1. Introduction

The education gender gap has been extensively studied worldwide. Historically, women have had lower overall educational attainment than men, at least partly because of the inequitable access. However, the consistent expansion of women's education after World War II has resulted in the reversal of the education gender gap in most countries (OECD 2015a).

The Program for International Student Assessment (PISA) is a standardised test of reading, mathematics, and science for individuals aged 15 years, conducted every 3 years. The test allows a rigorous comparison of students' achievements and educational equity across countries and time. Since the tests were first administered by the Organisation for Economic Co-operation and Development (OECD) in 2000, girls have outperformed boys in reading, and boys have marginally outperformed girls in mathematics, with science performance being equal (OECD 2015b).

One notable limitation of the PISA is the under-representation of developing countries. India withdrew from the test shortly after its debut in 2009, reporting that the test disadvantaged its students by not accounting for the socioeconomic context. Although China first participated in the 2009 round, by 2018, it limited its participation to a few provinces and municipalities in its most developed coastal region. This nonrepresentativeness of the sample makes interpreting its findings difficult.

Despite China's rapid progress in developing an educational system that is advancing in line with its fast-growing economy, the inequality in this system is increasing, drawing attention from policy makers and researchers. By international standards, there is substantial inequality in both the inputs and outputs of education in China, most notably along the urban–rural dimension because of the historical *hukou* (household registration) system. However, one important aspect of this under-researched issue is the gender achievement gap. A consensus is that girls outperform boys in language and underperform boys in math, but the source of the gender gap has been rarely explored, especially in developing countries. China is an interesting country to explore the heterogeneity in the gender gap, given the massive internal immigration, son preference, *hukou* status, and educational inequality. Compared with the previous literature highlighting the student–teacher gender match (Xu and Li 2018; Gong et al. 2018), rather than estimating the contribution of characteristics to the gender gap, this paper focuses on exploring the strong heterogeneity, both observed and unobserved, in gender differences in academic achievement.

Our contributions are threefold. First, this study is one of the few studies on student gender gaps in math, Chinese, and English that accounts for school selection, streaming, and variations in cognitive abilities by using the China Education Panel Survey (CEPS) baseline survey conducted in the academic year 2013–2014. CEPS is a nationally representative survey of students in grade 7 (approximately 13 years old) and grade 9 (approximately 15 years old) in junior high schools and has a design similar to that of the PISA. Specifically, CEPS has two important features from which empirical studies would benefit most. The survey implements a uniform cognitive test for all students, which is designed to test students' innate cognitive ability rather than academic achievements. Moreover, empirical research could benefit from the information on classroom randomisation in the data. Both principals and teachers are asked whether the students in the school or the relevant grade are randomly assigned into classes. We take advantage of the random design to account for biases arising from streaming within schools and use the uniform cognitive scores to capture the difference in cognitive skills.

Furthermore, this study differs from the literature that has used the same dataset, because we restrict the data to the sample by applying the strictest selection rules on randomisation. Regardless of sample selection, there is no significant difference by gender in cognitive ability within schools. Using either the full sample or the randomised sample according to the survey of principals only, girls outperform boys in exam scores in Chinese and English by over 0.6 and 0.48 standard deviations (SD), respectively, after controlling for cognitive ability. However, when we further restrict the sample to classes randomised according to all teachers' responses, the gender gap in Chinese decreases to 0.5 SD on the basis of randomised classes, suggesting that the streaming effect may account for 20% of the gender gap in Chinese. Our results suggest that not accounting for these important factors might result in upward biases in the estimated gender achievement gap, in favour of girls.

Second, we present the standard decomposition into observable and unobservable characteristics at the mean and along the whole unobserved talent distribution. To explore the contribution of different characteristics to the gender gap, our Blinder–Oaxaca (BO) decomposition results indicate that differences in the mean characteristics can explain at most one quarter of the gender gap in math and almost none of the gap in Chinese and English. This finding is supported by the DiNardo–Fortin–Lemieux (DFL) decomposition, which visualises the contribution of the observed characteristics through the construction of the counterfactual outcome distributions. The results indicate that the unexplained 'girl premium' accounts for virtually all the raw girl overachievement in Chinese and English and approximately three quarters of the girl advantage in math. Moreover, the unexplained ‘girl premium’ is monotonically decreasing in the conditional quantiles of the exam score distribution for all subjects, being particularly large at the lower tails, but vanishes at the top decile in math. After accounting for cognitive skill by using uniform cognitive scores, the results suggest that noncognitive skills may account for the persistent gender gaps in the three subjects over the distribution. This finding is consistent with those in the literature suggesting that boys may behave differently than girls in various dimensions of noncognitive skills.

Third, we conduct heterogenous analysis to explore possible channels through which noncognitive skills aggravate the disadvantage of boys. A considerable literature has documented that noncognitive skills have played substantial roles in explaining the gender gap among younger students between 11 and 12 years old (Golsteyn and Schils 2014; Fortin et al. 2014; Attanasio et al. 2020). The Big Five framework has often been applied when discussing the social and emotional skills as broadly defined, including openness to experience, conscientiousness, emotional stability, extraversion, and agreeableness (OECD, 2020). However, in the empirical literature of economics, estimating the causal effect of noncognitive skills is difficult because of the difficulty in obtaining data with reliable measurements of noncognitive skills, and the necessary controls of relevant characteristics (e.g. demographic factors, school area, whether living with parents, and having a female teacher).

In the Chinese context, it is important to consider the impact of massive urbanisation under the binary *hukou* divide over the past decades. Girls from rural areas may or may not migrate with their parents to urban cities, given the salient son preference in some rural areas. However, girls might compensate for the disadvantage by spending more hours on homework and private tutorials, partially reflecting variation in noncognitive skills across gender. Those ‘cultural’ or unobserved factors may affect their academic achievement through different educational resources and other factors affecting noncognitive skills. For instance, a female math teacher may make a stronger impact on the noncognitive skills of girls by praising and reducing subject gender stereotypes (Gong et al. 2018). Our findings imply that the gender gap is highly heterogeneous across socioeconomic groups and that noncognitive skills may play a predominant role in explaining the remaining gender difference in academic achievements, after accounting for school selection, streaming, and cognitive ability.

The remainder of the paper is organised as follows; Section 2 briefly reviews the literature; Section 3 presents the institutional background; Section 4 discusses the decomposition methods; Section 5 presents the data; Section 6 presents and interprets the main empirical results; and Section 7 concludes.

