RESEARCH   **Open Access**

CrossMark

The gender gap among school children in poor rural areas of western China: evidence from a multi-province dataset

Hua Zhou[1], Di Mo[2*], Chengchao Zhou[3], Alexis Medina[2], Yaojiang Shi[5], Linxiu Zhang[4] and Scott Rozelle[2]

Abstract

Background: The gender gap remains a major impediment in the path towards equality and it is especially wide in low-income countries. Up to the early 2000s, many studies documented extensive inequalities in China: girls had poorer health, less nutrition and less education than their male counterparts. The goal of this study is to examine whether the gender gap persists, given that China is now making the transition into the ranks of upper-middle income countries. We consider educational outcomes, mental and physical health status, as well as non-cognitive outcomes.

Methods: We draw on a dataset containing 69,565 observations constructed by combining data from 7 different school-level surveys spanning 5 provinces. The surveys were all conducted by the authors between 2008 and 2013 using uniform survey instruments and data collection protocols in randomly selected schools across western provinces in rural China. The sample children range in age from 9 to 14 years (with 79 % of the sample being aged 10 to 12). Our analysis compares rural girls with rural boys in terms of 13 different indicators.

Results: With the exception of anemia rates, the health outcomes of girls are equal to those of boys. Girls and boys are statistically identical in terms of weight-for-age, height-for-age, and prevalence of intestinal worm infections. Girls performed better than boys on five of six cognitive and educational performance indicators. Girls performed worse than boys on all mental health indicators. All estimates are robust to the inclusion of different age ranges, controlling for the level of household assets, ethnic minority status, as well as the addition of provincial dummies.

Conclusions: Our findings suggest that with the exception of non-cognitive outcomes, anemia and standardized math test scores, the gender gap in our study areas in China appears to be diminishing.

Background

The plight of girls in low-income countries—in terms of both health and education—has drawn attention from many researchers in a variety of fields. Empirical evidence has shown that girls in low-income countries often have worse health outcomes than their male counterparts. For example, excessive female mortality at young ages has been well documented in low-income countries such as India [1–4], China [5–8], Pakistan [9], and Bangladesh [10]. Even though the under-five mortality rates are decreasing in most developing countries, girls are not benefiting proportionally from these declines [8]. The rate

of infection with intestinal worms (*Ascaris lumbricoides*) in Laos is higher for girls aged 12–23 months than for boys of the same age (30 % versus 24 %) [11]. Similar gender gaps have been found with Giardia infections in Zambian preschools [12] and with soil-transmitted helminths (STHs) in the National Primary School of Malaysia [13].

Besides having worse health, girls have also been shown to have worse nutrition and education outcomes. Data from Brazil, the Gaza strip, and India all show a higher prevalence of anemia among girls than boys [14–16]. Among eighth grade students in Nepal, boys have higher mean math test scores compared with girls (13.1 versus 10.1) [17].

When countries become richer, however, this gender gap seems to close; the literature shows that in developed

* Correspondence: dimo@stanford.edu
[2]Freeman Spogli Institute for International Studies, Stanford University, Stanford, USA
Full list of author information is available at the end of the article

Zhou *et al. International Journal for Equity in Health* (2016) 15:162

Page 2 of 11

countries, the health, nutrition, and educational outcomes of girls are typically better than or equal to those of boys. For example, in the Netherlands, research shows no statistical difference between boys and girls (aged 7 to 12 years) with respect to height (HAZ), weight (WAZ), or Body Mass Index (BMI) [18]. Other researchers show that the BMI of girls is similar to or better than that of boys in other countries in Western Europe, including Portugal [19, 20], the UK [21], Wales [22], and the United States [23]. In terms of educational performance and attainment, girls in developed countries are often found to outperform boys. For example, girls have been shown to have higher levels of mathematics achievement in Hong Kong [24], and to have higher language scores in Finland [25] and the Netherlands [26].

Although many studies in developed countries show that gender gaps in education, health and nutrition seem to draw close as income levels rise, similar results have not been found in the case of mental health problems. For example, many studies in developed countries have found that adolescent girls are significantly more likely to experience low and moderate levels of depression, anxiety, and low self-esteem than adolescent boys [27–30]. Fewer studies have been conducted in developing countries. However, a study in Egypt, for example, found that Egyptian girls are twice as likely to suffer from depression as boys [28].

From the 1980s through the early 2000s, when China was still counted among the ranks of poor and lower middle-income countries, researchers found significant signs that young Chinese girls (aged 0–17 years) were not doing as well as their male peers [29–32]. In 2006-07, researchers found that boys still had higher mean hemoglobin (Hb) levels relative to girls (ages 12–14) [32], and the prevalence of iron-deficiency anemia was significantly higher among high school girls than among high school boys (23.4 % versus 17.2 %) [33]. The average school enrollment rate is higher for boys than for girls among rural children in both the 7–14 year age range and the 15–18 year age range [34].

One possible explanation for this gender gap in China may be cultural. There is a strong literature base showing that rural China has traditionally valued sons over daughters. To the extent that this is still persistent, parents—especially those in poorer households—might be more likely to allocate limited resources to the more valued children—in this case, sons [35, 36]. As a result, boys may be expected to have better health and educational outcomes than girls.

Over the past decade, since 2005, China has joined the ranks of middle-income countries and household incomes have risen accordingly [37]. During 2008 and 2013, China has also issued several national policies, such as the Free Lunch Scheme across the poor rural areas and the New Cooperative Medical Scheme [38, 39]. Though such policies are yet to be rigorously examined, it is plausible

that they can slowly contribute to narrowing the gender gap in China. Have girls "caught up" with boys in terms of health, nutrition, and educational outcomes? The evidence is mixed. On the one hand, some studies suggest that the gender gap is closing. In 2006 and 2007, one study found that WAZ measures of girls were similar to those of boys [40]. Another study in the same year found no significant gender differences in underweight preschoolers (aged 0 to 5 years old) [41].

On the other hand, since 2005, other research has documented a persistent gender gap. For example, Ren and Rammohan (2014) find that the gender gap in WAZ widened between 1991 to 2009 [42], while Chyi and Zhou (2014) find that primary and junior high school enrollment rates are still higher among boys than among girls [43]. In addition, there is evidence to suggest that girls still suffer from mental health issues disproportionately more than boys do [44, 45], although this research has not typically been based on empirical measures.

Unfortunately, nearly all of the available evidence since 2005 is based on either small sample sizes or cover a small geographical region, so it is difficult to draw any firm conclusions from the existing literature.

In this study, we aim to fill this gap in the literature. The goal of our study is to describe the health and nutrition status, the cognitive and educational performance, and the non-cognitive outcomes of girls, ages 9 to 14, in rural areas of western China and to compare their outcomes to those of their male counterparts. To do this, we draw on a dataset containing nearly 70,000 observations constructed by combining data from several of our own school-level surveys. Using these data, we are able to compare rural girls with rural boys in terms of 13 different health, nutrition, cognition, educational performance, and non-cognitive outcome variables.

