International Journal of Educational Development 39 (2014) 203–214

Contents lists available at ScienceDirect

International Journal of Educational Development

journal homepage: www.elsevier.com/locate/ijedudev

Gender differences of academic performance in compulsory education in rural Southwestern China

Xin Gong [a,*], Yanqing Ding [b], Mun C. Tsang [a]

[a] Teachers College, Columbia University, USA
[b] Graduate School of Education, Peking University, China

ARTICLE INFO

Keywords:
Gender difference
Gender gap
Academic performance
HLM
China

ABSTRACT

Based on data from the Southwest Basic Education Project, this paper provides an empirical estimation of gender gaps in Chinese and math among primary and lower-secondary school students in poor counties of Southwestern China. Major findings from 2-level HLM analysis by grade (3, 5, 7, and 9) include: (1) small positive gaps favoring girls in Chinese and negative gaps in math were found for all grades. (2) Gaps were larger in higher grades, girls' advantage in Chinese experienced a small increase over the 2006–2010 period, and gaps varied by province. (3) Being a minority and having a low SES interplayed with gender in some circumstances.

1. Introduction

Gender difference in student academic achievement has been a heated research and policy topic for decades, because of its significant wage effect for adult life (Christie and Shannon, 2001; Murnane et al., 2004; Rose, 2006) and the equity concern for reducing gender gap in education (Marks, 2008). Most studies found a positive gender gap (female better) in language art; and a traditional negative gap in STEM (Science, Technology, Engineering, and Mathematics) that might have decreased over time and across countries in recent years (Zhang and Tsang, 2012).

There are relatively few empirical studies of gender gap in academic achievement in China. Available evidence indicates that girls have higher achievements in Chinese and English, and the gender gap in math is mostly not significant (Hannum et al., 2008; Lai, 2010; Lu and Du, 2010; Zhang et al., 2010; Zhang and Tsang, 2012). However, some studies did find a negative gender gap in math (Turner, 1994; Wang et al., 2012). Among these empirical studies, there are relatively fewer rural studies than urban studies that fully focused on the gender gap issues. Similarly, a recent media discussion in China has drawn attention to the trend that girls are outperforming boys in every subject (including math), with a focus on urban schools (Beijing Youth Daily, 2012). The limited number of rural studies that looked at rural settings only reported the results of gender academic difference as a byproduct rather than the research focus; and their reported gaps are mostly positive for Chinese (Hannum et al., 2008; Liang and Du, 2011; Lu and Du, 2010) or for a total score (Brown and Park, 2002), while the gaps are either insignificant or negative for math (Liang and Du, 2011; Lu and Du, 2010; Sun et al., 2009; Wang et al., 2012). Moreover, few rural or urban studies examined gender differences over time and by ethnic groups. It is important to empirical estimate the gender gap with a rigorous model to test whether girls perform better than or similar to boys in poor rural areas of China. Also important is to examine whether gender difference varies by subject, across grades, over time, by ethnic groups, and by other characteristics.

Based on a 3-wave student and school level data from the Southwestern Basic Education Project (SBEP), this paper attempts to examine gender difference of primary and middle school students' academic performance in China's poor rural counties. SBEP is an education improvement project jointly initiated by the United Kingdom and China. The project began in 2006 and was completed in 2010, targeting selected schools in 27 poor counties of Guangxi, Sichuan, Guizhou and Yunnan. This dataset provides a unique opportunity to explore gender gaps in a region with a large variety of ethnic minorities. For example, Yunnan province has the most variety of ethnic groups among the provinces in China. There are 25 ethnic large minority groups (with a population over 5000) in Yunnan (Government of Yunnan Province, 2012), accounting for one third of the total population of the province. Guangxi and Guizhou have a large number of ethnic minority groups as well.

* Corresponding author at: 525 West 120th Street, Box 11, Macy 348, Center on Chinese Education, Teachers College, Columbia University. New York, NY 10027, USA. Tel.: +1 917 370 8233.
E-mail addresses: xg2143@columbia.edu, xg2143@tc.columbia.edu (X. Gong).

http://dx.doi.org/10.1016/j.ijedudev.2014.07.004

Besides, this dataset covers four grades (3, 5, 7, and 9), providing the opportunity to explore gender gap in academic performance during the compulsory education cycle.[1]

The next section of the paper is a brief literature review and the statement of key research questions. Section 3 describes the data sources and the estimation methods. Section 4 presents the empirical findings and Section 5 is a summary and conclusion.

2. Gender and academic performance: a brief review

There has been plenty of research in the English literature about the determinants of student academic performance, where education production function theory is used as the analytical framework. Similar interests also apply to China (An et al., 2007; Park and Hannum, 2001; Tsui, 2005; Liu and Lu, 2008; Zhang and Tsang, 2012). However, because of the lack of publicly released student achievement data, there are fewer studies on China.

Among the studies regarding student performance in China, earlier studies have found SES and related family factors to be key determinants apart from the usually unmeasured ability factor. For instance, SES was found to have a weak but positive correlation with academic performance in meta analysis (White, 1982; Zou, 1994). Liu and Lu (2008) identified a positive association between family SES and academic performance based on a rural sample in western China. Other important determinants include age (+, Sun et al., 2009), gender (e.g. Brown and Park, 2002; though not all studies confirmed gender to be a significant factor), minority status (−, Liang and Du, 2011), number of siblings (−, Liang and Du, 2011), family income or SES (+, Tsui, 2005; Liang and Du, 2011), migration status (−, Liang and Du, 2011), parental expectations (+, Tsui, 2005), education level of fathers (Zhao, 2005), boarding status (+, Du et al., 2010; −, Lu and Du, 2010), academic aspirations (+, An et al., 2007; Sun et al., 2009), and industriousness (+, An et al., 2007; Sun et al., 2009). Classroom level factors such as teacher's education level (Zhang et al., 2010; Du et al., 2010), and having local teachers (An et al., 2007), are also found to be important. However, fewer studies incorporated school level predictors due to the lack of relevant information. Besides, in terms of outcome subject and measures, most studies used test scores of Chinese and/or math, or a measure of total score, whereas only a few examined the determinants of academic performance in other subjects.

Regarding gender gap, for empirical studies outside China, findings regarding the gender factor have been consistent for language art achievement and mixed for math scores. For many countries, the gender gap is generally positive (female better) in language (reading) and negative (male better) in math (Bedard and Cho, 2010; Dwyer and Johnson, 1997; Kenney-Benson et al., 2006; Kimball, 1989). The positive gender gap in language seems to be enlarging over time while the negative gender gap in math is narrowing (National Center for Education Statistics, 2004; Holmlund and Sund, 2008; Marks, 2008). Also, the degree of decrease in the gender gap in math over the time period seems to vary by country (Bedard and Cho, 2010; Holmlund and Sund, 2008); the gender gap in math was actually found to be positive in some countries such as Hungary and Sweden (Schmidt and Kifer, 1989). More interestingly, some studies found that math gap occurred early in early elementary school years and grew with the grade, on the basis of longitudinal samples of children in the U.S. (Fryer and Levitt, 2010). The timing has not been clearly established in previous literature (Aunola et al., 2004), although Sohn (2012) argued that some general consensus of the U.S. literature on gender gap in math was reached, saying that "a consistent math gap in favor of males does not appear until adolescence; the gap widens as individuals grow older".

However, for China, empirical analysis of gender gap studies in primary and secondary education achievements is rather limited. Consider first studies of urban areas. Turner (1994) studied gender differences in mathematics performance among 235 Chinese middle school students in Wuhan. Relying on *T* test and test of variance, he reported no significant differences in the mean scores in the logic sub-test but significant differences in favor of the boys in space and numeracy sub-tests. In a recent paper based on data from 7235 students in the Dongcheng District of Beijing and a value-added ordinary least squares (OLS) estimation approach, Lai (2010) found that girls outperformed boys in Chinese and English in both the Middle School Graduation Exam (MSGE) and High School Entrance Exam (HSEE). In math, girls performed better in MSGE but boys performed better in HSEE. More recently, Zhang and Tsang (2012) studied gender differences in academic achievement of high-school graduates in Jinan city. Using student scores in the national college entrance examination (NCEE), they found a significant and positive gender gap in Chinese and English but no significant gap in math. Chen (2012), with a sample of NCEE takers in Jiangmen city of Guangdong Province, had similar findings.