2. Literature Review

The gender gap in educational attainment and test scores have been well documented in many countries. Overall, the gender gap in years of schooling or attainment of secondary or tertiary education has been closing or even reversing (Barro and Lee 2013, OECD 2015b). As a rigorous standardised test for 15-year-olds in reading, math, and science, results from the PISA highlight a persistent pattern of an education gender gap across subjects in the OECD countries since 2000: girls on average outperformed boys in reading by the equivalent of 1 year of school, while boys outperformed girls in math by approximately 3 months of school (OECD 2015b).

Golsteyn and Schils (2014) suggest that boys outperform girls in math and underperform in language in the Netherlands. Taking advantage of the availability of an IQ test in their data, they argue that boys have higher assertiveness and employ those skills more efficiently than girls do. Bertocchi and Bozzano (2019) review the growing literature on the education gender gap and highlight the role of traditional culture and gender stereotypes. Fortin et al (2014) find that girls outperform boys in high school from the 1980s to the 2000s and argue that the expectation for advanced education is the most important factor accounting for the large gap in achievement. Munir and Winter-Ebmer (2018) decompose worldwide PISA 2012 math and reading scores by using the Juhn–Murphy–Pierce approach (Juhn et al. 1993), which accounts for the unobserved talent distribution of students. They show that boys' scores increase more than those of girls over the distribution of talent, for both math and reading. Using PISA 2009, San Román and Rica (2012) demonstrate that girls outperform boys in both math and reading in more gender-equal societies. Moreover, they find that girls perform better in families where the mother works outside the home, which they interpreted as evidence of intergenerational transmission of gender role attitudes. Hermann and Kopasz (2019) highlight the importance of variation in educational policies across countries in explaining the educational gender gap. Using a difference-in-difference approach, they show that early tracking and more student-oriented teaching practices both favour girls.

Only a couple of studies have focused on the gender educational gap in China, both using the CEPS dataset. Xu and Li (2018) focus on the effect of the student–teacher gender match on the achievement gender gap in math, Chinese, and English, using a subsample in which the teacher assignment is plausibly random. They find that teacher's gender does not affect boys but that having a female teacher increases girls' academic performance and their self-perceived ability and interactions with the teacher, especially in math. Gong et al. (2018) use a similar multivariate regression approach and similar sample selection but focus on noncognitive outcomes. They show that having a female teacher not only improves girls' exam scores but also raises their mental status and learnings motivation through changing gender stereotypes.

A number of related studies have also focused on other dimensions of education inequalities in China, for example, *hukou* status, migration status, and school segregation, while controlling for gender. Zhao et al. (2017) document a large rural–urban student cognitive ability gap in favour of urban students in China, in the order of 0.4 SD, using CEPS. They show that half of the gap can be accounted for by differences in characteristics, using the standard BO decomposition. Notably, no evidence of a gender gap in cognitive ability has been presented after conditioning on observed characteristics. Liu and Chiang (2019) present evidence from CEPS that student learning motivations are significantly related to both family socioeconomic status (SES) and teacher–student interaction. Similarly, Duan et al. (2018) show that students from low-SES families benefited more from parental involvement in academic activities. However, in both cases, the coefficients of gender are not reported.

Wang et al. (2017a, 2017b) explore the impact of parental migration on rural children's educational outcomes by using a primary dataset of children either staying at home, without or with migrant parents, or staying with their migrant parents in Shanghai and Suzhou and attending private migrant schools. Their findings show very large negative effects of attending private migrant schools, after accounting for selection and observable differences in school quality. Unfortunately, gender is only part of the wider demographic controls, and the effect is not reported. Using CEPS, Wang et al. (2018) show that the high proportion of migrant students has a small positive effect on local students' Chinese exam scores in urban public middle schools in China, with a marginally larger effect for local boys than for local girls. However, whether the results are generalisable to the peer effects of all migrant students on local students remains unclear because migrant students in urban public schools are not representative of all migrant children in cognitive ability and educational attainment. Chen and Feng (2013) present direct evidence that migrant students in Shanghai who enrol in private migrant schools perform significantly worse than their public school counterparts, who in turn perform almost as well as the local non-migrant students, in math and Chinese examinations.

3. Institutional Background

China implemented 9-year compulsory education in 1986, comprising 6 years of primary education and 3 years of junior high school. In 2016, the country recorded a nearly 100% enrolment rate and a 93.4% completion rate of 9-year compulsory education (Ministry of Education, 2017).

Another important institutional feature of China is the *hukou* (household registration) system, which classifies individuals as having a rural or urban status at birth, usually according to the mother's *hukou* status. Chan (2009) and Meng (2012) extensively review the history of the *hukou* system and discuss its key role in China's labour market reforms. Education resources at the primary and secondary level are highly unequal in favour of urban residents in China.

Table 1 highlights the substantial gap in financial resources and teacher qualifications between urban and rural schools in China. Public spending (per student) on rural schools is approximately 80% of that spent on urban schools. The differences in books and computers are also notable, with rural schools having half as many computers as urban schools.

Tables 1: Differences between urban and rural schools

	Urban schools	Rural schools
School resources:		
Computers in school	176	86
Books in library	96,321	64,000
Public spending per student (yuan)	1,162	929
Teacher qualification distribution		
Below diploma	0.6%	3.1%
Diploma	13.3%	17.0%
Degree	83.0%	77.3%
Higher degree	3.1%	2.6%
Proportions of not living with parents		
<10%	13.0%	22.4%
10%–30%	75.9%	44.8%
30%–50%	9.2%	20.7%
>50%	1.9%	12.1%
Observations	54	58

Notes: CEPS baseline survey. Public spending excludes capital investment and teacher salaries.

Teacher quality is arguably the most important input in the education production function (Hanushek 2020). Table 1 shows that 86% and 80% of urban and rural school teachers hold at least a 4-year university degree, respectively. Table 1 also shows that approximately 10% of urban schools report a significant proportion (defined as 30% or more) of students not living with at least one parent, compared with one third of their rural counterparts. The prevalence of the so-called 'left-behind children', namely, children not living with at least one parent, highlights the disadvantage of rural children because of massive rural-to-urban migration and the *hukou* system, which excludes rural children from enrolling in urban schools.

