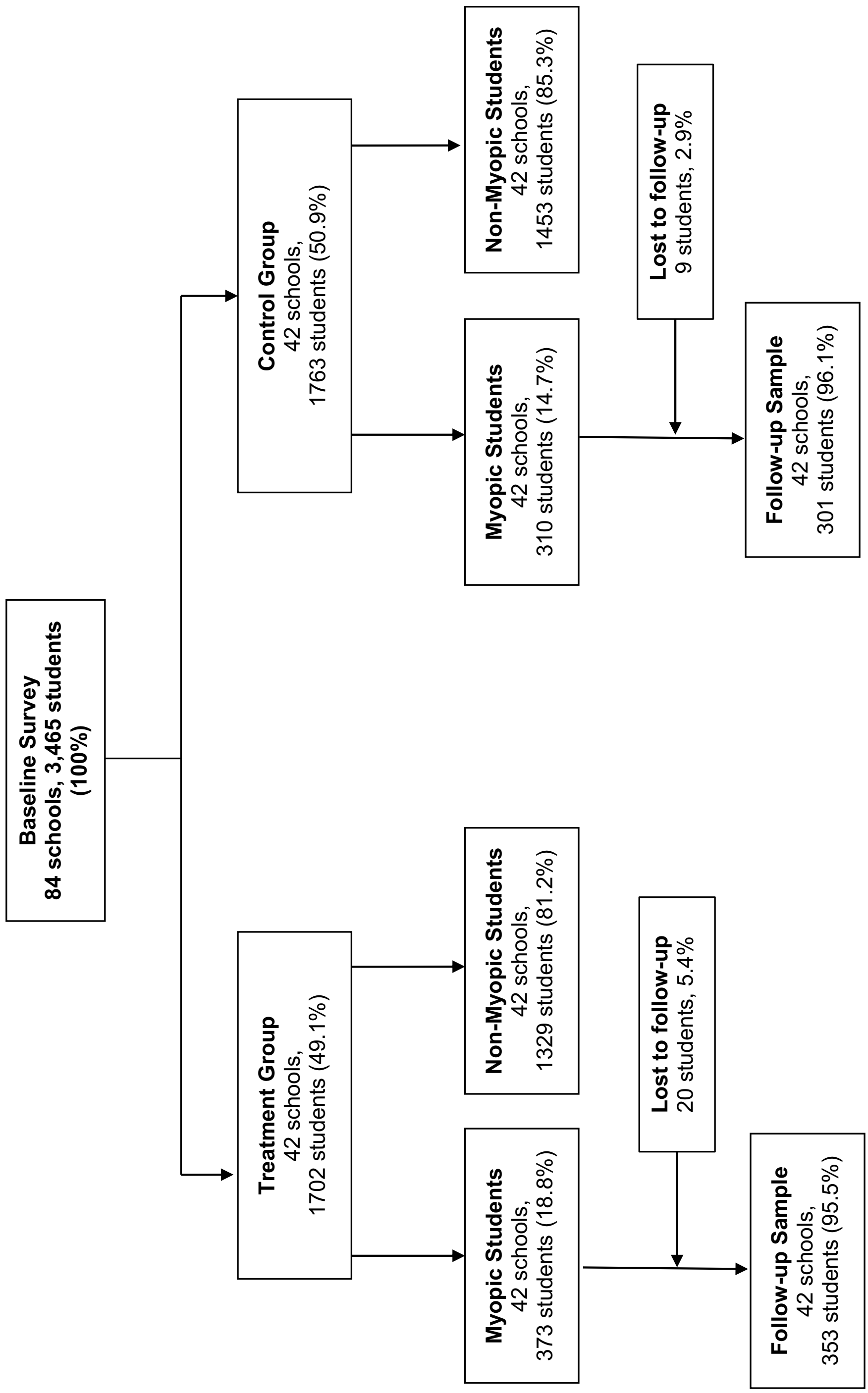


**Figure 1.** Experiment profile of western China program.

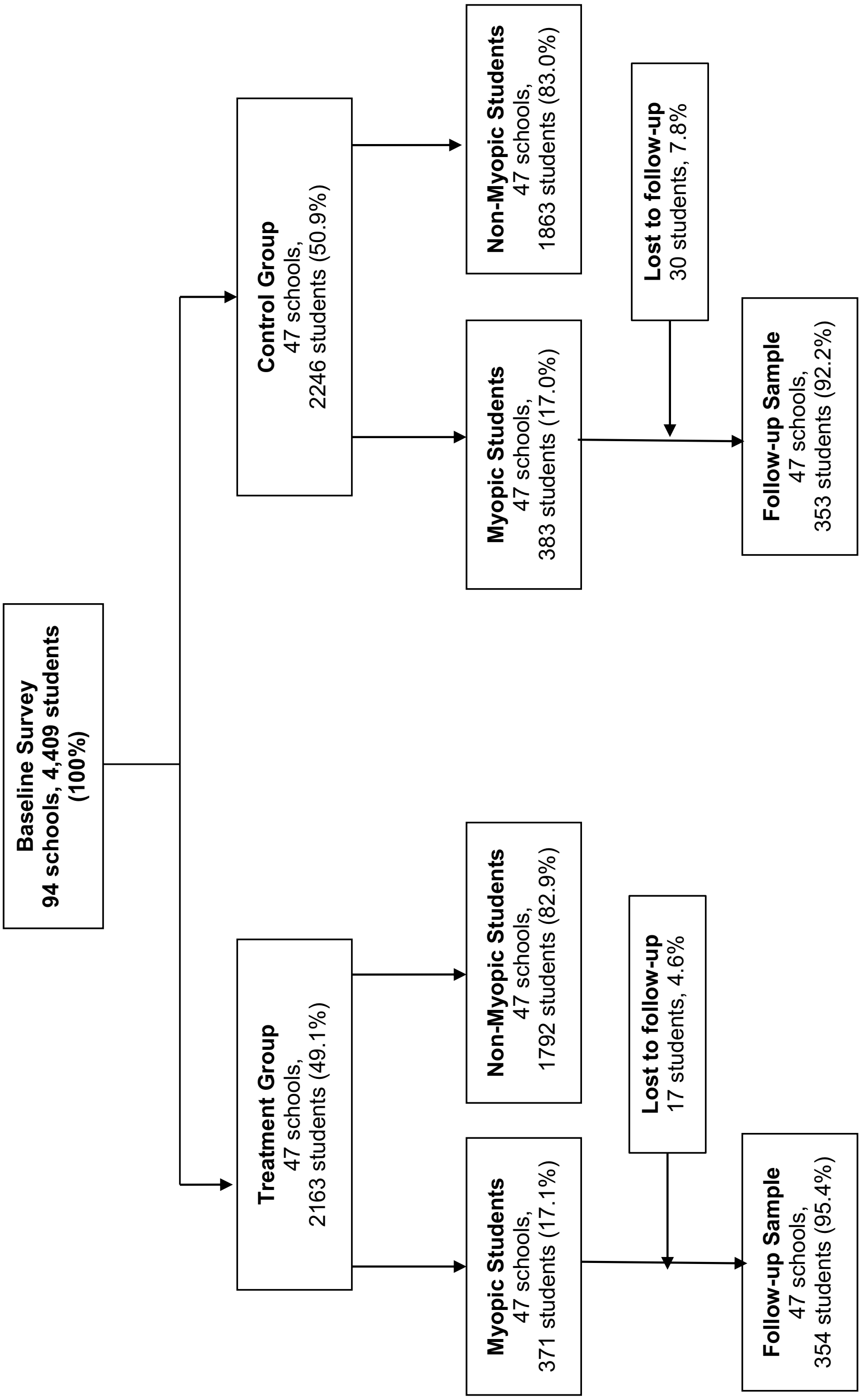


**Figure 2.** Experiment profile of eastern China program.

a description of the program were given to the parents. Students were also given in-class training about the importance of vision care and wearing eyeglasses. Based on each student's prescription, a pair of glasses was produced for each student. The eyeglasses for the Eastern China program were produced by the same manufacturing plant as were the glasses for the Western China program. The glasses were dispensed in the classroom about four weeks after baseline (similar to the timing that was followed in Western China the prior year).

