A Better Vision for Development:
Eyeglasses and Academic Performance in Rural Primary Schools in China

Paul Glewwe
University of Minnesota

Albert Park
Hong Kong University of Science and Technology, CEPR, and IZA

Meng Zhao
University of Tokyo

June, 2014

Abstract:

About 10% of primary school students in developing countries have poor vision, but very few of them wear glasses. Almost no research examines the impact of poor vision on school performance, and simple OLS estimates are likely to be biased because studying harder often adversely affects one's vision. This paper presents results from a randomized trial in Western China that offered free eyeglasses to rural primary school students. The most conservative of our preferred estimates, which averages across data from two counties, indicates that wearing eyeglasses for one academic year increased average test scores of students with poor vision by 0.16 standard deviations, equivalent to 0.3 additional years of schooling, but the preferred estimate that is based on the one county where data collection was more closely monitored suggests that wearing eyeglasses increased test scores by 0.41 or more standard deviations, equivalent to 0.9 additional years of schooling. We also find that the benefits are greater for under-performing students. A simple cost-benefit analysis suggests very high economic returns to wearing eyeglasses, raising the question of why such investments are not made by most families. We find that girls are more likely to refuse free eyeglasses, and that lack of parental awareness of vision problems, mothers' education, and economic factors (expenditures per capita and price) significantly affect whether children wear eyeglasses in the absence of the intervention.

Data collection for the Gansu Survey of Children and Families was supported by grants from The Spencer Foundation Small and Major Grants Programs (wave 1), by NIH Grants 1R01TW005930-01 and 5R01TW005930-02 (wave 2), and by a grant from the World Bank (wave 2). Travel was supported by the Center for International Food and Agricultural Policy at the University of Minnesota.

1. Introduction

Most economists agree that higher levels of education increase economic growth (Barro, 1991; Mankiw et al., 1992; Hanushek and Kimko, 2000; Krueger and Lindahl, 2001; Sala-i-Martin et al., 2004; Hanushek and Woessmann, 2008), raising incomes and the quality of life. Support for education is also strong among the international development community. Two of the eight Millennium Development Goals from the United Nations Millennium Summit in 2000 focus on education: all children should complete primary school, and gender equality should prevail at all education levels.

Yet school enrollment may have little effect on economic growth and individuals' incomes if children do not learn effectively while they are in school. Although economists and other social scientists have identified several education policies that increase school enrollment, much less is known about how to increase student learning (Glewwe and Kremer, 2006). Recently, randomized control trials have begun to produce valuable evidence on the effect of specific interventions on student learning (e.g Duflo, Hanna, and Ryan, 2012; Glewwe, Kremer, and Ilias, 2010; Glewwe, Kremer, and Moulin, 2009; Banerjee et al, 2007). These interventions, and most education policy reforms, have focused on improving the quality of schools and teachers—the supply side of education.

Much less attention has been paid to increasing students' capacity and motivation to learn, which often reflects decisions that parents make on their children's behalf. Researchers have found that health interventions, such as school meals and deworming programs, increase attendance and enrollment (Afridi, 2011; Vermeersch and Kremer, 2004; Miguel and Kremer, 2004), but did not find evidence that these school-level interventions increase learning as measured by test scores. One study did find that reducing iron-deficiency among children in a poor region of China raised math test scores (Luo et al., 2012). If learning can be improved significantly with low-cost investments, then it is important for policy formulation to understand why these investments are not being made. One possibility is lack of information; Jensen (2010) finds that simply informing students about the likely returns to further education increases years of schooling; our study also finds evidence that lack of information can help explain apparent suboptimal behavior.

This paper examines a health-related intervention with the potential to raise student learning in developing countries that, to date, has received little attention: providing eyeglasses to primary school students with vision problems. About 10% of primary school students in developing countries have vision problems. In almost all cases these problems can be corrected with properly fitted eyeglasses, but very few of these children actually have eyeglasses. This paper presents results of a randomized trial in Western China that offered free eyeglasses to children in grades 4, 5 and 6 living in two counties. It estimates the impact of being offered eyeglasses and, because one third of those offers were turned down, it also estimates the impact of wearing eyeglasses. The most conservative of our preferred estimates, which averages across two counties, indicates that offering free eyeglasses to students with poor vision increases average test scores by 0.11 to 0.16 standard deviations; however in one of the two experimental counties, test scores increased by 0.26 to 0.39 standard deviations. Similarly, our conservative estimate of the impact of actually wearing eyeglasses for one academic year is 0.16 to 0.22 standard deviations, which we estimate is equivalent to 0.3 to 0.5 additional years of schooling. However, in one of the two experimental counties, wearing eyeglasses increased test scores by at least 0.41 standard deviations, equivalent to 0.9 additional years of schooling. An increase of time in school of this size leads to higher life cycle wages, the present value of which easily exceeds the cost of the glasses.

These findings imply that households fail to make high-return investments, which raises an important policy-relevant question: What explains this failure? We study the determinants of children accepting the free eyeglasses offered by the project and also use a richer dataset on rural children from the same province to examine the determinants of wearing glasses absent the intervention. We find that both information failures, i.e. lack of awareness of vision problems, and credit constraints appear to be important factors.

The rest of the paper is organized as follows. Section 2 introduces relevant aspects of primary education in rural China, and reviews the literature on the extent of vision problems among primary school students in developing countries, and how those problems affect students' academic performance. Section 3 describes the randomized trial and the data collection, and the next section presents the methodology used to estimate program impacts. Section 5 investigates the possibility of selection bias when township pairs that did not implement the program correctly are excluded from the estimates, and examines whether the treatment and control townships are similar. The next three sections present the results, check their robustness, and investigate whether they vary by student characteristics, respectively. Section 9 presents estimates that explore why some children did not accept the free eyeglasses, and more generally why most children with poor vision do not wear eyeglasses absent the intervention. A final section summarizes the results and provides recommendations for further research.

2. Background and Literature Review

A. Primary Education in Rural China. China has achieved nearly universal primary school enrollment. In the year 2000, only four percent of adults aged 25 to 29 did not have any formal schooling (Hannum et al., 2008). The Law on Compulsory

Education passed in 1986 mandates that all children complete nine years of schooling—six years of primary school and three years of lower secondary school. Yet the rural poor and some minority populations continue to face difficulties in meeting this schooling goal (Hannum et al., 2008, Hannum, Park, and Cheng, 2007).

In rural areas of Western China, nearly all children attend the nearest public primary school, located in their village or a nearby village. Teachers are allocated to schools within the county by the county educational bureau, and the county government pays their salaries. Thus, disparities in primary school quality within counties are usually modest (Li et al., 2009), reducing the incentive to bypass the local school. In China, the Center for Disease Control under each county's Health Bureau is charged with conducting physical exams of all students, including eye exams. In principle, these exams should be conducted annually for all students, but budgetary and staff constraints cause many schools to conduct physical exams only once every two or three years. The results of these exams are given to teachers, who are expected to convey the information to parents.

B. Vision Problems and School Performance. Very little data exist on the vision problems of school-age children in developing countries. Bundy et al. (2003) report that about 10% of school-age (5-15 years old) children have refraction errors (myopia, hypermetropia, strabismus, amblyopia, and astigmatism), which constitute about 97% of these children's vision problems. Almost all refraction errors can be corrected with properly fitted eyeglasses, but most children with these problems in low income countries do not have glasses. Zhao et al. (2000) found that, in one district in Beijing, 12.8% of children age 5-15 years had vision problems, of which 90% were refraction errors. Only 21% of the children with vision problems had glasses. He et al. (2007) report that 36.8% of 13-year-olds and 53.9% of 17-year-olds in middle schools in a county in Guangdong, a rich southern province, had myopia, and that less than half had eyeglasses. Rural children with vision problems who live in poor, remote areas and attend primary schools are even less likely to have glasses, as will be shown below. In China, a commonly held (but mistaken) view is that wearing glasses causes children's vision to deteriorate faster.

The lack of data on children's vision problems in developing countries has led to very little research on the impact of poor vision on students' academic performance. Only two published studies exist. First, Gomes-Neto et al. (1997) found that primary school children in Northeast Brazil with vision problems had a 10 percentage point higher probability of dropping out, an 18 percentage point higher probability of repeating a grade, and scored 0.2 to 0.3 standard deviations lower on achievement tests. Yet these estimates could be biased; to the extent that some of these children wore glasses, their vision could be correlated with unobserved factors that affect learning, such as parental preferences for educated children. Also, even if none of them wore glasses, their vision can be affected by their home environment (e.g. lighting quality) and their daily activities, including time spent studying and doing homework. Thus their vision may be correlated with unobserved factors that directly affect school performance (e.g. hours studying), leading to biased estimates. Second, Hannum and Zhang (2012), using data from the Gansu Survey of Children and Families (described more below) and propensity score matching, find that for children with poor vision aged 13-16, wearing eyeglasses increases math and literacy test scores significantly (by 0.27 and 0.43 standard deviations) but does not increase language scores. Unfortunately, they could not fully address the problem of self-selection in wearing glasses; indeed, they show that wearing glasses tends to be associated with higher socio-economic status and greater academic achievement and engagement.

3. Project Description and Data Available

The lack of rigorous studies on the impact of providing eyeglasses to students with visual impairments in developing countries led to the implementation of the Gansu Vision Intervention Project in 2004 in Gansu Province in northwest China. This section describes the project and the data available to evaluate its impact.

A. The Gansu Vision Intervention Project. In 2004, a team of Chinese and international researchers, in cooperation with the Center for Disease Control of Gansu's Bureau of Health, implemented a randomized trial to examine the impact of providing eyeglasses to primary school students with poor vision in two counties (hereafter, County 1 and County 2). The project covered nearly all grade 4-6 students in primary schools in these two counties.

Gansu Province is in northwestern China. It is geographically diverse, including areas of the Loess Plateau, the Gobi desert, mountainous areas, and vast grasslands. In 2004, the year of the intervention, its population was 25.4 million, about three fourths of whom live in rural areas (National Bureau of Statistics, 2005). In 2004, Gansu ranked 30th out of 31 provinces in rural per capita disposable income, with only Tibet having a lower mean income (National Bureau of Statistics, 2005). Using official poverty lines, a World Bank report found that 22.7 percent of Gansu's rural population was poor in 1996, compared to 6.3 percent for China as a whole (World Bank, 2000).

The two counties were chosen for the study because they are typical rural counties in Gansu, are located within several hours drive of Lanzhou (the provincial capital), which facilitated monitoring by Gansu's Center for Disease Control (CDC), and had CDC staff available to implement the project. County 1 is a Tibetan minority autonomous district. Its population was 217,000 in 2004, 15% of whom were in urban areas. In the 2000 census, 63% of the county's population were Han Chinese and 30% were Tibetan. County 2 is more populous than County 1 and is located in a different municipality, but it has a similar land area. Its 2004 population was over 500,000, of whom 13% were in urban areas. Nearly all were Han Chinese. Both counties fall in the middle third of Gansu's counties in terms of GDP per capita in 2004, with County 1 being somewhat poorer than County 2.[1]

County 1 consists of 22 townships, of which 19 participated in the program. These 19 townships have 101 primary schools. Ten of County 1's 19 townships were randomly assigned to the program in 2004, and the other nine were the control group. County 2 consists of 23 townships, (including the county seat) of which 18 participated in the program. These 18 townships have 155 primary schools. Nine of the 18 townships were randomly assigned to the eyeglasses intervention in 2004, and the other nine were assigned to the control group. In both counties, the excluded townships include the county seat (the counties' main urban centers, where incomes are higher and eyeglasses are easier to obtain) and a few townships in sparsely populated remote locations.

Random assignment was implemented as follows. In each county, all included townships were ranked by income per capita in 2003. Starting with the first two townships (the two wealthiest), one was randomly assigned to be a treated township and the other was assigned to the control group; this was repeated for all subsequent township pairs. In County 1, the 19th township (the poorest) was not paired with another township; it was randomly assigned to the treatment group. In each township primary schools were either all assigned to the treatment group or all assigned to the control group.[2]

[1] All figures in this paragraph except County 1 census data are from Gansu Statistical Yearbook (2005). County 1 census data are from searching the county's name in Wikipedia (accessed Nov. 23, 2011).

[2] Primary schools with less than 100 students were excluded from the project to avoid high travel costs to a few very remote schools. Students in such schools are only 6% of primary students in the two counties.

A baseline survey was conducted at the end of the 2003-04 school year (June of 2004) to collect data on student characteristics, academic test scores, and visual acuity. Data were collected from both treatment and control schools from all students finishing grades 1-5 in June of 2004. The treatment school students who would be entering grades 4-6 in the fall of 2004 and had poor vision were offered free eyeglasses. In each county, an optometrist was hired later that summer to travel to each township to conduct more in-depth eye tests for students who accepted the offer and had permission from their parents. If poor vision was confirmed, they were prescribed appropriate lenses. Students had a limited choice of colors and styles for their eyeglass frames. The Gansu CDC then ordered all of the eyeglasses from a company with an established quality reputation. The fall semester of 2004 began on August 26th, and most of the eligible students who accepted the offer received eyeglasses by mid-September. Teacher monitoring reports and field visits by project staff suggested a high propensity to wear the eyeglasses. At the end of the 2004-05 school year (late June or early July of 2005), exam scores for the 2004 fall semester and the 2005 spring semester were collected.

Unfortunately, in 5 of the 18 control townships students were given eyeglasses because, after providing the eyeglasses in the treatment townships, the remaining funds were used to buy eyeglasses for students with poor vision in the paired control township. This occurred in two control townships in County 1[3] and three control townships in County 2. In addition, in another township pair in County 1 the desire to treat more children led to a "role reversal" in which no one in the treatment township was offered glasses while a significant share of children with poor vision in the control township were

[3] In a third control township in County 1, four children in the control group received glasses, but three of these four did not have poor vision. This control township is retained in the analysis. Excluding it and its matched pair (or including them but dropping these four students) has very little effect on the results.

offered glasses. Thus there are six pairs of townships in County 2 and six pairs (plus the poorest township, the one randomly assigned to be treated) in County 1 for which the randomization was done according to the plan. Although this failure was unfortunate, we undertake a number of robustness checks to demonstrate that it does not substantially weaken our ability to convincingly identify a positive impact of wearing eyeglasses on academic performance, as discussed in detail in Section 5.