4. Methodology:

To estimate the gender gap in student achievement while controlling for observed attributes, we first generate results based on ordinary least squares (OLS) as the benchmark.

$$y_{ij} = \alpha_i + \beta_i X_{ij} + \delta male_{ij} + school_j + \varepsilon_{ij} \tag{1}$$

where $male_{ij}$ is a dummy for being male i at school j and X_{ij} are controls for individual, family, and institutional characteristics. In this paper, the dependent variables include the educational outcomes as measured by midterm exams in math, Chinese, and English. The three subjects are compulsory in all Chinese secondary schools. Given that these midterm exams are set by schools independently, we standardise the exam scores within schools and grades to account for heterogeneity across schools. The rural *hukou* dummy indicates the student has a rural *hukou*. Notably, we allow for full interaction between a female dummy with both rural *hukou* and the gender of the subject teacher. Individual covariates include age, ability scores, grade, and number of siblings. Parent's background includes dummies of highest qualification of both parents. Following the literature, teachers' characteristics are also included as controls, for example, age, experience, qualification, and teaching certificate. We also include study hours both in and out of school, as well as dummies for being young or old for the relevant grade. We also control students' innate abilities as measured by the standardised cognitive ability test scores. A school identifier, $school_j$, is included to control for school fixed effect and distribution of the academic rank of teachers.

We control for both in-school and out-of-school study hours as measures of student effort: the former is defined as the time spent on tasks assigned by their class teachers (e.g., homework) and the latter is the time spent on tasks assigned by private tutors or off-school classes. In China, there is a significant variation in family spending on children's education, mostly through private tuition. Family background variables include parental highest qualification and a single child dummy.

BO decomposition

The literature and our results suggest a persistent gap in educational achievement between male and female students. One of the main aims of this paper is to evaluate how much of the gap can be attributed to the systematic differences in characteristics between male and female students.

We first undertake the **BO decomposition** as a benchmark. Under the structure of twofold BO decomposition, the total difference is decomposed into explained and unexplained parts.

$$D = E(Y_M) - E(Y_F) = \beta_M\{E(X_M) - E(X_F)\} + (\beta_M - \beta_F) * E(X_F) \tag{2}$$

where β_M and β_F are the coefficients for male and female, respectively. Equation (2) represents the crude difference between male and female students conditional on characteristics. The first term $\beta_M\{E(X_M) - E(X_F)\}$ represents the gender difference explained by observable difference. The second term in equation (2) represents the unexplained part by the group difference and unobserved factors. BO decomposition is developed to decompose the total difference into mean outcomes and relies on the restrictive linearity assumption between outcomes and characteristics.

DFL decomposition

The BO decomposition decomposes the total difference in mean outcomes and relies on the restrictive linearity assumption between outcomes and characteristics. To generalise the decomposition, we use **DFL** decomposition. DiNardo, Fortin and Lemieux (1996) relax the parametric restriction in BO and develop a semiparametric decomposition method to derive a counterfactual distribution of outcomes, known as DFL decomposition. In our case, DFL decomposition answers this question: what test scores would the male students achieve if they had attributes of female students? To answer this question, we reweight the boys to have the same attributes as female students.

The marginal distribution of outcome for male students can be expressed as

$$f_{g_1}(y) = \int_{x\in\Omega_x} f\big(y, x \big| g_x = 1, g_y = 1\big)\, dx = \int_{x\in\Omega_x} f\big(y\big|x, g_y = 1\big) f(x|g_x = 1) dx$$
$$f_{g_0}(y) = \int_{x\in\Omega_x} f\big(y, x \big| g_x = 0, g_y = 0\big)\, dx = \int_{x\in\Omega_x} f\big(y\big|x, g_y = 0\big) f(x|g_x = 0) dx \tag{3}$$

where Ω_x denotes the domain of characteristics and y denotes the exam scores and cognitive score. g_y and g_x represent the group structure effect and group characteristics, respectively. x is a set of observed covariates. By applying the law of iterated expectation, equation (3) describes the outcomes attributed to individual characteristics and the group-specific effect, where $g \in (0,1)$ represents girls and boys.

Notably, $f_{g_1}(y)$ is observed in the data, and one of the counterfactual distributions is $f^c_{g_1}(y)$ when boys have the same characteristics as girls, such that

$$f^{c}_{g_1}(y) = f\big(y;\ g_x = 0, g_y = 1\big) = \int f\big(y\big|x, g_y = 1\big)\, f(x|g_x = 0)dx =$$
$$\int f\big(y\big|x, g_y = 1\big)\, \psi_x(x) f(x|g_x = 1)dx \qquad (4)$$

where $\psi_x(x) = \frac{f(x|g_x = 0)}{f(x|g_x = 1)}$, under the strong assumption that the group-specific effect is the same, conditional on observed characteristics: $f\big(y\big|x, g_y = 1\big) = f\big(y\big|x, g_y = 0\big)$.

The weights are the key element in generating the counterfactual distribution and are designed to estimate by using Bayes' rule.

$$P(x|g_x = 0) = \frac{P(g_x = 0|x)dF(x)}{\int_x P(g_x = 0|x)dF(x)}$$

$$P(x|g_x = 1) = \frac{P(g_x = 1|x)dF(x)}{\int_x P(g_x = 1|x)dF(x)}$$

$$\psi_x(x) = \frac{P(g_x = 0|x)}{P(g_x = 1|x)} * \frac{P(g_x = 1)}{P(g_x = 0)}$$

The procedures for estimating the weights follow DiNardo et al. (1996). A probit model is used to estimate the propensity scores of being males.

Quantile decomposition

To allow for heterogeneity, we perform quantile decomposition to explore how the gap varies with unobserved factors. Following the recent development of quantile decomposition by Chernozhukov et al (2013) based on the Machado–Matta procedure (Mata and Machado, 2005), we construct the counterfactual distribution based on quantile regressions. Unlike the standard OB decomposition that only estimates the mean difference by using OLS, the conditional quantile decomposition could characterise both the mean difference and the dispersion of the outcome variable. We specify the θth quantile of the conditional distribution of y_i given X_i,

$$Q_\theta(y_i|X_i) = F^{-1}_{y|X}(\theta|X_i) = X_i\beta(\theta) \qquad (5)$$

where $\beta(\theta)$ is a vector of quantile regression coefficients and is estimated separately for each θ.