The students in the control group, as in the Western China program, were given only the prescription, and a letter was given to the parents. No additional training was provided to the students. As in the Western China program, students and their families were blind to the cluster RCT.

## Data collection

### Western China program

We conducted a total of two waves of surveys: one at baseline and one at endline. Each survey was administrated to all sample students in the 84 schools. The baseline and endline surveys were conducted in September 2012 (the beginning of fall semester) and in June 2013 (at the end of the spring semester), respectively.

*Baseline survey and eye examination*. Baseline surveys were administered to all students in the classroom in two blocks. In the first block of the survey, enumerators administered questionnaires to students in regard to age, gender, eyeglasses wear (whether they wore glasses prior to the intervention), and parental educational status.

In the second block, separate mathematics tests appropriate for children in the fourth and fifth grades were administered. Local educators assisted with the selection of questions from items developed for the Trends in International Mathematics and Science Study. The examination was timed (25 minutes) and proctored by two study enumerators at each school. Mathematics was chosen for testing to reduce the effect of home learning on performance and to better focus on classroom learning. For analysis, we normalise scores, using the control group's baseline distribution.

At the same time as the school survey, a two-step eye examination was administered to all treatment and control students. The eye examination teams were trained by the Zhongshan Ophthalmic Center at Sun Yat-sen University and followed a prescribed protocol. First, a team of two trained staff administered visual acuity (VA) screenings using Early Treatment Diabetic Retinopathy Study eye charts. Students who failed the visual acuity screening test (cut-off is defined by VA of either eye less than or equal to 6/12, or 20/40) were enrolled in a second vision test (refraction). The refraction was carried out at each school one to two days after the initial eye screening. This second vision test was conducted by a team of one optometrist, one nurse, and one assistant staff and involved cycloplegic automated refraction with subjective refinement to determine prescriptions for children's eligibility for eyeglasses (cut-off for myopia is $\leq -0.75$ D).

Several supplementary data collection instruments also were administered to parents, teachers, and school principals. Parents were asked to complete a survey form. The parent questionnaire forms comprised 13 questions that concerned the ownership of selected items, such as a television, toilet, or motorcycle. The questions are used to assess and measure the value of family assets. In addition, teachers in the sample classes were asked questions about their experience in teaching and teacher rank. School principals were asked to answer questions about whether they recruit contract teachers and the number of years that the school had been operating (or the school's history).

*Endline survey*. In May 2013, approximately seven months after eyeglasses were dispensed, an endline survey, which was the same as the baseline survey, was conducted. As in the baseline survey, the enumeration team collected post-treatment maths test scores. The test was timed and proctored in the same way as was done at the baseline at each school. As in the case of the baseline, we also normalised scores, using the control group's baseline distribution.

As we were conducting the endline survey, we also conducted unannounced spot checks to collect information on eyeglasses wear or student compliance. A team of two enumerators was sent into schools in advance of the rest of the survey team to conduct classroom checks. The enumerators stood outside of the classroom and counted the number of myopic students in the sample class who were wearing eyeglasses. The enumerators were given a list of the students diagnosed with myopia to record individual-level information on their compliance rate.

### *Eastern China program*

The Eastern China program was implemented during the school year, following the program in Western China. We conducted our baseline survey and eye examination in September 2013, and the endline survey was conducted in May 2014. The time in between the distribution of eyeglasses and the endline survey for the Eastern China program was (by design) almost exactly the same as in the Western China program.

## *Balance and attrition check*

The baseline characteristics of the students, families, teachers, and schools in our sample are shown in Appendix Table 1. The table presents information on key variables that we used in the study. Overall, we find that the baseline characteristics are well balanced across the treatment and control groups in both RCTs. Column 3 shows coefficients estimated by regressing each of the baseline characteristics on the treatment dummy for the Western China program. Only one of the 12 coefficients is statistically significant at the 5% level. Column 6 shows coefficients estimated by regressing each of the baseline characteristics on the treatment dummy for the Eastern China program. Only one of the 11 coefficients is significant at the 10% level (and none, at the 5% level).

The overall attrition rate between baseline and endline was about 4% in the Western China program and 6% in the Eastern China program. Such attrition rates are considered to be low in the RCT literature (Duflo et al. 2008). In Appendix Table 2, we show that student attrition is independent of treatment assignment and baseline characteristics in each program.

### *Statistical approach*

We estimate the treatment effect, using the ITT approach to measure the overall effect of a program (regardless of compliance). We compare the endline standardised maths test scores between nearsighted students in treatment schools and nearsighted students in the control schools, controlling for baseline standardised maths test scores. The ITT analysis was estimated using ordinary least squares (OLS) regression in two specifications. First, we use an unadjusted model:

$$Y_{1i} = \beta_0 + \beta_1 Z_j + \beta_2 y_{0i} + \varepsilon_i \quad (1)$$

where $Y_{1i}$ is the endline standardised maths scores for student $i$, treatment dummy $Z_i$ takes value of 1 if the school that the student attends is assigned to the treatment group, and $y_{0i}$ is the standardised maths score at the baseline. We account for clustering of the error term $\varepsilon_i$ at the school level.

Second, we use an adjusted model to improve the efficiency of the estimation. We build on the unadjusted model in equation (1) by adding a set of controls and strata-fixed effects:

$$Y_{1i} = \beta_0 + \beta_1 Z_j + \beta_2 y_{0i} + X_i'\gamma + \varphi_s + \varepsilon_i \quad (2)$$

where all variables and parameters are the same as those in equation (1), except that we added a vector of other baseline characteristics, $X_i$ (Table 1) and strata fixed effects $\varphi_j$. As in the unadjusted model, we account for clustering of the error term $\varepsilon_i$ at the school level.

**Table 2.** Baseline vision-screening results.

| Variable | Total (1) | Control Group (2) | Treatment Group (3) | Difference (3)–(2) (4) |
|---|---|---|---|---|
| Panel A. Western China Program | | | | |
| Number of students screened | 3,465 | 1,763 | 1,702 | |
| Number of students with myopia | 683 | 310 | 373 | |
| Prevalence of myopia (%) | 19.7 | 17.6 | 21.9 | 4.3* |
| | [39.8] | [38.1] | [41.4] | (2.3) |
| Among myopic students: | | | | |
| Severity of myopia (refractive power) | −1.617 | −1.601 | −1.630 | −0.028 |
| | [1.813] | [1.719] | [1.889] | (0.158) |
| Wore eyeglasses prior to treatment (1 = yes) | 0.183 | 0.184 | 0.182 | −0.002 |
| | [0.387] | [0.388] | [0.387] | (0.038) |
| Panel B. Eastern China Program | | | | |
| Number of students screened | 4,409 | 2,246 | 2,163 | |
| Number of students with myopia | 754 | 383 | 371 | |
| Prevalence of myopia (%) | 17.1 | 17.1 | 17.1 | 0.0 |
| | [37.7] | [37.6] | [37.7] | (1.7) |
| Among myopic students: | | | | |
| Severity of myopia (refractive rower) | −1.042 | −1.043 | −1.041 | 0.003 |
| | [1.937] | [2.018] | [1.853] | (0.170) |
| Wore eyeglasses prior to treatment (1 = yes) | 0.178 | 0.178 | 0.178 | 0.000 |
| | [0.383] | [0.383] | [0.383] | (0.034) |

*Note*. Columns 1 to 3 present the means, with standard deviations in brackets. Column 4 shows the coefficients estimated by regressing each variable on the treatment dummy with standard errors, in parentheses, clustered at the school level. The cut-off for myopia is defined by refractive power ≤−0.75 dioptres (D). Severity of myopia reports is refractive power as defined by spherical equivalent.
*$p < .10$, **$p < .05$, ***$p < .01$

**Table 3.** OLS estimators of impact on maths scores (ITT).

| | Western China Program | | Eastern China Program | |
|---|---|---|---|---|
| | Unadjusted | Adjusted | Unadjusted | Adjusted |
| Dep. Var: Endline Standardised Maths Score | (1) | (2) | (3) | (4) |
| Treatment (1 = yes) | 0.157* | 0.137* | 0.043 | 0.046 |
| | (0.072) | (0.074) | (0.708) | (0.673) |
| Baseline standardised maths score controlled | Yes | Yes | Yes | Yes |
| Additional controls | No | Yes | No | Yes |
| Mean of dep. var. in control group | 0.149 | | −0.618 | |
| Observations | 653 | 617 | 706 | 603 |
| $R^2$ | 0.420 | 0.438 | 0.087 | 0.155 |