B. Data. Four sources of data are used in the analysis: 1) school records on basic student characteristics and exam scores before and after the intervention; 2) results of health exams, including vision tests, conducted by the county CDC in each primary school before eyeglasses were provided; 3) information from optometrists' records on the students fitted for eyeglasses; and 4) data from the Gansu Survey of Children and Families (GSCF), a longitudinal study of children in rural Gansu that is separate from the Gansu Vision Intervention Project. The basic information in the school records include students' grades during the 2003-04 and 2004-05 school years, students' sex, ethnicity, and birthdate, and the occupation and education level of the head of each students' household (usually the father). Scores on exams (Chinese, math and science) given at the end of each semester in each grade were also collected.[4]

One important characteristic of the test score variables has important implications for the analysis: in many cases schools design their own exams, so the test scores may not be strictly comparable across schools. Given random assignment of townships to the treatment and control groups, this non-comparability of exams across schools does not cause biased estimates, but it does add noise to the data, akin to a school random effect, that must be addressed in the estimation. This is discussed in detail in Section 4.

[4] In some schools, these exam scores are averages of several exams, including an end of semester exam.

The school health data include whether a student wears glasses (and if so, the student's grade when he or she first wore them), students' height, weight and hemoglobin count, and at least one measurement of vision for each eye (students who received glasses have additional measurements related to fitting them with eyeglasses). In China, doctors usually conduct eye exams by asking patients to read (with one eye covered) a standard eye chart from five meters away. The chart has 12 rows of the letter E facing in different directions; the top row has very large E's, and each subsequent row has smaller E's. If a patient cannot read the first row, which corresponds to the worst eyesight, his or her vision is coded as 4.0. If he or she can read the first row but not the second, his or her vision is coded as 4.1, and so forth. If the patient can read the 10^{th} row, the normal level, his/her eyesight is coded as 5.0. A patient who can read all 12 rows is coded as 5.2. There is information from optometrists only for the children offered eyeglasses; it includes whether a child was fitted for eyeglasses, and if not, the reason eyeglasses were declined.

The Gansu Survey of Children and Families (GSCF) was conducted in rural areas of Gansu province. The GSCF was first conducted in 2000 for a random sample of two thousand children aged 9-12. A second wave (GSCF2) was conducted in 2004; all but 131 of the original 2000 children were re-interviewed.[5] In addition, 886 oldest younger siblings of the original 2000 children, if they were 8 years old or older, were also interviewed in 2004. GSCF2 collected detailed information on vision and wearing eyeglasses from both sets of children and their parents, as well as data on lighting conditions at home and at school, the cost and availability of eyeglasses, and many household and village characteristics. In addition to self-reported vision data, GSCF2 also

[5]The reasons children were not reinterviewed include: 108 had moved out of the counties where they had resided in 2000; 8 died; 4 were seriously ill; 2 had parents who divorced; 1 household refused to be re-interviewed; and 8 children were not reinterviewed for unknown reasons.

collected objective measurements of each child's eyesight through an eye exam, for both the originally sampled children and the 886 younger siblings, conducted by professionally trained staff from the Gansu CDC.

C. Descriptive Statistics. Table 1 presents descriptive statistics, first for all 37 townships and then for the 25 "compliant townships", that is the townships that fully complied with their random assignment. Among all 37 townships there were 28,789 students in grades 4-6 in 2004-05 in the two counties. Of these students, 14.5% (4,177) had poor vision in the sense that either the left eye or the right eye (or both) had a visual acuity score below 4.9.[6] Only 3.2% of the students in these counties with vision problems (134 out of 4,177) already had eyeglasses. Those with vision problems had test scores almost identical to those of students without problems (78.8% vs. 78.7% for Chinese, 79.1% vs. 79.0% for mathematics, and 80.6% vs. 80.8% for science) at the end of the spring 2004 semester (1-2 months before the program began). Very similar patterns are also seen for the 25 compliant townships.

The data in Table 1 suggest that vision problems have little effect on students' academic performance. Indeed, simple t-tests show, for both counties separately and when combined, that none of the above-mentioned small differences in test scores is significant. But this conclusion may be misleading because study habits may affect eyesight. In particular, several medical studies have shown that doing "near-work", that is spending many hours doing activities with the eyes focused on objects about 1 meter away, can cause myopia (e.g. Angle and Wissmann, 1980; Mutti et al., 2002). However,

[6] Although children with a visual acuity score of 4.9 in one or both eyes were also offered eyeglasses, only 6.8% (17 out of 249) accepted. In contrast, 56.5% of children (109 out of 193) with a visual acuity score of 4.8 in one or both eyes accepted the glasses. Since the exact cutoff point between good and poor vision is somewhat arbitrary, this suggests that the cutoff point for poor vision should be below 4.9, as opposed to below 5.0. Indeed, the low take-up rate for children with a visual acuity score of 4.9 makes it impossible to estimate the impact of providing eyeglasses to those children.

the medical evidence is not unanimous; Lu et al. (2009) find no relationship between near-work and myopia. Thus, it may be that students who study more are more likely to develop myopia, the most common refractive eye problem.

Indeed, the data available before the Gansu Vision Intervention Program was implemented suggest that studying does harm students' vision. Among the grade 1 children in the data, very few have poor vision (only 3.0% are classified as having a visual acuity score below 4.8 in one or both eyes), but this increases dramatically as children continue in school (7.7% of students in grade 3, and 16.5% in grade 5). Thus children's grade 1 test scores are unlikely to be seriously affected by vision problems because so few have poor vision, but differences in visual acuity among older students would reflect, in part, time spent studying. OLS regressions of current visual acuity (mean over both eyes) on average test scores (average over Chinese, math and science) in grade 1, controlling for school random effects, current grade, and parents' education on the sample children in grades 3-5 in the 2003-04 school year show a *negative* impact of higher test scores on vision that is almost significant at the 10% level (p-value = 0.116). Thus, simple comparisons of test scores across students with good vision and students with poor vision are likely to underestimate the negative impact of vision on student performance (since students with good vision tend to study less). Similarly, for the GSCF data, a probit regression of poor vision in 2004 on study time in 2000 (controling for sex, age, parental education and expenditure), yields a significantly (10% level) negative impact of study time on visual acuity four years later.

Table 2 presents information on how the Gansu Vision Intervention Project was implemented for the students with poor vision, for both the full sample of 37 townships and the compliant sample of 25 townships. Of the 4,177 students with poor vision in the full sample, 1,978 were in the program schools and almost all (except for one township, as explained above) were offered eyeglasses (those who already had eyeglasses were offered new ones), while most of the 2,199 in the control group were not offered glasses (except for students in six control townships, as explained above). Of the 1,978 students in the treatment townships, 1,384 (70.0%) accepted, while the other 594 declined (or, in one township, were not offered) eyeglasses. The main reasons given for declining the offer were objection of the household head (187) and refusal by the child (89). Similar patterns hold for the 25 townships in the compliant sample; in particular, only 69.8% of the students offered eyeglasses accepted them.[7]

4. Methodology

Almost all primary school age children in Gansu province are in school; the GSCF data from the year 2000 show that only 1.4% of children age 9-12 were not enrolled in school. Thus, provision of eyeglasses cannot raise school enrollment; the sole impact is on academic performance. The random assignment of schools to participate or not participate in the Gansu Vision Intervention Project greatly simplifies analysis of the impact of the project on student learning. To ease interpretation, all estimates in this paper use standardized test scores as the dependent variable; test scores are standardized by subtracting the mean and then dividing by the (student level) standard deviation, using the control schools' mean and standard deviation, separately for each subject and grade.[8]

[7] The number of students with bad vision in the control schools in County 1 falls from 703 in the full sample to 112 in the compliant sample. This reflects the fact that one of the three control townships in County 1 where the interventions was improperly implemented was very large, accounting for nearly half of the students in the 10 townships in County 1 that were assigned to the control group. See below for further discussion of the comparability of the compliant sample to the full sample.

[8] Strictly speaking, this could be misleading because each school administered a different test. Yet the test score data we have range from 10 to 100 points for each test, and the school level medians are close to 80

A. Estimating the Impact of Offering Eyeglasses. The simplest estimate of the program's impact on children in grades 4-6 with poor vision is a t-test that compares the mean test scores of the students with poor vision in the program schools with the same mean for the students with poor vision in the control schools. This estimates the impact of the *offer* of eyeglasses (intention to treat effect), not the impact of receiving eyeglasses.

This t-test can be calculated by regressing the (standardized) test score variable (T) on a constant term and a dummy variable for enrollment in a program school (P):

$$T = \alpha + \beta P + u \qquad (1)$$

where u is a residual that is uncorrelated with P due to randomized program assignment. Reflecting the sample design, all regressions include a dummy variable for each pair of townships within which randomization was done (not shown in equation(1)). See Bruhn and McKenzie (2009) for a justification of this approach.

Estimates of β in equation (1) use only students with poor vision. To obtain more precise estimates of β one can use an estimation method that adds students with good eyesight. Intuitively, this "double difference" method compares the difference in test scores of children with good vision and poor vision in treatment schools with the same difference for children in control schools. By focusing on within-school variation, this specification is less subject to bias from imperfect randomization of treatment across schools; non-random assignment causes bias only if treated and untreated schools differ systematically in the differential performance of children with good and poor vision. It

(between 72 and 87 for 80% of the schools, with smallest being 61). Moreover, the school level standard deviations are around 10 (between 7 and 14 for 80% of the schools). Thus schools already normalize to some extent, and our normalization is not forcing schools with very different distributions onto a similar scale. Indeed, as a robustness check we normalized test scores using within-school means and standard deviations; the results are very similar. However, complications do arise because the tests used are not strictly comparable across schools; this is discussed further below.

also has the advantage that it compares only children who took the same test, since it is based on within-school comparisons; issues arising from the use of different tests are discussed below.

The equation to be estimated for this specification is:

$$T = \alpha + \pi PV + \tau P + \beta PV^{*}P + u \qquad (2)$$

where PV is a dummy variable indicating poor vision. In this specification the program's impact on students with good vision (PV = 0) will be τ, which one would expect to be zero, and the program's impact on students with poor vision will be $\tau + \beta$, which equals β if, as expected, τ equals zero. The τ coefficient is also a check on the randomization; if the schools that participated in the program were better (worse) than average, then τ would be positive (negative).[9] Finally, the estimate of π measures the impact of poor vision on test scores, which one would expect to be negative. Yet this estimate will be biased toward zero because students who study more are likely to have worse vision. Fortunately, correlation between u and PV does not lead to bias in the estimate of the program impact (β),[10] and neither does random measurement error in PV (see Appendix I).

For equations (1) and (2), adding other explanatory variables may lead to more precise estimates. Several child and parent variables were tried, but none increased precision. In contrast, adding students' test scores in the spring of 2004, before eyeglasses were provided, greatly increases the precision of the estimated program impacts. Note also that when pre-program test scores are added as controls, bias occurs in estimates of

---

[9] Even if randomization was perfectly implemented, τ could be different from zero if there were spillover effects of the program onto children with good vision. This is investigated in Section 7.

[10] One way to see this is to assume that the correlation takes the form $u = \theta PV + \varepsilon$, where ε is uncorrelated with both PV and P. Then equation (2) becomes $T = \alpha + \pi PV + \tau P + \beta PV^{*}P + \theta PV + \varepsilon = \alpha + (\pi+\theta)PV + \tau P + \beta PV^{*}P + \varepsilon$; this regression will not yield unbiased estimates of π, but the estimate of β is still unbiased. More generally, in equation (2) u is not correlated with PV*P after conditioning on PV.

equation (2) only if treatment status is correlated with differences in *changes* in student performance across children with good and poor vision. This is investigated in Section 7.

A final issue is how to obtain correct standard errors for the estimates of program effects. Standard errors should allow for heteroscedasticity of unknown form, as well as for correlation in the error term (u) across students in the same schools, and even students in different schools in the same townships, since unobserved school or township characteristics could lead to correlated errors for students in the same school or township. Indeed, as explained above schools often use their own tests, as opposed to county or province tests; Appendix 2 shows that this generates a school level random effect.[11]

The best approach to address this correlation is to use covariance matrices that allow for "clustering" of the error terms (see Wooldridge, 2010, Chapter 20). Yet for this paper the standard clustering formula has two disadvantages. First, OLS estimation of equations (1) and (2) that allows for correlation of unknown form at the township level loses information, leading to less precise estimates because these covariance matrices do not distinguish between students in the same school and students in different schools in the same township. The school random effects described above imply that the correlation of the error terms is likely to be much stronger for the first set of students. To account for this differential correlation, we estimate specifications with school random effects, which distinguish between students in the same school and students in different schools, *and* we also allow for correlation of unknown form for the error terms of students in the same township. This specification is consistent even if the error terms in equations (1) and (2) do *not* follow the "classical" random effects form (see Wooldridge, 2010, pp.866-67).

[11] Appendix 2 also shows that the use of different tests still leads to a consistent estimate of β.

The second disadvantage is that covariance matrices that allow for clustering of the error terms are valid only as the number of clusters (number of townships) goes to infinity. Our estimates, which drop township pairs for which the randomization was improperly implemented, are based on 25 townships. Several authors have shown that these covariance matrices can be misleading when there are 30 or fewer clusters (see Cameron et al., 2008). To check whether our estimates have this problem, we also present p-values estimated using the wild bootstrap, as Cameron et al. (2008) suggest.