Following Hospido and Moral-Benito (2014), we implement quantile decomposition as follows:

First, we separately estimate the $\beta(\theta)$ for each gender. The results contain two vectors, $\beta^{g}(\theta)$ for two groups, where $g \in (0,1)$ represents female and male students.

Next, conditional on x_i, we estimate the outcomes and $\beta(\theta)$ for the two groups separately and obtain the estimated value of outcomes conditional on covariates. The conditional quantile functions are defined as follows:

$$\hat{q}_\theta^1 = X_i\beta^1(\theta) \text{ and } \hat{q}_\theta^0 = X_i\beta^0(\theta) \quad (6)$$

We then construct the conditional distribution function (CDF) as follows:

$$F_{Y_1}(q|X_i) = \int_0^1 1(X_i\beta^1(\theta) \leq q)d\theta \text{ and } F_{Y_0}(q|X_i) = \int_0^1 1(X_i\beta^0(\theta) \leq q)d\theta \quad (7)$$

where $F_{Y_1}(q|X_i)$ and $F_{Y_0}(q|X_i)$ are the conditional cumulative distribution function of boys and girls, given the random variable X evaluated at quantile q. On the basis of conditional CDF, we construct the unconditional distribution function by integrating over X_i for each group.

$$\hat{F}_{Y_1}(q|g = 1) = \int \hat{F}_{Y_1}(q|x)dF_X(x|g = 1) \text{ and } \hat{F}_{Y_0}(q|g = 0) = \int \hat{F}_{Y_0}(q|x)dF_X(x|g = 0) \quad (8)$$

After estimating the unconditional CDF, the unconditional quantile function can be expressed as follows:

$$\hat{q}_\theta^1 = \inf\{y: \hat{F}_{Y_1}(q|g = 1) \geq \theta\} \text{ } and \text{ } \hat{q}_\theta^0 = \inf\{y: \hat{F}_{Y_0}(q|g = 0) \geq \theta\} \quad (9)$$

Following the DFL reweighting method in equation 7, the counterfactual distribution of male students with female characteristics is achieved by reweighting the outcome distribution by integrating the characteristics of female students. The counterfactual quantile function can be expressed as

$$\hat{q}_\theta^C = \inf\{y: \int \hat{F}_{Y_0}(q|x)dF_X(x|g = 1) \geq \theta\}=\inf\{y: \hat{F}_{Y_0}(q|g = 1) \geq \theta\} \quad (10)$$

where $\hat{q}_\theta^C$ represents the counterfactual quantile distribution of male students with the characteristics of female students. The last step is to decompose the difference between two groups into

$$\Delta^{QD} = \hat{q}_\theta^1 - \hat{q}_\theta^0 = (\hat{q}_\theta^1 - \hat{q}_\theta^c) + (\hat{q}_\theta^c - \hat{q}_\theta^0) \quad (11)$$

The first and the second part on the right-hand side of equation (11) estimates the effects of coefficients (coefficients effect) and the effects of characteristics (characteristics effect), respectively. The coefficients effect estimates unobserved components after reweighting the boys with the same characteristics of girls. Likewise, the characteristics effect estimates the observed differences in characteristics.

5. Data and Sample

5.1 Description of CEPS

This study is based on the baseline survey of CEPS, a large-scale, nationally representative longitudinal survey starting with two cohorts: 7th and 9th graders. The baseline survey was conducted by the National Survey Research Centre(NSRC) at Renmin University of China in the academic year 2013–2014, with five different questionnaires to the sample students, parents, class headteachers, core subject teachers other than headteachers, and school principals. A class headteacher is a designated teacher with overall responsibility for a particular class and is responsible for establishing class rules, leading class actions, and providing nonacademic support. Moreover, the survey includes standardised cognitive ability tests for students in each grade and an internet-based personality test for all sample students and collects transcripts of important (midterm) examinations.

CEPS follows a stratified, multistage sampling design with probability proportional to size (PPS), randomly selecting a school-based, nationally representative sample of approximately 20,000 students in 438 classrooms of 112 schools in 28 county-level units in mainland China. The student questionnaire covers students' demographic characteristics, mobility and migration status, childhood experience, health status, household structure, parent–child interactions, in-school performance, extracurricular activities, relationships with teachers and peers, social behaviour development, and expectations for the future.

The parent questionnaire covers parents' demographic characteristics and lifestyles, parent–child interactions, educational environment and investment for child, community environment, parent–teacher interactions, and parents' perceptions of school education and expectations for the future of the child.

The questionnaires for headteachers and core subject teachers cover teachers' demographic characteristics, teaching experience, comments on students' behaviours, parent–teacher interactions, comparison between local and nonlocal students, perceptions of education, and degree of stress and job satisfaction.

The questionnaire for school principals covers their demographic characteristics, perceptions of education, their school's educational facilities, daily management, enrolment of students, statistics of the student body and staff body, and other school characteristics. Our classification of urban and rural schools is based on the location of the school (see Table A1 in the Appendix). The 53 urban schools include all downtown and peripheral areas of cities and county-towns regardless of the community type and represent 47.8% of schools in the sample. The rural schools include schools in rural areas, townships outside county-towns, and in urban–rural fringe areas. The proportion of rural *hukou* pupils in urban and rural schools are 36.1% and 72.0%, respectively.

5.2 Sample

The sample used in this paper comprises the cross-section of students in the CEPS baseline survey in the academic year 2013–2014. Various research has emerged since its introduction, taking advantage of the randomised assignment into classes on the basis of a principal's response. Notably, both teachers and principal are asked whether the grade and the school randomly distribute students into classes. A concern is inconsistency between teachers' and principal's responses. There are many classes in which the teachers argue that students are grouped by their course scores, but the corresponding principal states that the students are randomly assigned. To address the streaming of ability, we follow the literature and adopt the randomised design.

The first analytical sample includes all observations without a missing value, comprising 112 schools and 430 classes after excluding eight classes without the responses of teachers. We also exclude students who join the school after grade 8. Furthermore, the classes in grade 9 do not have information on whether they were regrouped for entering grade 8. Thus, we use the follow-up survey for grade 7 in 2013–2014 and obtain the information on which schools have reassigned the students when entering grade 8. Next, we exclude those schools in our main analytical sample, including eight classes in grade 9. Table A2 shows the roadmap of the sample selection. There are 112 schools and 417 classes used in our analytical sample without accounting for ability streaming.