Methods
Data
The data used for this study are aggregated from seven different school-level surveys that the authors and collaborators conducted between 2008 and 2013. Figure 1 shows the location of the five provinces covered by the dataset: Ningxia, Qinghai, Gansu, Shaanxi and Guizhou province. The provinces that are included in our sample are relatively poor provinces in China. The average GDP per capita (PPP adjusted) is $4,209 for Gansu, $11,736 for Qinghai, $12,469 for Ningxia, $13,636 for Shaanxi and $8,481 for Guizhou. These provinces all have a lower GDP per capita than the national average of $13,980 [46]. Table 1 provides the provinces, years, sample sizes and primary outcomes of the surveys that the paper uses.
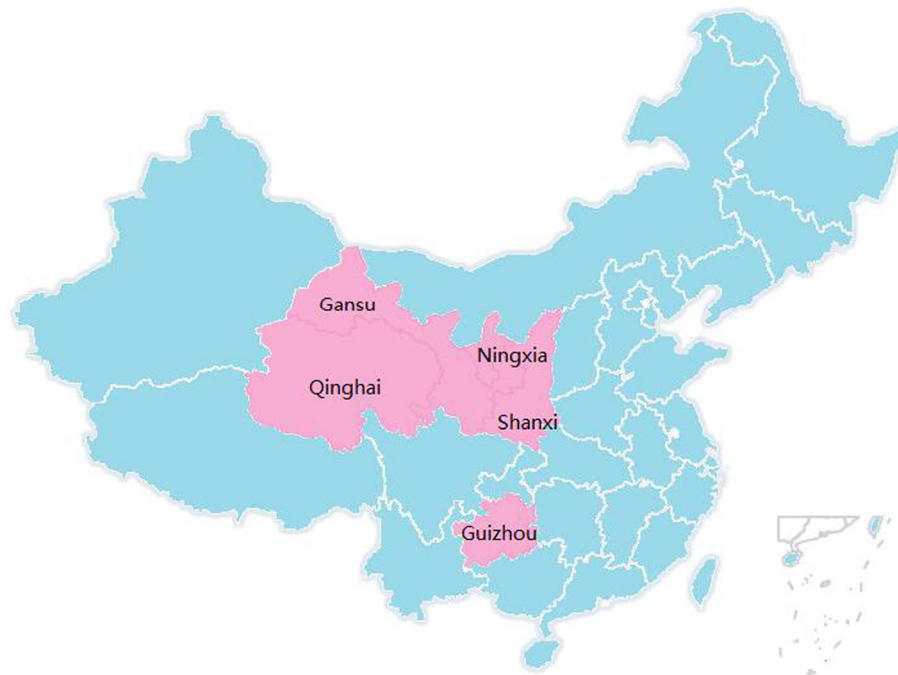
Zhou *et al. International Journal for Equity in Health* (2016) 15:162

Page 3 of 11

Fig. 1 Location of five study provinces in China

Sample selection

The seven surveys all employed random sampling strategies that were nearly uniform across studies. First, we obtained a list of all counties in each of the five provinces. Second, we randomly selected sample counties from those meeting our study criteria. Third, using official records, we created a list of all primary (and/or secondary) schools in the sample counties. Fourth, we randomly selected schools from the resulting sampling frame. Finally, within each of the randomly selected schools we randomly selected students (or classes of students) for inclusion in the studies. For more information on the sampling procedure, please refer to Additional file 1.

Data collection and outcome measures

We use a mega-dataset that is pooled from seven previously collected surveys. Each of the surveys was designed independently to study a particular topic, such as anemia, worms, and educational performance of the rural population in China [44, 47–57]. This explains why some of the

Table 1 Description of surveys and datasets

(1)	(2)	(3)	(4)	(5)	(6)
Project number	Province	Year	Sample size	Age	Primary outcome variables
1	Shaanxi	2008	4,055	9–14	Standardized math test score, Math course grades given by teachers, Mental Health Test, Hemoglobin
2	Shaanxi	2009	6,013	9–13	Standardized math test score, Math course grades given by teachers, Mental Health Test, Self-esteem, Self-efficacy, Hemoglobin
3	Ningxia, Qinghai	2009	15,169	9–14	Standardized math test score, Math course grades given by teachers, Mental Health Test, Hemoglobin
4	Gansu, Qinghai, Shaanxi	2011	32,915	9–14	Standardized math test score, Math course grades given by teachers, Standardized Chinese test score, Chinese course Grades given by teachers, Hemoglobin
5	Ningxia	2011	1,800	11	Standardized math test score, Math course grades given by teachers, Mental Health Test, Self-esteem, Hemoglobin
6	Gansu	2010	5,306	9–14	Standardized math test score, Math course grades given by teachers, Mental Health Test, Self-esteem, Self-efficacy, Hemoglobin
7	Guizhou	2013	4,307	9–11	WAZ, HAZ, STH, Standardized math test score, Math course grades given by teachers, Working memory, Processing speed, Hemoglobin

Zhou *et al. International Journal for Equity in Health* (2016) 15:162

Page 4 of 11

age ranges of the participants differ slightly in two of the seven surveys (Table 1). Nevertheless, since most of the surveys measured math test scores, all of the surveys targeted children that were ages 9 to 14 – so that participants are old enough to take the administered standardized tests. In fact, nearly four out of every five students (79 %) of our sample are between the ages of 10 and 12.

Even though each survey initially served as an independent study, the procedures were almost uniform across the seven surveys. Hence, this should serve as a point of strength for the study and its use of the pooled dataset. In total, the dataset includes information on 1005 schools from 71 rural inland counties in the five study provinces. During the baseline survey we collected data on students' basic demographic information, including gender and age. Table 2 provides definitions for the key variables used in the paper.

The primary outcomes examined in this study include 13 measures of health, nutrition, cognition, educational performance, and non-cognitive outcomes. To measure health and nutrition, we have data on weight-for-age z-scores (WAZ), height-for-age z-scores (HAZ), infection with soil-transmitted helminths (STH), and anemia prevalence. We collected two measures of cognition: working memory, and cognitive processing speed. We collected four measures of educational performance: scores from standardized tests of math and Chinese language, and grades given by teachers in math and Chinese language courses. We collected three measures of non-cognitive

outcomes: scores on a Mental Health Test, a self-esteem scale, and a self-efficacy scale.