B. IV Estimates of the Impact of Receiving Eyeglasses. The methods presented thus far estimate the impact of being offered eyeglasses, not the impact of receiving them. In general, the former impact will be less than the latter because students who did not accept the offer of eyeglasses will not benefit from them. OLS estimates of the benefit of receiving eyeglasses may be biased if parents and/or students who accept the eyeglasses differ in unobserved ways from those who decline the offer. For example, parents of students who accept the offer may be more interested in education and so may do other things that raise their children's test scores.

Instrumental variable (IV) estimation can be used to obtain consistent estimates of the impact of receiving eyeglasses. In particular, one can estimate the impact of treatment on the treated using the same equations presented above, replacing P (offer of eye-glasses) with "G", actually receiving eyeglasses.[12] While G may be correlated with the residual, P can be used as an instrument for G; by definition, P is uncorrelated with u, and also has strong explanatory power for G. Note that G = 1 not only for students who

[12] Strictly speaking, the IV estimates are local average treatment effects (LATE), i.e. estimates of the impact of wearing glasses for those students that were induced by the program to wear eyeglasses. Yet since very few students had eyeglasses before the program, LATE estimates are very close to the impact of receiving eyeglasses on those who actually received them (impact of the treatment on the treated).

accepted the glasses in the program schools but also for the few students who already had glasses, in either the program or control schools.

IV estimates of equation (1) are straightforward; one need only replace P with G and use P as an instrument for G. There is one complication with IV estimates of equation (2); replacing P with G in that equation yields $T = \alpha + \pi PV + \tau G + \beta PV*G + u$. Although one can be in a program school if one does not have poor vision, it makes little sense to wear glasses if one does not have poor vision, so that $G = 0$ whenever $PV = 0$, and thus G and PV*G are perfectly correlated.[13] Thus the IV estimates of equation (2) exclude the τG term. Finally, note that IV estimation is valid even if the randomization was not implemented as planned. As long as the *plan* was randomized then the instrument is valid if it has explanatory power for having eyeglasses.

A final complication with IV estimation is that the findings of Cameron et al. (2008) regarding the wild bootstrap have not been verified for IV estimation, so we are unable to report them.

5. Checks for Treatment/Control Balance and for Selectivity into Compliant Sample

Before presenting estimates of program impacts, we first check whether the treatment and control townships are similar, which should be the case since the treatment was randomly assigned. We then check whether the compliant sample suffers from problems of selection bias. This check is important because we prefer the compliant sample estimates to the full sample estimates, for two reasons: 1. The intention to treat (ITT) estimate for the full sample almost certainly underestimates the ITT effect of the

[13] This correlation is not exactly equal to one in the data (it is 0.86), but this is the case only because a very small percentage of students report wearing classes even though they have good vision.

program; and 2. The IV estimates of the impact of providing eyeglasses are more precisely estimated for the 25 compliant townships. For completeness, we report all results for both the compliant sample and the full sample.

A. Balance for the Full Sample. Table 3 presents evidence as to whether the treatment and control groups among the full set of 37 townships were similar before the program was implemented, which should hold given that the treatment was randomly assigned. The results are presented separately by county; combining both counties yields very similar results. The treatment and control means for 10 different variables are given in the first and second columns, respectively, and the third column reports the difference, with asterisks indicating statistical significance. Beginning with County 1, when all 10,208 students are compared, only two of the ten differences are significant, and only at the 10% level, and none of the bootstrapped p-values is significant. When comparing only the 1,552 students with poor vision, two out of nine are significant, but again only at the 10% level. These two differences are more significant when bootstrapped p-values are used, yet when all 19 bootstrapped p-values are compared there is one significant at the 5% level and another at the 1% level, which is not very different from random chance.

The bottom half of Table 3 presents the same comparisons between the treatment and control groups for County 2. When all 18,581 children are compared, none of the four test scores that are the focus of this study have statistically significant differences, and among all 10 variables only one (a dummy variable for poor vision) is significantly different at the 5% level. One statistically significant variable out of ten could be due to random chance. Moreover, recall that when the number of clusters is around 30 or fewer asymptotic approximations may lead to incorrect inference. Thus the last column of

Table 3 presents p-values based on the wild bootstrap (in addition, it includes school random effects for consistency with the results presented below). None of the 10 p-values indicates statistical significance at conventional levels, which indicates that the randomization was properly implemented.

The last nine rows of Table 3 focus on the 2,625 County 2 children in the full sample who have poor vision. Of the four test score variables, one is significant at the 10% level, which raises little concern, but the visual acuity variable is significantly different at the 1% level, and this is also the case for the bootstrapped p-values. However, of the 19 bootstrapped p-values for County 2 in Table 3 this is the only one significant at the 5% or 1% level, which could well be due to random chance. More importantly, there is no evidence of differences in any of the four test score variables.

We conclude that the differences between the treatment and control groups in Table 3 are consistent with random assignment. In fact, since we performed the random assignment ourselves we know that it was indeed random. For future reference, we note that among students with poor vision in both counties it appears that, by random chance, visual acuity is somewhat lower in the treatment group than in the control group.

B. Comparison of the Full and Compliant Samples. Twelve (six pairs) of the original 37 townships did not correctly implement the program. In five pairs, a large percentage of students with poor vision (ranging from 34% to 72%) in the control group schools were offered – and accepted – glasses, and in the sixth pair none of the students with poor vision in the treatment township were offered glasses, while 33% of the students with poor vision in the control township were offered, and accepted, eyeglasses. This provision of eyeglasses to one third of the control group will almost certainly lead to underestimation of the impact of the program on student learning. More specifically, it will underestimate the intention to treat (ITT) effect (the impact of offering eyeglasses to students with poor vision) of implementing the program in all schools in the two counties.[14] In contrast, estimates limited to the 25 clusters that followed random assignment will yield unbiased estimates *if there are no systematic differences between the full sample and the 25 clusters where the program was correctly implemented.*

Table 4 examines whether key variables have systematic differences between the 25 compliant townships and 12 "non-compliant" townships, separately for the two counties. (Estimates that combine both counties yield similar results.) The results for County 1 in the top half of the table show almost no statistically significant results and suggest no difference between the compliant and non-compliant sample. Of the 19 differences for County 2 in the third column of Table 4, none is significant at the 10% level and two are significant at the 5% level; this is very close to what one would expect if the true values all equal zero. Moreover, the p-values based on the wild bootstrap (with school random effects) in the last column of Table 4 indicate that these two significant differences in the means of the treatment and control townships may be significant only at the 10% level. Overall, there is no sign of selection bias in Table 4.

Another check of whether restricting the sample to the 25 compliant townships leads to selection bias is to focus on the 18 townships in the full sample that were randomly assigned to be control townships. The question is whether there are some school or student characteristics among the 105 schools in these 18 townships that make some of these schools more likely to provide glasses (more likely to violate their random

[14] One could argue that the full sample accurately estimates one kind of ITT effect, namely the effect of the program as it was actually implemented, i.e. when random assignment was not followed for one third of the sample. Yet almost all of the violations of random assignment involved offering eyeglasses to the control group, and in a "real" implementation of the program there would be no control group since all children with vision problems should be eligible. Thus such an estimate would be of little use for policy decisions.

assignment) than other schools; the differences examined in Table 4 included both treatment and control townships and so may have masked differences in the control townships. For example, if control schools in wealthier townships, or with better educated parents, or with fewer ethnic minority students, are more likely to pressure officials to provide their children with eyeglasses, that would suggest sample selection.

This was investigated using school level regressions for these 105 schools. Simple regressions (available from the authors upon request) show that none of the above-mentioned variables has any influence on which control schools were (mistakenly) provided eyeglasses. The only two variables with predictive power were that control schools with a higher percentage of students with poor vision were more likely to be provided glasses, and that control schools paired (at the township level) with treatment schools that tended to have fewer children with poor vision were also more likely to be provided glasses. This is consistent with field reports that non-compliance was due to local officials wanting to use funds left over after providing eyeglasses to all children in the treatment school decided to provide eyeglasses to students with poor vision in the control townships.

C. Balance for the Compliant Sample. A final check is to assess whether the treatment and control townships are well balanced for the 25 compliant townships. Results are reported in Table 5. Comparing the differences for all County 1 students, two of the ten differences are significantly different at the 5% level, and the same is true for the bootstrapped p-values. While this is more significance than one would expect, which is one variable significant at the 10% level, this level of significance could also be due to random chance. When the sample is limited to students with poor vision in County 1, one variable, visual acuity, is significant at the 1% level in both the third and fourth columns, and one variable, ethnic minority, is significant at the 10% level in column 4 but not in column 3. Thus the only remarkable difference among these students is the same as that seen in Table 3 which compares the treatment and control samples that we know were randomly sampled.

The bottom half of Table 5 presents the same comparisons for County 2. When all students are compared, none of these 10 differences is significant, which indicates a well-balanced random assignment. When the sample is limited to children with poor vision, the only variable with a significant difference is visual acuity, which is significantly different at the 1% level. Yet, as in Table 3 this could well be due to random chance. The bootstrapped p-values in the fourth column of Table 5 show the same levels of statistical significance seen in the third column. Overall, the results in Table 5 are broadly consistent with random assignment, and again since we implemented the randomization we know that the assignment was indeed random.

A last check for selection bias is to compare the difference in means of the baseline variables between the treatment and control townships for the compliant townships with the same difference for the non-compliant townships. The main worry is that differences in initial test scores between treatment and control townships in the compliant townships are systematically different from the same differences in non-compliant townships. For example, the impact of the program could be underestimated if the randomization plan was violated primarily in township pairs where the treatment township had higher test scores than the control township (because this behavior implies that, in the compliant townships, the treated townships would tend to have lower test scores than the control townships). We find that the difference in these two differences is not statistically significantly different from zero for any of the test scores, nor for the average test score.[15]

To summarize this section, checks for balances and differential attrition indicate the following. First, the results in Table 3 indicate that the full sample of 37 townships is generally well balanced between the treatment and control townships. Second, the results in Table 4 show no evidence of pre-program differences between compliant and non-compliant townships, which suggests that selection bias should not affect results based on the compliant townships alone. Finally, balance checks for the 25 compliant townships are also generally consistent with good balance between the treatment and control townships; the main possible exception is differences in visual acuity, but for the most part this simply reflects a feature of the full sample of 37 townships and thus is unlikely to result from selection bias. Most importantly, there is no difference in test scores in any of these checks, so we conclude from Tables 4 and 5 that estimates based on the compliant sample are unlikely to be affected by selection bias.

6. Estimates of Program Impact

This section presents estimates, for both the 25 compliant townships and the full sample of 37 townships, of the impact of the Gansu Vision Intervention Project on the test scores of students in grades 4-6 in the spring semester of 2005. Thus these results measure the impact of the project after one academic year. As explained above, all test scores have been normalized separately for each subject and grade. To increase precision, all estimates condition on pre-intervention (spring 2004) test scores.

[15] Results available from the authors upon request. Among other covariates, only ethnic minority is unbalanced for the compliant sample, with the difference between treatment and control townships also being statistically different from the difference in noncompliant townships.

A. Estimates of the Impact of Being Offered Eyeglasses. Table 6 presents estimates of equations (1) and (2) using both the sample from the 25 townships that implemented the program correctly and the full sample of 37 townships.[16] Results are reported for County 1, County 2, and the pooled sample. All estimates include school random effects and allow for heteroscedasticity, including correlation (clustering) among observations within the same township, of unknown form for the individual level error term. Coefficients on lagged test scores, constants, and strata indicators are suppressed to avoid clutter; the coefficents on lagged test scores for the compliant sample are presented in Appendix Table 1.

Interestingly, the estimated impacts are much stronger in County 1 than in County 2. For County 1, the estimated impact on test scores using the sample of children with poor vision is 0.39 standard deviations, which is highly statistically significant even when assessed using wild bootstrap (p-value of 0.026). Adding children with good vision as an additional control group, the estimated impact is 0.26 standard deviations and statistically significant. In contrast, the estimated impact for County 2 is 0.08 (0.07) standard deviations and statistically insignificant for the sample of poor vision (all) children,.

Given the strong impacts for County 1, for that county we further investigate the impacts on tests of different subjects (none of the subject-specific impacts is statistically significant in County 2). We find that the impact is greatest for Chinese, 0.41 (0.29) standard deviations using the sample of poor vision (all) children, and smallest for science, 0.26 (0.15) standard deviations for the sample of poor vision (all) children. Using non-bootstrapped standard errors, all of the impacts on subject-specific test scores

[16] Estimates of equation (1) classify students whose worst eye has a visual acuity score of 4.9 as having good vision. Yet recall that such children were offered glasses, and 17 out of 249 accepted them. Those 17 are excluded from the regression. Dropping all 249 of these children from the sample does not affect the results.

are statistically significant; using the wild bootstrap, estimated impacts are significant for the Chinese test (for both samples, children with poor vision and all children) and the science test (only for the sample of all children).

If one believes that the results for County 1 reflect the true potential of offering eyeglasses to children with poor vision, then the County 1 estimates can be regarded as the preferred estimates. If, however, one gives equal weight to the insignificant impacts found in County 2, then it is best to estimate impacts pooling the data from both counties. Much depends on our assessment of the reasons for not finding significant impacts in County 2. When we pool the sample, the estimated treatment effect on all test scores is 0.16 (0.11) standard deviations using the sample of poor vision (all) children. The estimates are still statistically significant using the non-bootstrapped standard errors, but using the wild bootstrap the impact is significant at conventional levels only for the sample of all children (p-value of 0.048, compared to p-value of 0.214 using the sample of children with poor vision).

Thus far, we have presented our preferred estimates using the compliant sample only. For transparency, in the bottom half of Table 6 we also report the results using the full sample of 37 townships. Due to non-compliance we expect these estimates to significantly underestimate the true program impacts. For this reason, they should be considered lower bounds for the true ITT effects. The estimated impacts will undoubtedly be smaller than those found for the compliant sample because adding one "treatment" township that did not offer the treatment and six "control" townships that offered the treatment to at least some students will reduce the gap in test scores between townships that were randomly assigned to be treated and those randomly assigned to be controls (assuming that the treatment has an effect).