*Note*. Maths scores are standardised, using data from the control group at baseline within each program. Additional controls in Columns 2 and 4 include the remaining 11 baseline variables listed in Table 1 (in addition to the baseline standardised maths score, which is already controlled for in regression). Standard errors are clustered at the school level. *p*-values of coefficients are reported in parentheses.
*$p < .10$, **$p < .05$, ***$p < .01$

## Results

### *Prevalence*

#### *Western China program*

Among the 3,465 children for whom we conducted eye examinations at 84 schools in the Western China program, 683 (20%) children were identified as having poor vision (Table 2). The prevalence is 18% in the control schools and 22% in the treatment schools. Although the difference is statistically significant at the 10% level, we believe that this is by chance. This is because there are more students in treatment schools who have mildly poor vision (that is, students that are counted as having poor vision because they just pass the threshold). As can be seen in the table,

**Table 4.** Spillover effects on students without vision problems.

| | Western China Program | | Eastern China Program | |
|---|---|---|---|---|
| Dep. Var: Endline Standardised Maths Score | (1) | (2) | (3) | (4) |
| Treatment (1 = yes) | 0.141* | 0.129 | 0.055 | 0.058 |
| | (0.088) | (0.107) | (0.505) | (0.449) |
| Baseline standardised maths score controlled | Yes | Yes | Yes | Yes |
| Additional controls | No | Yes | No | Yes |
| Observations | 2,668 | 2,513 | 3,322 | 2,810 |

*Note*. Maths score are standardised, using data from the control group at baseline within each program. Additional controls in Columns 2 and 4 include the same sets of baseline variables listed in Table 1 (in addition to the baseline standardised maths score, which is already controlled for in regression), with two differences. The first is that we did not control for baseline eyeglasses ownership, as those students without myopia, by definition, did not have eyeglasses. The second is that we replaced the measure of myopia (again, by definition, those students who were not categorised as myopic did not have measures of myopia severity) with visual acuity (which we did for all the students, both with and without myopia, as a measure of vision function at baseline). *p*-values of coefficients, clustered at the school level, are reported in parentheses.
**p* < .10, ***p* < .05, ****p* < .01

**Table 5.** Compliance rate at baseline and endline.

| | Total | Control Group | Treatment Group | Difference (3)–(2) |
|---|---|---|---|---|
| Variable | (1) | (2) | (3) | (4) |
| Panel A. Western China Program | | | | |
| Number of students with myopia | 683 | 310 | 373 | |
| Wore eyeglasses at baseline (1 = yes) | 0.183 | 0.184 | 0.182 | −0.002 |
| | [0.387] | [0.388] | [0.387] | (0.038) |
| Wore eyeglasses at endline (1 = yes) | 0.385 | 0.262 | 0.490 | 0.228*** |
| | [0.487] | [0.447] | [0.501] | (0.067) |
| Panel B. Eastern China Program | | | | |
| Number of students with myopia | 754 | 383 | 371 | |
| Wore eyeglasses at baseline (1 = yes) | 0.178 | 0.178 | 0.178 | 0.000 |
| | [0.383] | [0.383] | [0.383] | (0.034) |
| Wore eyeglasses at endline (1 = yes) | 0.453 | 0.238 | 0.675 | 0.437*** |
| | [0.498] | [0.427] | [0.469] | (0.053) |

*Note*. Columns 1 to 2 present means with standard deviations reported in brackets. Column 3 shows coefficients estimated by regressing each variable on the treatment dummy, with standard errors, in parentheses, clustered at the school level.
**p* < .10, ***p* < .05, ****p* < .01

there is no difference between students in treatment and control schools in terms of the severity of myopia (a continuous variable).

After determining that one in six students in rural primary schools was myopic, we wondered how many of these students wore eyeglasses prior to our interventions. According to our baseline data, the share of myopic students who wear glasses is low. Only 18% of nearsighted students were wearing eyeglasses at the time of the baseline. This percentage is identical across treatment (18%) and control schools (18%). Clearly, poor vision is a problem in rural China. Further, few in those areas have their vision problems corrected.

### *Eastern China program*

The prevalence of poor vision and the proportion of students with poor vision who wore eyeglasses at the baseline in the Eastern China program are almost same as those in the Western China program. Among the 4,409 children for whom we conducted eye examinations at 94 schools in the Eastern China program, 754 (17%) children were identified as having poor vision. The prevalence of poor vision was identical across treatment (17%) and control schools (17%).

Among those with poor vision, we find a similarly low percentage of students who wore eyeglasses prior to our interventions. About 18% of myopic students in the Eastern China program wore eyeglasses at baseline. The percentage is identical across treatment (18%) and control schools (18%).

**Table 6.** Heterogeneity of ITT treatment effects by baseline teacher and school characteristics (pooled sample).

| Dep. Var: Endline Standardised Maths Score | Heterogeneity by Teaching Experience | Heterogeneity by Teacher Rank | Heterogeneity by Contract Teacher | Heterogeneity by School History | Heterogeneity by Four Variables | Heterogeneity by low schooling quality index |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (5) |
| Treatment (1 = yes) | 0.218* | 0.123 | 0.289 | 0.146 | 0.425** | 0.287** |
| | (0.054) | (0.255) | (0.124) | (0.197) | (0.030) | (0.020) |
| Treatment × teacher experience less than 3 yrs dummy | −0.481** | | | | −0.460** | |
| | (0.019) | | | | (0.030) | |
| Teacher experience less than 3 years dummy | −0.021 | | | | −0.028 | |
| | (0.888) | | | | (0.850) | |
| Treatment × teacher without rank | | −0.196 | | | −0.116 | |
| | | (0.398) | | | (0.622) | |
| Teacher without rank dummy dummy | | −0.298** | | | −0.342** | |
| | | (0.049) | | | (0.023) | |
| Treatment × recruitment of contract teachers | | | −0.242 | | −0.198 | |
| | | | (0.262) | | (0.368) | |
| Recruitment of contract teachers dummy | | | −0.059 | | −0.097 | |
| | | | (0.741) | | (0.584) | |
| Treatment × school history less than 10 yrs | | | | −0.192 | −0.101 | |
| | | | | (0.369) | (0.643) | |
| School history less than 10 yrs dummy | | | | −0.398*** | −0.410*** | |
| | | | | (0.010) | (0.007) | |
| Treatment × low schooling quality index | | | | | | −0.440*** |
| | | | | | | (0.003) |
| Low schooling quality index | | | | | | −0.434** |
| | | | | | | (0.023) |
| Additional controls (coefficients not shown) | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1220 | 1220 | 1220 | 1220 | 1220 | 1220 |
| $R^2$ | 0.377 | 0.369 | 0.369 | 0.369 | 0.380 | 0.380 |

*Note*. Samples of two programs were pooled together in the heterogeneity analysis. Four baseline teacher and school characteristics with significant differences between the two programs were included (teaching experience, teacher age, homeroom teacher turnover, school history). Additional controls include the 11 baseline variables, except the variable interacted with the treatment dummy in Table 1. Standard errors are clustered at the school level. *p*-values of coefficients are reported in parentheses.
*$p < .10$, **$p < .05$, ***$p < .01$

In summary, we find that the prevalence of poor vision and the proportion of students with poor vision who wore eyeglasses pre-treatment is almost identical across the two study sites. Specifically, about one in six students is myopic. This implies that myopia is pervasive among primary school-aged children across regions and school systems in China. Among myopic students, however, only one in six reported wearing eyeglasses. This result is found in both the Eastern and Western China program schools.[2]

To interpret our findings in the context of school-aged children in China (i.e., students in primary and middle schools, aged approximately 6 to 15 years), a recent meta-study documents the prevalence of poor vision and the proportion of myopic children who received vision care and used eyeglasses (He et al. 2015). He *et al*. used data from eight cross-sectional surveys conducted in seven provinces of China and found that about 25% of school-aged children have poor vision; and, among those who were identified as having poor vision, the use of eyeglasses is only about 14%. In other words, our schools are similar to those reviewed in He *et al*. Poor vision is a common problem, and there are few students who are being treated for their condition.