For rural China, most of the relevant studies are based on data for western China and mostly for primary education. First, we identified three studies that used data from the Gansu Survey of Children and Families (GSCF) to explore predictors of academic achievement in rural Gansu, a province in northern China; all of them targeted 9–12 year olds and therefore primary school students; and their findings on the gender differences are mixed. Note that in these Gansu studies, gender is mostly a control variable in a typically OLS regression model. At a closer look, An et al. (2007)'s study of teaching quality and student outcomes used a sample of 1926 primary school children in the GSCF Survey of 2000 and found that gender was insignificant in the achievement regression (the achievement measure is a combined score of math and language scores for children aging 9–12 in primary grades) but positive in the industriousness equation. Using the updated GSCF Survey in 2004, Sun et al. (2009) reported no gender difference in Chinese but a small girls' disadvantage in math at the primary school level for children aged 9–12 years old. Both of these two studies employed OLS; they focused on family and school factors of academic performance in general, not gender gap. Using the GSCF data in 2000 and 2007 and for the sample of children aging 9–12, Hannum et al. (2008) found a positive gender gap in Chinese and no significant gap in mathematics. The focus of this paper is on the family predictors of gender gap. At least for Gansu, their findings suggest that "rural parental educational attitudes and practices toward boys and girls are more complicated and less uniformly negative for girls than commonly portrayed" (p. 3).

Second, based on OLS modeling for a sample of about 4000 grade 4 and grade 5 students in 75 rural primary schools of Qinghai and Ningxia (close to Gansu) in 2009, Wang et al. (2012) found a female disadvantage in math; such an negative gender gap in math still existed for home-staying girls (not statistically significant for school boarding girls) after controlling for nutrition intake (anemia) and anxiety level. This study paid direct attention to the gender gap issue.

Third, there have been three rural studies that involve provinces in Southwestern China, mostly of which involve primary education. Although they only reported findings on gender academic gaps as byproducts, which are additional to their main research questions (e.g. school merger, teacher effectiveness or poverty),

[1] In China, compulsory education consists of primary school and lower-secondary school years (from grade 1 to grade 9). The legal school starting age is either six (e.g. in Yunnan) or seven (e.g. in some places of Guizhou), but poor families may send their kids to school later because of financial constraints or other reasons.

their findings are informative. First, an early study by Brown and Park (2002) utilized a 1997 survey of household and studied 472 school aged children (age 5.5–17) in poor rural counties of six provinces in different parts of China (including Sichuan and Guizhou of Southwestern China). They found that girls in lower-secondary school outscored boys by 0.2–0.7 standard deviation in a standardized overall examination score, i.e., average scores of the most recent language and math exams, while the gender academic gap at the primary school level was not significantly different from zero. Second, another two studies utilized 2-level hierarchical linear model (HLM) to analyze the predictors of academic performance. In Lu and Du (2010)'s analysis on the school merger movement in rural education, a female academic advantage in Chinese was found but there was no significant difference in math, based on a sample of 787 rural primary school students in Guangxi, who were in grade 4 in 2006 and were in grade 6 in 2008. Using a student-class 2-level value-added HLM and a sample of 3326 students in 123 rural primary schools in five provinces of western China (Gansu, Guangxi, Ningxia, Sichuan and Yunnan), Liang and Du (2011) reported findings of gender gap similar to those of Lu and Du (2010).

In summary, for China, most of the studies found a positive gender gap for language arts (Chinese and English); findings for English all come from urban studies. For math, the China studies found either a negative gender gap or no significant gap. Negative gender gaps in math were sometimes reported in the rural studies while most urban studies documented insignificant gender difference in math. Most rural studies only use gender as a control variable for their other research questions. There is an insufficiency of studies that fully focus on the gender gap issue in lower-secondary education in rural China. There is a lack of studies of gender gap variation across grade-level and over time. Note also that studies that cover students from multiple grades have not separated the estimation of the gender gap by grade. They do not explore the interaction between gender and other variables to see if the gaps differ by province, by certain student, family or school characteristics.

Using data on students in four grades in compulsory education (primary and lower-secondary education) in poor rural counties in three Southwestern provinces in China over three years (2006, 2008, and 2010), this study addresses three research questions: (1) Are there significant gender gaps in Chinese and mathematics achievement in compulsory education? (2) Does the gap differ across grade and vary over time and by province? and (3) Does the gap interplay with other risk factors such as being an ethnic minority or having a low SES? Does it differ by some selected school level predictors? In addressing these research questions with two-level HLM estimation, this study attempts to contribute to the literature on gender gap in academic performance in China by carrying out analyses not done or not adequately covered in previous studies: (1) it covers three Southwestern provinces for which there is little research on this subject; (2) it examines the gender gap in four grades of the compulsory-education cycle and the analysis will be done separately by grade; and (3) for each of the two subjects and for each grade, it tracks changes in gender gap over a period of four years.

3. Data and methods

3.1. Data source and structure

The data for this study comes from three consecutive surveys conducted in SBEP schools located in 27 poor counties of Southwestern China (Guangxi, Sichuan, Guizhou, and Yunnan). SBEP was a joint project between UK and China "to support the Chinese government's target of Nine Year Compulsory Education by increasing government capacity to implement effective programs to target the most disadvantaged children".[2] The project consisted of five components: aid for poor students, teacher training, school management, monitoring and evaluation, and social development and institution development. There were three major goals: improving equitable access, raising education quality, and strengthening education management. The project was initiated in September 2006 and ended in March 2011. Accordingly, as part of the output of the project, monitoring surveys (both for students and principals) were phased in at three different times (waves): (1) baseline survey in September of 2006; (2) midterm survey in October of 2008; and (3) concluding survey in October 2010. In total, the three-times-survey cover four grades: 3rd and 5th grade in primary school, 7th grade and 9th grade in lower-secondary school.

For this study, we focus on the randomly sampled SBEP schools in Guangxi, Guizhou and Yunnan provinces while excluding Sichuan and non-SBEP schools. Corresponding to the aim of the survey, more than 98% of the schools were SBEP schools that received SBEP interventions, and these SBEP schools were randomly sampled from the pool of SBEP schools in the 27 project counties. Sichuan observations are excluded because there are no Sichuan observations in the third wave and the excluded observations are only a small portion of the data.[3] In addition, the reason for not including non-SBEP schools is that the non-SBEP schools (less than 2%) were not randomly selected. All students in the selected grades of the selected schools were then asked to participate in the surveys. Usually there were 1 or 2 classes in the primary grades, and more classes in the 7th and 9th grades. Overall, the final operational sample contains 31,590 students for the 3rd grade, 25,604 students for the 5th grade, 7959 students for the 7th grade and 7557 students for the 9th grade. The number of schools is about 400 for the primary school sample (3rd grade and 5th grade), and it is close to 100 for the lower-secondary school sample.

Standardized tests for Chinese and mathematics by grade were administrated to the sampled students in different grades. The tests were designed by national and provincial experts and then administered by the provincial department of education. Although the tests were not the same for each wave of data, they had been standardized for across-wave comparison. Additionally, there is rich student level and school level information in the student and principal surveys available for constructing predictors for these scores. For example, at the student level, apart from the commonly controlled student characteristics (age, gender, ethnic background) and family characteristics (sibling, SES), the SBEP surveys also include information on disability, orphanage, boarding status and boarding aid. School-level information includes principals' characteristics (degree, experience, certificate, etc.), school size, and school quality indicators (e.g. percentage of qualified teacher in the school).

Overall, the response rate is high. The response rate at the school level is very high and the schools did a good job in distributing and collecting the surveys. Also, the matching rate for the surveys of basic information (student and principal survey) and performance tests is over 90%. In addition, the proportion of missing data is small for most of the variables, except for a few predictors. For example, for the grade 3 Chinese sample, most of

[2] This description was cited from the website of Cambridge Education, a partner with the British Council, which led the technical assistance for the project. Source: http://www.cambed.com/International/Internationalpresence/Internationalprojects/ChinaSouthwestBasicEducationProjectSBEP.aspx. Retrieved on July 2013.

[3] The portions of Sichuan students are 4.30% for 2006 and 6.98% for 2008, relatively lower than the other three provinces. Reasons may include: there are fewer SBEP schools in Sichuan; the schools are in remote areas and smaller in scale; and no Sichuan schools participated in the concluding survey of 2010.

the variables have missing rates lower than 5%; variables with higher missing rates are sibling (16.06%), Aid (12.31%), schools' participation in SDP (School Development Project) (6.89%), proportion of senior teachers (28.95%), proportion of trained teachers (11.81%), proportion of boarding students (32.80%), proportion of minority students (9.59%).[4] Considering that there are differential missing rates for different variables in this dataset, we decide to employ multiple imputation (MI) and the dummy flag approach as the two major missing data treatment approaches (Allison, 2002).[5]

Comprehensive review from the above implies that this dataset has a two-level structure: student level and school level. We examined the data structure by subject-grade: each subject-grade combination has a contact sample. Each sample contains three different student cohorts. They entered the targeted grade in three different survey years. Moreover, since there are cases of school attrition and the entering of new schools in new waves, the number of schools and the composition of schools vary by survey years. However, the majority of schools stay across the three waves.[6] We pooled 3 waves of data for all observed students in the available schools, so as to make use of the full information of the dataset (including the benefit of using a large sample size), to allow for school fixed effects, and to obtain a clear estimate of changes in test score over time for the same grade and same subject. Therefore our estimates of the gender gaps refer to the average gender gaps for students in SBEP schools in Guangxi, Guizhou and Yunnan over three waves (2006, 2008 and 2010).