Compared to the estimates for County 1 using the compliant sample (0.39 and 0.26 standard deviations for the samples of poor vision children and all children), the estimated impact on test scores is smaller, as expected: 0.16 (0.08) standard deviations for the sample of poor vision (all) children. Nonetheless, we still find some statistically significant impacts of the offer to provide eyeglasses in that county. Using non-bootstrapped standard errors, the estimated impacts using the sample of poor vision children are marginally significant (at 10% significance level) for the average test score and significant at the 5% level for Chinese and math scores. Using the wild-bootstrap, impacts are significant for math test scores at the 5% (10%) level for the sample of poor vision (all) children and nearly significant at the 10% level for Chinese scores for the sample of poor vision children. None of the other impacts is statistically significant, including all estimates for County 2 and the pooled sample; however the fact that quite a few of the estimated impacts remain statistically significant in County 1 despite substantial downward bias is notable.

The finding that eyeglasses had a much larger impact in County 1 than in County 2 is somewhat surprising and merits further discussion. One would not expect the true impacts of wearing eyeglasses to be very different in these two counties given their similar school and home environments. Measured impacts can be weakened by problems in project implementation or data collection, but it is more difficult to imagine plausible reasons why such problems would lead to overestimation of impacts.[17] Consider first whether differences in how the data were collected in the two counties could explain the

[17] Manipulation of data to find stronger positive impacts is implausible in our context, as pre-treatment test scores were collected from school records before eyesight was tested or eyeglasses provided, and the post-treatment test scores were also collected from school records without reference to the treatment status of individual children (test scores for all students were collected from administrative records and matched later to pre-treatment test scores). Moreover, there is no obvious incentive for those who collected the data in County 1 to exaggerate program benefits.

lack of impact in County 2. In both counties, the county-level CDC was in charge of the project. In County 2, with the support of the county Education Bureau, test score and other data were collected in a highly decentralized way, with electronic Excel files sent to schools to be filled in by teachers or school officials and then returned to the CDC. In contrast, in County 1 all the data were collected in person by a relatively small number (6 to 8) of well-trained county CDC staff who participated in training sessions led by the authors. These staff visited each school, filling in paper forms, which were later entered into computers. County 2's decentralized approach was loosely monitored, which likely led to more variable data quality. The teachers were busy and had little incentive to do the data collection carefully. In particular, accurately matching post-treatment test scores to long lists of students is tedious, painstaking work. They may not have had a good understanding of the project's objectives, especially those in the control townships. In contrast, County 1 CDC professional staff were directly trained by the authors and accountable to the county CDC Director and provincial CDC staff for the quality of their work; they thus were highly motivated and did their work conscientiously. The County 1 CDC has an exceptional reputation in Gansu province for the quality of their work (unlike County 2), which is one reason why that county was chosen as a study site. Indeed, when the original data files were received, many more problems were found in the County 2 data.

It is also possible that other aspects of project management were superior in County 1, for example in monitoring whether students continued to wear their eyeglasses, replacing lost or damaged glasses, and ensuring that eyeglasses in fact were given to students with the poorest vision. County 2 is a much more populous county than County 1, so that monitoring the project was more difficult for the CDC there. Because implementation was highly decentralized, it is much more likely that eyeglasses were distributed in ways inconsistent with the criteria set by the project, and that other monitoring and data collection were performed poorly and/or inconsistently.

B. IV Estimates of the Impact of Wearing Eyeglasses. This subsection presents IV estimates of the impact of wearing eyeglasses for one year on student test scores, for both the 25 compliant townships and the full set of 37 townships (Table 7).[18] Note that, in contrast to the ITT estimates, the full sample is not expected to systematically underestimate the impact since the instrument used (the planned randomization) is a valid instrument for both samples. Yet the compliant sample is still preferred because it presents more precise estimates; the first stage relationship between receiving glasses and being qualified for glasses based on the initial randomization is much stronger than for the full sample due to lack of compliance in the full sample. In the regressions including only children with poor vision the R^2 of the first stage regression is 0.495, the t-statistic for the program township variable is 19.41, and the F-test for the explanatory power of that variable (after excluding all other covariates) is very high (283.05). In contrast, for the full sample the R^2 coefficient is much lower, 0.305, the t-statistic for the program townships variable is only 7.72, and the F-statistic for the explanatory power of that

[18] Some students had worn eyeglasses for more than one year; of the 1,245 children with glasses, 199 had obtained them on their own, of whom 94 obtained them one year ago, 85 obtained them two years ago, and 20 obtained them 3 or 4 years ago, so only 105 of the 1,245 children had them for over one year. Recall that only 59 children in the sample with bad vision had glasses; thus 140 of the 199 children who report having obtained eyeglasses on their own do not appear to have had bad vision. This could reflect a mis-diagnosis that led their parents to obtain glasses for them, or measurement error either in the visual acuity variables or the variable indicating wearing eyeglasses. Measurement error in reported wearing eyeglasses does not imply inconsistency since that variable is instrumented. Measurement error in visual acuity could lead to selection bias in the regressions that include only children with poor vision, but the direction of bias is ambiguous; children with poor vision who wear glasses and were mistakenly dropped from the sample probably have relatively mild vision problems, so excluding them removes children whose benefit from having glasses is modest, leading to upward bias in the estimated impact of glasses, yet including children with good vision who do not wear glasses will increase the test scores of children without glasses, leading to downward bias. Measurement error in visual acuity (PV) is unlikely to cause bias in estimates that have both poor vision and good vision children; that is, the analysis in Appendix I extends to IV estimation.

variable is only 14.22, which is not much higher than the conventional cut-off point indicating weak instruments.

The IV estimation results are presented in Table 7. As expected, all of the estimated impacts are larger than the ITT estimates, because they reflect the impact of actually receiving eyeglasses rather than just being offered them. Similar to the ITT estimates, for County 1 when using the compliant sample all of the estimated impacts are statistically significant (top half of Table 7).[19] In particular, the impacts on all test scores are significant at the 1% level for both the sample of poor vision children and all children. The estimated impact on all test scores is 0.68 (0.41) standard deviations for the sample of poor vision (all) children, and the subject specific impacts for poor vision (all) children are 0.72 (0.46) standard deviations for Chinese, 0.47 (0.33) standard deviations for math, and 0.45 (0.23) standard deviations for science.

Similar to the ITT estimates, the impact estimates are much smaller and statistically insignificant for County 2. Pooling the compliant samples for the two counties, the impacts on test scores are 0.22 (0.16) standard deviations for the sample of poor vision (all) children, and still significant at the 5% level.

Next, we turn to the IV estimates for the full sample of 37 townships. Results are presented in the bottom half of Table 7. Most notably, we find that the positive impacts in County 1 for the sample of children with poor vision remain large and statistically significant (at the 1% level), and that the estimated impacts are generally similar to those for the compliant sample. The estimated impact on all test scores is somewhat smaller (0.54 standard deviations) than for the compliant sample (0.68 standard deviations), with

[19] As noted above, the IV estimates' standard errors are not bootstrapped. Yet for the (preferred) compliant sample note that ITT estimates' statistical significance is generally robust to using the bootstrap, and the statistical significance of the non-bootstrapped IV estimates and non-bootstrapped ITT estimates is similar.

the impact on Chinese and math test scores being greater (0.75 and 0.56 standard deviations, compared to 0.72 and 0.47 standard deviations for the compliant sample) and the impact on science test scores being smaller (close to zero compared to 0.44 standard deviations for the compliant sample). These results are important, because they show that for the county where strong impacts are found, despite the compliance problems, a regression specification using an identification strategy that relies on the original randomized design still estimates significant impacts of magnitudes similar in size to estimates using the compliant sample.

One weaker result using the full sample of townships in County 1 is that we do not find significant impacts using the IV specification for the full sample. However, note that the impact estimates for all test scores, Chinese test scores, and math test scores are all greater in magnitude using the full sample than when using the compliant sample (although science tests have smaller impacts), suggesting that focusing on the compliant sample is not leading to estimates that are systematically biased upward, and that the insignificant IV estimates for the sample of all children are driven by weaker power rather than by smaller impacts.

For county 2, just as for the ITT results, the estimated impacts are small and statistically insignificant using the full sample. When we combine the samples for the two counties, we also get imprecise estimates that are not statistically different from zero, in contrast to the significant results for the compliant sample, undoubtedly due to the weaker power of the estimates using the full sample.

In summary, focusing on the more precise estimates based on the 25 compliant townships, our estimates indicate that wearing eyeglasses for 8-9 months raises grade 4-6 students' test scores in County 1 by 0.41 standard deviations or more of the distribution of test scores, which is a large impact for such a short time. If we consider the small and statistically impact of eyeglasses in County 2 to be a valid estimate of the treatment effect in that county, then our preferred impact estimate using the pooled sample is 0.16 to 0.22 standard deviations, which we consider to be a lower bound estimate of the true treatment effect. One can express the treatment effect in terms of an equivalent gain from additional time in school. The 2000 GSCF administered identical Chinese and math tests to children in grades 4, 5 and 6. Relatively few of the children were in grade 6, so we focus on grades 4 and 5. The mean test scores of grade 5 students were 0.37 standard deviations higher in Chinese and 0.51 standard deviations higher in math than the mean scores of grade 4 students. Comparing the average gains on these two tests (0.44) with the estimated gains from wearing glasses of 0.41 standard deviations in County 1 and 0.16 to 0.22 in the two counties together, the impact of wearing glasses is equivalent to 0.9 years of schooling in County 1 and an additional one third to one half of a year in school for the pooled sample. Put another way, providing eyeglasses to students with poor vision raised learning per year of school by 90 percent in County 1, or increased it by 33 to 50 percent when the two counties are combined.

7. Robustness Checks

The estimates in Section 6 rely on assumptions that could be challenged. First, the estimates that compare children with poor vision to those with good vision assume that providing the former glasses does not affect the latter's test scores. Second, all estimates that compare children with good and poor vision assume that changes in test scores over time would have been similar for both groups in the absence of the program. This section checks the validity of these assumptions.

Consider first whether children with good vision were affected by the program. They could have benefited if their teachers spent less time helping students with poor vision, or if they learned from their now better performing classmates with poor vision. If this were the case, estimates of equation (2) would underestimate the true impact of the program on students with poor vision, since comparing students with poor vision to those with good vision ignores positive spillovers of the program onto the latter. Conversely, teachers may have been distracted from general teaching by the need to monitor students who were given glasses, or may have given more attention to those now better-motivated students. This would lead to overestimation of the program's impact on the students who were offered glasses. Finally, if the intervention affects students with good vision, these impacts should be included when evaluating the total social benefits of the intervention.

Table 8 presents estimates similar to those for equation (1) in Table 6, except that the sample includes only students with good vision, instead of students with poor vision. None of the six estimated program impacts is either large or statistically significant; they range from -0.044 to 0.006, none has a t-statistic above 0.5, and the wild bootstrap p-values are 0.69 or larger. Averaging over all three tests, the estimated effect is very small, -0.022, with a 95% confidence interval ranging from -0.148 to 0.104, ruling out impacts of 0.11 or above. Finally, estimates that allow the program's impact to vary by the proportion of children with bad vision in a student's grade in his or her school (spillovers should be larger in classrooms where more children received eyeglasses) also show no effect of any kind (not shown in Table 8). We conclude that there is no evidence of sizeable spillover effects, and thus that the estimates in Section 6 are unlikely to be biased due to spillovers.

Next, perhaps the checks for pre-program differences in Table 5 are insufficient for estimates that compare 2005 test scores, conditional on 2004 scores, across students with good vision and poor vision, because even if the test score *levels* in the spring of 2004 were similar across the treatment and control groups it is possible that the *changes* over time differ across those groups. This is examined in Table 9, which re-estimates the specification in Table 6 that includes students with poor vision and with good vision, but does so using data from one year earlier. If the relative changes in test scores for these two groups of students were sufficiently different in the treatment and control schools, one would find a "program effect" even before the program began. Yet there is no evidence of such an effect; using the average over all three tests the estimated "program effect" is only -0.02, and insignificant. Thus we find no evidence of possible bias in estimates of (2), which compare children with poor vision to children with good vision.

8. Heterogeneous Treatment Effects

The impact of providing eyeglasses may vary for different types of students. This section examines such variation by students' visual acuity and initial (2004) test scores. For brevity, the results are not disaggregated by county.

Perhaps the most obvious dimension along which the impact of eyeglasses would vary is by students' visual acuity; students with particularly bad vision should benefit the most from this intervention. This is examined in the first column of Table 10. Among students with poor vision (visual acuity below 4.9) are students with very poor vision, which we define as visual acuity below 4.4. Using this definition, about 20% of students with poor vision have very poor vision. The first set of results in Table 10 includes only children with poor vision; it finds a positive program impact but no additional impact on children with very poor vision. Indeed, the additional impact point estimate is negative, though it is far from significant. Adding students with good vision to the regression (i.e.

estimating equation (2)) gives a similar result. Thus there is no evidence that children with very poor vision benefit more from the program. This could reflect compensatory behavior by some children with poor vision, e.g., sitting in the front of the classroom, that diminishes the impact of the degree of poor vision on academic performance.

Another possible dimension of program heterogeneity is by students' initial performance; students with poor vision *and* relatively low academic performance may benefit more (in terms of improved learning) than students with poor vision whose academic performance is average or above average. This is examined in the second column of Table 10. When only students with poor vision are included, the impact is lower for students with higher initial (2004) test scores, but the expected negative coefficient on the interaction is insignificant. Yet when students with good vision are added to the regression the (triple) interaction effect is somewhat larger and more precisely estimated, so that it is statistically significant at the 1% level (with a bootstrapped p-value of 0.014). Recalling that the average 2004 score was normalized to zero, these estimates imply that average students experience an increase of 0.11 standard deviations, while below average students (defined as those whose 2004 average test score was one standard deviation below the mean) had a gain of 0.27 standard deviations, and above average students (those whose 2004 average score was one standard deviation above the mean) experienced a small loss of 0.06 standard deviations.[20] Thus providing eyeglasses appears to lead to more equitable educational outcomes among students with poor vision.