## Effect of providing eyeglasses

### Western China program

In this section, we focus on the subsample of students with poor vision only. We report the impact of the vision program using the ITT model, as seen in Table 3. When estimating the impact, using the unadjusted model from equation (1), we find a positive and significant treatment effect on standardised maths test scores in the Western China program. Controlling for baseline standardised maths scores, providing subsidised free eyeglasses and training improved the standardised maths score of myopic students at the endline by 0.16 standard deviations compared to the control group. The impact is statistically significant at the 10% level.

The results are almost identical when using the adjusted model from equation (2). Adding baseline characteristics to the model as controls increases the magnitude of the estimate coefficient to 0.14 standard deviations. The impact also is statistically significant, also at the 10% level.

### Eastern China program

In contrast to the moderately large, significant, and robust effects of the interventions in the Western China program, we find no impacts of the treatment in the Eastern China program when using the ITT model. When using the unadjusted model from equation (1), the point estimate of the impact, using the coefficient on the treatment variable, is only 0.04 (Table 3, Column 3). Its *p*-value (0.708) means that the coefficient is not statistically different from zero. When we use the adjusted model from equation (2), by adding baseline characteristics as controls, the point estimate slightly increases to 0.05 but is still statistically insignificant ($p$-value = 0.673; Table 3, Column 4).

In summary, the same intervention in the two school settings leads to different results. In the Western China program (in poor rural public schools), providing subsidised free eyeglasses and training in vision care improved the standardised maths scores of myopic students at the endline after one year. The results are robust regardless of whether we used the adjusted or unadjusted model. In contrast, when we implemented the same intervention of free eyeglasses and training to nearsighted children in the migrant communities of the Eastern China program, we found no improvement of test scores.

## Spillover effect on students without vision problems

We then identified the spillover effects of the program (i.e., the impact of providing eyeglasses to myopic students on students without vision problems within the same school). There are two reasons that we might expect to find positive spillovers among those students without vision problems. First, it could be that, when large shares of myopic students wear glasses and improve their vision, teachers might be able to spend less time with helping myopic students and shift more of their efforts to helping those without vision problems. Second, it also could be the case that, after getting glasses, myopic students would bother their non-myopic classmates less, as they could follow the class more closely and focus more on the materials being taught in class.

We repeated our analysis in previous section, using the sample from non-myopic students, and found evidence to support the idea that there are positive spillovers, at least in the Western China program (Table 4). Students in the treatment schools without myopia increased their scores by 0.14 standard deviations more than did their counterparts in the control schools (unadjusted model, $p$-value = 0.088). Although positive coefficients are found in both the unadjusted and adjusted models in both settings, the coefficients are, in general, not statistically significant, except for the one in the unadjusted model for the Western China sample (significant at the 10% level).

## Mechanisms

In this section, we examine three hypotheses about why the impact of our provision of free eyeglasses differed between the Western China program and the Eastern China program schools. In the first subsection, we determine whether the absence of compliance in Eastern China might have been the reason that there was no measured impact (Hypothesis 1). In the second subsection, we review the results presented in the Research Setting section above and assess whether the difference in program impacts might have been due to the fact that there were fundamental differences in the nature of the students and families between the Western and Eastern China program schools (Hypothesis 2). Finally, we review the results from the Research Setting section and perform heterogeneous analysis to assess whether differences in teacher and school characteristics between the Western and Eastern China program schools might explain the differences in the impact analysis results (Hypothesis 3).

### *Hypothesis 1: compliance as related to impact*

With no meaningful impact in the Eastern China program, we wondered whether the difference in the results in the ITT analysis were caused by implementation failure as well as whether the Eastern China program was implemented in a way that nearsighted students were less likely to wear their eyeglasses than were nearsighted students in the Western China program. To address Hypothesis 1, we present the results of our surveys that examined compliance between the treatment and control schools in wearing eyeglasses at the endline of both programs (Table 5).

#### *Western China program*

In the Western China program, as discussed earlier, the share of nearsighted students that wore eyeglasses at baseline was well balanced across treatment and control schools. The share was identical across treatment (18%) and control schools (18%).

After the implementation of the program in the treatment schools, we find that the compliance at endline was not complete. According to the results of our unannounced in-school visits, although 100% of nearsighted students were given free eyeglasses in our treatment group, only 49% of nearsighted students wore them. The share of students who wore eyeglasses in the control schools during the year of our RCT also rose slightly to 26% (from a baseline share of 18% due to non-project reasons and to the fact that we sent prescriptions home to the parents of the control students). In sum, after one academic year of program implementation, the intervention increased the share of students who wore eyeglasses by 23 percentage points. The difference in the share of nearsighted students at the baseline who wore glasses at the endline, although much lower than if there were perfect compliance, was statistically significant at the 1% level.

#### *Eastern China program*

Similar patterns were observed in the Eastern China program, albeit (and a bit surprisingly) with a higher rate of compliance. As shown earlier, the share of nearsighted students who wore eyeglasses was balanced across treatment and control schools. The share was identical across treatment (18%) and control schools (18%). After the intervention, the share of nearsighted students at baseline who wore their glasses in the treatment schools increased to 68%, while the share in control schools rose slightly to 24% (almost identical to the rise of the share that was observed in the control schools of the Western China program). In other words, the intervention in the Eastern China program increased the share of nearsighted students who wore their glasses by 44 percentage points. The difference was significant at the 1% level. Note, too, that the rate of compliance (68%) and the increase in compliance (44 percentage points) are higher in the Eastern China than in the Western China program, with a compliance rate of 44% and an increase in compliance of 19 percentage points.

When comparing baseline characteristics of compliers in the two programs, we find that the compliers in Eastern China were about 0.89 D in spherical equivalent less myopic than their counterparts in Western China (*p*-value <0.01). Taking this result at face value could imply a self-selection of more myopic students into the Western China program, which, in turn, would generate the larger and significantly positive impact in the Western China program. As described in the Research Setting section, however, a difference of 0.89 D in spherical equivalent should not be considered a big difference from a clinical perspective.

In summary, although imperfect compliance was present in both programs, the overall increase in compliance was actually higher in the Eastern China program. This suggests that the significant impacts found in the Western China, but not in the Eastern China, programs were not caused by project implementation issues.

### *Hypothesis 2: differences in individual and family characteristics*

The Research Setting section contains a discussion of the differences in the individual and family characteristics between students in the Western and Eastern China programs. It is possible that, if students and their families from the Eastern China program were from more vulnerable subpopulations than those from the Western China program, the differences in the observed impacts of the eyeglass program between the Western and Eastern China interventions would be due to the differences in the individual and family characteristics. As noted, however, the families of students from the Eastern China program schools have (observable) qualities that favour higher education outcomes, as the parents were better educated and wealthier and had more resources available for family investments.

In addition, we conducted a heterogeneity analysis based on three additional student characteristics and one parental characteristic: baseline test score, severity of the vision problem, whether a student already had eyeglasses, and the level of parental education. We interacted each of these three baseline student and parental characteristics with the treatment dummy (Appendix Table 3). Although the signs of the regression coefficients in three of the four interaction terms support the hypothesis for effect heterogeneity, i.e., a larger effect for students with lower baseline test scores and those with worse vision and a smaller effect for students with bad vision who already have eyeglasses, none of the coefficients are statistically significant (p-value = 0.707, 0.655 and 0.214, respectively).

Thus, based on these two points, it does not appear that the absence of an impact in the Eastern China program schools, compared to the positive impact in the Western China program schools, was due to differences in individual and family characteristics.