3.2. Estimation methods

We employ a 2 level HLM analysis for the determinants of academic performance by grade and by subject for three major reasons: (1) as mentioned above, the data is featured in clustered sampling (from school to grade), and the student sample is grouped within schools. (2) The data has detailed school level information, which permits modeling of level 1 intercept and level 1 predictors slopes (student level) as outcomes in level 2 equations. (3) As compared to OLS, Hierarchical linear model (HLM) allows for the analyses of data at different levels simultaneously (Bryk and Raudenbush, 1992; Raudenbush and Bryk, 2002).[7]

School selection bias by student ability is not likely to be a problem in poor rural areas of China and is not treated in this study. Particularly, there is little school selection bias in the primary education grades (3rd and 5th), as all primary schools are neighborhood schools. There might be some concern about non-random school choice in the secondary (especially upper-secondary) school grades, for which donating fees and other hidden channels allow for school selection. However, in lower-secondary grades of poor counties in remote areas, school selection by students is less severe because there are high traveling costs to commute more in these areas,[8] and because poor families cannot afford paying for school selection.

As the main analysis, we used the intercept-as-outcome model specification to make comparison across grades. Some other specifications including a slope-as-outcome model in which slope equations for gender, SES and aid are modeled in level 2, are also tried. Estimates for the gender gap are similar (available upon request). We conducted HLM estimation for each of the eight samples (four grades, two subjects). The equations of the intercept-as-outcome 2-level HLM model are as follows:

$$\begin{aligned}\text{Level 1}: \quad & y_{ij} = \beta_0 + \beta_1 \text{girl} + \beta_2 C_{ij} + \beta_3 F_{ij} + \beta_4 FS_{ij} + \beta_5 TS_{ij} \\ & \quad + \delta_1 \text{Time 1} + \delta_2 \text{Time 2} + \delta_3 \text{Guangxi} + \delta_4 \text{Guizhou} + \varepsilon_{ij} \\ \text{Level 2}: \quad & \beta_0 = \pi_{00} + \pi_0 \text{Principal}_j + \gamma_0 \text{other\_S}_j + u_{0j}\end{aligned} \tag{1}$$

where subscript i indicates student unit and j indicates school unit.

In the level-1 (student level) equation, y stands for standardized test scores of Chinese or Math; girl is the dummy for gender; C (child characteristics) indicates other child characteristics such as age, minority, minority language (speaking a minority language at home), disability; F (family characteristics) includes number of sibling, Orphan, and SES (family socio-economic status); FS (family–school related factors) includes boarder, aid for boarding, and school–home distance; TS (student reported classroom experience[9]) includes group learning (experienced group learning or not) and No_TECH (1 if the classroom has no technical equipment); Time 1 and Time 2 are the two time dummies (students in the Time 3 year (2010) as the reference group); and Guangxi, Guizhou are the provincial dummies (Yunnan as the reference province).

In level-2 (school level), the intercept from level-1 equation is modeled against a number of school level predictors. Among them, Principal indicates a set of principal predictors: P_female (being a female principal), P_minor (being a minority principal), P_age (age of principal), P_degree (principal education degree), P_years (years of being a principal), P_certi (having a principal certificate), and P_teach (still teach students when being a principal); and Other_S (other school characteristics), including S_dis (the distance of the school to township center), Equip (available school equipment), Tenrol (school enrollment of that time), SDP (whether the school is a School Development Project[10] participant), Tea_Q_R (percentage of qualified teachers in the school), Tea_S_R (percentage of senior teachers in the school), Tea_AM_R (percentage of Arts/Music/PE teachers in the school), Tea_T_R (percentage of teachers who received SBEP training), School SES, Boarder_R (percentage of boarding students), Minority_R (percentage of minority students), and SBEP (status of being a SBEP school).

The definitions and measures of all variables are presented in Table 1. The coefficient of "girl" represents the gender gap, and the coefficients of the two time dummies reflect the changes of scores for an average student in the sample.

Apart from above, in order to further explore the changes of gender gap over time and the sources of gender gap, interaction

[4] Samples for other grades have similar missing patterns, with the only difference being that samples for lower secondary grades have comparatively lower missing rates of major variables than the samples for primary school grades.

[5] However, MI results are not available for grade 7 and grade 9 because the corresponding imputations were not successful due to data and model complexity. So we will mainly report the results from the dummy flag approach.

[6] For instance, in the third grade sample, 77.19% of the schools participated in all three waves, and an additional 13.41% showed up in two waves. Other samples have a portion of 58.54–77.85% showed up in all three times as well.

[7] Specifically, compared to pooled OLS which ignored clustering and nesting, the advantages of using HLM include: (1) take into account within school correlation/clustering (e.g. the dependence of student performance among students in the same class/grade/school, because of peer effect and other usually unmodeled or unmeasured factors); (2) get the right standard error, accurately accounting for uncertainty in prediction and estimation (the violation of the independence of errors assumption may bias the standard error of the regression coefficients); (3) explicitly model between group difference, as compared to being simply ignored or analyzed simply (with dummies or two level predictors) or separately (subsample analysis); (4) allow for heterogeneous effects of X (child and family characteristics) by school.

[8] In these areas, students usually have to travel many hours to attend a nearest middle school.

[9] We also have variables reflecting teacher–student interaction, such as frequencies of answering questions in the classroom (FreeQ), making your own comments (FreeE), time needed to finish your homework (Homework), teacher's help for your homework (T_help), award and punishment, etc. However, they are not included into the model because they are suspicious of high endogeneity.

[10] SDP (School Development Project) is the symbolic initiative of SBEP, aiming to achieve its goal of improving school management. It involves training for school principals to improve their capability of managing schools.

terms are added to the level-1 equation to see if gender difference varies by survey year, by province, by minority status, and by low SES status. For instance, we add girl*Time dummies to analyze the year by year difference of the gender gap.[11] The year by year difference may reflect some influence from the SBEP interventions. In a similar vein, provincial difference in gender gap is estimated by adding girl*provincial dummies. The association between gender and two risk factors (minority and low SES) were also investigated by adding interactions into the level-1 equation. The two interactions, i.e., boy (girl) *minority and boy (girl)* low_SES, represent double risks. Here minority has the same measure, whereas low_SES is defined as having a family SES lower than sample average. Note also that for the Chinese scores, a ‘‘boy'' dummy may replace the ‘‘girl'' dummy in case where the gender gap is positive to capture a risk rather than an advantage. The equation set is as follows:

$$\begin{aligned}\text{Level 1}: \quad & y_{ij} = \beta_0 + \beta_1\text{girl} + \beta_2 C_{ij} + \beta_3 F_{ij} + \beta_4 FS_{ij} + \beta_5 TS_{ij} \\ & + \delta_1\text{Time1} + \delta_2\text{Time2} + \delta_3\text{Guangxi} + \delta_4\text{Guizhou} \\ & + \beta_6\text{girl} * \text{Time1} + \beta_7\text{girl} * \text{Time 2} + \beta_8\text{girl} \\ & * \text{Guangxi} + \beta_9\text{girl} * \text{Guizhou} + \beta_{10}\text{girl(boy)} \\ & * \text{minority} + \beta_{11}\text{girl(boy)} * \text{low\_SES} + \varepsilon_{ij} \\ \text{Level 2}: \quad & \beta_0 = \pi_{00} + \pi_0\text{Principal}_j + \gamma_0\text{other\_S}_j + u_{0j} \end{aligned} \tag{2}$$

In addition, a non-random varying gender coefficient model will be used to explore the varying of gender gap by 3 selected school characteristics, namely, female principal, school size and school SES. They are expected to be more likely to generate impacts on the gender coefficient.[12] The equation set is listed below.

$$\begin{aligned}\text{Level 1}: \quad & y_{ij} = \beta_0 + \beta_1\text{girl} + \beta_2 C_{ij} + \beta_3 F_{ij} + \beta_4 FS_{ij} + \beta_5 TS_{ij} \\ & + \delta_1\text{Time1} + \delta_2\text{Time2} + \delta_3\text{Guangxi} + \delta_4\text{Guizhou} + \varepsilon_{ij} \\ \text{Level 2}: \quad & \beta_0 = \pi_{00} + \pi_0\text{Principal}_j + \gamma_0\text{other\_S}_j + u_{0j} \\ & \beta_1 = \pi_{10} + \pi_{11}P\_\text{female}_j + \pi_{12}\text{Tenrol}_j + \pi_{13}\text{SchoolSES}_j \end{aligned} \tag{3}$$

Finally, it is important to note that our HLM's null model (one way ANOVA with random effects) results indicated that the intra-class correlation coefficient (ICC) is 0.28–0.46.[13] This means that at least 28% of the total variation of scores came from the school level, which confirmed the necessity to conduct a HLM analysis.