[20]These calculations use the fact that, for students with poor vision, the average 2004 test score was -0.16, and the standard deviation was 0.97. So the impact for an average student is $0.081 - 0.166 \times (-0.16) = 0.108$.

9. If the Benefits Are So Large, Why Do Some Children not Wear Eyeglasses?

The eyeglasses provided by the Gansu Vision Intervention Project cost about 120 yuan (about $15 U.S.). As explained above, their estimated impact on learning after only one year of use is equivalent to one third to one half of a year of schooling, which should lead to higher wages when a student enters the workforce. De Brauw and Rozelle (2007) estimate that each year of schooling in rural China increases wages by about 9.3% for those under 35 years old. Our own estimates of a Mincerian wage function using data on family members aged 15 to 35 from Wave 2 (in 2004) of the GSCF yield a much lower estimate: 4.6%. The GSCF data also indicate that a wage earner age 15-25 who completes lower middle school (grade 9) earned about 710 yuan per month, or about 8,520 yuan per year. Using the most conservative estimate of the impact of schooling on wages, and assuming that the program effect is equivalent to an increase of only one third of a year of schooling, the program should increase such a wage earner's annual income by 128 yuan per year (8520*0.33*0.0456). Assuming that this person finishes grade 9 and then works 40 years before retiring, the present discounted value (PDV) of this increase in wages easily exceeds the cost of glasses; even using a 10% discount rate the PDV will be 830 yuan, and using a more reasonable 5% discount rate yields a PDV of 1,834 yuan.

These large benefits from wearing eyeglasses, relative to their cost, combined with many refusals of free eyeglasses[21] and very infrequent wearing of eyeglasses absent the intervention, point to a failure to make what appears to be a high-return investment. A better understanding of the causes of this failure has important policy implications. Is the cost of obtaining good quality eyeglasses too high, especially for the poor who may

[21] As explained above, only about 70% of the students with poor vision in the program schools accepted the eyeglasses, even though they entailed no cost. The stated reasons for not accepting them are not very informative, the two most common being "child refused" and "household head refused" (see Table 2).

be credit-constrained? Even if eyeglasses are offered at no cost, and at a nearby location, parents may hesitate because accepting the offer may be thought to entail an obligation to purchase new glasses in later years should the original pair be lost or broken, or should the child's prescription need to be updated.

Alternatively, parents may simply be unaware of their children's vision problems, or may incorrectly believe that eyeglasses will weaken their children's eyes or that poor vision has little effect on learning at a young age. Even if parents are advised that their children need eyeglasses, they may doubt this advice, or think that their children's vision problems are minor and not worth having their child fitted for eyeglasses. Alternatively, parents may view eyeglasses as useful only for schooling, and may have low educational aspirations for their children. Parents may also be influenced by community norms on the value of eyeglasses, or of education. To explore these possibilities, we investigate which children accept the eyeglasses offered by the project, and we also use the 2004 GSCF data to estimate the determinants of wearing glasses in the absence of an intervention.

Table 11 presents probit estimates of the factors associated with accepting the eye-glasses offered in the project schools. The first variable to check is students' visual acuity; children with minor vision problems have less reason to wear glasses, while those with serious problems would benefit more. As expected, better visual acuity (average over both eyes) has a highly significant negative impact on accepting glasses.[22] The standard deviation of the visual acuity variable is 0.234, so raising visual acuity by a standard deviation reduces the probability of accepting eyeglasses by 11.2 percentage points (0.234×0.479).

[22] Other specifications were tried. For example, parents may feel that a child whose average visual acuity is below 4.8 does not need glasses if one of the eyes has normal visual acuity. Yet regressions using the acuity of the best eye, or the worst eye, or the difference between the two eyes had no added explanatory power.

One unexpected result is that girls are less likely to accept eyeglasses than boys: 73.6% of boys accepted them, compared to 66.0% of girls. The regression results show that girls have an 8.1 percentage point lower probability of receiving glasses, a highly significant difference. The reasons for this are unclear. The stated reasons for not accepting eyeglasses are similar for boys and girls. Anecdotal evidence suggests that girls may worry more than boys that wearing glasses makes them less attractive.

Four other factors significantly affect the probability of accepting eyeglasses. First, those children with poor vision who already wore eyeglasses (49 of 1528) were 17.5 percentage points more likely to accept new ones. This is unsurprising given that they were already convinced of the need for glasses, and many may have needed a new prescription. Two other results are that children in households headed by a schoolteacher or a village cadre were less likely to accept glasses. These effects are significant at the 1% and 10% levels, respectively, and are very large, with schoolteachers' children 22.0 percentage points less likely to accept them, and village cadres' children 33.6 percentage points less likely to accept them. Perhaps these local authority figures decline program benefits to avoid being perceived as manipulating the program for personal benefit. Alternatively, it may be, rather ironically, that these authority figures had serious doubts about the merits of eyeglasses. Fourth, students in wealthier townships were more likely to accept the eyeglasses offered; a one standard deviation increase in average township income raises the probability of accepting glasses by 7.1 percentage points. Perhaps the residents of wealthier townships are more accustomed to both children and adults wearing eyeglasses.

Finally, four plausible factors had no significant impact on accepting the offered eyeglasses. First, and rather surprisingly, more educated parents were no more likely to accept them (indeed, the point estimate is slightly negative). Second, students' initial test scores had no effect. Third, the main ethnic minority in these two counties, Tibetans (about 14.5% of the students), were less likely to accept the eyeglasses, but this effect is statistically insignificant. Finally, there was no difference in acceptance by grade level.

Further insights can be obtained from examining the 2004 GSCF data. We examine 925 children who were in primary school (and between the ages of 8 and 15) in that survey, 413 of whom were the "index" children from the 2000 GSCF and 512 of whom were younger siblings of those index children. These data contain much more information, including vision-related information, than do the school records of the students who participated in the Gansu Vision Intervention Project.

Evidence for the hypothesis that many parents are unaware of their children's vision problems is seen in the 2004 GSCF. Mothers were asked to assess their children's vision using five categories, from very good to very bad. As seen in Table 12, the vast majority (86%) of mothers of children with good vision, as measured by trained optometrists, correctly report that their children had good or very good vision. Yet 82% of those whose children's vision was only fair, and 62% of mothers whose children's vision was poor, also stated that their children's vision was good or very good.

Similar findings are found when these children were asked if they had vision problems, as seen in Table 13. Children with good vision or with fair vision rarely report vision problems (difficulty seeing the blackboard in school, trouble doing homework due to poor vision, and eye pain at home when studying in dim light). Children with poor vision are much more likely to report problems – 30.4% cite difficulty seeing the blackboard in school, 26.1% report trouble doing homework due to poor vision, and

29.0% cite eye pain at home when studying in dim light – yet for each of these problems about 70% of students with poor vision report not experiencing the problem.

One can apply regression analysis to the 2004 GSCF data to examine almost all of the hypotheses presented at the beginning of this section. Of the 925 children in primary school who were 8-15 years old in that survey, 23 (2.5%) report wearing glasses. The following data in the 2004 GSCF survey are useful for assessing these hypotheses: mothers' and fathers' assessments of their children's vision; mothers' estimates of the cost of eyeglasses and of the distance to the nearest locality where eyeglasses are sold; parents' reports of whether they wear eyeglasses; community literacy rates; and parental aspirations for their children's education.

Table 14 reports the results of five regressions. Given the very small share of children wearing glasses, the marginal effects of changes in the explanatory variables are small in percentage terms (but large relative to the base percentage wearing glasses). Nonetheless, the results are highly suggestive regarding the factors affecting the decision to wear eyeglasses. We focus on the coefficients that are statistically significant, and report the marginal effects of the fullest specification in the last column of Table 16.[23]

The first regression (Column 1) is the most parsimonious. It shows that children's visual acuity has a highly significant negative impact on the probability of having glasses, as expected. Unlike the Table 11 results, child sex has no effect, and older children are more likely to report having eyeglasses; while this latter result may seem to reflect that older children have more vision problems, the regression already controls for visual acuity and so this result may reflect more parental acceptance of eyeglasses for older children.

[23] Regressions that include parental aspirations are excluded from Table 16; adding those variables reduces the sample size, and the results are generally insignificant.

Mothers', but not fathers', education has a strong positive impact on having eyeglasses. Finally, households with greater per capita expenditures are more likely to purchase eyeglasses for their children, conditional on these other factors. This is similar to the results of Hannum and Zhang (2012) who, using the GSCF data, find that household wealth levels in the year 2000 were significantly positively associated with wearing glasses at age 13-16 in 2004.

Column 2 in Table 14 considers whether lack of awareness of children's vision problems reduces their probability of having eyeglasses. Mothers who think that their children have poor vision are much more likely to obtain glasses for them, but fathers' assessments have no significant effect; the coefficient is positive but lower than that for mothers. Note that the estimated impact of the child's actual visual acuity weakens (from -1.33 to -0.86); this suggests that mothers' perceptions matter more than actual acuity, and is consistent with many mothers not knowing their children's visual acuity.[24]

The next set of results examines whether perceived price and distance dissuade some parents from obtaining eyeglasses for their child. Price has the expected negative effect, and is significant at the 10% level. In contrast, the effect of distance is small, insignificant, and in an unexpected direction. Adding an interaction between price and log per capita expenditure did not produce a statistically significant coefficient.

The mothers or fathers of 37 (4%) of the 925 children in the sample report that they themselves wear eyeglasses; such parents presumably understand their benefits. The fourth column in Table 14 shows that, controlling for all the variables discussed thus far, having a parent who wears eyeglasses has a strong positive effect on the probability that a

[24] A related issue is whether some parents think that providing eyeglasses will increase the deterioration in their children's vision. There is no information in the 2004 GSCF on this attitude, but in the 2007 GSCF a new sample of mothers was asked, and about 25% opined that glasses would worsen their child's vision.

child has eyeglasses. Indeed, when this variable is added the impact of the child's actual visual acuity falls (in absolute value) to -0.67 and becomes statistically insignificant.

Finally, the last column in Table 14 examines whether community characteristics have effects beyond those of parent and child characteristics. The community literacy rate, an indicator of the value placed on education by, and the general socio-economic status of, the community, has a significant positive impact on a child's probability of having glasses.

10. Summary and Conclusion

Vision problems create a potential barrier to learning for about 10% of primary school age children in both developed and developing countries. Fortunately, almost all childhood vision problems are easily corrected by correctly fitted eyeglasses. In developed countries such as the U.S., public programs such as Medicaid and the Children's Health Insurance Program pay for children's eye exams and some NGOs provide free eyeglasses to children from poor families.[25] In contrast, in developing countries very few children with vision problems have eyeglasses, especially at the primary level, and these children are rarely assisted by either public or private organizations.

This paper examines the impact of providing eyeglasses to school age children with poor vision in rural areas of Gansu province, one of China's poorest provinces. A randomized control trial was implemented in 37 townships of two counties in Gansu, which included about 29,000 children in 251 schools, of whom about 12% had poor vision. Our preferred estimates, which are based on the 25 compliant townships, indicate that offering eyeglasses to children with poor vision increases their test scores (averaged

[25] In the U.S., NGOs providing free eye exams or glasses include Vision USA, Lions Clubs, New Eyes for the Needy, Sight for Students, Helen Keller Foundation International, and Essilor Vision Foundation. Helen Keller Foundation International runs programs on a very limited scale in a number of developing countries.

over three subjects) by 0.26 to 0.39 standard deviations of the distribution of those test scores in County 1, and by 0.11 to 0.16 standard deviations for the two counties combined.

For about one third of these children, either they or their parents refused the offer of free eyeglasses, which suggests that the impact of actually wearing eyeglasses is about 50% higher than these estimates. Indeed, instrumental variable estimates of the impact of wearing eyeglasses yield estimates of between 0.41 to 0.68 for County 1, small and statistically insignificant for County 2, and 0.16 and 0.22 standard deviations for the two counties together. Depending on what one believes about the reasons for lower impacts in County 2, one can consider the estimates for County 1 or the pooled estimates to be preferred. However, given that the economic, school, and environmental conditions do not differ dramatically in the two counties, it seems unlikely for the true impacts of wearing eyeglasses in school to diverge so greatly, suggesting that explanations related to the implementation of the project or data collection are more likely explanations for the large difference in results. Note in particular that measured impacts can be weakened by problems in project implementation or data collection, yet it is more difficult to imagine plausible reasons why such problems would lead to overestimation of impacts. Even the estimates that are averages over the two counties are rather large effects, equivalent to one third to one half of a year of schooling. Simple calculations suggest that the benefits in terms of higher wages greatly outweigh the costs. Thus our results indicate that providing eyeglasses is a low cost and easily implementable intervention that could improve the academic performance of a substantial proportion of primary (and secondary) school students in developing countries.

Our finding that providing eyeglasses to children with poor vision increases their test scores is not particularly surprising. Perhaps the more important question is to understand why parents do not obtain eyeglasses for their children. Our estimates suggest that parental misperceptions, especially the impression that their children's eyesight is adequate, play a major role. There is also evidence that low income and perceived high prices may be contributing factors. Yet more research, perhaps including additional randomized trials, is needed to understand parents' decision making with respect to providing glasses for their children, and to design policies that will ensure that all school age children in developing countries who need eyeglasses will in fact have them.

References

Afridi, Farzana. 2011. "The Impact of School Meals on School Participation: Evidence from Rural India", *Journal of Development Studies* 41(11): 1636-1656.