The second analytical sample is built for regression analysis by taking advantage of the design of randomisation. On the basis of our first analytical sample, we follow the literature to select a randomised class in each school on the basis of its principal's response, restricting our sample to 70 schools with 251 classes. Last, we restrict our analytical sample on the basis of both teachers' and principal's responses on randomisation, being aware of the inconsistency. The final analytical sample comprises 30 schools and 109 classes.

Table 2 presents summary statistics of key variables by gender and *hukou* status. Conditional on *hukou* status, we observe persistent gender gaps in favour of girls in raw exam scores across all subjects. Moreover, the gender difference in the standardised cognitive ability test scores is tiny and statistically insignificant within each *hukou* status. What is more surprising is the significant gap in favour of urban students in the standardised cognitive ability scores, which are meant to be independent of the subject knowledge required for the grade. Although the differences in exam scores are also in favour of urban students, these are not based on standardised exams.

As expected, there are substantial gaps between urban and rural *hukou* students in parental educational expenses, the probability of being a single child, study hours, and teacher ranks, reflecting the deep urban–rural divide in China. In particular, rural students are much less likely to be a single child than rural students, because of the laxer implementation of the family policy in rural areas, especially when the firstborn is a girl. However, there is little gender difference in teacher ranks when holding *hukou* status constant, reflecting that mixed-sex schooling is universal and that school admission is almost always gender-blind.

Table 2: Summary statistics by *hukou* status and gender

	Urban *hukou*		Rural *hukou*	
VARIABLES	**Female**	**Male**	**Female**	**Male**
Academic outcomes:				
Math (Raw)	86.08	82.70	77.05	74.32
Chinese (Raw)	89.88	83.11	85.11	77.59
English (Raw)	92.54	81.69	82.35	69.42
Ability (standardised)	0.22	0.20	-0.08	-0.07
Key controls:				
Parental educational expenses	2,348	1,984	520	455
Single child	0.61	0.66	0.21	0.34
Time in homework by teacher	5.95	5.53	5.86	5.09
Time in homework by parents	2.21	1.99	1.72	1.40
Time in tutorial	2.02	1.91	0.91	0.75
Grade 3 Teacher	0.02	0.03	0.03	0.03
Grade 2 Teacher	0.73	0.76	0.97	0.97
Grade 1 Teacher	1.39	1.37	1.28	1.22
Senior Teacher	0.74	0.68	0.50	0.48
Observations	3,793	3,712	3,828	3,892

Notes: N = 15,225. Teacher rank shows the average numbers of teachers in the three core subjects at each rank, adding up to three for each student. Time is measured as average hours per week including weekends. Grade 3 is the entry rank for qualified teachers; Senior Teacher is the highest rank.

6. Empirical Results

6.1. OLS (Ordinary Least Squares) Benchmark

Compared with the *hukou* status, which is a fundamental determinant of socioeconomic outcomes in China, the urban–rural location of the school is to a large extent an outcome of parental choice. To avoid the endogeneity of school type, we focus on *hukou* status, which is predetermined, in the main analysis. However, we explore the heterogeneous effects of our main findings by assessing urban or rural schools separately in the sensitivity analysis. Moreover, in the regression analysis, we also control for school fixed-effects, to allow for time-invariant school effects.

Based on three different samples, the OLS results of the standardised exam scores in the three subjects and the standardised cognitive ability test scores are presented in Table 3. The first sample includes all observations having non-missing values in the survey; the next two subsamples are based on a randomised sample given the principal's responses or both the principal's and teachers' responses.

Table 3 shows that holding all other factors constant, girls outperform boys by more than 0.6 and 0.48 an SD in Chinese and English, respectively, based on the full sample. The advantage in math for girls is 0.03 SD, which is statistically insignificant. Rural *hukou* status *per se* is associated with a 0.05 SD higher math score and 0.06 SD lower English score. For rural *hukou* girls, the rural disadvantage in English is more than offset by a 0.12 SD premium. However, the results might be biased because of streaming between classes. Thereby, we test the gender gap using restricted samples to eliminate the streaming effect. On the basis of the strictest sample, we find that girls again outperform boys in both Chinese and English by approximately 0.5 SDs, but the estimate of Chinese is 20% smaller than using the full sample. However, what is reassuring is that neither the gender of the student nor rural *hukou*, on their own or interacted, have a significant effect on the cognitive test score, which is supposed to be a measure of ability independent of one's subject knowledge.

Table 3: OLS results based on three samples

	(1) Math	(2) Chinese	(3) English	(4) Cognitive
All sample (N = 15,225):				
Female	0.029	0.599***	0.484***	-0.009
	(0.03)	(0.04)	(0.05)	(0.02)
Rural *hukou*	0.047*	0.025	-0.062**	0.004
	(0.02)	(0.03)	(0.03)	(0.02)
Female # Rural *hukou*	0.016	0.020	0.122***	-0.018
	(0.03)	(0.03)	(0.03)	(0.03)
Randomised sample given principal response (N = 8,706):				
Female	-0.000	0.598***	0.476***	-0.019
	(0.05)	(0.04)	(0.07)	(0.02)
Rural *hukou*	0.007	-0.018	-0.091**	-0.018
	(0.03)	(0.03)	(0.04)	(0.03)
Female # Rural *hukou*	0.056	0.053	0.134***	0.028
	(0.04)	(0.04)	(0.04)	(0.04)
Randomised sample given principal and teachers response (N = 3,638):				
Female	0.017	0.508***	0.484***	0.014
	(0.07)	(0.08)	(0.10)	(0.03)
Rural *hukou*	-0.021	-0.077	-0.163**	0.014
	(0.05)	(0.06)	(0.07)	(0.04)
Female # Rural *hukou*	0.106	0.114	0.161**	0.045
	(0.07)	(0.07)	(0.07)	(0.05)

Notes: Regression results are consistent with the sample in the last panel, using strictest sample selection on randomisation. Robust standard errors in parentheses. ***, **, and * indicate significance at 10%, 5%, and 1%, respectively.

Although the specification in Table 3 includes a comprehensive set of control variables, we might be reasonably concerned with the endogeneity of some of the controls. For instance, more study hours and higher expenditure on private tuition might be an indicator for struggling with the subject rather than simply higher effort. Therefore, using the strictest random sample, in Table 4, we test the robustness of our mains findings with respect to successively adding controls for the student's individual characteristics, parents' characteristics, and learning effort and resources.