4. Results

4.1. Descriptive statistics and raw gender gaps

Table 2 presents the means of the variables in the three years. As seen in the dependent variable row of Table 2, the means of raw test scores for the sampled students were all around 50, in all grades and in both subjects. About 49% of the students are girls. This type of gender-balanced enrollment could be seen as an achievement for schools in these poor rural counties. The percentage of minority students was high across grades (41–59% compared to the national average of around 9%), reflecting the population's ethnic structure. Group teaching was widely adopted (70%), but there were about 20% of students who were in classrooms without any technical equipment. At the school level, female principals were rare. Most of the principals did teaching at the same time of being the principal. The two lower secondary grades had larger school sizes and higher ratios of boarding students.

Table 3 presents the raw gender gaps by grade and by subject calculated from raw test scores, by time and province. Generally, it shows a positive gender gap in Chinese and a negative one in math. The gap seems to increase over time for Chinese for grade 5 and grade 7. But overall the growth direction tends to be mixed for math. In addition, there are some provincial differences.

We also computed raw gender gaps by selected student and school characteristics (descriptive table not shown but available upon request). One interesting finding is on SES, high SES is related to a deduction in girls' academic advantage in Chinese while exacerbating girls' academic disadvantage in math. So was the direction of effect for average school SES.

4.2. Significant predictors of academic performance

The impacts of level 1 and level 2 predictors were identified from the estimated coefficients of the intercept-as-outcome model given in Eq. Equation set 1. Results based on the dummy flag missing data treatment technique are described below and are not shown in tables due to page limit (available upon request).

We have identified several significant student level predictors for Chinese performance, including age (−), girl (+), being a minority student (−), number of sibling (−), SES (+), receiving boarding aid (−), being a boarder (−), group teaching (+), Time 1 (−) and Time 2 (−) for all grades. Distance from home to school had a negative impact on student academic performance for most grades. In addition, disability (−), and orphan (−) were statistically significant only for the two primary grades. Similarly, Guizhou SBEP students had a relatively lower average Chinese score for primary school students (0.2–0.3 standard deviation). Most of the signs were as expected and consistent with previous studies (An et al., 2007; Sun et al., 2009; Lu and Du, 2010). It is also important to note that group learning experience showed a positive impact on Chinese performance, which is part of the rationale for the SBEP intervention on teachers and might somewhat reflect SBEP's impact on changing the way of teaching and learning. For the unexpected negative impact of receiving boarding aid, we infer that it might come from the fact that the recipient status of boarding aid was related to low SES and other disadvantage factors. Additionally, school SES exhibited a positive impact on Chinese scores, and the impact spanned across grade 3, grade 5 and grade 9.

Similar patterns were identified for the math performance model.[14] For example, being a minority student, SES, distance, receiving boarding aid, group teaching have the same signs of impacts as in the Chinese performance model, across the four grades.

4.3. Findings on the average gender gap

As the focus of this paper, the estimates of gender gap by grade-subject are shown in the first two rows in Table 4. To ensure

[11] To some degree, this is generating the same effect of estimating the slope equation for ''girl'' where Time dummies are controlled.

[12] Other school level variables were also tested but found not to be statistically significant.

[13] The corresponding ICC for the 3rd grade Chinese sample, the 3rd grade math sample, the 5th grade Chinese sample, the 5th grade math sample, 7th grade Chinese sample, the 7th grade math sample, the 9th grade Chinese and the 9th grade math sample are 28.24%, 30.83%, 28.34%, 33.58%, 45.02%, 46.31%, 35.71%, 37.96%, respectively.

[14] The math findings are different from the Chinese findings in four ways: (1) the impact of being older was not consistently negative (not significant for grade 3 and grade 9). (2) Girls performed worse than boys in math. (3) The coefficient of SES was not statistically significant for grade 9 students. (4) There were no significant provincial differences in math performance. (5) School level predictors of average math scores for the school included the number of equipment (−) and the ratio of minority students (+) for primary schools. These finding are generally consistent with those reported in previous studies on China.

Table 1
Measures of the variables.

Variable			Measurement
Dependent variable	T_score		Standardized test scores (mean = 0; standard deviation = 1), 0–100
Student level	*C*	Age	Continuous: 7–13 for grade 3; 10–14 for grade 5; 12–16 for grade 7; 14–18 for grade 9
		Girl	Dummy: 1 for girl; 0 for boy
		Minority	Dummy: 1 for minority; 0 for non-minority
		Minority_language	Dummy: 1 for minority language; 0 otherwise
		Disability	Dummy: 1 if disable; 0 otherwise
	F	Orphan	Dummy: 1 if orphan; 0 otherwise
		Sibling	Continuous: the number of siblings for the student, 0, 1, 2...
		SES	Ordered categorical: created by combining the number of listed assets in the family (including: clothing, food, and pocket money; and the ownership of telephone, house type, and car), standardized into 0–12 levels, low to high
	FS	Boarder	Dummy: 1 if board in school; 0 otherwise
		Aid	Dummy: 1 if receiving boarding aid; 0 otherwise
		Distance	Categorical: 1–4, from close to far
	TS	Group	Dummy: 1 if the teacher for the class used a group teaching method in the class
		no_tech	Dummy: 1 if there is no technical equipment for the classroom
	Time dummies	Time 1	Survey time 1, referring to Year 2006
		Time 2	Survey time 2, referring to Year 2008
		Time 3	Survey time 2, referring to Year 2010
	Provincial dummies	Yunnan	1 if in Yunnan province; 0 otherwise
		Guangxi	1 if in Guangxi province; 0 otherwise
		Guizhou	1 if in Guizhou province; 0 otherwise
School level	Principal predictors	P_age	Continuous: the principal's age
		P_female	Dummy: the principal's gender, 1 if female; 0 otherwise
		P_minority	Dummy: the principal's ethnic background, 1 if minority
		P_degree	Ordered categorical: principal's degree level: order from 1 to 4, high level to low level (college, secondary college, secondary technical school, high school and lower)
		P_years	Continuous: principal's years of working the school
		P_certi	Dummy: 1 if the principal is certified; 0 otherwise
		P_teach	Dummy: 1 if the principal is teaching while being the principal
	Other_S	S_distance	Ordered categorical: 1–4, distance to the township center, from close to far
		Equip	Continuous: number of equipments (0–10), including projector, video camera, DVD, teaching aids, computer, internet, etc.
		Tenrol: school size	Continuous: number of students in the school
		SDP	Dummy: 1 if the school participated in the SDP (School Development Plan); 0 otherwise
		School SES	Continuous: constructed based on student family SES, range: 1.477–11.802
		Tea_Q_R	100*ratio of qualified teachers
		Tea_S_R	100*ratio of senior teachers
		Tea_T_R	100*ratio of trained teachers
		Boarder_R	100*ratio of boarding students
		Minority_R	100*ratio of minority students
		Tea_AM_R	100*ratio of arts, music and PE teachers

robustness, two missing data treatment techniques are used for the analysis; they are the dummy flag method and the multiple imputation method.

The results indicate that for Chinese, girls performed better than boys across all four grades, ranging from 0.059 to 0.086 of a standard deviation. The gap was larger in lower secondary grades. For math, however, the gender gap was consistently negative across the four grades, ranging from −0.085 to −0.022 of a standard deviation. The largest gap was seen in the 7th grade, and the impact was marginally significant at 10% level for grade 9. There was almost no difference in the gender gap estimates from the two missing data treatment approaches. It should be noted that the estimated gender differences was small in magnitude, less than 0.1 of a standard deviation, or one point of raw score; also, compared to other student background variables in our study (e.g. SES), the size of the gender gap was relatively small. However, the size of the gender gap was similar in magnitude to that of being a minority or speaking a minority language at home.