Children's height and weight were measured and recorded by trained nurses from local provincial-level hospitals. The children were measured in light clothing without shoes. Weight was measured with a calibrated electronic scale. Body height was measured using a standard tape measure. WAZ scores were calculated using a SAS program for the 2000 Center for Disease Control growth chart for children aged 0–20 years [58]. Physical indicators of height were used to construct HAZ scores using WHO AnthroPlus, a software application of the WHO Reference 2007 for children aged 5 to 19 years old that is used to monitor the growth of school-aged children and adolescents [59].

Data on STH infections were collected from 2179 school-aged children in May, 2013 in Guizhou province. Children were considered positive for STH infection if either one of their stool samples tested positive for one or more types of STHs, including *Ascaris lumbricoides* (Ascaris), *Trichuris trichuria* (whipworm), and *Ancylostoma duodenale* or *Necator americanus* (hookworm).

Hemoglobin concentrations (Hb) were measured onsite using a Hemocue Hb 201+ finger prick system. Per WHO guidelines, we use an Hb cutoff of 115 g/L for children aged 9 to 11 years and 120 g/L for those aged 12 to 14 years [60].

To measure cognition, two cognitive measures were generated using the Wechsler Intelligence Scale for Children,

Table 2 Variable definitions

(1) Variable number	(2) Variable	(3) Description
1	WAZ	Weight-for-age z-score
2	HAZ	Height-for-age z-score
3	Anemia rate	Hb < 115 g/L, if age > =9 and < =11 (1 = yes, 0 = no); Hb < 120 g/L, if age > =12 and < =14 (1 = yes, 0 = no)
4	STH infection	Child is infected with any of the three types of STH: *Ascaris*, hookworm, or *Trichuris* (1 = yes, 0 = no)
5	MHT	Mental Health Test. The purpose of the test was to measure the level of each student's anxiety. The test is scored out of 90 points, where a lower (higher) score corresponds to lower (higher) anxiety.
6	Self-esteem score	Standardized score on Rosenberg Self-esteem Scales I & II (SES)
7	Self-efficacy score	Standardized score on General Self-efficacy Scale (GSE)
8	Standardized math test score	Math test score that is standardized by subtracting the mean and dividing by standard deviation within each wave of survey of each project.
9	Math course grades given by teachers	Normalized grades given by teachers
10	Standardized Chinese test score	Math test score that is standardized by subtracting the mean and dividing by standard deviation within each wave of survey of each project.
11	Chinese course grades given by teachers	Normalized grades given by teachers
12	Working memory index (WMI)	Standardized score on the working memory module of the WISC-IV
13	Processing speed index (PSI)	Standardized score on the processing speed module of the WISC-IV

Zhou *et al. International Journal for Equity in Health* (2016) 15:162

Page 5 of 11

Fourth Edition (WISC-IV): working memory and processing speed [61]. The working memory index (WMI) is assessed through two core subtests: the Digit Span subtest and the Letter Number Sequencing subtest. The processing speed index (PSI) is assessed through two core subtests: the coding subtest and the symbol search subtest. Raw scores obtained from these subtests are converted to age-scaled index scores using tables of norms from the official WISC-IV administration and scoring manual for China.

To measure educational performance, students were administered a standardized Chinese language test and a standardized math test. Our enumeration team monitored the tests, carefully proctored them in order to minimize cheating, and strictly enforced the time limits. Scores on both standardized tests were normalized (with mean zero and standard deviation equal to one) [62]. In addition, enumerators collected year-end Chinese language and math grades given by teachers. We normalized these grades by school.

To measure non-cognitive outcomes, we used a psychological test of well-being, the Mental Health Test (MHT), to measure the mental health of children. The test is a variation of the Children's Manifest Anxiety Scale (CMAS). The CMAS is a scale that has been widely used in the United States and other developed countries for more than a decade as a screening and clinical tool. Professor Zhou Bucheng of East China Normal University developed the MHT scale used in this study. Researchers have used this test extensively across China to measure the mental health of grade school students in urban contexts. The purpose of the test is to measure the level of each student's anxiety. The test is scored out of 90 points, where a lower (higher) score corresponds to lower (higher) anxiety. The test results can be broken down into eight subcategories, each of which represents a specific aspect of anxiety: school performance, social relationships, loneliness, self-punishment, sensitivity, physical symptoms, fear, and impulsiveness. A score of greater than 8 on any subpart is considered clinically high and indicates a need for treatment. A total score of 65 or higher indicates high risk for mental health problems and an urgent need for professional intervention [44, 63].

Self-esteem was assessed using the Rosenberg Self-esteem Scales I & II (SES) [64]. The SES consists of 10 statements, each of which is scored as a Likert scale.

Self-efficacy was assessed using the General Self-efficacy Scale (GSE) originally created by Schwartzer and Jerusalem (1995) in the 1970s [65]. We use the Chinese language version of the GSE, developed in 1995 [66]. The scale was created to predict how individuals cope with and/or adapt after experiencing stressful events.

Table 3 shows individual and family characteristics by gender. The average age of girls and boys in our sample is 11 years old. On average, girls in our sample are slightly

Table 3 Comparisons of the individual and family characteristics of girls and boys in rural areas of Western China

	Girl	Boy	*P*-value for the difference between girls and boys
Age (years)	10.95	11.02	<0.01
Ethnicity (1 = ethnic minority; 0 = otherwise)	0.31	0.32	0.01
Household size	5.42	5.05	<0.01
At least one parent takes migrant job (1 = yes; 0 = otherwise)	0.54	0.55	0.34
Education of mom (1 = at least primary education; 0 = otherwise)	0.52	0.52	0.72
Education of dad (1 = at least primary education; 0 = otherwise)	0.70	0.69	<0.01

Data source: Authors' data

younger than boys (the difference is 0.07 years or 25 days; *P*-value < 0.01). About 30 % of children in our sample are from ethnic minority groups and boys in our sample are 1 % more likely to be an ethnic minority than girls (*P*-value = 0.01). The average household size of girls (5.42) and boys (5.05) in our sample are significantly different (*P*-value < 0.01). More than half of the sample students have at least one parent that has migrated for work, though no significant difference is found between girls and boys (*P*-value = 0.34). Also, about half of the students' mothers have at least a primary school education, though no significant difference was found between the genders (*P*-value = 0.72). The education level of fathers in our sample is only slightly higher than that of mothers (by 1 percentage point). In the case of the sample fathers, we do find a significant difference between boys and girls in our sample (*P*-value < 0.01), as 70 % of the fathers of girls have at least a primary school education, while 69 % of the fathers of boys do. In summary, then, although there are differences in the individual and family characteristics of boys and girls, the magnitude of these differences is rather small. In the following section, we consider whether these differences impact our estimates.

Results

Overall, health and nutrition indicators for the full sample are poor (Table 4). The WAZ and HAZ of the sample children are -0.67 and -0.97, respectively. The STH infection rate is 41.9 %, and the anemia prevalence is 15.9 %.

