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

Additional file 1

A. Sampling Procedure

This study of the gender gap in rural China uses seven surveys that span five western provinces. As mentioned in the Methods section of the study, these surveys were previously collected by the authors and collaborators to study a variety of topics including intestinal worms, educational performance, anemia, etc. All of these surveys, however, were collected in poor rural parts of China, and all of them targeted students aged 9 to 14. More importantly, the sampling protocol was uniform across the seven surveys. The following provides an insight into the specific steps of the sampling procedure:

- For each province, randomly selecting counties that meet a threshold of poverty requirements (For example, the county cannot be the home of the province government, which would entail that it is more urban and/or wealthier).
- Compiling a list of all schools within each county. To ensure representation of rural schools, the protocol then excluded schools that:
  - Have less than 200 or 400 students (depending on the province)
  - Have less than six grades (complete elementary school).
- Randomly selecting class (or classes) and then randomly selecting students from each selected school/class.
- Collecting the needed variables (e.g. Health indicators, academic scores, non-cognitive assessments) using the appropriate methods (e.g., utilizing nurse practitioners to collect Hb measure; etc.).

In addition to Table 1 in the original manuscript, we also include a summary of the surveys and a list of the published studies based on each survey (Table A1).

Additional File 1
Gender gap in rural western China

Table A1. Published studies based on the seven collected surveys

Provinces	Year	Number of students	Studies
Shaanxi	2008	4,055	Luo R, Kleiman-Weiner M, Rozelle S, Zhang L, Liu C, Sharbono B, Shi Y, Yue A, Martorell R, Lee M: **Anemia in rural China's elementary schools: prevalence and correlates in Shaanxi province's poor counties**. *Ecology of food and nutrition* 2010, **49**:357–372. Luo R, Shi Y, Zhang L, Liu C, Rozelle S, Sharbono B, Yue A, Zhao Q, Martorell R: **Nutrition and educational performance in rural China's elementary schools: Results of a randomized control trial in Shaanxi Province**. *Economic development and cultural change* 2012, **60**:735–772. 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**:640–647. Luo R, Wang X, Zhang L, Liu C, Shi Y, Miller G, Rozelle S, Yu E, Martorell R: **Alarmingly high anemia prevalence in Western China**. *Southeast Asian Journal of Tropical Medicine and Public Health42(5)* 2011. 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 Economic Studies* 2012, **58**:385–404.
Shaanxi	2009	6,013	Chen X, Yi H, Zhang L, Mo D, Chu J, Rozelle S: **Do poor students benefit from China's Merger Program? Transfer path and educational performance**. *Asia Pacific Journal of Education* 2014, **34**:15–35. Yaojiang Shi, Fang Chang, Xiaoqing Su, Renfu Luo, Linxiu Zhang, Scott Rozelle: **Parental training, anemia and the impact on the nutrition of female students in China's poor rural elementary schools**. *China Ag Economic Review* 2012, **4**:151–167. Luo R, Wang X, Zhang L, Liu C, Shi Y, Miller G, Rozelle S, Yu E, Martorell R: **Alarmingly high anemia prevalence in Western China**. *Southeast Asian Journal of Tropical Medicine and Public Health42(5)* 2011. 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 Economic Studies* 2012, **58**:385–404.

Table A1 (Continued). Published studies based on the seven collected surveys

Provinces	Year	Number of students	Studies
Ningxia, Qinghai	2009	15,169	Luo R, Zhang L, Liu C, Zhao Q, Shi Y, Miller G, Yu E, Sharbono B, Medina A, Rozelle S, Martorell R: **Anaemia among Students of Rural China's Elementary Schools: Prevalence and Correlates in Ningxia and Qinghai's Poor Counties**. *J Health Popul Nutr* 2011, **29**:471–485. 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**. *BMJ* 2012, **345**:e4809. 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**. *Economics of Education Review* 2013, **37**:1–12. Luo R, Wang X, Zhang L, Liu C, Shi Y, Miller G, Rozelle S, Yu E, Martorell R: **Alarmingly high anemia prevalence in Western China**. *Southeast Asian Journal of Tropical Medicine and Public Health42(5)* 2011.
Gansu	2010	5,306	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 Economic Review* 2013, **24**:165–176.
Gansu, Qinghai, Shaanxi	2011	32,915	Yang Y, Wang H, Zhang L, Sylvia S, Luo R, Shi Y, Wang W, Rozelle S: **The Han-Minority Achievement Gap, Language, and Returns to Schools in Rural China**. *Econ Dev Cult Change* 2015, **63**:319–359.
Ningxia	2011	1,800	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 Development* 2014, **64**:766–780. 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 Economic Studies* 2012, **58**:385–404.
Guizhou	2013	4,307	Liu C, Luo R, Yi H, Zhang L, Li S, Bai Y, Medina A, Rozelle S, Smith S, Wang G, Wang J: **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.