Results also indicate some increases in gender gap over grades. For example, based on the dummy-flag approach, the increase in the positive gap in Chinese between grade 3 and grade 7 was between 0.020 and 0.027 of a standard deviation. The increase in gender gap in math (in absolute magnitude) over grades was also small but larger than that for Chinese. The larger increase in math gap between the third grade and the seventh grade might be related to the growth pattern of boys' math capability.[15]

As a note, we also conducted OLS estimates of the gender gaps, which were somewhat different but similar to the above findings from HLM. For example, the OLS estimate of the gender gap for the grade 3 Chinese sample was 0.057, about 10% lower than the HLM estimate of 0.065.[16]

Linking our findings to the literature, studies of gender gap at the primary school level in rural China generally found a positive gender gap in Chinese but either a negative or no gender gap in math, and our study adds to this literature with similar findings. For example, our finding of a positive gender gap in Chinese at the primary level in Southwestern rural China is similar to that of the study of Guangxi in the same region by Lu and Du (2010). Although our finding of a negative gender gap in math at the primary level is

[15] It should also be stressed that changes in gender gap over grades could possibly reflect changes of student composition (i.e., the cohort effect). However, our data does not allow for separation of the grade and composition effects.

[16] In this example, the *R* square of OLS model is 0.165, compared to HLM's two pseudo *R*-square measures: 0.108 for the proportion of variance explained at level 1, and 0.189 for the proportion of variance explained at level 2. Yet the measures in HLM are not directly comparable to the OLS *R* square.

Table 2
Selected mean statistics for four grades by subject (Chinese and math) and by grade.

Subject		Chinese			Math		
Grade		Grade 3	Grade 7	Grade 9	Grade 3	Grade 5	Grade 7
Sample size (*N*)		31,198	8114	7624	31,590	25,604	7959
Variable		Mean					
Dependent variable	T_score	49.630	48.049	49.423	47.289	48.757	52.358
Student level	Age	9.912	13.583	15.600	9.892	11.841	13.616
	Girl	0.490	0.486	0.494	0.490	0.488	0.489
	Minority	0.415	0.602	0.573	0.415	0.423	0.604
	Minority language	0.268	0.405	0.345	0.268	0.260	0.406
	Disability	0.025	0.034	0.024	0.025	0.023	0.034
	Orphan	0.013	0.006	0.003	0.013	0.007	0.006
	Sibling	2.512	2.073	2.136	2.512	2.472	2.063
	SES (0–12)	5.045	5.276	5.757	5.034	5.285	5.276
	Boarder	0.096	0.645	0.629	0.096	0.122	0.645
	Aid	0.103	0.446	0.447	0.105	0.126	0.444
	Distance (1–4)	1.608	2.465	2.502	1.612	1.595	2.466
	Group	0.706	0.740	0.663	0.708	0.733	0.740
	no_tech	0.202	0.225	0.173	0.202	0.240	0.222
	Time 1	0.380	0.319	0.398	0.377	0.421	0.322
	Time 2	0.353	0.334	0.308	0.353	0.405	0.337
	Time 3	0.267	0.347	0.294	0.269	0.173	0.341
	Yunnan	0.338	0.332	0.316	0.334	0.316	0.333
	Guangxi	0.162	0.372	0.324	0.161	0.174	0.368
	Guizhou	0.500	0.297	0.361	0.505	0.510	0.299
School level	P_age	38.349	39.606	39.573	38.384	37.977	39.598
	P_female	0.039	0.019	0.022	0.039	0.046	0.019
	P_minority	0.411	0.586	0.565	0.412	0.424	0.589
	P_degree (1–4)	2.282	1.598	1.614	2.281	2.280	1.601
	P_years	6.450	5.600	5.829	6.482	6.037	5.602
	P_certi	0.713	0.820	0.836	0.716	0.728	0.817
	P_teach	0.909	0.896	0.893	0.909	0.914	0.899
	S_distance	1.776	1.232	1.217	1.773	1.763	1.237
	Equip	5.375	7.162	7.121	5.395	5.362	7.149
	Tenrol: school size	671.688	1016.418	1077.520	673.132	675.519	1015.604
	SDP	0.592	0.581	0.591	0.595	0.593	0.578
	School SES	5.130	5.412	5.701	5.123	5.254	5.411
	Tea_Q_R	87.083	93.963	93.590	87.261	87.577	93.920
	Tea_S_R	21.193	5.531	4.924	21.240	20.585	5.489
	Tea_T_R	54.655	72.543	69.977	54.649	52.862	72.066
	Boarder_R	19.419	49.943	48.490	19.387	19.493	50.268
	Minority_R	41.274	54.838	52.235	41.262	42.780	54.891
	Tea_AM_R	2.764	3.449	3.695	2.744	2.685	3.446

Table 3
Raw gender gap for Chinese and Math by time and province.

Subject by year and province			Primary school		Lower-secondary school	
			Grade 3	Grade 5	Grade 7	Grade 9
Chinese	2006	All	0.2125	0.4685	−0.2784	0.0334
		Yunnan	0.3227	1.2937	0.1017	−0.0578
		Guangxi	−1.8282	−2.7836	−1.5254	0.1735
		Guizhou	0.5618	0.5873	0.3918	0.7688
	2008	All	1.2441	0.4784	1.2903	2.0931
		Yunnan	0.5358	1.2025	−0.0793	0.1445
		Guangxi	2.7321	2.9259	2.0518	3.5399
		Guizhou	1.1572	−0.6725	1.8791	2.1661
	2010	All	0.8436	1.2838	1.3453	0.2905
		Yunnan	0.7974	0.6876	1.8167	0.6314
		Guangxi	3.1825	3.0767	2.0634	1.6513
		Guizhou	0.0785	0.8531	−0.0786	−0.6444
Math	2006	All	−0.6349	−0.1974	−0.7074	−0.4449
		Yunnan	−1.0015	−0.6664	−2.1309	−2.8651
		Guangxi	0.2397	0.1458	−0.1275	1.8389
		Guizhou	−1.0063	−0.3096	−0.8232	−1.6373
	2008	All	−0.4836	−1.2142	−1.6686	−0.0373
		Yunnan	−1.0496	−0.9825	−3.5193	−1.9105
		Guangxi	−0.4075	0.2926	−0.5808	0.8802
		Guizhou	−0.2106	−1.8214	−0.9005	0.8142
	2010	All	−0.3189	0.3926	−0.3196	−0.6429
		Yunnan	−0.1925	−0.0278	0.2274	−1.1141
		Guangxi	1.2793	1.7297	−0.0735	−0.0964
		Guizhou	−0.8822	−0.0116	−1.4551	−0.6216

Note: the differences were computed based on the raw scores of non-missing units.

Table 4
Gender difference in academic outcomes across grades: Chinese and math.

Models		Chinese				Math			
		Grade 3	Grade 5	Grade 7	Grade 9	Grade 3	Grade 5	Grade 7	Grade 9
Dummy flag	Intercept-as-outcome	0.065***	0.071***	0.085***	0.076***	−0.024***	−0.028***	−0.085***	−0.034*
		(0.009)	(0.010)	(0.017)	(0.018)	(0.009)	(0.010)	(0.017)	(0.019)
	Complete model with best fitness	0.059***	0.067***	0.086***	0.081***	−0.029***	−0.032***	−0.082***	−0.040*
		(0.009)	(0.011)	(0.017)	(0.022)	(0.009)	(0.011)	(0.021)	(0.023)
MI	Intercept-as-outcome	0.066***	0.072***	–	–	−0.022**	−0.024**	–	–
		(0.009)	(0.010)			(0.009)	(0.010)		

Notes: (a) For lower-secondary grades, MI is not available due to the complexity of missing data structure and the complexity of the imputation models (which are necessary). (b) The complete models with best fitness: for different sample by grade and subject matter, HLM models differ in terms of equation settings and variables in the equation. Mean of the level 1 variable is controlled in the slope equation for this variable in level 2. The reported results are from the models with best fitness and at least estimable (some models are not estimable due to singularity and non-convergence). (c) Complete results from the intercept-as-outcome models (and some alternative models) are not presented due to page limitation, available upon request.
* Stands for significant level of 10%.
** Stands for significant level of 5%.
*** Stands for significant level of 1%.

different from Lu and Du (2010),[17] but consistent with the finding from Sun et al. (2009) and Wang et al. (2012), which targeted primary education in northern rural China (Gansu, Ningxia and Qinghai). It is also interesting to note that research on primary education in Gansu in northern China reported somewhat different results. For example, Sun et al. (2009) found no significant gender gap in Chinese but Hannum et al. (2008) did. Sun found a small negative gap in math but Hannum reported no gap in math. All these four studies examined gender gap at the primary level, not lower-secondary education.

At the same time, our findings enrich the gender gap literature for lower-secondary education in rural China by showing a positive gender gap in Chinese and negative one in math for junior school students in Southwestern China's poor counties. Except for Brown and Park (2002), whose sample involved some children enrolled in lower-secondary schools, there have been no other studies of gender gap in rural China at the lower-secondary level. Brown and Park (2002) documented that girls score higher on a composite measure of test scores in junior secondary school, and it is consistent to our subject specific findings to a large degree.