### *Hypothesis 3: differences in teacher and school quality*

Although we find little support for Hypotheses 1 and 2 in terms of explaining the differences in impact in the Western and Eastern China program schools, we do believe that there is support for Hypothesis 3. The analysis in the Research Setting section, in which we describe the differences between the two RCT research settings, suggests that the difference in impacts of the two vision care experiments might stem from variations in the quality of schooling across the two contexts. In short, some combination of less teaching experience, fewer teachers who have a credentialed rank, more recruitment of contract teachers, and a shorter school history (which we believe, at least in part, measures overall school quality) in the Eastern China program schools may be significantly reducing the quality of education being provided in private migrant schools as well as the impact of the vision care intervention. Perhaps the most plausible explanation for our findings is that, even though the vision care program helped to improve the vision of nearsighted students (by providing free glasses and getting students to wear them), the better-seeing, vision-enhanced students did not increase their learning outcomes because the quality of the schooling was so low in the first place.

In the rest of this subsection, we discuss heterogeneity analysis. We pool the sample of the two programs and conduct heterogeneity analysis in an attempt to show how much of the measured difference in the program impacts across the two contexts can be explained by four variables that measure teacher- and school-quality variables (Table 6). Specifically, we created four new variables (or interaction terms) by interacting measures of teaching experience, teacher rank, recruitment of contract teachers, and schooling history with the treatment variable. We then add these interaction terms to equation 2 (the adjusted model) one at a time (Columns 1–4) and as a group (Column 5).

The results in Table 3 show a difference in the estimated program impact of 0.09 (the coefficient in the Western China program is 0.14 with a *p*-value of 0.074, and the coefficient in the Eastern China program is 0.05 with a *p*-value of 0.673). Table 6, Column 1 shows how much of the differential program impact across two contexts is explained by teacher experience alone. The negative sign on the coefficient of the treatment and teaching experience interaction variable supports the hypothesis that the program's impact on the educational performance of students with poor vision is lower for those taught by teachers with less than three years of teaching experience. The interaction coefficient, −0.48, is statistically significant at the 5% level (Table 6, Column 1). As shown in Table 1, schools in the Eastern China program had 11% more teachers with less than three years of teaching experience relative to schools in the Western China program. Taken together, these findings imply that teacher experience alone appears to explain about half (56%) of the difference in the program impacts in the two schooling contexts.[3]The large contribution of teaching experience is in keeping with the existing literature that shows that teaching experience is one of the most important factors for the quality of teaching (Rockoff 2004).

Table 6, Columns 2–4 present the results in regard to how much of the differential program impacts across the Western and Eastern school settings is explained by each of the other three teacher- and school-quality variables (when added one by one). In each column, the point estimate of interaction term is negative. A negative interaction term, given the way that we defined the variables, suggests that the program's impact on the educational performance of students with poor vision is lower for those in those schools that are assessed as lower quality. Specifically, the three interaction coefficients are −0.20 in the case of the coefficient of interaction between treatment and teacher without teacher rank, −0.24 in the case of the coefficient of interaction between treatment and recruitment of contract teachers, and −0.19 in the case of the coefficient of interaction between treatment and school history of less than 10 years. Although all variables are negative, it is important to note that none of the interaction coefficients is statistically significant (at the 10% level).

Table 6, Column 5 presents the results in regard to how much of the differential program impact across two contexts is explained by the four teacher and school characteristics jointly. The point estimates of all four interaction terms are negative, which, taken together, support Hypothesis 3. The results suggest that the lower teaching experience, the lower share of teachers with formal ranks, the larger share of schools with contract teachers, and the shorter school histories of the schools in the Eastern China program schools all could contribute to the absence of impact in the Eastern China program. The joint test of the four interaction terms was statistically significant at the 1% level (*p*-value = 0.009). As in the results in Columns 1–4, only the coefficient on the teaching experience interaction variables is statistically significant at the 5% level. Its point estimate of −0.46 is similar to the point estimate (−0.48) in the regression that added the teaching experience interaction term alone. This also means that, by using this approach (and relying on only statistically significant coefficients) to identify effects of teaching quality, we can explain about half of the gap between the measured impact of the vision care program in the Western and Eastern program schools.

In the last column, a principal component analysis is conducted by generating an index of (low) quality of schooling based on the four dummies. We find a consistent pattern of effect heterogeneity by schooling quality: a significantly negative interaction term of the low schooling quality index and treatment dummy (p-value = 0.023) indicating a positive relationship between schooling quality and program impact.

Based on these findings, we can draw the conclusion that, although low rates of compliance in wearing eyeglasses (Hypothesis 1) and systematic differences in the characteristics of individuals and their families (Hypothesis 2) do not explain the absence of an impact in the Eastern China program, the differences in the quality of the schooling (Hypothesis 3) can account for most of the difference.

## Conclusion

Poor vision is the most common impairment to affect school-aged children in the developing world, comprising half of all disabilities among children (Congdon et al. 2008). This paper provides the first empirical examination of the causal link between the provision of vision care and educational performance of school-aged children in two different school settings in a developing context. We take advantage of two large-scale RCTs of school-based eyeglasses promotion programs in four provincial-level regions of China. One program was conducted in public schools in a rural area in Western China. The other program was conducted in private schools of migrant communities in Eastern China. We found that 17 to 20% of the students were suffering from poor vision; fewer than one in five nearsighted students were wearing glasses.

Overall, our results show a significant impact of 0.14 standard deviations when providing free eyeglasses and training to students with poor vision in the Western China program sites. It should be noted that this is a very large effect on learning. In contrast, we find no impacts of the program in the Eastern China schools. By comparing teacher experience, teacher rank, recruitment of contract teachers, and school history, our analysis suggests that it is plausible that the lack of a measurable impact in the Eastern China program is due to the poorer quality of schools in the migrant communities. Hence, the policy lesson from this part of the study is that, before a lot of time and effort are put into programs that get children to school, provide greater nutrition, or provide glasses (or some other input), it is imperative that the quality of the school, overall, is improved.

In support of the conclusion that the eyeglasses intervention failed to raise student academic performance due to the poor quality of schooling in migrant schools, previous impact evaluation work in migrant schools may be useful to examine. A review of the literature found two other RCTs that sought to raise the academic performance of students in China's migrant schools. One program provided teacher training (English language instruction for English teachers; Zhang et al. 2013), and another provided remedial tutoring to students through a computer-assisted learning (CAL) program (Mo et al. 2012). The results of these two interventions are consistent with our conclusions. The teacher training program was not successful due, in no small part, according to the authors, to the poor quality of teachers in migrant schools. The CAL program, however, was successful. In this case, however, the CAL program provided an alternative way for students to learn. In other words, CAL worked because it was substituting for poor quality schools/teaching, while, in our study, teacher training and eyeglasses were used to complement the quality of instruction in the schools. Therefore, following the main argument of this paper, when schools are of poor quality, simply providing eyeglasses is not enough to generate a meaningful impact on student academic performance, as they are only a complement to good teaching/school quality.

Although the poor quality of migrant schools is not a new result, our findings show that those interested in improving schools and student learning need to consider whether the school setting of their interventions is providing a quality education. In particular, our results contribute to the existing systematic reviews that show that health interventions consistently do not generate significant increases in learning (Evans and Popova 2016). If the goal is to increase the learning of children by giving glasses to those who have poor vision, then doing so in a setting such as that in the migrant communities of Eastern China is not productive.

Before drawing final conclusions, we have to acknowledge an important limitation regarding the interpretation of the nature of the differences in the results between the two study settings. Specifically, although the analysis found that the vision care program had a significant impact on student academic performance in Western China but not in Eastern China, at conventional levels of statistical significance we cannot reject the hypothesis that the two programs have the same impact. The results from the pooled sample regression show that the difference between these two estimates of treatment effects is not statistically significant from zero in either the unadjusted or adjusted models (p-value = 0.289, 0.523, respectively, in Appendix Table 4). Nevertheless, by documenting the differential impacts of almost identical interventions in two different contexts, and explaining the mechanism behind this difference, our results have important implications for the design, targeting, evaluation, and upscaling of interventions in developing countries. In short, our results show that the importance of context needs to be considered to generate meaningful results and translate the results into effective policies.