B. Additional analysis

The main analysis of this study compares the health, academic, and non-cognitive outcomes of the boys and girls in our sample. The findings indicated that girls tend to have more overall anxiety than boys; that except for anemia, girls have comparable outcomes to boys in terms of health; and that they outperform boys in educational and cognitive outcomes, with the exception being the standardized math test scores. For purposes of brevity, we picked three outcomes: anemia, standardized math test scores, and standardized Chinese test scores, and conducted a regression analysis to enhance the descriptive comparison. We regressed the outcomes on the gender variable while controlling for individual and family characteristics and province-fixed effects (Table 7). The regression yielded consistent results.

This section conducts further analysis to test the robustness of these findings. First, we regress the outcome indicators on the gender dummy variable, like Table 7, this time, however, excluding the older and younger students from the sample (Table B1). These are students who are either younger than 10 or older than 12 years old – and constitute 21% of the sample. When comparing the coefficients of the gender variable (Table 7 and Table B1, row 1), we find that there is only a 1 percentage point difference for the Chinese test score outcomes. All the other magnitudes and signs, however, are identical. These findings suggest that our estimates of the gender gap are not sensitive to students age.

Table B1. OLS regression results of the gender gap in the health and educational outcomes for students of age 10-12

	Anemia	Standardized math test	Standardized Chinese test
	(1)	(2)	(3)
Gender (1=male;0=female)	-0.02***	0.21***	-0.08***
	(0.00)	(0.02)	(0.02)
Age (years)	0.03***	-0.03***	-0.04***
	(0.00)	(0.01)	(0.01)
Ethnicity (1=ethnic minority; 0=otherwise)	-0.00	-0.19***	-0.29***
	(0.01)	(0.04)	(0.04)
Household size	0.00***	-0.02***	-0.03***
	(0.00)	(0.01)	(0.01)
Mother's education (1=at least primary education; 0=otherwise)	-0.01**	0.04***	0.02
	(0.00)	(0.01)	(0.01)
Father's education (1=at least primary education; 0=otherwise)	-0.00	0.17***	0.14***
	(0.00)	(0.01)	(0.01)
At least one parent takes migrant job (1=yes; 0=otherwise)	0.01***	-0.05***	-0.01
	(0.00)	(0.02)	(0.02)
Household asset (1=higher than median; 0= otherwise)	-0.02***	0.05***	-0.01
	(0.01)	(0.02)	(0.02)
Province dummies	Yes	Yes	Yes
Constant	-0.20***	0.17	0.46***
	(0.04)	(0.12)	(0.15)
Observations	21,527	15,222	10,911
R-squared	0.010	0.057	0.052

Note: 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

Second, we test whether our original results are sensitive to the level of wealth. Like the first exercise, we run the main regression while removing the variable for household assets (Table B2). We compare the results to our estimates in (Table 7, row 1) and find that excluding the household assets from the regression does not change either the magnitudes or the signs of the gender coefficient. Thus, we conclude that our results are not sensitive to the household wealth.