Compared to findings from studies with urban sample, our results for rural Southwestern China shared the same conclusion of a positive gender gap in Chinese language achievement. But one important difference is that we found a negative gender gap in math whereas insignificant math gaps were reported in the urban sample studies. Our finding of a negative gender gap in math may be due to the fact that students in our sample are from the poor rural counties in Southwestern China, where boy preference still pervades and therefore girls still lack the necessary resources and support for the development of math skills as compared to urban China (Hannum, 2005; Hannum et al., 2008; Lin and Qin, 2010; Tsui, 2002; Wang, 2005; Wang et al., 2012).[18] Comparatively, the insignificant gender gap in math in urban schools may imply that parental support for girls, girls' better non-cognitive skills and their better compliance with the modern school environment in urban areas have helped the girls to develop comparable achievement in math (Lai, 2010).

4.4. Heterogeneous gender gaps

By adding interaction terms with the girl dummy into the level-1 equation (see Eq. Equation set 2), we further explored the gender gap by province, by year, by minority and by low family SES status. The results are shown in Table 5. For Chinese (Panel A), fifth grade female students' academic advantage was smaller in 2006 and 2008 than 2010; this means that the positive gender gap in Chinese was increasing over time. For fifth grade, the gender gap in Chinese also varied by province in that the gap for Guizhou province was significantly smaller than that for Yunnan province. For most grades, being a minority student did not impose double risks for the boys, except in grade 3, where a minority boy achieved a statistically significant 0.04 standard deviation (about half a score point) lower than a non-minority girl. But being in a low-SES family did not worsen boys' disadvantage in Chinese language achievement.

The pattern is somewhat different for math (Panel B of Table 5). In grade 5 and 7, female students' math disadvantage was larger in the previous two waves than the 2010 wave (when SBEP project was completed). This means that the negative gap has been narrowed. Another important finding is that, for each of the four grades, the negative gender gap in math for Guangxi students was significantly smaller in magnitude from that for Yunnan province. Such difference was not seen between Guizhou students and Yunnan students for each of the four grades. To explore more about provincial differences in the gender gap in math, we conducted separate estimation of the gender gap in math by province. As shown in Table 6, the positive math gap is largest for Yunnan, and somewhat smaller for Guizhou; but for Guangxi, the math gap is statistically positive for grade 3 students while insignificant for other grades. Overall, results consistently indicated that gender gap in math varied significantly with provincial location. In terms of double risks, low-SES instead of minority played a minor role for fifth grade students: girls in poor families have experienced a smaller negative math achievement gap than girls in high SES families, though it was only statistically significant at 10% level. At the same time, the interaction between the minority dummy and the girl dummy was not statistically significant in any grade.

By estimating Eq. Equation set 3, we have the results on gender gap by three selected school level predictors presented in Table 7.

[17] This difference may be related to differences in economic development status and differences in the degree of boy preference between rural areas of Gansu province and southwestern China. Note also that Hannum et al. (2008) found there was little evidence of gender gap in economic investment in education for the Gansu sample. This may partially explain why the gender gap in math is insignificant in most of the Gansu studies. Comparatively, there might still be gender imbalanced education investment in the three southwestern provinces under our study.

[18] While controversy exists as to whether disadvantage in math for girls is truly innate rather than environmentally developed, traditionally, girls are often thought to have genetic disadvantage in math and receive lower expectations from others in their math capability (Gaulin and Hoffman, 1988; Berenbaum et al., 2008). Such stereotype often leads to girls' lack of self-confidence in math achievement. At the same time, girls in rural families usually get fewer resources for development than boys due to boy preference. Without changes in gender ideology and in the gender imbalanced educational investment, girls in the poor areas of rural China may still perform worse in math than boys.

Table 5
Gender differences by year, by province and by multi-risks (dummy flag, intercept-as-outcome model).

Variables		Grade 3	Grade 5	Grade 7	Grade 9
Panel A: Chinese					
Estimate of the girl slope in level 1	Before adding interactions	0.065***	0.071***	0.085***	0.076***
		(0.009)	(0.010)	(0.017)	(0.018)
	This model (a)	0.050**	0.178***	0.157***	0.032
		(0.022)	(0.027)	(0.041)	(0.045)
Interactions' coefficients	girl*2006	−0.016	−0.138***	0.050	0.042***
		(0.024)	(0.029)	(0.041)	(0.045)
	girl*2008	0.015	−0.064**	−0.138***	0.134
		(0.024)	(0.028)	(0.043)	(0.047)
	girl*Guangxi	0.042	−0.039	−0.047	−0.015
		(0.030)	(0.031)	(0.043)	(0.047)
	girl*Guizhou	−0.003	−0.065***	−0.033	−0.057
		(0.021)	(0.022)	(0.043)	(0.044)
	boy*minority	−0.043**	−0.023	−0.0006	−0.050
		(0.02)	(0.021)	(0.037)	(0.038)
	boy*low_ses	0.018	−0.013	0.038	0.033
		(0.017)	(0.018)	(0.030)	(0.032)
Panel B: Math					
Estimate of the girl slope in level 1	Before adding interactions	−0.024***	−0.028***	−0.085***	−0.034*
		(0.009)	(0.010)	(0.017)	(0.019)
	This model (a)	−0.045**	0.013	−0.063	−0.155***
		(0.022)	(0.028)	(0.042)	(0.047)
Interactions' coefficients	girl*2006	−0.005	−0.100***	−0.084*	0.073
		(0.024)	(0.029)	(0.043)	(0.047)
	girl*2008	−0.040*	−0.122***	−0.123***	−0.005
		(0.023)	(0.028)	(0.041)	(0.050)
	girl*Guangxi	0.102***	0.108***	0.096**	0.122**
		(0.030)	(0.031)	(0.043)	(0.049)
	girl*Guizhou	0.017	0.013	0.069	0.064
		(0.021)	(0.023)	(0.043)	(0.046)
	girl*minority	0.017	0.021	−0.034	0.041
		(0.020)	(0.021)	(0.037)	(0.040)
	girl*low_ses	0.011	0.035*	0.024	0.012
		(0.017)	(0.018)	(0.031)	(0.034)

Notes: (a) these results are based on the dummy flag missing data treatment technique in the modified intercept-as-outcome model of Eq. Equation set 2. (b) The base group is boys in Yunnan of 2010.
* Stands for significant level of 10%.
** Stands for significant level of 5%.
*** Stands for significant level of 1%.

First, very interestingly, being in a school with a female principal increased girls' academic advantage in third-grade Chinese, whereas the female principal's impact was negative for grade 7 math. Second, at the primary level, larger school size showed a decreasing effect on the positive gender gap in Chinese and it also exacerbated the gender gap in math for grade 5. Third, school SES was positively associated with gender gaps in Chinese for the two primary grades, and this applied to the math gap for grade 3 as well. Fourth, the consistent sign of the interactions for school size and school SES cast different implications for the two subjects. If we aim to reduce the positive gender gap (or in other words, reduce boys' disadvantage) for Chinese, larger class size and lower school SES are relevant. However, if we aim to narrow the gender gap in math, smaller school size and higher school SES appear to be preferable.

Table 6
Gender difference in math by province (dummy flag, intercept-as-outcome model).

Sample		Grade 3	Grade 5	Grade 7	Grade 9
Full sample		−0.024***	−0.028***	−0.085***	−0.034*
		(0.009)	(0.010)	(0.017)	(0.019)
3 Subsamples	Yunnan	−0.067***	−0.069***	−1.249***	−0.123***
		(0.014)	(0.015)	(0.260)	(0.026)
	Guangxi	0.050**	0.037	−0.034	0.029
		(0.021)	(0.026)	(0.029)	(0.029)
	Guizhou	−0.033***	−0.025**	−0.073**	−0.016
		(0.01)	(0.012)	(0.034)	(0.026)

Notes: (a) sample size varies by provincial subsample. For example, for grade 3, proportions of Yunnan, Guangxi and Guizhou students are 33%, 16%, and 51%, respectively. Sample compositions are similar in other grades. (b) The model used for this sub group analysis by province is basically the same as listed in Eq. Equation set 1, except removing the two provincial dummies.
* Stands for significant level of 10%.
** Stands for significant level of 5%.
*** Stands for significant level of 1%.

4.5. Robustness checks

To check the robustness of the gender gap estimates, we have employed three alternative schemes as follows: (1) including in the model one additional control, i.e. grade size. Since all students in the sampled grade have been sampled, we can use the number of sampled students in the dataset in each grade as the grade size (ranging from 5 to 80 for the third grade). (2) Adding some variables with a high missing rate (e.g. whether parents were out of home for work) and some variables previously considered to be endogenous but may be important for the understanding of the education process (student expectation, frequency in raise/answer questions in the classroom) in the controls. We find that estimates of the coefficient of the "girl" dummy are robust across the above two schemes and the baseline specification used earlier in this study. (3) As already shown in Table 5, our gender gap estimates are robust across two different missing data treatment methods (dummy flag and MI).