## Note

1. By the time that our study was conducted, there were 592 nationally designated poor counties out of 2,851 county-level divisions in China.
2. Some readers may wonder why such a high share of students with myopia are in school without glasses, but this issue is beyond the scope of this paper. Other research has identified several factors that seek to account for this phenomenon: (1) China's school system in rural areas almost never has the funding for or knowledge of health (including vision) problems that allows districts or schools to invest in a vision-screening program. Thus, less than one-third of school-aged children have ever had a vision exam at school (Bai *et al*. 2014); (2) The county-level health system is not set up to provide quality vision care. On average, each county hospital (which serves populations of up to 500,000) is staffed with two to three eye doctors. Moreover, these so-called doctors, on average, have only one year of education beyond high school. Likewise, although there are private sector optometry retailers, opticians have little formal training and frequently fail to provide either accurate prescriptions or high-quality eyeglasses (Zhou et al. 2014). In addition, the number of opticians in a county is usually limited, and there little opportunity to reach out into remote villages; (3) Finally, there are many myths about glasses in rural China that affect the beliefs of many rural parents and teachers. The most common is that young students (in primary school) should not wear glasses, as wearing glasses only makes one's eyes get worse (Ma et al. 2015).
3. These calculations use the fact that the differential effect by teachers' teaching experience is 0.05 standard deviations (0.48 * 0.11) and accounts for 56% of the total difference in program impact of 0.09 S.D. (0.14–0.05).

## JEL Codes: C93, I12, I21, O15

## Acknowledgments

We are indebted to Steve Boucher for his insightful comments. We also are grateful for helpful comments from seminar participants in CES North America in Michigan, especially Jessica Leight and Xiaoyang Ye. We would also like to acknowledge our colleagues, Matt Boswell, Hongmei Yi, Xiaopeng Pang, Xiaobing Wang, Huai Zhou, and the 300 enumerators from the Center for Experimental Economics in Education, Shaanxi Normal University; Center for Chinese Agricultural Policy, Chinese Academy of Sciences; Renmin University of China; and Tongji University. We give special thanks to staff from Zhongshan Ophthalmic Center, Sun Yat-sen University, for their invaluable guidance and advice. We also are grateful for financial and technical support from OneSight, Luxottica-China, Essilor, and CLSA. All errors are our own.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This project was supported by the 111 Project (B16031) and OneSight.

## Notes on contributors

*Xiaochen Ma* is a lecturer at China Center for Health Development Studies, Peking University. He is a health economist who focuses on creating and testing innovative strategies to delivery better quality of care for underserved communities in low- and middle-income countries.

*Huan Wang* is a Research Scholar at Stanford University. Her research focuses on assessing educational quality and identifying effective educational programs and policies to improve student outcomes in rural China. By conducting large-scale randomized controlled trials, she evaluates the impact of social emotional learning on reducing dropouts in rural junior high schools, the impact of independent reading on student performance, and the effect of vision care programs on learning and schooling path.

*Yaojiang Shi* is a Professor of Economics at the Center for Experimental Economics in Education (CEEE) at Shaanxi Normal University in Xi'an, China. His research focuses on experimental economics, rural education, and rural development.

*Sean Sylvia*, PhD, is an Assistant Professor in the Department of Health Policy and Management at UNC's Gillings School of Global Public Health. As a health and development economist, Dr. Sylvia's research focuses on designing and evaluating innovative approaches to improve the delivery of health services in developing countries. In past and ongoing projects, he has studied the design of performance-based incentives for providers, the implementation of school-based health and nutrition programs, community health worker interventions to improve early childhood health and development, and the measurement of and interventions to improve the quality of primary care in low-resource settings.

*Lei Wang*, PhD is an Associate Professor in the International Business School at Shaanxi Normal University. Dr. Wang's research focuses on health and nutrition programs to improve early childhood health and development using experimental research design methods.

*Yiwei Qian* is a PhD candidate in Economics from the University of Southern California. He specializes in Development Economics and Applied Microeconomics. His research primarily focuses on developmental interventions on emerging human capital challenges in low-income contexts and related econometrics topics in Program Evaluation.

*Scott Rozelle* is the Helen F. Farnsworth Senior Fellow and the co-director of Stanford Center on China's Economy and Institutions in the Freeman Spogli Institute for International Studies and Stanford Institute for Economic Policy Research at Stanford University. His research focuses almost exclusively on China and is concerned with: agricultural policy, including the supply, demand, and trade in agricultural projects; the emergence and evolution of markets and other economic institutions in the transition process and their implications for equity and efficiency; and the economics of poverty and inequality, with an emphasis on rural education, health and nutrition.

## ORCID

Huan Wang http://orcid.org/0000-0003-3219-7226

## References

Acemoglu, D. 2010. “Theory, General Equilibrium, and Political Economy in Development Economics.” *Journal of Economic Perspectives* 24 (3): 17–32. doi:10.1257/jep.24.3.17.

An. X. “Teacher salaries and the shortage of high-quality teachers in China's rural primary and secondary schools.” Chinese Education & Society 51, no. 2 (2018): 103–116

Bai, Y., H. Yi, L. Zhang, Y. Shi, Xi. Ma, N. Congdon, Z. Zhou, M. Boswell, and S. Rozelle. 2014. “An Investigation of Vision Problems and Vision Care System in Rural China.” *Southeast Asian Journal of Tropical Medicine and Public Health* 45:1464–1473.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert, and C. Udry. 2015. “Development Economics. A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries.” *Science* 348 (6236): 6236. DOI:10.1126/science.1260799.

Barrett, C., and M. Carter. 2010. “The Power and Pitfalls of Experiments in Development Economics: Some Non-Random Reflections.” *Applied Economic Perspectives and Policy* 32 (4): 515–548. doi:10.1093/aepp/ppq023.