Table 4 suggests that the results in Table 3 now corresponding to Specification 4 are highly robust. The finding of a substantial 'girl premium' and a further premium of girls with rural *hukou* for English remains statistically significant even in the most parsimonious specification. The magnitude of the effects is also largely insensitive to the various sets of controls for Chinese and English. Further, we find that the magnitudes of coefficients are not associated with the inclusion of additional variables, implying that the randomisation between classes is effective.

Table 4: Sensitivity tests with respect to various controls

	Specification 1	Specification 2	Specification 3	Specification 4
Math				
Female	0.070	0.029	0.022	0.017
	(0.07)	(0.07)	(0.07)	(0.07)
Rural *hukou*	-0.089	-0.056	-0.023	-0.021
	(0.06)	(0.05)	(0.05)	(0.05)
Female*Rural *hukou*	0.106	0.107	0.111	0.106
	(0.07)	(0.07)	(0.07)	(0.07)
Chinese				
Female	0.510***	0.521***	0.515***	0.508***
	(0.07)	(0.07)	(0.07)	(0.08)
Rural *hukou*	-0.131*	-0.106*	-0.080	-0.077
	(0.07)	(0.06)	(0.06)	(0.06)
Female*Rural *hukou*	0.128*	0.116	0.119*	0.114
	(0.07)	(0.07)	(0.07)	(0.07)
English				
Female	0.422***	0.489***	0.490***	0.484***
	(0.11)	(0.10)	(0.10)	(0.10)
Rural *hukou*	-0.251***	-0.210***	-0.168**	-0.163**
	(0.08)	(0.07)	(0.07)	(0.07)
Female*Rural *hukou*	0.160**	0.163**	0.169**	0.161**
	(0.08)	(0.07)	(0.07)	(0.07)
Cognitive scores				
Female	0.030	0.016	0.011	0.014
	(0.03)	(0.03)	(0.03)	(0.03)
Rural *hukou*	-0.023	-0.024	0.017	0.014
	(0.05)	(0.05)	(0.04)	(0.04)
Female*Rural *hukou*	0.034	0.039	0.044	0.045

	(0.05)	(0.05)	(0.05)	(0.05)
Observations	3,638	3,638	3,638	3,638
Set of control variables:				
Individual characteristics		+	+	+
Parents characteristics			+	+
Learning effort and resources				+

Notes: Robust standard errors in parentheses. ***, **, and * indicate significance at 10%, 5%, and 1%, respectively.

Table A3 in the Appendix further checks the robustness of our preferred specification in Table 3, with respect to school type and family structure. The results suggest that the girls' premia are larger in rural schools than in urban schools. However, living with their parents has a greater influence on the girls' academic achievements than that of boys. Among children not living with their parents, girls do not perform better than boys in Chinese.

Both school types and family structure are subject to concerns of endogeneity; thus, advanced econometrics methods and availability of instrumental variables are required to overcome this. Given the insensitivity of our main findings along both dimensions, we again use the preferred specification for the benchmark BO decomposition in the following. Notably, by allowing for school fixed-effects, we effectively control for school type by only relying on within-school variation in the exam and cognitive test scores. Moreover, the quantile decomposition in the next subsection manages heterogeneity in unobservables by using a semiparametric approach.

Table A4 in the Appendix tests the robustness of our preferred specification with respect to teacher's gender, student's cognitive ability, and parents' education. Girls benefit significantly more in general performance from a female teacher in math than in Chinese and English. Girls having cognitive scores above the median also have a larger advantage than girls with lower cognitive scores do. Further, there are smaller differences across three courses between girls and boys from well-educated families. The evidence implies that although there is less gender gap among the most educated families, higher cognitive skills enlarges the gender gap.

6.2. Blinder-Oaxaca Decomposition

The BO decomposition is commonly used to decompose the mean differences in outcomes between two groups into those arising from differences in characteristics (the explained) and different coefficients (the unexplained). In our context, it indicates how much the achievement gap in favour of girls could be closed if boys could behave and be treated as girls are.

Table 5 presents the BO decomposition for the differences in the means of the standardised exam scores in the three key subjects and the standardised cognitive ability test scores under three sets of controls. In column 1, which only allows for individual characteristics, almost all the differences in exam scores are explained by the unexplained components for all key subjects. When we further control for parental characteristics, the explained components still do not contribute to explaining the gender gap in exam scores in any subject. It is only when we additionally allow for study hours and school effects in column 3 that the gender gap explained by differences in observed characteristics become statistically significant for English at the 5% level. In relative terms, the differences in characteristics account for approximately 18%, 4%, and 4% of the gender gap for math, Chinese, and English respectively, but only the explained component is significant for English.

Table 5: Blinder–Oaxaca decomposition with Different Sets of Controls

	Specification 1	Specification 2	Specification 3
Math:			
Difference	0.175***	0.175***	0.175***
	(0.03)	(0.03)	(0.03)
Explained	0.020	0.022	0.032*
	(0.02)	(0.02)	(0.02)
Unexplained	0.155***	0.152***	0.143***
	(0.03)	(0.03)	(0.03)
Chinese:			
Difference	0.547***	0.547***	0.547***
	(0.03)	(0.03)	(0.03)
Explained	0.011	0.015	0.021
	(0.01)	(0.01)	(0.01)
Unexplained	0.536***	0.533***	0.527***
	(0.03)	(0.03)	(0.03)
English:			
Difference	0.517***	0.517***	0.517***
	(0.03)	(0.03)	(0.03)
Explained	0.008	0.013	0.023
	(0.01)	(0.01)	(0.01)
Unexplained	0.508***	0.503***	0.494***
	(0.03)	(0.03)	(0.03)
Cognitive scores:			
Difference	0.079**	0.079**	0.079**
	(0.03)	(0.03)	(0.03)
Explained	0.046**	0.050**	0.046**
	(0.02)	(0.02)	(0.02)
Unexplained	0.033	0.030	0.033
	(0.03)	(0.03)	(0.03)
Specifications:			
Individuals	+	+	+
Parents		+	+
Learning and school effect			+

Observations	3,638	3,638	3,638

Notes: Robust standard errors in parentheses. ***, **, and * indicate significance at 10%, 5%, and 1%, respectively.

Our results show that even when we allow for observed learning and school effects, all the differences in characteristics at best account for a marginal or small proportion of the overall gender gap in exam scores across all subjects within schools. This finding implies that much of the gender gap in academic achievement is from the unexplained components, namely, differences in the coefficients.