With the exception of anemia, there are no statistically significant differences between boys and girls in terms of health or nutrition (Table 5). Boys in our sample have slightly higher WAZ (*P*-value = 0.40) and HAZ (*P*-value = 0.27) scores, however, the differences are not statistically

Zhou *et al. International Journal for Equity in Health* (2016) 15:162

Page 6 of 11

Table 4 Outcomes for the average children in rural areas of western China

(1) Outcome number	(2) Outcomes	(3) Sample size	(4) Unit	(5) Value
1	WAZ	1,450	Z-score	-0.67
2	HAZ	4,223	Z-score	-0.97
3	Anemia rate	50,598	%	15.87
4	STH infection rate	2,179	%	41.85
5	MHT	30,529	0–90 points	36.35
	Learning anxiety	30,529	0–15 points	7.94
6	Self-esteem	13,119	10–40 points	25.25
7	Self-efficacy	11,319	10–40 points	25.69
8	Working memory	4,305	45–150 points	78.49
9	Processing speed	4,305	45–160 points	87.33

See Table 2 for complete definitions of all variables. We have sufficient power for our tests due to the large sample size of the dataset we pooled together from seven surveys. For a two-sided test with a significance level of 0.05 and when the effect estimator has a limiting normal distribution, we have 80 % power to detect a MDE (minimum detectable effect) of 0.16 of WAZ, 0.10 of HAZ, 1 percentage points of anemia rate, 6 percentage points of STH infection rate, 0.48 points of MHT, 0.16 points of self-esteem score, 0.31 points of self-efficacy score, 0.03 SD of standardized math test score, 0.06 SD of math course grades given by teachers, 0.05 SD of standardized Chinese test score, 0.04 SD of Chinese course grades given by teachers, 0.86 points of WMI and 1.14 points of PSI [78, 79]

In this table, we only include outcome measure of variables that have standalone significance. We do not include (for example) standardized test scores of math and Chinese language scales since these scores are only used to compare the relative performance of children of different genders
Data source: Authors' data

significant. Similarly, the STH infection rate is 3 percentage points higher among boys (43.31 %) than among girls (40.16 %), but the difference is again statistically insignificant (*P*-value = 0.14). Anemia prevalence among girls (16.3 %), however, is statistically higher than that among boys (15.5 %, *P*-value = 0.01), although the raw difference is less than one percentage point.

In terms of cognition and educational performance, girls outperform boys on five of the six outcome measures. Girls score higher than boys on both the working memory (*P*-value = 0.02) and processing speed scales (*P*-value < 0.01, Table 5). Girls also score higher than boys on the standardized Chinese language test (*P*-value < 0.01), and are given higher grades by teachers in both math (*P*-value = 0.05) and Chinese language (*P*-value < 0.01) classes. For the latter class, in particular, girls outperform boys by about one third of a standard deviation. On the other hand, boys (0.07 SD) outperform girls (-0.08 SD) on the standardized math test (*P*-value < 0.01).

Average sample scores from the three non-cognitive assessments; the MHT (36.35 out of 90), the self-esteem test (25.25 on a scale from 10 to 40), and the self-efficacy test (25.69 on a scale from 10 to 40) show that poor performance is common among both genders in the sample (Table 4). Girls, however, perform consistently worse than boys on all three assessments (Tables 5 and 6). Out of the three tests, the overall MHT score (Table 5) shows the largest difference between the two groups, amounting to 2.44 points (*P*-value < 0.01) between girls (37.64) and boys (35.20). Put differently, girls score 6.93 % higher than boys, indicating that they have *more* anxiety than their male counterparts. Table 6

Table 5 Comparison of health, nutrition, cognition, educational status, and non-cognitive outcomes of girls and boys in rural areas of western China

(1) Outcomes number	(2) Outcomes	(3) Sample Size	(4) Unit	(5) Boys	(6) Girls	(7) P-value
1	WAZ	1,450	Z-score	-0.64	-0.69	0.40
2	HAZ	4,223	Z-score	-0.95	-0.99	0.27
3	Anemia prevalence	50,598	%	15.46	16.32	0.01
4	STH infection rate	2,179	%	43.31	40.16	0.14
5	MHT	30529	0–90 points	35.20	37.64	<0.01
6	Self-esteem	13,119	10–40 points	25.35	25.13	<0.01
7	Self-efficacy	11,319	10–40 points	25.87	25.49	<0.01
8	Standardized math test	49168	SD	0.07 ± 0.01	-0.08 ± 0.01	<0.01
9	Math course grades given by teachers	9532	SD	-0.02 ± 0.01	0.02 ± 0.01	0.05
10	Standardized Chinese test	13,707	SD	-0.04 ± 0.01	0.04 ± 0.01	<0.01
11	Chinese course grades given by teachers	15,753	SD	-0.14 ± 0.01	0.15 ± 0.01	<0.01
12	Working memory	4,305	45–150 points	78.16	78.88	0.02
13	Processing speed	4,305	45–160 points	86.25	88.6	<0.01

See Table 2 for complete definitions of all variables. SD = standard deviation
Data source: Authors' data

Zhou *et al. International Journal for Equity in Health* (2016) 15:162

Page 7 of 11

Table 6 Comparison of categorical breakdown of the MHT score between girls and boys in rural areas of western China

(1)	(2)	(3)	(4)	(5)	(6)
Outcome number	Outcomes	Unit	Boys	Girls	P-value
1	Learning anxiety	0–15 points	7.71	8.19	<0.01
2	Personal anxiety	0–10 points	3.87	4.09	<0.01
3	Loneliness anxiety	0–10 points	2.82	2.76	0.01
4	Self-Blaming tendency	0–10 points	5.03	5.42	<0.01
5	Sensitivity tendency	0–10 points	4.58	4.72	<0.01
6	Body anxiety	0–15 points	4.92	5.12	<0.01
7	Phobia anxiety	0–10 points	4.95	3.70	<0.01
8	Impulsive tendency	0–10 points	2.57	2.39	<0.01
9	Total	0–90 points	35.20	37.64	<0.01

Data source: Authors' data

shows the subcategories of the MHT, where the scores differences between the two genders are small, but precise.[1]

In order to check if the individual, family and geographical differences between girls and boys confound our findings regarding the gender gap, we regressed three of our outcome variables (anemia, standardized math and Chinese test scores) on the gender variable while controlling for individual and family characteristics and provincial dummies (Table 7). We find that the regression results are largely in line with the descriptive results in our paper, indicating that the individual, family and geographical differences between girls and boys do not affect the descriptive findings on the gender gap. For example, the regression results show that girls are 2 percentage points more likely to be anemic than boys

(*P*-value < 0.01). This finding that there is a slightly higher anemia prevalence among girls in our sample than among boys is consistent with the descriptive analysis. The regression results also show that girls perform better than boys in the standardized Chinese test by 0.09 SD (*P*-value < 0.01), which is similar to the 0.08 SD (*P*-value <0.01) margin found from the descriptive analysis. The regression results are also consistent with the descriptive analysis in showing that boys on average score higher on the standardized math test than girls.