Table 7
Gender differences by selected school level predictors: principal's gender, school size and school SES (dummy flag, intercept-as-outcome model).

Variables		Grade 3	Grade 5	Grade 7	Grade 9
Panel A: Chinese					
Non-randomly varying female slope	Intercept	0.066***	0.076***	0.086***	0.077***
		(0.010)	(0.010)	(0.017)	(0.018)
	P_femal	0.108**	0.041	−0.043	−0.014
		(0.046)	(0.046)	(0.130)	(0.127)
	Tenrol: school size	−0.00003*	−0.00006***	0.00002	0.000001
		(0.00002)	(0.00002)	(0.00003)	(0.00004)
	School SES	0.023***	0.007	0.0006	0.014
		(0.006)	(0.006)	(0.016)	(0.015)
Proportion of variance explained at each level	Level-1	10.90%	11.39%	7.01%	16.11%
	Level-2	18.85%	35.54%	30.28%	39.70%
Reliability estimates of random effects	Intercept	0.945	0.945	0.971	0.959
The number of level-1 units		31,198	25,834	8114	7624
The number of level-2 units		392	366	100	99
Non-missing rate		100%	100.00%	100.00%	100.00%
Panel B: Math					
Non-randomly varying female slope	Intercept	−0.020**	−0.020**	−0.077***	−0.033*
		(0.009)	(0.010)	(0.017)	(0.019)
	P_femal	0.004	0.010	−0.251**	−0.032
		(0.045)	(0.047)	(0.127)	(0.133)
	Tenrol: school size	−0.00002	−0.00007***	0.00002	0.00005
		(0.0002)	(0.00002)	(0.00004)	(0.00004)
	School SES	0.017***	0.016***	0.018	0.0009
		(0.006)	(0.006)	(0.016)	(0.016)
Proportion of variance explained at each level	Level-1	12.18%	13.53%	3.42%	49.49%
	Level-2	19.61%	25.32%	31.99%	17.46%
Reliability estimates of random effects	Intercept	0.952	0.948	0.97	0.968
The number of level-1 units		31,590	25,604	7959	7557
The number of level-2 units		292	366	100	99
Non-missing rate		100.00%	100.00%	100.00%	100.00%

Notes: (a) these results are based on the dummy flag missing data treatment technique in the modified intercept-as-outcome model of Eq. Equation set 3. (b) The base group is boys in Yunnan of 2010.
* Stands for significant level of 10%.
** Stands for significant level of 5%.
*** Stands for significant level of 1%.

5. Conclusions

This paper has estimated the gender gap in Chinese and math achievements among primary and lower-secondary school students in the poor rural counties of Southwestern China, based on data from the Southwest Basic Education Project (SBEP). It seeks to address several gaps or weaknesses in research on this subject in China, namely, the limitedness of studies on rural areas, in Southwestern China, and on lower-secondary education, as well as the lack of analyses of heterogeneous gender gaps and how gender gaps vary over time.

Results from our 2-level HLM models indicate that, in the SBEP schools of rural Southwestern China's poor counties, there was a positive gender gap favoring girls in Chinese while the math gap was negative favoring boys across four grades (3, 5, 7, 9) in both the primary and lower-secondary level.

The positive gender gap in Chinese is consistent to other studies of rural China. Also interestingly, while most of the other rural studies reported an insignificant gender gap in math, our study found a negative gender gap in math, which is consistent to Wang et al. (2012) only. This may imply that girls in these areas are still less invested in education than boys in the remote mountainous poor counties of rural Southwestern China, in comparison to Gansu province. Compared to studies of urban China, our study is similar in reporting a positive gender gap in Chinese; the difference lies in that our study found a negative gender gap in math while the finding was mixed in studies of urban China. This might be related to the fact that there is no or very limited boy preference in urban China. In short, in comparison to previous studies, our study confirms the positive gender gap in Chinese but has a different finding on the gender gap in math. Our study implies that girls' academic disadvantage in math may be more severe in rural Southwestern China, and it should be further examined in the future.

Moreover, our study went beyond previous studies to explain gender gaps across grades, over time, by province, and the interaction between gender and other characteristics. First, it finds that the gender gaps were larger in higher grades for Chinese and math. This might be related to a possible occurrence of differential drop out rate by gender in which "academic weak girls are more likely to drop out in primary schools while most boys continue on to junior secondary schools", as found in Brown and Park (2002, p. 523). Our paper cannot test this story directly because we only have information of the same child for a single year; but our finding of the increased gender gap in favor of girls across grade may be somewhat related to this possible occurrence. Second, we also found that female student's advantage in Chinese language increased a little bit over the three waves, while their disadvantage in math became smallest in 2010. The Southwestern Basic Education Project, therefore, might have increased girls' academic achievement to a larger degree as compared to boys, at least in the three provinces under analysis. Third, gender gap varied by geographical location: there was a significant difference in fifth-grade Chinese between Guizhou and Yunnan; and for all four grades in math, the gender gap for Guangxi was smaller than that for Yunnan. Fourth, some student and school factors were found to be associated with the gaps. In grade 3, being a minority imposed double risk for boys in Chinese, and having a low family SES caused double risks for girls in fifth-grade math. Smaller school size and higher School SES were found to increase gender gap in Chinese for the two primary grades, while similar but slightly different interaction effects only applied in the math gap for grade 3.

Overall, the average gender gaps in Chinese and math for the Southwestern region as a whole are all less than one-tenth of a standard deviation and are thus small. However, our study also finds that average gender gaps do vary significantly among the three provinces; and the gender gaps are significantly larger for Yunnan province. Given the economic, cultural, and geographical diversities among different provinces and regions in China, our findings imply that the seriousness of the gender gaps in academic achievement is likely to differ by provinces and regions. Our findings from the SBEP schools may have implications for many schools that are moving toward that direction, as the Chinese government has been investing heavily with over 10 billion dollars in the past four years to improve education quality along with the implementation of the national plan for education reform and development (2010–2020) (China Daily, 2014).

This study has several limitations. First, because of the lack of data, we have not controlled for prior test scores of the students, so the models used in this article are not value-added. We interpreted the gender gap of academic performance for a particular grade in our estimation as the accumulated gap over the schooling years till the grade of the survey. Second, we do not have information on teacher quality at the class level (while we do have some school level quality indicators). Third, although we have precluded possible endogeneity that could be generated by school selection (as long as it is not related to gender), academically weak girls may drop out earlier than boys (or the other way). This concern is needed be explored further in future studies to fully understand the meaning of the empirically estimated gender gap in cross-sectional data, even though our data combines three cohorts of cross-sectional data. This can be done in school or household survey data that also collect outcomes for drop out students. Fourth, our study is not designed to answer questions such as the reasons behind the gender gap, which may be better investigated using a mix of quantitative and qualitative research methods. Further research may be conducted to address these limitations, to explain the sources of the gender gap, and to explore interventions for reducing such gender gaps.

Acknowledgements

This study comes from a research project in Center on Chinese Education of Teachers College, Columbia University. The authors would also like to thank China Scholarship Council, for its financial support for Xin Gong.