Angle, John, and David Wissmann. 1980. "The Epidemiology of Myopia." *American Journal of Epidemiology* 111(2):220-228.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics*, 122(3):1235-1264.

Barro, Robert. 1991. "Economic Growth in a Cross-Section of Countries," *Quarterly Journal of Economics.* 106(2):407-443.

Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics*. 1(4):200-232.

Bundy, Donald, Arun Joshi, Megan Rowlands and Yung-Ting Kung. 2003. "EnVISIONing Education in Low Income Countries." The World Bank. Washington, D.C.

Cameron, Colin, Jonah Gelbach and Douglas Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90(3):414-427.

de Brauw, Alan, and Scott Rozelle. 2007. "Returns to Education in Rural China" in Emily Hannum and Albert Park, eds., *Education and Reform in China* (London: Routledge).

Duflo, Esther, Rema Hanna and Stephen Ryan. 2012. "Incentives Work: Getting Teachers to Come to School," *American Economic Review* 102(4):1241-1278.

Gansu Statistical Yearbook. 2005. Beijing: China Statistics Press.

Glewwe, Paul, and Michael Kremer. 2006. "Schools, Teachers, and Education Outcomes in Developing Countries," in E. Hanushek and F. Welch, eds., *Handbook of the Economics of Education*. North Holland.

Glewwe, Paul, Michael Kremer, and Nauman Ilias. 2010. "Teacher Incentives". *American Economic Journal: Applied Economics* 2(3):205-227.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya" *American Economic Journal: Applied Economics* 1(1):112-135.

Gomes-Neto, João., Eric Hanushek, Raimundo Leite and Roberto Frota-Bezzera. 1997. "Health and Schooling: Evidence and Policy Implications for Developing Countries." *Economics of Education Review*. 16(3):271-283.

Hannum, Emily, Jere Behrman, Meiyan Wang, and Jihong Liu. 2008. "Education in the Reform Era", in L. Brandt and T. Rawski, eds., *China's Great Economic Transformation*. New York: Cambridge University Press.

Hannum, Emily, Albert Park, and Kai-ming Cheng. 2007. "Introduction: Market Reforms and Educational Opportunity in China", in Emily Hannum and Albert Park, eds. *Education and Reform in China* (Oxford: Routledge).

Hannum, Emily, and Yuping Zhang. 2012. "Poverty and Proximate Barriers to Learning: Vision Deficiencies, Vision Correction and Educational Outcomes in Rural Northwest China". *World Development* 40(9):1921-1931.

Hanushek, Eric, and Dennis Kimko. 2000. "Schooling, Labor Force Quality, and the Growth of Nations," *American Economic Review* 90(5):1184-1208.

Hanushek, Eric, and Ludger Woessmann. 2008. "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature* 46(3):607-668.

He, Mingguang, Wenyong Huang, Yingfeng Zheng, Li Huang & Leon B. Ellwein. 2007. "Refractive Error and Visual Impairment in School Children in Rural Southern China." *Ophthalmology* 114(2):374-382.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling," *Quarterly Journal of Economics*, 125(2):515-548.

Krueger, Alan, and Mikael Lindahl. 2001. "Education for Growth: Why and for Whom?" *Journal of Economic Literature* 39(4):1101-1136.

Li, Wen, Albert Park, and Sangui Wang. 2009. "School Equity in Rural China", in E. Hannum and A. Park, eds., *Education and Reform in China*. New York: Routledge.

Lu, Bei, and others. 2009. "Associations between Near Work, Outdoor Activity and Myopia among Adolescent Students in Rural China: The Xichang Pediatric Refractive Error Study: Report Number 2." *Archives of Ophthalmology* 127(6):769-775.

Luo, Renfu and others. 2012. "Nutrition and Educational Performance in Rural China's Elementary Schools: Results from a Randomized Control Trial In Shaanxi Province". *Economic Development and Cultural Change* 41(4):1012-1025.

Mankiw, N. Gregory, David Romer and David Weil. 1992. "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* 107(2):407-437.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impact on Education and Health in the Presence of Treatment Externalities", *Econometrica* 1(1):159-217.

Mutti, Donald, and others. 2002. "Parental Myopia, Near Work, School Achievement, and Children's Refractive Error." *Investigative Ophthalmology and Visual Science* 43(12):3633-3640.

National Bureau of Statistics. 2005. *Statistical Yearbook of China* (Beijing: China Statistical Press).

Sala-i-Martin, Xavier, Gernot Doppelhofer and Ronald Miller. 2004. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*. 94(4): 813-835.

Vermeersch, Christel and Michael Kremer. 2004. "School Meals, Educational Achievement, and School Competition: Evidence from a Randomized Evaluation". *World Bank Policy Research Working Paper No. 3523.* The World Bank. Washington, D.C.

World Bank. 2000. *China: Overcoming Rural Poverty.* The World Bank. Washington, D.C.

Wooldridge, Jeffrey. 2010. *Econometric Analysis of Cross Section and Panel Data*, second edition. Cambridge, MA: MIT Press.

Zhao, JiaLiang, and others. 2000. "Refractive Error Study in Children: Results from Shunyi District, China." *American Journal of Ophthalmology.* 129(4):427-435.

Table 1: Descriptive Statistics by County

	All 37 Townships			25 Compliant Townships		
	County 1	County 2	Both Counties	County 1	County 2	Both Counties
Number of children in grades 4-6 in 2004-05	10,208	18,581	28,789	6,130	12,772	18,902
Children with vision problems	1,552	2,625	4,177	787	1,742	2,529
	(15.2%)	(14.1%)	(14.5%)	(12.8%)	(13.6%)	(13.4%)
Of which:						
Had glasses already	67	67	134	22	36	58
	(4.3%)	(2.6%)	(3.2%)	(2.8%)	(2.1%)	(2.3%)
Did not have glasses	1,485	2,558	4,043	765	1,706	2,471
	(95.7%)	(97.4%)	(96.8%)	(97.2%)	(97.9%)	(97.7%)
Test scores in spring 2004 (before the intervention):						
Students without vision problems:						
Chinese	78.6	79.0	78.8	78.6	79.0	78.9
Mathematics	78.6	79.4	79.1	79.0	79.2	79.1
Science	80.2	80.7	80.6	80.6	80.8	80.7
Students with vision problems:						
Chinese	77.9	79.2	78.7	77.1	78.7	78.2
Mathematics	77.5	79.8	79.0	76.8	79.2	78.5
Science	80.2	81.1	80.8	80.2	80.8	80.6

Notes:

1. Vision problem is defined as a visual acuity score < 4.9 in one or both eyes. As explained in the text, although the 298 children in the full sample for whom one or both eyes had a score of 4.9 were offered glasses, only 18 (6.0%) accepted the glasses, so the analysis focuses on children for whom one or both eyes had a score of less than 4.9.

Table 2: Implementation of Gansu Vision Intervention Project

	All 37 Townships			25 Compliant Townships		
	County 1	County 2	Both Counties	County 1	County 2	Both Counties
Students in grades 4-6 in 2004-05 w/ vision problems	1,552	2,625	4,177	787	1,742	2,529
Of which:						
In control schools	703	1,496	2,199	112	889	1,001
In program schools	849	1,129	1,978	675	853	1,528
Students in program schools who:						
Accepted the offer to receive glasses	521	863	1,384	417	649	1,066
Did not accept the offer to receive glasses	328	266	594	258	204	462
Reasons given for not accepting glasses:						
Household head refused	65	122	187	54	91	145
Child refused	43	46	89	42	38	80
Cannot adjust to glasses	61	0	61	58	0	58
Mixed astigmatism	12	0	12	11	0	11
Optometrist not available	30	13	43	27	7	34
Pathological change in fundus	36	43	79	33	30	63
Eye problem cannot be corrected by glasses	5	0	5	5	0	5
Astigmatism	1	0	1	1	0	1
Vision not correctable	0	22	22	0	19	19
Child is handicapped	0	2	2	0	2	2
Missing	75	18	93	27	17	44

Notes:

1. Vision problem is defined as a visual acuity score < 4.9 in one or both eyes.

Table 3: Pre-Program Differences between Treatment and Control Groups (All 37 Townships, separately by county)

Variable	*Treatment Mean*	*Control Mean*	*Difference*	*Random Effects p-values for Differences based on Wild Bootstrap*
		All Children, County 1		
Chinese test	-0.103	-0.244	0.141	0.816
Math test	-0.097	-0.257	0.160	0.814
Science test	-0.027	-0.261	0.234	0.438
Average test	-0.092	-0.308	0.216	0.666
Ethnic minority	0.418	0.508	-0.090	0.432
Visual acuity	5.05	5.04	0.01	0.890
Poor vision	0.132	0.185	-0.053	0.858
Male	0.533	0.553	-0.020	0.332
Head yrs educ	8.53	7.46	1.07*	0.270
Age	9.82	10.12	-0.29*	0.294
		Children with Poor Vision Only, County 1		
Chinese test	-0.228	-0.221	-0.006	0.558
Math test	-0.245	-0.180	-0.065	0.884
Science test	-0.053	-0.211	0.158	0.322
Average test	-0.212	-0.248	0.035	0.632
Ethnic minority	0.379	0.445	-0.066	0.462
Visual acuity	4.60	4.64	-0.04*	0.0.14**
Male	0.485	0.486	-0.0005	0.690
Head yrs educ	7.89	7.22	0.67*	0.008***
Age	10.21	10.22	-0.002	0.372
		All Children, County 2		
Chinese test	-0.143	-0.045	-0.098	0.496
Math test	-0.110	0.019	-0.129	0.536
Science test	-0.087	-0.004	-0.083	0.362
Average test	-0.137	-0.012	-0.125	0.452
Ethnic minority	0.020	0.009	0.011	0.763
Visual acuity	5.03	5.01	0.02	0.692
Poor vision	0.108	0.184	-0.076**	0.124
Male	0.527	0.534	-0.007	0.664
Head yrs educ	8.88	8.67	0.20	0.484
Age	11.13	11.19	-0.06	0.836
		Children with Poor Vision Only, County 2		
Chinese test	-0.161	-0.047	-0.113	0.500
Math test	-0.126	0.081	-0.208*	0.442
Science test	-0.136	0.051	-0.187	0.104
Average test	-0.171	0.034	-0.206	0.318
Ethnic minority	0.017	0.008	0.009	0.912
Visual acuity	4.49	4.62	-0.13***	0.002***
Male	0.496	0.470	0.026	0.684
Head yrs educ	8.94	8.66	0.28	0.376
Age	11.34	11.32	0.02	0.368

Statistical significance of mean differences is based on regressions that account for clustering at the township level.

Asterisks denote statistical significance: * 10% level, ** 5% level, *** 1% level.

Table 4: Pre-Program Differences between Compliant and Non-Compliant Townships (separately by county)

Variable	*Compliant Mean*	*Non-Compliant Mean*	*Difference*	*Random Effects p-values of differences based on Wild Bootstrap*
		All Children, County 1		
Chinese	-0.161	-0.147	-0.014	0.224
Math	-0.139	-0.180	0.041	0.348
Science	-0.082	-0.159	0.077	0.460
Average	-0.155	-0.197	0.042	0.292
Ethnic minority	0.404	0.522	-0.118*	0.054*
Poor vision	0.128	0.187	-0.059*	0.362
Visual acuity	5.05	5.03	0.01	0.662
Male	0.545	0.534	0.011	0.796
Head yrs educ	8.44	7.67	0.77	0.708
Age	9.87	10.02	-0.16	0.062
		Children with Poor Vision Only, County 1		
Chinese	-0.302	-0.148	-0.154	0.470
Math	-0.271	-0.160	-0.111	0.584
Science	-0.134	-0.116	-0.018	0.740
Average	-0.286	-0.171	-0.115	0.530
Ethnic minority	0.356	0.464	-0.108	0.148
Visual acuity	4.62	4.62	0.01	0.994
Male	0.499	0.472	0.027	0.420
Head yrs educ	7.75	7.42	0.32	0.636
Age	10.22	10.21	0.01	0.932
		All Children, County 2		
Chinese	-0.102	-0.096	-0.006	0.572
Math	-0.082	0.009	-0.090	0.274
Science	-0.047	-0.058	0.011	0.730
Average	-0.094	-0.059	-0.035	0.496
Ethnic minority	0.007	0.034	-0.027	0.694
Poor vision	0.136	0.152	-0.016	0.598
Visual acuity	5.02	5.02	0.00	0.968
Male	0.531	0.530	0.001	0.962
Head yrs educ	9.01	8.31	0.70**	0.092*
Age	11.06	11.36	-0.30**	0.072*
		Children with Poor Vision Only, County 2		
Chinese	-0.145	0.001	-0.146	0.566
Math	-0.063	0.101	-0.164	0.506
Science	-0.062	0.036	-0.098	0.730
Average	-0.110	0.056	-0.165	0.648
Ethnic minority	0.005	0.025	-0.020	0.524
Visual acuity	4.58	4.54	0.04	0.362
Male	0.479	0.485	-0.005	0.818
Head yrs educ	8.98	8.39	0.60	0.200
Age	11.26	11.46	-0.20	0.134

Statistical significance is based on regressions that account for township level clustering. Strata dummy variables are excluded; they are perfectly collinear with the compliant dummy variable. Asterisks denote statistical significance: * 10% level, ** 5% level, *** 1% level.