We conduct a few tests to examine the robustness of the above results. First, in order to check whether the gender difference is sensitive to student age, we compare the regression of gender gap for the whole sample students (Table 7) with the regression including only the students of age 10 to 12, who constitute the majority of the sample (Table B1 in Additional file 1). We find that excluding the students aged under 10 and over 12 almost does not change the estimated gender gap.

Second, to assess whether the estimated gender gap is sensitive to the wealth levels of families, we run two types of regressions to estimate the gender gap: one with and one without controlling for the family's level of household assets (Table 7 and Table B2 in Additional file 1). We find that controlling for household assets (versus not controlling for household assets) does not change the estimated gender gap.

Finally, we examine whether the gender difference comes from across or within provinces. We compare the regression of gender gap including the provincial dummies (Table 7) with the regression without the provincial dummies (Table B3 in Additional file 1) and find that the gender gap we observe comes from within provinces

Table 7 OLS regression results of the gender gap in the health and educational outcomes

	Anemia (1)	Standardized math test (2)	Standardized Chinese test (3)
Gender (1 = male;0 = female)	-0.02*** (0.00)	0.21*** (0.01)	-0.09*** (0.02)
Age (years)	0.01*** (0.00)	-0.06*** (0.01)	-0.07*** (0.01)
Ethnicity (1 = ethnic minority; 0 = otherwise)	-0.00 (0.01)	-0.19*** (0.03)	-0.31*** (0.04)
Household size	0.00** (0.00)	-0.02*** (0.00)	-0.03*** (0.01)
Mother's education (1 = at least primary education; 0 = otherwise)	-0.01** (0.00)	0.03*** (0.01)	0.02* (0.01)
Father's education (1 = at least primary education; 0 = otherwise)	-0.00 (0.00)	0.16*** (0.01)	0.14*** (0.01)
At least one parent takes migrant job (1 = yes; 0 = otherwise)	0.01** (0.00)	-0.05*** (0.01)	0.00 (0.02)
Household asset (1 = higher than median; 0 = otherwise)	-0.02*** (0.00)	0.05*** (0.02)	0.01 (0.02)
Province dummies	Yes	Yes	Yes
Constant	-0.01 (0.02)	0.40*** (0.08)	0.79*** (0.10)
Observations	26,919	19,101	13,645
R-squared	0.007	0.061	0.060

Each column presents the OLS regression results from regressing the outcome variables (i.e. anemia, standardized math test scores and standardized Chinese test scores) on the gender variable while controlling for the individual variables and family variables. Provincial dummies are included in all regressions. Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$

Zhou *et al. International Journal for Equity in Health* (2016) 15:162

Page 8 of 11

instead of across provinces. For more details on any of these tests, please refer to Additional file 1.

In summary, girls have the same nutritional and health outcomes as boys, except for anemia. In academics, except for standardized math score, girls do better than boys. The gender gap, however, is still very much evident when it comes to non-cognitive outcomes. Girls in our sample have lower self-esteem, lower self-efficacy and more overall anxiety than their male counterparts. Further analysis shows that, overall, our estimates are robust; the evidence for the narrowing gender gap is not sensitive to the removal of the relatively younger and older students from our sample. Neither is the evidence sensitive to the level of household assets or provincial heterogeneity. These robustness checks further suggest that the gender gap in China has indeed become modest.[2]

Discussion

The primary aim of this study has been to document whether rural Chinese girls are worse off than their male counterparts in terms of health, nutrition, cognition, educational performance, and non-cognitive outcomes. Overall, our findings suggest that with a few exceptions, the gender gap in our study areas in China appears to be modest.

We appear to be documenting a shifting pattern in China. In studies in the 1990s and early 2000s, many research teams documented gender gaps in the terms of many health, nutrition and cognitive variables. However, in our study (using data from 2008 to 2013), the gender gap among school-aged children, as it pertains to physical health and academic performance, appears to be taking a similar trajectory to that in developed countries. With the exception of non-cognitive outcomes, anemia and standardized math test scores, the gender gap in our study areas in China appears to be diminishing. While it is beyond the scope of our paper to identify the exact mechanism behind this shift, it may be due to some combination of rising incomes [67–69], falling levels of fertility, and better employment opportunities off the farm [70]. The evidence suggests that Chinese families, even those in relatively poor rural areas, are treating girls and boys on a more equal basis.

Even more interestingly, the gender gap for cognition and educational performance has not only disappeared, but has also reversed. Our data show that girls perform better on standardized tests of cognition and Chinese language than do boys, and also receive higher grades in math and Chinese language classes. The only academic area in which boys still outperform girls is on standardized math tests. On the one hand, our standardized math test results are consistent with the literature base that shows that boys outperform girls in math, even in the case of developed countries [71, 72]. Explanations

for why boys tend to perform better in math are complicated, but it is commonly thought that societal expectations may exert an influence on the relative academic performance of male and female students. On the other hand, a possible reason for the differences between the standardized math test gender gaps and the teacher-graded math test gaps may be that the results reflect the teachers' own biases and their expectations of students. Studies have shown that when teachers grade student tests, scoring is likely to be affected by their perceptions of student performance or their preferences toward male or female students [73]. For example, if teachers perceive that girls are higher-performing students or if they favor female students, girls may receive higher marks than their actual performance on a test deserves.

These gains, however, may have come at a cost. Girls are lagging behind boys in terms of their non-cognitive outcomes. Girls have higher levels of learning anxiety and lower levels of both self-esteem and self-efficacy. Our results are consistent with the international literature which finds that girls are more likely to suffer from mental health problems, to exhibit higher levels of depression or anxiety and to have lower levels of self-esteem during adolescence than boys. In summary, our results show that girls outperform (or exhibit no significant differences from) boys in most health and educational outcomes with the exception of a number of non-cognitive outcomes.

This study has a number of strengths. First, the large size of the aggregated sample, comprising seven different datasets, is much larger (n > 69,000) than that of similar studies. This gives the study a high degree of statistical power and considerable external validity—at least for relatively poor and rural regions of western China. Second, all of the observations were collected by a single research team that used a common sampling strategy. The data collection instruments and the enumeration protocols were both standardized, allowing us to take advantage of the full aggregated sample in our analyses.