Table B2. OLS regression results of the gender gap in the health and educational outcomes for students of age 10-12 (without controlling for household asset)

	Anemia	Standardized math test	Standardized Chinese test
	(1)	(2)	(3)
Gender (1=male;0=female)	-0.02***	0.21***	-0.09***
	(0.00)	(0.01)	(0.02)
Age (years)	0.01***	-0.06***	-0.07***
	(0.00)	(0.01)	(0.01)
Ethnicity (1=ethnic minority; 0=otherwise)	-0.00	-0.18***	-0.31***
	(0.01)	(0.03)	(0.04)
Household size	0.00**	-0.02***	-0.03***
	(0.00)	(0.00)	(0.01)
Mother's education (1=at least primary education; 0=otherwise)	-0.01***	0.03***	0.02*
	(0.00)	(0.01)	(0.01)
Father's education (1=at least primary education; 0=otherwise)	-0.00	0.16***	0.14***
	(0.00)	(0.01)	(0.01)
At least one parent takes migrant job (1=yes; 0=otherwise)	0.01***	-0.05***	0.00
	(0.00)	(0.01)	(0.02)
Province dummies	Yes	Yes	Yes
Constant	-0.02	0.42***	0.80***
	(0.02)	(0.08)	(0.10)
Observations	26,919	19,101	13,645
R-squared	0.006	0.060	0.060

Note: 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

Third, we test whether the gender difference in the outcome indicators stem from within or across the five provinces in our sample. We run the main regression without the province dummies (Table B3) and compare the results to the regression in Table 7, which includes the province fixed effects. The consistent signs and magnitudes of the outcomes similarly suggest that regional heterogeneity does not influence the effect of gender on the outcome indicators.

Therefore, we conclude that the observed gender differences stem from within provinces and not across them.

Table B3. OLS regression results of the gender gap in the health and educational outcomes (without province dummies)

	Anemia	Standardized math test	Standardized Chinese test
	(1)	(2)	(3)
Gender (1=male; 0=female)	-0.02***	0.21***	-0.09***
	(0.00)	(0.01)	(0.02)
Age (years)	0.01***	-0.06***	-0.07***
	(0.00)	(0.01)	(0.01)
Ethnicity (1=ethnic minority; 0=otherwise)	0.02**	-0.16***	-0.09***
	(0.01)	(0.02)	(0.03)
Household size	0.00**	-0.02***	-0.04***
	(0.00)	(0.00)	(0.01)
Mother's education (1=at least primary education; 0=otherwise)	-0.00**	0.04***	0.03***
	(0.00)	(0.01)	(0.01)
Father's education (1=at least primary education; 0=otherwise)	-0.00	0.16***	0.13***
	(0.00)	(0.01)	(0.01)
At least one parent takes migrant job (1=yes; 0=otherwise)	0.01**	-0.04***	0.01
	(0.00)	(0.01)	(0.02)
Household asset (1=higher than median; 0= otherwise)	-0.02***	0.05***	0.04**
	(0.00)	(0.02)	(0.02)
Constant	-0.02	0.41***	0.80***
	(0.02)	(0.08)	(0.10)
Observations	26,919	19,101	13,645
R-squared	0.006	0.056	0.039

Note: 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 not included in any of the regressions. Robust standard errors in parentheses; *** p<0.01, ** p<0.05

Finally, we consider whether the observed gender difference is sensitive to the ethnic minority status (Table B4). We compare the indicator outcomes between boys and girls after excluding ethnic minorities from our sample. The results mirror the comparison when ethnic minorities were included (Table 5), suggesting that the gender gap does not depend on ethnicity.

Table B4. Comparisons of health, nutrition, educational status, and non-cognitive outcomes of girls and boys in rural China (excluding ethnic minority students)

Outcomes number	Outcomes	Girl	Boy	P-value
1	WAZ	-0.55	-0.50	0.79
2	HAZ	-0.78	-0.78	0.99
3	Anemia prevalence	0.15	0.13	<0.01
4	STH infection rate	.37	.44	.34
5	MHT	38.52	35.89	<0.01
6	Self-esteem	25.01	25.35	<0.01
7	Self-efficacy	25.04	25.45	0.01
8	Standardized math test	-0.08	0.07	<0.01
9	Math course grades given by teachers	0.02	-0.02	0.05
10	Standardized Chinese test	0.04	-0.04	<0.01
11	Chinese course grades given by teachers	0.15	-0.14	<0.01
12	Working memory	88.60	86.25	<0.01
13	Processing speed	78.88	78.16	0.02

Data source: Authors' data