Bauchet, J., J. Morduch, and S. Ravi. 2015. "Failure Vs. Displacement: Why an Innovative Anti-Poverty Program Showed No Net Impact in South India". *Journal of Development Economics* 116: 1–16. DOI:10.1016/j.jdeveco.2015.03.005.
Brown, P., and A. Park. 2002. "Education and Poverty in Rural China." *Economics of Education Review* 21 (6): 523–541. doi:10.1016/S0272-7757(01)00040-1.
Bruhn, M., and D. McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal. Applied Economics* 1 (4): 200–232.
Chen, X. F. Lai, H. Yi, L. Zhang, J. Chu, and S. Rozelle. 2015. "Schooling Institutions and Academic Achievement for Migrant Children in Urban China." *Migration & Development* 4 (2): 1–13. DOI:10.1080/21632324.2014.986868.
China National Bureau of Statistics. 2008–2014. *Statistical Yearbook*. Beijing, China. China: China State Statistical Press.
"China National Development and Reform Commission." 2015. *Report on Population and Social Development*. http://www.sdpc.gov.cn/gzdt/201507/t20150731_743695.html
Chu, J., P. Loyalka, J. Chu, Q. Qu, Y. Shi, and G. Li. 2015. "The Impact of Teacher Credentials on Student Achievement in China". *China Economic Review* 36: 14–24. DOI:10.1016/j.chieco.2015.08.006.
Congdon, N. Y. Wang, Y. Song, K. Choi, M. Zhang, Z. Zhou, Z. Xie, *et al*. 2008. "Visual Disability, Visual Function, and Myopia among Rural Chinese Secondary School Children: The Xichang Pediatric Refractive Error Study (X-pres)—report 1." *Investigative Ophthalmology & Visual Science* 49 (7): 2888–2894. DOI:10.1167/iovs.07-1160.
Culter, D., and A. Lleras-Muney. 2014. "Education and Health: Insights from International Comparisons." *Encyclopedia of Health Economics* 1: 232–245.
Deaton, A. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48 (2): 424–455. doi:10.1257/jel.48.2.424.
Duflo, E., R. Glennerster, and M. Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." *Handbook of development economics* 4 (2007): 3895–3962.
Elliott, D., and C. Howell-Duffy. 2015. "'Non-tolerances' and the Science of Prescribing Spectacles." *Optometry in Practice* 16 (4): 131–144.
Engberg, J., B. Gill, G. Zamarro, and R. Zimmer. 2012. "Closing Schools in a Shrinking District: Do Student Outcomes Depend on Which Schools are Closed?" *Journal of Urban Economics* 71 (2): 189–203. DOI:10.1016/j.jue.2011.10.001.
Evans, D., and A. Popova. 2016. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." *The World Bank Research Observer* 31 (2): 242–270. doi:10.1093/wbro/lkw004.
Ganimian, A., and R. Murnane. 2016. "Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Impact Evaluations." *Review of Educational Research* 86 (3): 719–755. doi:10.3102/0034654315627499.
Glewwe, P., Eric A. Hanushek, Sarah D. Humpage, and R. Ravina. 2013. "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010." In *Education Policy in Developing Countries*, edited by P. Glewwe, *et al*., 13–64. Chicago: University of Chicago Press.
Haveman, R., and B. Wolfe. 1995. "The Determinants of Children's Attainments: A Review of Methods and Findings." *Journal of Economic Literature* 33 (4): 1829–1878.
He, M. F. Xiang, Y. Zeng, J. Mai, Q. Chen, J. Zhang, W. Smith, K. Rose, and Ian G. Morgan. 2015. "Effect of Time Spent Outdoors at School on the Development of Myopia among Children in China: A Randomized Clinical Trial." *The Journal of the American Medical Association* 314 (11): 1142–1148. DOI:10.1001/jama.2015.10803.
Krishnaratne, S., and H. White. 2013. *Quality Education for All Children? What Works in Education in Developing Countries*. New Delhi: International Initiative for Impact Evaluation. Working Paper 20.
Kwong, J. 2004. "Educating Migrant Children: Negotiations between the State and Civil Society." *The China Quarterly* 108 (180): 1073–1088. doi:10.1017/S030574100400075X.
Lai, F., C. Liu, R. Luo, L. Zhang, X. Ma, Y. Bai, B. Sharbono, and S. Rozelle. 2014. "The Education of China's Migrant Children: The Missing Link in China's Education System". *International Journal of Educational Development* 37: 68–77. DOI:10.1016/j.ijedudev.2013.11.006.
Li, T., L. Zhang, and S. Rozelle. 2014. "Encouraging Classroom Peer Interactions: Evidence from Chinese Migrant Schools". *Journal of Public Economics* 111: 29–45. DOI:10.1016/j.jpubeco.2013.12.014.
Ma, X. N. Congdon, H. Yi, Z. Zhou, X. Pang, M. E. Meltzer, Y. Shi, M. He, Y. Liu, and S. Rozelle. 2015. "Safety of Spectacles for Children's Vision: A Cluster-Randomized Controlled Trial." *American Journal of Ophthalmology* 160 (5): 897–904. DOI:10.1016/j.ajo.2015.08.013.
Maluccio, J., J. Hoddinott, J. R. Behrman, R. Martorell, A. R. Quisumbing, and Aryeh D. Stein. 2009. "The Impact of Improving Nutrition during Early Childhood on Education among Guatemalan Adults." *Economic Journal* 119 (537): 734–763. DOI:10.1111/j.1468-0297.2009.02220.x.
McEwan, P. 2015. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Review of Educational Research* 85 (3): 353–394. doi:10.3102/0034654314553127.
Miguel, E., and M. Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217. doi:10.1111/j.1468-0262.2004.00481.x.
Mo, D., H. Yi, L. Zhang, Y. Shi, S. Rozelle, and A. Medina. 2012. "Transfer Paths and Academic Performance: The Primary School Merger Program in China." *International Journal of Educational Development* 32 (3): 423–431. DOI:10.1016/j.ijedudev.2011.11.001.

Peters, J., J. Langbein, and G. Roberts. 2016. "Policy Evaluation, Randomized Controlled Trails, and External Validity—A Systematic Review". *Economics Letters* 147: 51–54. DOI:10.1016/j.econlet.2016.08.013.

Pritchett, L., and J. Sandefur. 2015a. "Learning from Experiments When Context Matters." *American Economic Review* 105 (5): 471–475. doi:10.1257/aer.p20151016.

Pritchett, L., and J. Sandefur. 2015b. "Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix." *Journal of Globalization and Development* 4 (2): 161–197.

Robinson, B., and W. Yi. 2008. "The Role and Status of Non-Governmental ('daike') Teachers in China's Rural Education." *International Journal of Educational Development* 28 (1): 35–54. doi:10.1016/j.ijedudev.2007.02.004.

Rockoff, J. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *The American Economic Review* 94 (2): 247–252. doi:10.1257/0002828041302244.

Vermeersch, C., and M. Kremer. 2005. *School Meals, Educational Achievement, and School Competition: Evidence from a Randomized Evaluation*. Washington, DC: World Bank Publications. Policy Research Working Paper, Report No. WPS3523.

Wang, C. 2002. "*Minban* Education: The Planned Elimination of the "People-managed" Teachers in Reforming China." *International Journal of Educational Development* 22 (2): 109–122. doi:10.1016/S0738-0593(00)00079-1.

Wang, L. M. Li, S. Zhang, Y. Sun, S. Sylvia, E. Yang, G. Ma, L. Zhang, D. Mo, and S. Rozelle. 2018. "Contract Teacher and Student Achievement in Rural China: Evidence from Class Fixed Effects." *Australian Journal of Agricultural and Resource Economics* 62 (2): 299–322. DOI:10.1111/1467-8489.12250.

Wang, X., R. Luo, L. Zhang, and S. Rozelle. 2017. "The Education Gap of China's Migrant Children and Rural Counterparts." *The Journal of Development Studies* 53 (11): 1865–1881. DOI:10.1080/00220388.2016.1274395.

Yiu, L., and J. Adams. 2013. "Reforming Rural Education: Understanding Teacher Expectations for Rural Youth." *The China Quarterly* 216: 993–1017. doi:10.1017/S0305741013001136.

Zhang, L., F. Lai, X. Pang, H. Yi, and S. Rozelle. 2013. "The Impact of Teacher Training on Teacher and Student Outcomes: Evidence from a Randomised Experiment in Beijing Migrant Schools." *Journal of Development Effectiveness* 5 (3): 339–358. DOI:10.1080/19439342.2013.807862.

Zhao, D., and B. Parolin. 2012. "School Mapping Restructure in Rural China: Achievements, Problems and Implications." *Asia Pacific Education Review* 13 (4): 713. doi:10.1007/s12564-012-9231-z.

Zhou, Z., J. Zeng, X. Ma, X. Pang, H. Yi, Q. Chen, M. E. Meltzer, M. He, S. Rozelle, and N. Congdon. 2014. "Accuracy of Rural Refractionists in Western China." *Investigative Ophthalmology & Visual Science* 55 (1): 154–161. DOI:10.1167/iovs.13-13250.

**Table 1.** Baseline characteristics by program treatment assignment.

| | Western China Program | | | | Eastern China Program | | | |
|---|---|---|---|---|---|---|---|---|
| | Control Group | Treatment Group | Difference (2)–(1) | *n* | Control Group | Treatment Group | Difference (6)–(5) | *n* |
| Baseline Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Age (years) | 10.919 | 10.855 | −0.064 | 682 | 10.955 | 10.930 | −0.026 | 753 |
| | [0.881] | [0.989] | (0.127) | | [0.967] | [0.873] | (0.097) | |
| Male (1 = yes) | 0.523 | 0.509 | −0.013 | 683 | 0.509 | 0.496 | −0.013 | 754 |
| | [0.500] | [0.501] | (0.039) | | [0.501] | [0.501] | (0.034) | |
| Severity of myopia (refractive power) | −1.601 | −1.630 | −0.028 | 683 | −1.043 | −1.041 | 0.003 | 753 |
| | [1.719] | [1.889] | (0.158) | | [2.018] | [1.853] | (0.170) | |
| Wore eyeglasses at baseline (1 = yes) | 0.184 | 0.182 | −0.002 | 683 | 0.178 | 0.178 | −0.000 | 731 |
| | [0.388] | [0.387] | (0.038) | | [0.383] | [0.383] | (0.035) | |
| Baseline maths score | 0.220 | 0.394 | 0.168 | 683 | 0.141 | 0.167 | 0.027 | 753 |
| | [1.034] | [0.945] | (0.107) | | [0.985] | [1.028] | (0.137) | |
| One or both parents with high school education or higher (1 = yes) | 0.186 | 0.214 | 0.028 | 670 | 0.307 | 0.312 | 0.005 | 727 |
| | [0.390] | [0.411] | (0.033) | | [0.462] | [0.464] | (0.039) | |
| Household asset value (1,000 RMB) | 29.059 | 29.875 | 0.816 | 656 | 33.910 | 38.460 | 4.550a | 697 |
| | [34.385] | [34.468] | (5.258) | | [31.465] | [34.014] | (2.692) | |
| Teaching experience less than 3 yrs (1 = yes) | 0.341 | 0.143 | −0.199aa | 83 | 0.362 | 0.340 | −0.021 | 93 |
| | [0.480] | [0.354] | (0.093) | | [0.486] | [0.479] | (0.099) | |
| Teacher without rank (1 = yes) | 0.000 | 0.000 | 0.000 | 83 | 0.447 | 0.362 | −0.085 | 93 |
| | [0.000] | [0.000] | (0.000) | | [0.503] | [0.486] | (0.102) | |
| Recruitment of contract teachers (1 = yes) | 0.610 | 0.738 | 0.128 | 83 | 0.894 | 0.957 | 0.064 | 93 |
| | [0.494] | [0.445] | (0.103) | | [0.312] | [0.204] | (0.054) | |
| School history less than 10 yrs (1 = yes) | 0.122 | 0.095 | −0.027 | 83 | 0.511 | 0.457 | −0.054 | 93 |
| | [0.331] | [0.297] | (0.069) | | [0.505] | [0.504] | (0.105) | |