References

Allison, P., 2002. Missing Data. Sage Publications, Thousand Oaks, CA.
An, Xuehui, Hannum, E., Sargent, T., 2007. Teaching quality and student outcomes: academic achievement and educational engagement in rural northwest China. China Int. J. 5 (2) 209–334.
Aunola, K., Leskinen, E., Lerkkanen, M., Nurmi, J., 2004. Developmental dynamics of math performance from preschool to grade 2. J. Educ. Psychol. 96 (4) 699–713.
Bedard, K., Cho, I., 2010. Early gender test score gaps across OECD countries. Econ. Educ. Rev. 29 (3) 348–363, http://dx.doi.org/10.1016/j.econedurev.2009.10.015.
Beijing Youth Daily, 2012, July. Nvsheng quanmian chaoyue nansheng ling ren you. (Girls' Better Performance in Every Subject is Worrisome).Beijing Youth Daily. Retrieved from: http://bjyouth.ynet.com/3.1/1207/05/7254774.html.
Berenbaum Sheri, A., Martin, C.L., Hanish, L.D., Briggs, P.T., Fabes, R.A., 2008. Sex differences in children's play. In: Becker, J.B., Berkley, K.J., Geary, N., Hampson, E., Herman, J., Young, E. (Eds.), Sex Differences in the Bra: From Genes to Behavior. Oxford University Press, New York (Chapter 14).
Brown, P.H., Park, A., 2002. Education and poverty in rural China. Econ. Educ. Rev. 21, 523–541.
Bryk, A.S., Raudenbush, S.W., 1992. Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods. Sage Publications, Newbury Park, CA.
Chen, Y., 2012. Putong gaozhong jiaoyu touru jixiao pinggu shizheng yanjiu: jiyu jiangmenshi 47 suo gaozhong diaocha de duoceng shuju.(An empirical study on education performance evaluation of high schools: based on the multi-level data from a survey about 47 high schools in Jiangmen). shuxue de shijian yu renshi (Math. Pract. Theory).42 (9) 248–257 (in Chinese).
China Daily, 2014, February. China Spends Billions on Rural Education. China Daily. Retrieved from: http://www.chinadaily.com.cn/business/2014-02/14/content_17282518.htm.
Christie, P., Shannon, M., 2001. Educational attainment and the gender wage gap: evidence from the 1986 and 1991 Canadian censuses. Econ. Educ. Rev. 20 (2) 165–180.
Du, P., Zhao, R., Zhao, D., 2010. Xibu wushengqu nongcun xiaoxue jisuosheng de xueye chengji yu xuexiao shiyingxing yanjiu.(A study on the academic achievement and school adaptability of rural elementary school boarding students from five provinces and autonomous regions in western China). jiaoyu xuebao (J. Educ. Stud.).6 (6) 84–91 (in Chinese).
Dwyer, C.A., Johnson, L.M., 1997. Grades, accomplishments and correlates. In: Willingham, W.A., Cole, N.S. (Eds.), Gender and Fair Assessment. Psychology Press, Erlbaum, NJ, pp. 127–156.
Fryer, R.G., Levitt, S.D., 2010. An empirical analysis of the gender gap in mathematics. Am. Econ. J. Appl. Econ. 2 (2) 210–240, http://dx.doi.org/10.1257/app.2.2.210.
Gaulin, S., Hoffman, H., 1988. Evolution and development of sex differences in spatial ability. In: Betzig, L., Mulder, M.B., Turke, P. (Eds.), Human Reproductive Behavior: A Darwinian. Cambridge University Press (Chapter 7).
Government of Yunnan Province, 2012. Yunnan sheng gaikuang(Summary of the State of Yunnan, China).. (in Chinese). Retrieved from: http://www.yn.gov.cn/yn_yngk/yn_sqgm/201201/t20120116_2914.html.
Hannum, E., 2005. Market transition, educational disparities, and family strategies in rural China: new evidence on gender stratification and development. Demography 42 (2) 275–299.
Hannum, E.C., Kong, P.A., Zhang, Y., 2008. Family Sources of Educational Gender Inequality in Rural China: A Critical Assessment, China Human Capital Projects: Gansu Survey of Children and Families Papers. University of Pennsylvania.
Holmlund, H., Sund, K., 2008. Is the gender gap in school performance affected by the sex of the teacher? Labor Econ. 15, 37–53, http://dx.doi.org/10.1016/j.labeco.2006.12.002.
Kenney-Benson, G., Pomerantz, E., Ryan, A., Patrick, H., 2006. Sex differences in math performance: the role of children's approach to schoolwork. Dev. Psychol. 42, 11–26.
Kimball, M.M., 1989. A new perspective on women's math achievement. Psychol. Bull. 105, 198–214.
Lai, F., 2010. Are boys left behind? The evolution of the gender achievement gap in Beijing's middle schools. Econ. Educ. Rev. 29, 383–399, http://dx.doi.org/10.1016/j.econedurev.2009.07.009.
Liang, W., Du, Y., 2011. Jiyu xuesheng xueye chengji de jiaoshi zhiliang pingjia – laizi zhongguo xibu nongcun xiaoxue de zhengju.(Estimating teacher effectiveness from students' academic achievement: Evidence from rural areas in western China). Beijing daxue jiaoyu pinglun (Peking Univ. Educ. Rev.).9 (3) 105–120 (in Chinese).
Lin, W., Qin, Y., 2010. Fumu de nanhai pianhao duiyu ertong xuexi zhuangkuang de yingxiang jiqi zuoyong qudao.(The impact of parental son preference on children's educational attainments and its mechanism). Econ. Sci. 4, 116–128 (in Chinese).
Liu, X., Lu, K., 2008. Student performance and family socioeconomic status results from a survey of compulsory education in western China. Chin. Educ. Soc. 41 (5) 70–83, http://dx.doi.org/10.2753/CED1061-1932410505.
Lu, K., Du, Y., 2010. The effect of layout adjustment of rural schools on student achievement: analysis based on the two-level value-added model. Tsinghua J. Educ. 31 (6) 64–73.
Marks, G.N., 2008. Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. Oxf. Rev. Educ. 341, 89–109.
Murnane, R.J., Willett, J.B., Levy, F., 2004. The growing importance of cognitive skills in wage determination. In: Henry, C.M. (Ed.), Race, Poverty, and Domestic Policy. Yale University Press, New Haven, CT.
National Center for Education Statistics (NCES), 2004. NAEP 1999 Trends in Academic Progress. NCES, Washington, DC, USA. Retrieved from: http://nces.ed.gov/nationsreportcard/pdf/main1999/2000469.pdf.
Park, A., Hannum, E., 2001. Do teachers affect learning in developing countries? Evidence from matched student–teacher data from China. Department of Education Paper prepared for the conference on rethinking social science research on the developing world in the 21st century, Park City, Utah, June 7–11.
Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd ed. Sage, Newbury Park, CA.
Rose, H., 2006. Do gains in test scores explain labor market outcomes? Econ. Educ. Rev. 26 (4) 430–446.
Schmidt, W.H., Kifer, E., 1989. Exploring relationships across systems population A. In: Robitaille, D.F., Garden, R.A. (Eds.), The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics. Pergamon Press, Oxford.
Sohn, K., 2012. A new insight into the gender gap in math. Bull. Econ. Res. 64 (1) 0307–3378.

Sun, Z., Liu, Z., Sun, B., 2009. Jiating, xuexiao yu ertong de xuexi chengji (Children's achievements and their families and schools: a study based on the rural areas in Gansu province). Beijing shifan daxue xuebao (shehui kexue ban) (J. Beijing Normal Univ. (Soc. Sci.)).5, 103–115 (in Chinese).

Tsui, Ming, 2002. The only child and educational opportunity for girls in urban China. Gend. Soc. 16 (1) 74–92, http://dx.doi.org/10.1177/0891243202016001005.

Tsui, M., 2005. Family income, home environment, parenting and mathematics achievement of children in China and the United States. Educ. Urban Soc. 37 (3) 336–355, http://dx.doi.org/10.1177/0013124504274188.

Turner, R., 1994. Sex differences in mathematical performance among Chinese middle school students. In: Bell, G., Wright, B., Leeson, N., Geake, J. (Eds.), Proceedings of the 17th Annual Conferences of the Mathematics Education Research Group of Australasia (Merga): Challenges in mathematics education: Constraints on Construction, vol. 1(2) pp. 649–666.

Wang, S., Zeng, J., Shi, Y., Luo, R., Zhang, L., 2012. Xibu rongcun diqu xiaoxuesheng jiankang yu jiaoyu xingbie chayi yanjiu.(Research on the determinants of gender difference in health and education for primary school students in western poor areas). nongcun jishu jingji (Rural Technol. Econ.).6, 4–14 (in Chinese).

Wang, W., 2005. Song preference and educational opportunities of children in China – "I wish you were a boy!" Gend. Issues 22 (2) 3–30.

White, K.R., 1982. The relationship between socioeconomic status and academic achievement. Psychol. Bull. 91 (3) 61–81.

Zhang, W., Xin, T., Kang, C., 2010. Jiaoshi bianliang dui xiaoxue sinianji shuxue chengji de yingxiang.(The impact of teacher variables on the fourth-grade math achievement: a value-added study). Jiaoyu xuebao (J. Educ. Stud.).6 (2) 69–76 (in Chinese).

Zhang, Y., Tsang, C.M., 2012. Gender gap in the National College Entrance Exam performance in China Paper presented at the International Forum on Economics of Education in Beijing, China on November 2–4, pp. 2012.

Zhao, L., 2005. Guanyu xiaoxue sheng xueye chanchu yingxiang yinsu deshizheng tansuo—yi jianhua shiyan xiaoxue weili. (Empirical Research on the Determinant Factors of Elementary School Students' Academic Performance—Evidence from Jianhua Experimental Elementary School). (Master's degree dissertation)-Beijing Normal University (in Chinese).

Zou, F., 1994. Jiating shehui jingji diwei yu xueye chengjiu zhi guanxi: hese fenxi.(The relationship between family social status and academic performance: meta-analysis). Jiaoyu yanjiu zixun (Educ. Res. Consult.).2 (3) 38–47 (in Chinese).