Table 5: Pre-Program Differences between Treatment and Control Groups (25 townships where randomization was correctly implemented, separately by county)

Variable	*Treatment Mean*	*Control Mean*	*Difference*	*Random Effects p-values for Differences based on Wild Bootstrap*
		All Children, County 1		
Chinese test	-0.133	-0.336	0.203	0.543
Math test	-0.112	-0.308	0.196	0.608
Science test	-0.032	-0.386	0.354	0.472
Average test	-0.112	-0.416	0.304	0.868
Ethnic minority	0.388	0.495	-0.106	0.016**
Visual acuity	5.04	5.09	-0.05**	0.025**
Poor vision	0.130	0.122	0.008	0.163
Male	0.542	0.561	-0.019	0.864
Head yrs educ	8.62	7.45	1.17	0.461
Age	9.78	10.35	-0.57**	0.286
		Children with Poor Vision Only, County 1		
Chinese test	-0.314	-0.223	-0.091	0.994
Math test	-0.296	-0.115	-0.181	0.551
Science test	-0.111	-0.279	0.168	0.522
Average test	-0.292	-0.250	-0.042	0.906
Ethnic minority	0.359	0.339	-0.019	0.071*
Visual acuity	4.61	4.70	-0.09***	0.006***
Male	0.496	0.514	-0.017	0.774
Head yrs educ	7.82	7.34	0.48	0.183
Age	10.16	10.54	-0.37	0.232
		All Children, County 2		
Chinese test	-0.058	-0.157	0.100	0.462
Math test	-0.079	-0.085	0.005	0.730
Science test	0.001	-0.107	0.108	0.555
Average test	-0.055	-0.141	0.86	0.561
Ethnic minority	0.004	0.010	-0.006	0.795
Visual acuity	5.01	5.04	0.02	0.290
Poor vision	0.121	0.155	-0.03	0.784
Male	0.530	0.531	-0.001	0.586
Head yrs educ	9.23	8.73	0.50	0.301
Age	10.97	11.18	-0.21	0.388
		Children with Poor Vision Only, County 2		
Chinese test	-0.105	-0.184	0.079	0.714
Math test	-0.101	-0.027	-0.075	0.784
Science test	-0.082	-0.044	-0.038	0.698
Average test	-0.117	-0.103	-0.014	0.894
Ethnic minority	0.001	0.009	-0.008	0.794
Visual acuity	4.50	4.65	-0.15***	0.006***
Male	0.503	0.457	0.046	0.418
Head yrs educ	9.18	8.79	0.39	0.372
Age	11.25	11.27	-0.03	0.120

Statistical significance of mean differences is based on regressions that account for clustering at the township level. Asterisks denote statistical significance: * 10% level, ** 5% level, *** 1% level.

Table 6: Estimated Program Effect After One Year: ITT Results Using Compliant Sample

	Dependent Variables					
	Average Test Scores			Subject Scores (County 1 only)		
Explanatory Variables	Both Counties	County 1	County 2	Chinese	Math	Science
Equation (1): Compliant Sample, Only Students with Poor Vision						
Treatment Township (β)	0.158**	0.393***	0.079	0.413***	0.269**	0.259**
	(0.078)	(0.125)	(0.094)	(0.124)	(0.124)	(0.114)
	[0.214]	[0.026]	[0.624]	[0.080]	[0.162]	[0.156]
Sample Size	2,474	732	1,742	745	733	732
Equation (2): Compliant Sample, All Students						
Poor Vision (π)	-0.022	-0.121**	-0.016	-0.162**	-0.116	-0.038
	(0.030)	(0.059)	(0.034)	(0.064)	(0.119)	(0.042)
Treatment Township (τ)	-0.013	-0.022	-0.028	0.005	-0.047	-0.017
	(0.064)	(0.130)	(0.077)	(0.119)	(0.074)	(0.127)
Poor Vision×Treatment	0.109**	0.257***	0.073	0.289***	0.212*	0.146***
Township (β)	(0.049)	(0.065)	(0.068)	(0.075)	(0.128)	(0.051)
	[0.048]	[0.026]	[0.374]	[0.060]	[0.128]	[0.024]
Sample Size	18,504	5,736	12,768	5,788	5,744	5,742
Equation (1): Full Sample, Only Students with Poor Vision						
Treatment Township (β)	-0.049	0.159*	-0.073	0.217**	0.162**	0.002
	(0.077)	(0.095)	(0.086)	(0.093)	(0.065)	(0.110)
	[0.676]	[0.208]	[0.616]	[0.108]	[0.048]	[0.966]
Sample Size	4,093	1,468	2,625	1,491	1,469	1,468
Equation (2): Full Sample, All Students						
Poor Vision (π)	0.023	0.040	0.010	0.033	-0.016	0.074*
	(0.022)	(0.044)	(0.027)	(0.047)	(0.027)	(0.043)
Treatment Township (τ)	-0.122**	-0.068	-0.117*	0.010	-0.083*	-0.093
	(0.054)	(0.084)	(0.065)	(0.080)	(0.049)	(0.079)
Poor Vision×Treatment	0.041	0.075	0.022	0.064	0.103**	0.024
Township (β)	(0.040)	(0.055)	(0.054)	(0.064)	(0.047)	(0.054)
	[0.330]	[0.294]	[0.692]	[0.346]	[0.086]	[0.636]
Sample Size	28,271	9,694	18,577	9,785	9,709	9,707

Notes: 1. Coefficients on lagged test scores, constant terms, and strata dummy terms are not shown (to reduce clutter).

2. Standard errors in parentheses; wild bootstrap p-values in brackets. All models include school random effects and allow for heteroscedasticity, including correlation (clustering) among observations within the same township, of unknown form for the individual level error term. Asterisks denote statistical significance: * 10% level, ** 5% level, *** 1% level.

Table 7: Effect of Eyeglasses After One Year: IV Results

Explanatory Variables	*Dependent Variable* Average Test Scores: Both Counties	County 1	County 2	Subject Scores (County 1 only): Chinese	Math	Science
Equation (1): Compliant Sample, Only Students with Poor Vision						
Has Eyeglasses (β)	0.224**	0.677***	0.105	0.715***	0.465**	0.447**
	(0.110)	(0.196)	(0.121)	(0.206)	(0.203)	(0.175)
Sample Size	2,474	732	1,742	745	733	732
Equation (2): Compliant Sample, All Students						
Poor Vision (π)	-0.023	-0.122**	-0.015	-0.159**	-0.115	-0.037
	(0.030)	(0.059)	(0.033)	(0.063)	(0.116)	(0.041)
Has Eyeglasses (β)	0.156**	0.411***	0.093	0.456***	0.334*	0.231***
	(0.071)	(0.101)	(0.087)	(0.114)	(0.201)	(0.080)
Sample Size	18,503	5,735	12,768	5,787	5,743	5,787
Equation (1): Full Sample, Only Students with Poor Vision						
Has Eyeglasses (β)	-0.099	0.542**	-0.126	0.752***	0.562***	0.009
	(0.161)	(0.214)	(0.150)	(0.274)	(0.215)	(0.368)
Sample Size	4,093	1,468	2,625	1,491	1,469	1,468
Equation (2): Full Sample, All Students						
Poor Vision (π)	0.002	-0.353	0.008	-0.359	-0.491	0.059
	(0.043)	(0.537)	(0.038)	(0.455)	(0.589)	(0.362)
Has Eyeglasses (β)	0.082	0.746	0.027	0.736	0.911	0.049
	(0.096)	(0.896)	(0.090)	(0.778)	(0.994)	(0.610)
Sample Size	28,270	9,693	18,577	9,784	9,708	9,706

Notes: 1. Coefficients on lagged test scores, constant terms, and strata dummy terms are not shown (to reduce clutter).

2. Standard errors in parentheses. All models include school random effects and allow for heteroscedasticity, including correlation (clustering) among observations within the same township, of unknown form for the individual level error term. Asterisks denote statistical significance: * 10% level, ** 5% level, *** 1% level.

3. The instrumental variable for having eyeglasses is a dummy variable that indicates that one has been selected into the program *and* has poor vision.

Table 8: Estimated Program Effects for Students with Good Vision (Compliant Sample)

	Dependent Variable					
	Average Test Scores			Subject Scores (County 1 only)		
Explanatory Variables	Both Counties	County 1	County 2	Chinese	Math	Science
Equation (1): School Random Effects, Only Students with Good Vision						
Treatment Township (β)	-0.022	-0.023	-0.038	0.006	-0.044	-0.020
	(0.065)	(0.130)	(0.079)	(0.118)	(0.073)	(0.127)
	[0.856]	[0.896]	[0.762]	[0.966]	[0.685]	[0.906]
Sample Size	16,030	5,004	11,026	5,043	5,011	5,010

Notes: 1. Coefficients on lagged test scores, constant terms, and strata dummy terms are not shown (to reduce clutter).

2. Standard errors in parentheses; wild bootstrap p-values in brackets. All models include school random effects. All models allow for heteroscedasticity, including correlation (clustering) among observations within the same township, of unknown form for the individual level error term. Asterisks denote statistical significance: * 10% level, ** 5% level, *** 1% level.

Table 9: Falsification Test: "Effect" of Program in 2003-04 School Year (Compliant Sample)

	Dependent Variable (2004 test scores)					
	Average Test Scores			Subject Scores (County 1 only)		
Explanatory Variables	Both Counties	County 1	County 2	Chinese	Math	Science
Equation (2): School Random Effects, All Students						
Poor Vision (π)	0.019	-0.015	0.023	-0,045	0.031	-0.026
	(0.015)	(0.052)	(0.015	(0.058)	(0.075)	(0.042)
Treatment Township (τ)	-0.010	-0.080*	0.012	-0.124***	-0.125**	0.052
	(0.029)	(0.047	(0.035)	(0.029)	(0.053)	(0.046)
Poor Vision×Treatment	-0.018	-0.011	-0.002	0.034	-0.068	0.013
Township (β)	(0.031)	(0.062)	(0.041)	(0.076)	(0.087)	(0.050)
	[0.576]	[0.828]	[0.980]	[0.664]	[0.558]	[0.778]
Sample Size	18,600	5,830	12,770	5,831	5,830	5,831

Notes: 1. Coefficients on lagged test scores, constant terms, and strata dummy terms are not shown (to reduce clutter).

2. Standard errors in parentheses; wild bootstrap p-values in brackets. All models include school random effects. All models allow for heteroscedasticity, including correlation (clustering) among observations within the same township, of unknown form for the individual level error term. Asterisks denote statistical significance: * 10% level, ** 5% level, *** 1% level.

Table 10: Interaction Effects Between Program, Visual Acuity and 2004 Test Scores

Explanatory Variables	*Dependent Variable: Average Test Score*	
Equation (1): School Random Effects, Only Students with Poor Vision		***N* = 2,474**
Treatment Township (β)	0.173**	0.138*
	(0.078) [0.130]	(0.080) [0.236]
Very Poor Vision	0.053	--
	(0.040)	
Very Poor Vision × Treatment Township	-0.083	--
	(0.070) [0.282]	
Avg. Test Score 2004 × Treatment Township	--	-0.104
		(0.090) [0.284]
Equation (2): School Random Effects, All Students *N* = 18,478		
Poor Vision (π)	-0.026	-0.013
	(0.033)	(0.030)
	-0.011	-0.007
Treatment Township (τ)	(0.064)	(0.062)
Poor Vision × Treatment Township (β)	0.120**	0.083*
	(0.052) [0.054]	(0.044) [0.092]
Very Poor Vision	0.030	--
	(0.053)	
Very Poor Vision × Treatment Township	-0.065	--
	(0.078) [0.410]	
2004 Avg. Test Score × Treatment Township	--	0.033
		(0.078)
2004 Avg. Test Score × Poor Vision	--	0.067**
		(0.031)
2004 Avg. Test Score × Poor Vision		-0.163***
× Treatment Township		(0.054) [0.014]

Notes: 1. Coefficients on lagged test scores, constant terms, and strata dummy terms are not shown (to reduce clutter).

2. Standard errors in parentheses; wild bootstrap p-values in brackets. All models allow for heteroscedasticity, including correlation (clustering) among observations within the same township, of unknown form for the individual level error term. Asterisks denote statistical significance: * 10% level, ** 5% level, *** 1% level.

Table 11: Probit Estimates of Factors Associated with Accepting Eyeglasses

Variable	Mean	Coefficient	Marginal Effects
Average visual acuity	4.550	-1.424***	-0.479**
		(0.563)	(0.203)
Female	0.498	-0.242***	-0.081***
		(0.059)	(0.019)
Had glasses before program began	0.032	0.653*	0.175*
		(0.382)	(0.078)
Household head is a teacher	0.017	-0.585***	-0.220**
		(0.236)	(0.095)
Household head is village leader (cadre)	0.017	-0.880*	-0.336*
		(0.484)	(0.183)
Township per cap. income, 2003 (yuan/yr)	1519.1	0.00040**	0.00013**
		(0.00020)	(0.00006)
Head years of schooling	8.59	-0.015	-0.005
		(0.020)	(0.007)
Test score, spring 2004 (avg. for 3 subjects)	-0.190	-0.012	-0.004
		(0.069)	(0.023)
Tibetan	0.144	-0.022	0.007
		(0.156)	(0.052)
Grade in 2003-2004 (3, 4 or 5)	4.26	-0.076	-0.026
		(0.125)	(0.043)
Observations		1497	

Notes: 1. Constant term is not shown (to reduce clutter).

2. Standard errors are in parentheses. The specification allows for both heteroscedasticity and clustering at the township level of unknown form. Asterisks denote statistical significance: * 10% level, ** 5% level, *** 1% level.

3. The sample consists of all children in the program schools in grades 4-6 in 2004-05 who were deemed to have poor vision (one or both eyes with visual acuity below 4.9).

Table 12: Mother's Assessment of Vision and Actual Visual Acuity (children age 8-15 who were enrolled in primary school in 2004)

Measured Acuity	*Mother's Assessment*					
	Very Bad	Bad	Fair	Good	Very good	Don't Know
Good (≥ 5.0)	1	4	92	251	367	4
Fair (4.8-4.99)	0	0	18	29	52	0
Poor (< 4.8)	1	7	17	14	29	1

Source: Gansu Survey of Children and Families.