The study has several limitations. First, given the nature of the sample, it is not possible to extrapolate our findings to China's non-poor areas. However, given the fact that our study areas are among the poorest in China, and the fact that the international literature suggests that the gender gap is narrower in richer areas, we can be fairly assured that the overall status of girls in China has improved relative to that of boys over the past decade. Second, although the paper compares girls and boys on a number of different outcomes, we are unable to identify the exact cause of any observed differences. Third, although most outcomes measures are derived from datasets collected in multiple provinces, the data for some outcome measures (such as WAZ, HAZ, Working memory index and Processing speed index) are only available from one province (Guizhou Province). As a result, the findings for these

Zhou *et al. International Journal for Equity in Health* (2016) 15:162

Page 9 of 11

outcomes measures may have limited generalizability to other areas of China.

The results in this paper should not be construed to mean that children in China's poor areas are not vulnerable. Indeed, the absolute levels of health, nutrition, cognition, educational performance, and non-cognitive outcomes among both girls and boys are still low. In comparison with international standards, children in rural China—both boys and girls—are shorter and lighter. Over 40 % are infected with intestinal worms. Other work has documented the poor levels of educational performance among rural children relative to urban children [74]. Perhaps a more accurate interpretation of the results of this paper is that *all* children in the rural areas of western China are vulnerable. They all require extra care, attention and resources.

From a policy perspective, our results may indicate the success of existing programs that aim to improve outcomes for rural Chinese girls. Additional research should focus on measuring the impact of such programs. If positive impacts from such programs can be identified, perhaps such programs should be expanded to cover boys as well as girls.

Increasing the number of female students who complete higher levels of education and achieving greater gender equality in education is important for boosting female labor force participation [75]. Improving female labor market outcomes is part of a process that can ensure strong, sustainable, balanced economic growth for China in the future. Beyond economic efficiency, higher levels of gender equality in education and in the workforce can contribute to eliminating discriminatory social institutions and cultural norms, and in turn increase the happiness and well-being of women and society as a whole [76]. Although our results show that rural areas in western China has made huge progress toward narrowing the gender gap in childhood and adolescence, gender inequalities still persist. For example, the fields of study typically chosen by young men and women appear to be perpetuating gender labor market segregation [77]. The government has an important role to play in promoting gender equality both in schools and in the workforce.

Conclusion

Our findings suggest that with the exception of non-cognitive outcomes, anemia and standardized math test scores, the gender gap in our study areas in China appears to be diminishing. With the exception of anemia rates and mental health indicators, the health outcomes of girls are equal to those of boys. Girls and boys are statistically identical in terms of weight-for-age, height-for-age, and prevalence of intestinal worm infections. Girls performed better than boys on five of six cognitive and educational performance indicators. With the exception of math standardized test score, girls score higher than boys on the working memory, the processing speed scales, the standardized Chinese language test, math and Chinese grades given by teachers.

To our knowledge, this paper is one of the first studies to document a shifting pattern of gender gap in rural China by using large-scale first-hand survey data. In studies in the 1990s and early 2000s, many research teams documented gender gaps in the terms of many health, nutrition and cognitive variables. However, in our study (using data from 2008 to 2013), the gender gap among school-aged children, as it pertains to physical health and academic performance, appears to be diminishing and taking a similar trajectory to that in developed countries.

Endnotes

[1]In order to check whether the gender difference is sensitive to the ethnic minority status, we compare the outcomes between the two gender groups excluding ethnic minorities from our sample (Table B4 in Additional file 1). We find that girls are still not statistically significantly different from boys in terms of WAZ, HAZ and STH infection rates. Also, girls are slightly more likely to be anemic than boys (P-value < 0.01), outperform boys on five out of the six education and cognition outcome measures, and perform worse than boys on all three non-cognitive outcomes, outcomes that are all consistent with results for the whole sample in Table 5. Because of this, we do not believe that ethnicity is a factor affecting the results on gender gaps among our sample.

[2]By looking at gender differences in our outcomes across years, we show that the patterns in the gender gap over time are largely consistent with the analytical results of our paper. For example, girls score higher than boys on the standardized Chinese language test (P-values < or =0.01) in 2011 and 2012. Boys outperform girls on the standardized math test in all years from 2009 to 2012 (P-values < 0.01). Additionally, girls tend to perform worse than boys in non-cognitive outcomes. For example, girls score higher than boys on the Mental Health Test from 2009 to 2011, indicating girls are more anxious than boys (P-values <0.01). Also, girls, on average, had lower measures of self-esteem than boys from 2010 to 2012 (P-values < or =0.01). Consistent with the results in the paper, anemia levels are higher for girls than for boys in 2011 and 2012 (P-values <0.05), although the difference is not as significant as in the years from 2009 and 2010.

Zhou *et al. International Journal for Equity in Health* (2016) 15:162

Page 10 of 11

Availability of data and material
Study level data are available upon request. All authors have given informed consent for data sharing.

Authors' contributions
ZH and DM performed the statistical analysis and drafted the manuscript. CZ participated in the statistical analysis. YS and LZ participated in the design of the study and organized the data collection. AM participated in the data collection and drafting the manuscript. SR participated in the design of the study, advised the statistical analysis and drafted the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
All of the individual studies were approved by the Stanford University Institutional Review Board (IRB) and relevant Chinese authorities. The caregivers of all participants provided informed oral consent, and the children themselves provided oral assent. The investigation was conducted after all consents and assents were obtained.

Author details
[1]Schools of Economics & Management, Tongji University, Shanghai, China. [2]Freeman Spogli Institute for International Studies, Stanford University, Stanford, USA. [3]School of Public Health, Shandong University, Jinan, China. [4]Center for Chinese Agricultural Policy, IGSNRR, Chinese Academy of Sciences, Beijing, China. [5]Center for Experimental Economics in Education, Shaanxi Normal University, Xi'an, China.