The last subpanel of Table 5 also reveals that the raw gender gap in the standardised cognitive test score is statistically significant (0.079 SD). The explained components together explain a statistically significant (0.046 SD) of the overall gender gap when we only allow individual characteristics. The contribution of the observable characteristics does not change much when we further allow for parental variables or learning and school effect. However, the unexplained components are never statistically significant.

Table A5 in the Appendix shows the OB decomposition corresponding to the last column of Table 5, by grouping characteristics into school-related, individual-related, and parents-related variables. Factors attributed to schools and teachers reduce the gender gap by 0.052–0.057 SD, while individual characteristics of students including learning hours and cognitive abilities increase the gender gap by 0.029–0.041 SD. Notably, the effects of the highest qualifications of both parents are neutral in effect on the gender gap of all subjects.

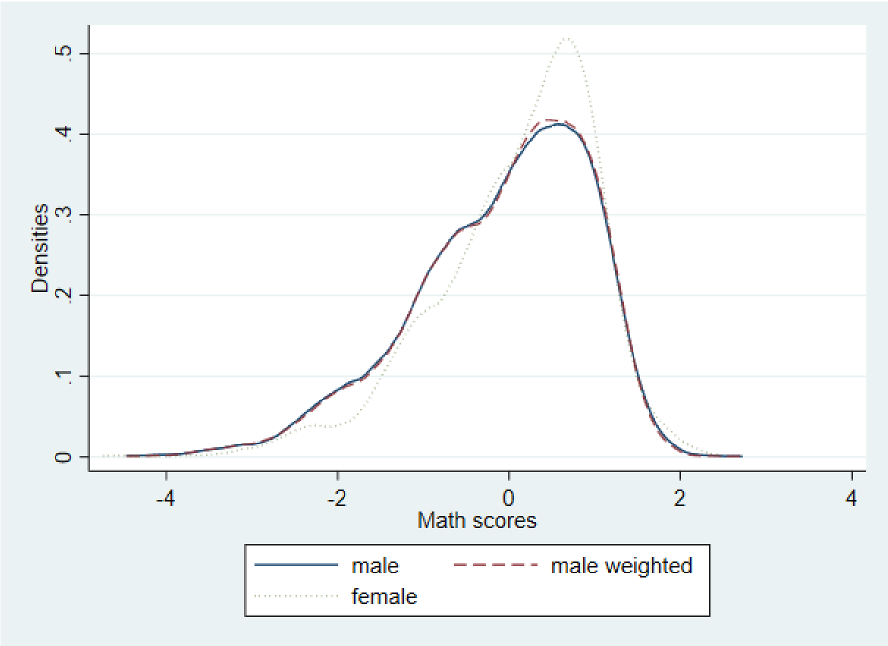
6.3. DiNardo-Fortin-Lemieux Decomposition

The **DFL** decomposition (DiNardo, Fortin and Lemieux (1996)) is effectively a semiparametric alternative to the aforementioned BO decomposition presented. Using graphs, DFL displays the distribution of the counterfactual outcome, alongside the unconditional distributions, visualising the relative contributions of observable characteristics to the raw gender gap.

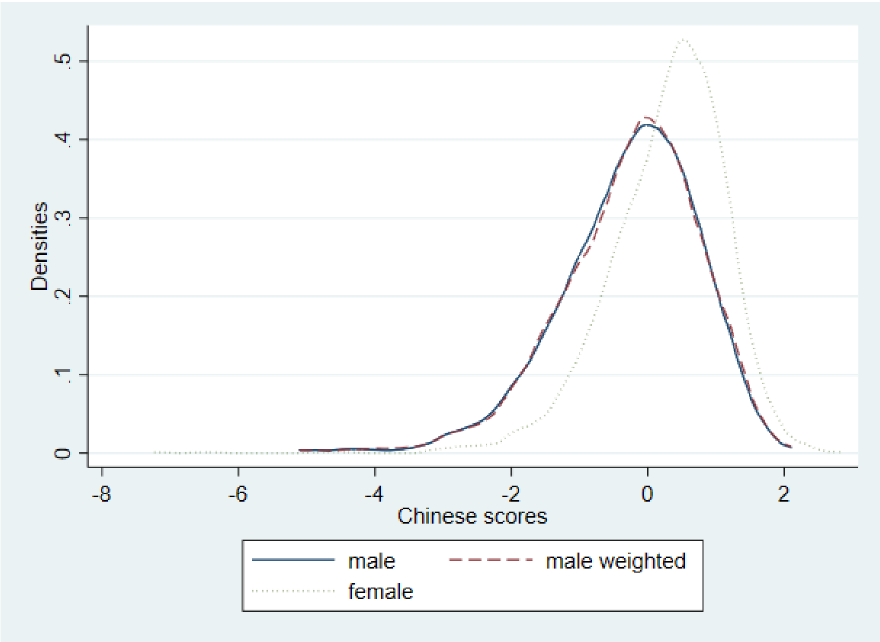
Figure 1 presents the DFL decomposition for standardised exam scores in Chinese, math, and English, respectively. The counterfactual male outcomes are constructed through reweighting to derive the test scores they would achieve if they had the same attributes as female students. That the actual and counterfactual male exam scores are almost identical to each other indicates that differences in observable characteristics account for little of the raw gender in those two subjects. These findings are consistent with the benchmark OB decomposition results.