*Note*. Rows 1–7 present individual-level results. Rows 8–11 present school-level results. Columns 1, 2, 5, and 6 include means, with standard deviations reported in brackets. Column 3 and 7 show coefficients estimated by regressing each of the baseline characteristics on the program dummy, with standard errors, in parentheses, clustered at the school level (except the last four variables, which use school-level data). Severity of myopia is the refractive power defined by spherical equivalent. Baseline maths scores are derived from the standardised maths score pooling data from two programs at the baseline. Household assets value is calculated by summing the value, as reported in the China Rural Household Survey Yearbook (Department of Rural Surveys, National Bureau of Statistics of China 2013), of items on a list of 13 owned by the family.

a$p < .10$, **$p < .05$, ***$p < .01$

**Table 2.** Baseline characteristics of non-missing observations by program treatment assignment.

| | Western China Program | Eastern China Program |
|---|---|---|
| Dep. Var: Drop out between baseline and endline | | |
| Treatment (1 = yes) | 0.028 | −0.027 |
| | (0.123) | (0.367) |
| Age (years) | 0.027[aa] | 0.019 |
| | (0.012) | (0.275) |
| Male (1 = yes) | −0.009 | −0.019 |
| | (0.522) | (0.272) |
| Severity of myopia (refractive power) | −0.001 | 0.002 |
| | (0.792) | (0.580) |
| Wore eyeglasses at baseline (1 = yes) | 0.007 | 0.006 |
| | (0.748) | (0.801) |
| Baseline maths score | −0.002 | 0.023 |
| | (0.774) | (0.134) |
| One or both parents with high school education or higher (1 = yes) | 0.034 | −0.005 |
| | (0.111) | (0.844) |
| Household asset value (1,000 RMB) | 0.000 | 0.000 |
| | (0.854) | (0.340) |
| Teaching experience less than 3 yrs (1 = yes) | −0.013 | −0.026 |
| | (0.337) | (0.446) |
| Teacher without rank (1 = yes) | omitted | 0.050 |
| | (.) | (0.272) |
| Recruitment of contract teachers (1 = yes) | −0.003 | 0.062 |
| | (0.897) | (0.101) |
| School history less than 10 yrs (1 = yes) | 0.019 | −0.029 |
| | (0.360) | (0.438) |
| | 0.028 | −0.027 |
| *Test for joint significance* | 0.291 | 0.526 |
| *N* | 644 | 645 |

*Note*. Rows 1–7 present individual-level results. Rows 8–11 present school-level results. Severity of myopia is the refractive power defined by spherical equivalent. Baseline maths scores are derived from the standardised maths score pooling data from two programs at the baseline. Household assets value is calculated by summing the value, as reported in the China Rural Household Survey Yearbook (Department of Rural Surveys, National Bureau of Statistics of China 2013), of items on a list of 13 owned by the family.

a*p* < .10, ***p* < .05, ****p* < .01

**Table 3.** Heterogeneity of ITT treatment effects by baseline student and parent characteristics (pooled sample).

| Dep. Var: Endline Standardised Maths Score | Heterogeneity by Baseline Maths Scores (1) | Heterogeneity by Severity of Myopia (2) | Heterogeneity by Baseline Eyeglasses (3) | Heterogeneity by Parents' Education (4) |
|---|---|---|---|---|
| Treatment (1 = yes) | 0.127 | 0.080 | 0.117 | 0.086 |
| | (0.243) | (0.427) | (0.243) | (0.380) |
| Baseline score above median (1 = yes) | 0.753aaa | | | |
| | (0.000) | | | |
| Treatment ×Baseline score above median | −0.051 | | | |
| | (0.707) | | | |
| Treatment effect for students with higher baseline score | 0.076 | | | |
| | (0.547) | | | |
| More severe myopia (1 = yes) | | 0.036 | | |
| | | (0.689) | | |
| Treatment × More severe myopia | | 0.052 | | |
| | | (0.655) | | |
| Treatment effect for students with more severe myopia | | 0.132 | | |
| | | (0.304) | | |
| Had eyeglasses at baseline (1 = yes) | | | 0.052 | |
| | | | (0.580) | |
| Treatment × Had eyeglasses at baseline | | | −0.152 | |
| | | | (0.214) | |
| Treatment effect for students with eyeglasses at baseline | | | −0.035 | |
| | | | (0.794) | |
| Parents with high school education or above (1 = yes) | | | | 0.015 |
| | | | | (0.864) |
| Treatment × Parents with high school education or above | | | | 0.014 |
| | | | | (0.907) |
| Treatment effect for parents with high school education or above | | | | 0.100 |
| | | | | (0.470) |
| Additional controls (coefficients not shown) | Yes | Yes | Yes | Yes |
| Observations | 1220 | 1220 | 1220 | 1220 |
| $R^2$ | 0.34 | 0.368 | 0.368 | 0.368 |

*Note*. Samples of two programs were pooled together in the heterogeneity analysis. Three baseline student characteristics and one parental characteristic were included (baseline score above median, severity of myopia, baseline eyeglasses ownership, parental education level). Additional controls include the 11 baseline variables, except the variable that was interacted with the treatment dummy in Table 1. Standard errors are clustered at the school level. *p*-values of coefficients are reported in parentheses.

a$p < .10$, **$p < .05$, ***$p < .01$

**Table 4.** Treatment effects by program location (pooled sample).

| | | Pooled Sample | |
|---|---|---|---|
| | Dep. Var: Endline Standardised Maths Score | (1) | (2) |
| (1) | Treatment (1 = yes, treatment effect for Western China Program) | $0.202^{aa}$ | $0.142^{*}$ |
| | | (0.029) | (0.092) |
| (2) | Program Location (1 = yes, Eastern China Program) | $-0.733^{aaa}$ | $-0.707^{aaa}$ |
| | | (0.000) | (0.000) |
| (3) | Treatment a Program Location | −0.165 | −0.096 |
| | | (0.289) | (0.523) |
| (4) | Treatment effect for the Eastern China Program | 0.037 | 0.046 |
| | | (0.125) | (0.122) |
| (5) | Baseline standardised maths score controlled | Yes | Yes |
| (6) | Additional controls | No | Yes |
| (7) | Observations | 1359 | 1220 |

*Note*. The interaction term between treatment and program location in row 3 captures the difference between the impact sizes of the two programs. The treatment effect for the Eastern China Program in row 4 is the sum of rows 1 and 3. Additional controls in Column 2 include the remainder of the 11 baseline variables listed in Table 1 (in addition to baseline standardised maths score, which is already controlled for in the regression). Standard errors are clustered at the school level. *p*-values of coefficients are reported in parentheses. Maths score are standardised, using data in the control group at baseline within each program.
a$p < .10$, **$p < .05$, ***$p < .01$