References
1. Bhargava A. Family planning, gender differences and infant mortality: evidence from Uttar Pradesh, India. J Econom. 2003;112:225–40 [Analysis of Data on Health: 2].
2. Murthi M, Guio A-C, Drèze J. Mortality, fertility, and gender bias in India: a district-level analysis. Popul Dev Rev. 1995;21:745–82.
3. Pandey A, Sengupta PG, Mondal SK, Gupta DN, Manna B, Ghosh S, Sur D, Bhattacharya SK. Gender differences in healthcare-seeking during common illnesses in a rural community of West Bengal, India. J Health Popul Nutr. 2002;20:306–11.
4. Asfaw A, Lamanna F, Klasen S. Gender gap in parents' financing strategy for hospitalization of their children: evidence from India. Health Econ. 2010;19:265–79.
5. Coale AJ. Excess female mortality and the balance of the sexes in the population: an estimate of the number of "Missing Females". Popul Dev Rev. 1991;17:517–23.
6. Li S, Zhu C, Feldman MW. Gender differences in child survival in contemporary rural China: a county study. J Biosoc Sci. 2004;36:83–109.
7. Banister J. Shortage of girls in China today. J Popul Res. 2004;21:19–45.
8. Sex Differentials in Childhood Mortality. United Nations, Department of Economic and Social Affairs, Population Division; 2011. [ST/ESA/SER.A/314]. http://news.stanford.edu/news/2011/november/china-nutrition-policy-112211.html
9. Hazarika G. Gender differences in children's nutrition and access to health care in Pakistan. J Dev Stud. 2000;37:73–92.
10. D'Souza S, Chen LC. Sex differentials in mortality in rural Bangladesh. Popul Dev Rev. 1980;6:257–70.
11. Kounnavong S, Vonglokham M, Houamboun K, Odermatt P, Boupha B. Soil-transmitted helminth infections and risk factors in preschool children in southern rural Lao People's Democratic Republic. Trans R Soc Trop Med Hyg. 2011;105:160–6.
12. Siwila J, Phiri IGK, Enemark HL, Nchito M, Olsen A. Intestinal helminths and protozoa in children in pre-schools in Kafue district, Zambia. Trans R Soc Trop Med Hyg. 2010;104:122–8.
13. Al-Mekhlafi H, Surin J, Atiya AS, Ariffin WA, Mohammed Mahdy AK, Che Abdullah H. Pattern and predictors of soil-transmitted helminth reinfection among aboriginal schoolchildren in rural Peninsular Malaysia. Acta Trop. 2008;107:200–4.
14. Bagni UV, Luiz RR, da Veiga GV. Overweight is associated with low hemoglobin levels in adolescent girls. Obes Res Clin Pract. 2013;7:e218–29.
15. Abudayya A, Thoresen M, Abed Y, Holmboe-Ottesen G. Overweight, stunting, and anemia are public health problems among low socioeconomic groups in school adolescents (12–15 years) in the North Gaza Strip. Nutr Res. 2007;27:762–71.
16. Muthayya S, Thankachan P, Zimmermann MB, Andersson M, Eilander A, Misquith D, Hurrell RF, Kurpad AV. Low anemia prevalence in school-aged children in Bangalore, South India: possible effect of school health initiatives. Eur J Clin Nutr. 2007;61:865–9.
17. Pangeni KP. Factors determining educational quality: student mathematics achievement in Nepal. Int J Educ Dev. 2014;34:30–41.
18. van der Niet AG, Hartman E, Smith J, Visscher C. Modeling relationships between physical fitness, executive functioning, and academic achievement in primary school children. Psychol Sport Exerc. 2014;15:319–25.
19. Lopes L, Santos R, Pereira B, Lopes VP. Associations between gross motor coordination and academic achievement in elementary school children. Hum Mov Sci. 2013;32:9–20.
20. Tester J, Sharma S, Jasik CB, Mietus-Snyder M, Tinajero-Deck L. Gender differences in prediabetes and insulin resistance among 1356 obese children in Northern California. Diabetes Metab Syndr Clin Res Rev. 2013;7:161–5.
21. Hopkins ND, Stratton G, Tinken TM, McWhannell N, Ridgers ND, Graves LEF, George K, Cable NT, Green DJ. Relationships between measures of fitness, physical activity, body composition and vascular function in children. Atherosclerosis. 2009;204:244–9.
22. Lee K, Valeria B, Kochman C, Lenders CM. Self-assessment of height, weight, and sexual maturation: validity in overweight children and adolescents. J Adolesc Health. 2006;39:346–52.
23. Elgar FJ, Roberts C, Tudor-Smith C, Moore L. Validity of self-reported height and weight and predictors of bias in adolescents. J Adolesc Health. 2005;37:371–5.
24. Guo J, Marsh HW, Parker PD, Morin AJ, Yeung AS. Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: a multi-cohort study. Learn Individ Differ. 2015;37:161–8.
25. Mullola S, Jokela M, Ravaja N, Lipsanen J, Hintsanen M, Alatupa S, Keltikangas-Järvinen L. Associations of student temperament and educational competence with academic achievement: the role of teacher age and teacher and student gender. Teach Teach Educ. 2011;27:942–51.
26. Golsteyn BH, Schils T. Gender gaps in primary school achievement: a decomposition into endowments and returns to IQ and non-cognitive factors. Econ Educ Rev. 2014;41:176–87.
27. Scott S. Fortnightly review: aggressive behaviour in childhood. BMJ. 1998;316:202–6.
28. Afifi M. Depression in adolescents: gender differences in Oman and Egypt. 2006.
29. Coale AJ, Banister J. Five decades of missing females in China. Demography. 1994;31:459–79.
30. Gu B, Roy K. Sex ratio at birth in China, with reference to other areas in East Asia: what we know. | POPLINE.org. Asia-Pac Popul J. 1995;10:17–42.
31. Arnold F, Zhaoxiang L. Sex preference, fertility, and family planning in China. Popul Dev Rev. 1986;12:221–46.
32. Ren XS. Sex differences in infant and child mortality in three provinces in China. Soc Sci Med. 1995;40:1259–69.
33. Shi Z, Lien N, Kumar BN, Dalen I, Holmboe-Ottesen G. The sociodemographic correlates of nutritional status of school adolescents in Jiangsu Province, China. J Adolesc Health. 2005;37:313–22.

Zhou *et al. International Journal for Equity in Health* (2016) 15:162

Page 11 of 11

34. Song L, Appleton S, Knight J. Why do girls in rural China have lower school enrollment? World Dev. 2006;34:1639–53.

35. Das Gupta M, Zhenghua J, Bohua L, Zhenming X, Chung W, Hwa-Ok B. Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea. J Dev Stud. 2003;40:153–87.

36. Behrman JR, Deolalikar AB. The intrahousehold demand for nutrients in rural south India: Individual estimates, fixed effects, and permanent income. J Hum Resour. 1990;25(4):665–96.

37. Chen Q, Goh C, Sun B, Xu LC. Market Integration in China. World Bank Policy Research Working Paper. Washington, DC: World Bank; 2011.

38. Gorlick A. Stanford research on nutrition influences policy change in China. 2011. http://news.stanford.edu/news/2011/november/china-nutrition-policy-112211.html.

39. Yang W. China's new cooperative medical scheme and equity in access to health care: evidence from a longitudinal household survey. Int J Equity Health. 2013;12:20.

40. Sawadogo PS, Martin-Prével Y, Savy M, Kameli Y, Traissac P, Traoré AS, Delpeuch F. An infant and child feeding index is associated with the nutritional status of 6-to 23-month-old children in rural Burkina Faso. J Nutr. 2006;136:656–63.

41. Svedberg P, et al. Child malnutrition in India and China. 2007.

42. Ren W, Rammohan A, Wu Y. Is there a gender gap in child nutritional outcomes in rural China? China Econ Rev. 2014;31:145–55.