Table 13: Children's Reports of Vision Problems, by Actual Visual Acuity (children age 8-15 who were enrolled in primary school in 2004)

Measured Visual Acuity	*Child Reports of Vision Problems*		
	Difficulty seeing blackboard (%)	Trouble doing homework due to poor vision (%)	Felt pain in eyes when studying at home in dim light (%)
Good (≥ 5.0)	8.9	7.0	19.4
Fair (4.8-4.99)	7.8	7.8	20.6
Poor (< 4.8)	32.0	25.3	30.7

Source: Gansu Survey of Children and Families.

Table 14: Determinants of Student Wearing of Eyeglasses
(from 2004 Gansu Survey of Children and Families)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	*Coefficient Estimates from Probit Specification*					*Marginal Effects*
Visual acuity	-1.329***	-0.858**	-0.852**	-0.670	-0.783*	-0.0196
(best eye)	(0.374)	(0.412)	(0.420)	(0.423)	(0.417)	(0.0120)
Female	-0.228	-0.210	-0.228	-0.145	-0.155	-0.00396
	(0.169)	(0.177)	(0.173)	(0.173)	(0.173)	(0.00417)
Age (years)	0.0849**	0.0902***	0.0861**	0.101***	0.101***	0.00253**
	(0.0339)	(0.0341)	(0.0346)	(0.0375)	(0.0386)	(0.00106)
Father's education	-0.0210	-0.0114	-0.0104	-0.000136	0.00228	5.69e-05
	(0.0325)	(0.0341)	(0.0334)	(0.0365)	(0.0375)	(0.000930)
Mother's education	0.0919***	0.0895***	0.0936***	0.0851**	0.0722*	0.00180*
	(0.0316)	(0.0326)	(0.0331)	(0.0364)	(0.0372)	(0.00105)
Log p.c. expend.	0.553***	0.515***	0.530***	0.497**	0.463**	0.0116**
	(0.184)	(0.190)	(0.189)	(0.204)	(0.201)	(0.00562)
Mother assessment		0.898***	0.887***	0.858***	0.809***	0.0508
of child's vision		(0.326)	(0.329)	(0.307)	(0.298)	(0.0364)
Father assessment		0.603	0.605	0.558	0.542	0.0247
of child's vision		(0.380)	(0.385)	(0.357)	(0.364)	(0.0251)
Estimated cost of			-0.00366*	-0.00375*	-0.00451**	-0.000113
Glasses			(0.00215)	(0.00218)	(0.00225)	(7.37e-05)
Estimated distance			0.00164	0.000704	0.000869	2.17e-05
to buy glasses			(0.00422)	(0.00392)	(0.00388)	(9.75e-05)
Parent wears				0.955***	0.962***	0.0678
Glasses				(0.322)	(0.318)	(0.0439)
Village literacy rate					1.300**	0.0325*
					(0.645)	(0.0171)
Constant	-0.392	-2.710	-2.668	-3.679	-3.890	
	(2.360)	(2.732)	(2.751)	(2.731)	(2.898)	
Observations	921	921	921	921	921	921

Notes: 1. Standard errors are in parentheses. The specification allows for both heteroscedasticity and clustering at the village level of unknown form. Asterisks denote statistical significance: * 10% level, ** 5% level, *** 1% level.

2. Estimated cost of glasses and distance to buy glasses are medians of parental reports.

Appendix Table 1: Estimated Program Effect: ITT Results Using Compliant Sample

	Dependent Variables					
	Average Test Scores			Subject Scores (County 1 only)		
Explanatory Variables	Full Sample	County 1	County 2	Chinese	Math	Science
Equation (1): School Random Effects, Only Students with Poor Vision						
2004 Chinese Score	0.222***	0.071**	0.304***	0.024	0.055	0.094**
	(0.046)	(0.032)	(0.044)	(0.033)	(0.034)	(0.039)
2004 Math Score	0.215***	0.105**	0.258***	0.099**	0.083	0.065*
	(0.030)	(0.044)	(0.038)	(0.039)	(0.063)	(0.038)
2004 Science Score	0.150***	0.028	0.185***	0.028	0.012	0.031
	(0.028)	(0.021)	(0.044)	(0.018)	(0.021)	(0.022)
Treatment Township (β)	0.158**	0.393***	0.079	0.413***	0.269**	0.259**
	(0.078)	(0.125)	(0.094)	(0.124)	(0.124)	(0.114)
	[0.214]	[0.026]	[0.624]	[0.080]	[0.162]	[0.156]
Sample Size	2,474	732	1,742	745	733	732
Equation (2): School Random Effects, All Students						
2004 Chinese Score	0.209***	0.088***	0.267***	0.093*	0.052***	0.064***
	(0.023)	(0.028)	(0.020)	(0.048)	(0.010)	(0.016)
2004 Math Score	0.212***	0.101**	0.261***	0.109***	0.109**	0.025
	(0.027)	(0.044)	(0.037)	(0.014)	(0.054)	(0.041)
2004 Science Score	0.194***	0.115***	0.217***	0.083***	0.075***	0.121***
	(0.019)	(0.036)	(0.027)	(0.029)	(0.020)	(0.038)
Poor Vision (π)	-0.022	-0.121**	-0.016	-0.162**	-0.116	-0.038
	(0.030)	(0.059)	(0.034)	(0.064)	(0.119)	(0.042)
Treatment Township (τ)	-0.013	-0.022	-0.028	0.005	-0.047	-0.017
	(0.064)	(0.130)	(0.077)	(0.119)	(0.074)	(0.127)
Poor Vision×Treatment	0.109**	0.257***	0.073	0.289***	0.212*	0.146***
Township (β)	(0.049)	(0.065)	(0.068)	(0.075)	(0.128)	(0.051)
	[0.048]	[0.026]	[0.374]	[0.060]	[0.128]	[0.024]
Sample Size	18,504	5,736	12,768	5,788	5,744	5,742

Appendix 1: Impact of Measurement Error in the Poor Vision (PV) Variable

This appendix shows that random measurement error in the poor vision variable will not lead to underestimation of program effects in estimation of equation (2). Let PV* be the (unobserved) true indicator of whether a child has poor vision, and let PV denote the observed value of that variable. Thus PV = PV* + ε, where ε is the measurement error. Since both PV and PV* are dummy variables, ε will clearly be correlated with PV*, so this is not classical measurement error.

Assume that both types of measurement error (PV = 0 when PV* = 1 and PV = 1 when PV* = 0) occur with the same frequency, denoted by θ where θ is assumed to be less than 0.5. Thus there are three possibilities:

Frequency	*Value of PV**	*Value of PV*	*Value of ε*
1-2θ (no error)	1 or 0	Same as PV*	0
θ	0	1	1
θ	1	0	-1

The assumption that both types of errors occur with the same frequency is plausible if the error in the underlying visual acuity variable is random and the density of the distribution of that variable is similar on both sides of the cutoff point (4.8), and the latter is approximately correct.

Measurement error alters equation (2) as follows:

$$T = \alpha + \pi PV^* + \tau P + \beta PV^*P + u$$

$$= \alpha + \pi(PV - \varepsilon) + \tau P + \beta(PV - \varepsilon)*P + u$$

$$= \alpha + \pi PV + \tau P + \beta PV*P + (u - \pi\varepsilon - \beta\varepsilon P)$$

Bias in OLS estimation of β will be primarily determined by whether the interaction term PV*P is correlated with the composite error term (u – πε – βεP). Focusing on bias due to measurement error, this will be determined by whether PV*P is correlated with πε + βεP. Covariance formulas imply Cov(πε + βεP, PV*P) = πCov(ε, PV*P) + βCov(εP, PV*P). The following derivations show that Cov(ε, PV*P) = Cov(εP, PV*P) = θE[P]:

$$\begin{aligned}
Cov(\varepsilon, PV*P) &= E[(\varepsilon - E[\varepsilon])(PV*P - E[PV*P])] \\
&= E[\varepsilon(PV*P - E[PV*P])] \\
&= E[\varepsilon PV*P] - E[\varepsilon]E[PV*P] \\
&= E[\varepsilon PV*P] \\
&= Prob[\varepsilon = 0]\times E[\varepsilon PV*P|\ \varepsilon = 0] + Prob[\varepsilon = -1]\times E[\varepsilon PV*P|\ \varepsilon = -1] + Prob[\varepsilon = 1]\times E[\varepsilon PV*P|\ \varepsilon = 1] \\
&= (1 - 2\theta)\times 0 + \theta\times(-1\times 0\times E[P]) + \theta\times(1\times 1\times E[P]) \\
&= 0 + 0 + \theta E[P] > 0
\end{aligned}$$

$$\begin{aligned}
Cov(\varepsilon P, PV*P) &= E[(\varepsilon P - E[\varepsilon P])(PV*P - E[PV*P])] \\
&= E[(\varepsilon P - E[\varepsilon]\times E[P])(PV*P - E[PV*P])]
\end{aligned}$$

$$= E[\varepsilon P(PV^*P - E[PV^*P])]$$
$$= E[\varepsilon PV^*P^2 - E[\varepsilon P]E[PV^*P])]$$
$$= E[\varepsilon PV^*P - E[\varepsilon]E[P]E[PV^*P])]$$
$$= E[\varepsilon PV^*P]$$
$$= \theta E[P] > 0$$

Thus we have:

$$Cov(\pi\varepsilon + \beta\varepsilon P, PV^*P) = \pi Cov(\varepsilon, PV^*P) + \beta Cov(\varepsilon P, PV^*P) = (\pi + \beta)\theta E[P]$$

This derivation has two implications. First, as measurement error decreases ($\theta \rightarrow 0$), this correlation goes to zero and so bias goes to zero. Second, it is reasonable to assume that $\pi = -\beta$. Quite simply, if we expect a child with poor vision to score lower on tests by a factor of π (note that $\pi < 0$), then providing that child with glasses should remove the problem, which implies an impact of the same magnitude but in the opposite direction ($\beta > 0$). This thus implies that the two terms in the bias tend to cancel each other out. Note that the fact that π may be estimated with bias does not matter for this derivation, which is based on the true underlying value of π, not an estimate of π.

Appendix 2: Implications of Having Different Tests in Each School

Suppose that a "common" test (the same test) had been implemented in all the schools. For such a test, denote the mean and standard deviation of students' scores as:

μ^c = mean test score over all schools for common test
σ^c = standard deviation of distribution over all schools for common test

Each school (denoted by s) would also have had its own mean and standard deviation if all of its students had taken this common test. They can be denoted as:

μ_s^c = mean test score of school s if it had used the common test
σ_s^c = standard deviation of students in school s if they had taken the common test

Clearly, the μ_s^c for each school are related to μ^c as follows:

$$\mu^c = \sum_{s=1}^{S} w_s \mu_s^c$$

where S is the total number of schools and w_s is the share of students in school s.

For the tests that were actually used, denote the means and standard deviations as:

μ_s^a = mean test score of school s for the actual test taken
σ_s^a = standard deviation of students in school s for the actual test taken

For each school, normalize both the common test and the actual test at the school level:

$\dfrac{t_{is}^{c} - \mu_{s}^{c}}{\sigma_{s}^{c}}$ = normalized (at school level) score of student i in school s on common test: n_{is}^{c}

$\dfrac{t_{is}^{a} - \mu_{s}^{a}}{\sigma_{s}^{a}}$ = normalized (at school level) score of student i in school s on actual test: n_{is}^{a}

Assume that these two normalizations are equal to each other. That is, assume that t_{is}^{c} and t_{is}^{a} have the same "shape" but a different mean and variance. This assumption implies that $n_{is}^{c} = n_{is}^{a}$, which further implies that:

$$\frac{t_{is}^{c} - \mu_{s}^{c}}{\sigma_{s}^{c}} = \frac{t_{is}^{a} - \mu_{s}^{a}}{\sigma_{s}^{a}}$$

$$t_{is}^{c} = (\sigma_s^{c}/\sigma_s^{a})\times t_{is}^{a} + \mu_s^{c} - (\sigma_s^{c}/\sigma_s^{a})\mu_s^{a}$$

Consider a simple regression equation using the standardized common test:

$$t_{is}^{c} = \alpha + \beta P_s + u_{is}$$

The above derivations imply that we can express this relationship as:

$$(\sigma_s^{c}/\sigma_s^{a})\times t_{is}^{a} + \mu_s^{c} - (\sigma_s^{c}/\sigma_s^{a})\mu_s^{a} = \alpha + \beta P_s + u_{is}$$

$$(\sigma_s^{c}/\sigma_s^{a})\times t_{is}^{a} = \alpha + \beta P_s + u_{is} + [(\sigma_s^{c}/\sigma_s^{a})\mu_s^{a} - \mu_s^{c}]$$

$$t_{is}^{a} = \alpha(\sigma_s^{a}/\sigma_s^{c}) + \beta(\sigma_s^{a}/\sigma_s^{c})P_s + (\sigma_s^{a}/\sigma_s^{c})u_{is} + [\mu_s^{a} - (\sigma_s^{a}/\sigma_s^{c})\mu_s^{c}]$$

The term in brackets is a standard school random effect. The main concern is that the program effect, β, is multiplied by school specific ratio, $(\sigma_s^{a}/\sigma_s^{c})$. Because one can choose any scale for our hypothetical common test, one can choose it so that, on average (averaging across all S schools), $(\sigma_s^{a}/\sigma_s^{c}) = 1$. Thus the plim of $\beta(\sigma_s^{a}/\sigma_s^{c})$ over all S schools will be β. In fact, for any given school the estimated impact could be different, even if the same (common) test had implemented over all schools; one could specify a random coefficients model even if all schools had used the same test. Yet it is standard practice to estimate a specification that assumes the same impact, and doing so estimates the average impact over all schools. Likewise, by regressing the actual test score (instead of a score on a hypothetical common test) on a constant and P_s, one estimates the average of $\beta(\sigma_s^{a}/\sigma_s^{c})$, which is β. A final point is that regressions that compare good vision and poor vision students in the same school are comparing students on the same test, and so there is no problem of comparing students on different tests. Even so, it is still useful to have random effect to account for other unobserved school-level factors.