43. Chyi H, Zhou B. The effects of tuition reforms on school enrollment in rural China. Econ Educ Rev. 2014;38:104–23.

44. Zhang L, Kleiman-Weiner M, Luo R, Shi Y, Martorell R, Medina A, Rozelle S. Multiple micronutrient supplementation reduces anemia and anxiety in rural China's elementary school children. J Nutr. 2013;143:640–7.

45. Wang X, Liu C, Zhang L, Yue A, Shi Y, Chu J, Rozelle S. Does financial aid help poor students succeed in college? China Econ Rev. 2013;25:27–43.

46. World Economic Outlook: China April 2015 Edition [https://www.imf.org/external/pubs/ft/weo/2015/01/pdf/c2.pdf].

47. Luo R, Shi Y, Zhang L, Zhang H, Miller G, Medina A, Rozelle S. The limits of health and nutrition education: evidence from three randomized-controlled trials in rural China. CESifo Econ Stud. 2012;58(2):ifs023.

48. Mo D, Luo R, Liu C, Zhang H, Zhang L, Medina A, Rozelle S. Text messaging and its impacts on the health and education of the poor: evidence from a field experiment in rural China. World Dev. 2014;64:766–80.

49. Miller G, Luo R, Zhang L, Sylvia S, Shi Y, Foo P, Zhao Q, Martorell R, Medina A, Rozelle S. Effectiveness of provider incentives for anaemia reduction in rural China: a cluster randomised trial. 2012.

50. Kleiman-Weiner M, Luo R, Zhang L, Shi Y, Medina A, Rozelle S. Eggs versus chewable vitamins: Which intervention can increase nutrition and test scores in rural China? China Econ Rev. 2013;24:165–76.

51. Luo R, Shi Y, Zhang L, Liu C, Li H, Rozelle S, Sharbono B. Community service, educational performance and social responsibility in Northwest China. J Moral Educ. 2011;40:181–202.

52. Liu C, Zhang L, Luo R, Rozelle S, Loyalka P. The effect of primary school mergers on academic performance of students in rural China. Int J Educ Dev. 2010;30:570–85.

53. Mo D, Yi H, Zhang L, Shi Y, Rozelle S, Medina A. Transfer paths and academic performance: the primary school merger program in China. Int J Educ Dev. 2012;32:423–31.

54. Sylvia S, Luo R, Zhang L, Shi Y, Medina A, Rozelle S. Do you get what you pay for with school-based health programs? Evidence from a child nutrition experiment in rural China. Econ Educ Rev. 2013;37:1–12.

55. Wong HL, Luo R, Zhang L, Rozelle S. The impact of vouchers on preschool attendance and elementary school readiness: a randomized controlled trial in rural China. Econ Educ Rev. 2013;35:53–65.

56. Loyalka P, Liu C, Song Y, Yi H, Huang X, Wei J, Zhang L, Shi Y, Chu J, Rozelle S. Can information and counseling help students from poor rural areas go to high school? Evidence from China. J Comp Econ. 2013;41:1012–25.

57. Liu C, Luo R, Yi H, Zhang L, Li S, Bai Y, Medina A, Rozelle S, Smith S, Wang G, et al. Soil-transmitted Helminths in Southwestern China: a cross-sectional study of links to cognitive ability, nutrition, and school performance among children. PLoS Negl Trop Dis. 2015;9:e0003877.

58. Organization WH, et al. WHO AnthroPlus for personal computers Manual: Software for assessing growth of the world's children and adolescents. Geneva: WHO; 2009.

59. Control centers for D, Prevention, et al. A SAS Program for the 2000 CDC Growth Charts (ages 0 To < 20 Y). 2014.

60. WHO, UN University, UN Children's Funds. Iron Deficiency Anaemia: Assessment, Prevention, and Control - A guide for programme managers. 2001.

61. Wechsler D. Wechsler intelligence scale for children—Fourth Edition (WISC-IV) administration and scoring manual. San Antonio: Psychol Assoc; 2003.

62. Mo D, Swinnen J, Zhang L, Yi H, Qu Q, Boswell M, Rozelle S. Can one-to-one computing narrow the digital divide and the educational gap in China? The case of Beijing migrant schools. World Dev. 2013;46:14–29.

63. Wang H, Yang C, He F, Shi Y, Qu Q, Rozelle S, Chu J. Mental health and dropout behavior: a cross-sectional study of junior high students in northwest rural China. Int J Educ Dev. 2015;41:1–12.

64. Blascovich J, Tomaka J. Measures of self-esteem. Meas Personal Soc Psychol Attitudes. 1991;1:115–60.

65. Schwarzer R. General Perceived Self-Efficacy in 14 Cultures. 1999.

66. Schwarzer R, Bä\s sler J, Kwiatek P, Schröder K, Zhang JX. The assessment of optimistic self-beliefs: comparison of the German, Spanish, and Chinese versions of the general self-efficacy scale. Appl Psychol. 1997;46:69–88.

67. Lincove JA. Determinants of schooling for boys and girls in Nigeria under a policy of free primary education. Econ Educ Rev. 2009;28:474–84.

68. Tansel A. Determinants of school attainment of boys and girls in Turkey: individual, household and community factors. Econ Educ Rev. 2002;21:455–70.

69. Lee YD. Do families spend more on boys than on girls? Empirical evidence from rural China. China Econ Rev. 2008;19:80–100.

70. Yuezhong MGYLY. Research on handling sex ratio unbalance at birth in China: stage, characteristics, mechanism and prospects [J]. Chin J Popul Sci. 2007;3:010.

71. Lavy V, Sand E. On the Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases. Cambridge: National Bureau of Economic Research; 2015.

72. Niederle M, Vesterlund L. Explaining the gender gap in math test scores: THE role of competition. J Econ Perspect. 2010;24:129–44.

73. Spencer SJ, Steele CM, Quinn DM. Stereotype threat and women's math performance. J Exp Soc Psychol. 1999;35:4–28.

74. Wang X, Liu C, Zhang L, Luo R, Glauben T, Shi Y, Rozelle S, Sharbono B. What is keeping the poor out of college?: Enrollment rates, educational barriers and college matriculation in China. China Agric Econ Rev. 2011;3:131–49.

75. Sinha N, Raju D, Morrison A. Gender Equality, Poverty and Economic Growth. SSRN Scholarly Paper. Rochester: Social Science Research Network; 2007.

76. Kabeer N, Natali L. Gender equality and economic growth: is there a win-win? IDS Work Pap. 2013;2013:1–58.

77. Guo C, Tsang MC, Ding X. Gender disparities in science and engineering in Chinese universities. Economics of Education Review. 2010;29(2):225–35.

78. Djimeu EW, Houndolo D. Power calculation for causal inference in social science: sample size and minimum detectable effect determination. 3ie Working Paper. 2016. 26.

79. Samii C. Will your study have enough power? 2015.