Figure 1: DFL decompositions

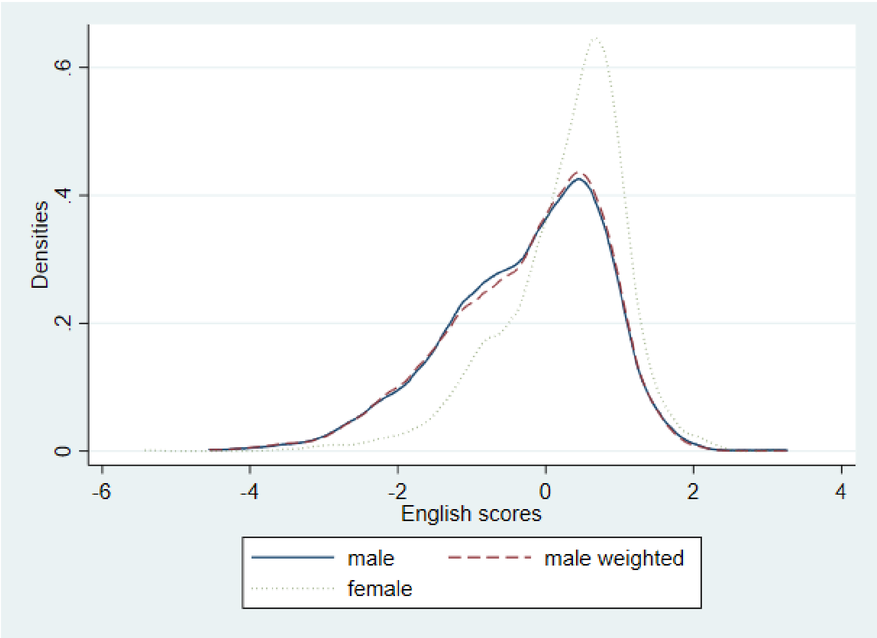
Panel A: Standardised math exam scores



Panel B: Standardised Chinese exam scores



Panel C: Standardised English exam scores



6.4. Quantile Decomposition

Both OLS and the BO decomposition focus on mean effects. However, we might expect the effect of key variables to be heterogeneous along the distribution of the residuals. Table 6 highlights the gender gap in exam scores in various subjects by *hukou* status, for selected percentiles along two such important dimensions: cognitive ability test scores and study hours. Table A6 in the Appendix presents the raw exam scores and the standardised cognitive ability scores at selected quantiles, by gender and *hukou* status and as a whole.

Broadly, there are two patterns in the heterogenous effect: a diminishing trend and a U-shaped trend. The U-shaped trend seems to be more pronounced with learning hours' percentiles rather than ability percentiles.

For most of the distributions by ability percentile, the gender gap is diminishing in the percentiles. For example, for urban students, the gender gaps in English at the 10th percentile are approximately 13 points; this is reduced to 11 points at the median and to 9 points at the 90th percentile. By contrast, the gender gaps by learning time percentiles and student achievement are nonmonotonic and sometimes may present a U-shaped trend, such as math scores for rural students.

Table 6: Gender achievement gap by percentiles of cognitive ability and learning hours

	Math scores		**Chinese scores**		**English scores**	
	Urban	**Rural**	**Urban**	**Rural**	**Urban**	**Rural**
By Ability Percentiles:						
10th	5.254	2.455	9.233	10.94	13.122	13.762
30th	3.136	5.706	5.687	7.877	11.302	15.767
50th	2.588	3.565	6.715	9.01	10.719	14.57
70th	0.288	1.646	5.875	6.166	9.28	13.145
90th	3.096	2.422	6.446	6.722	9.319	10.129
By Learning Hours percentiles:						
10th	10.3	7.579	10.895	10.169	17.711	17.513
30th	3.336	-0.709	7.396	5.17	12.623	11.271
50th	2.223	-3.239	5.643	4.025	9.564	5.719
70th	0.27	0.773	5.088	8.053	9.373	12.694
90th	-0.197	5.399	5.496	8.056	7.861	11.774
Observations	7,505	7,720	7,505	7,720	7,505	7,720

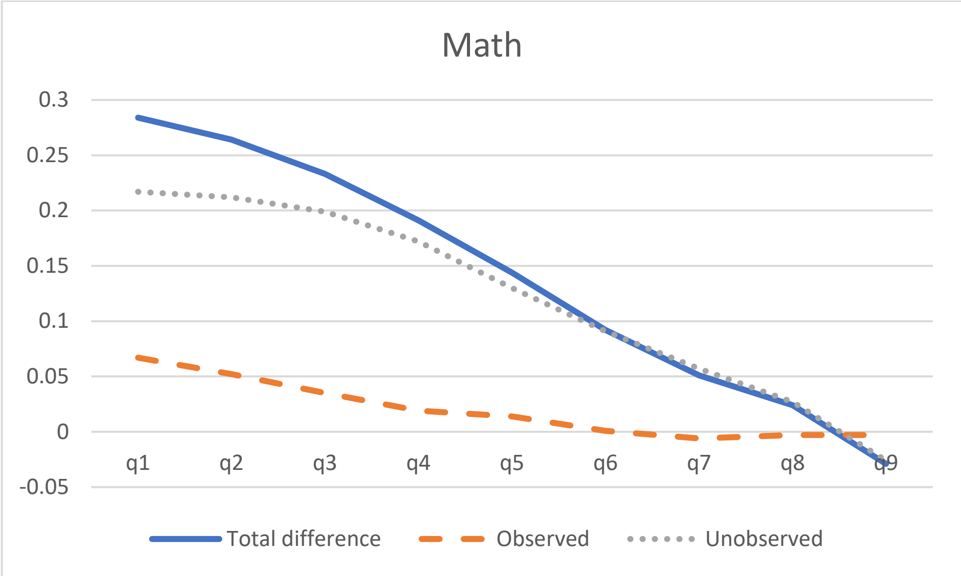
These patterns are indicative of the strong heterogeneous effects of cognitive ability and study hours. In general, we might also expect the effects of other variables such as parental education and educational expenditure to be heterogeneous. Table A7 in the Appendix presents mean cognitive ability scores and education expenditures, by gender, school, *hukou*, and highest parental educational attainment. The quantile decomposition is a semiparametric method based on the estimation of the conditional distribution by quantile regression. The conditional distribution is then integrated over the range of covariates to allow the decomposition of changes in the conditional quantile distribution of outcomes into the explained effects because of differences in characteristics and the unexplained effects because of differences in coefficients; in our context, we label the latter as ‘girl premium’.

Figure 2 visualises the quantile decomposition results for the three subject exam scores and the cognitive ability test score, using the preferred specification in Table 3, which controls for individual and parental characteristics, and learning and school effects, and school fixed-effects. Consistent with the BO decomposition, for Chinese and English, almost all of the gender achievement gap in favour of girls across quantiles is explained by the unexplained ‘girl premium’. By contrast, regardless of the quantiles, approximately one quarter of the gender gap in math is explained by differences in characteristics. What is new in the quantile decomposition is that the girl premium is diminishing in the quantiles. At the lower tail of the conditional exam scores’ distribution, there are very large girl premiums, especially in English. This unexplained girl premium diminishes further up the conditional ‘ability’ distribution, and vanishes altogether at the top decile in math.

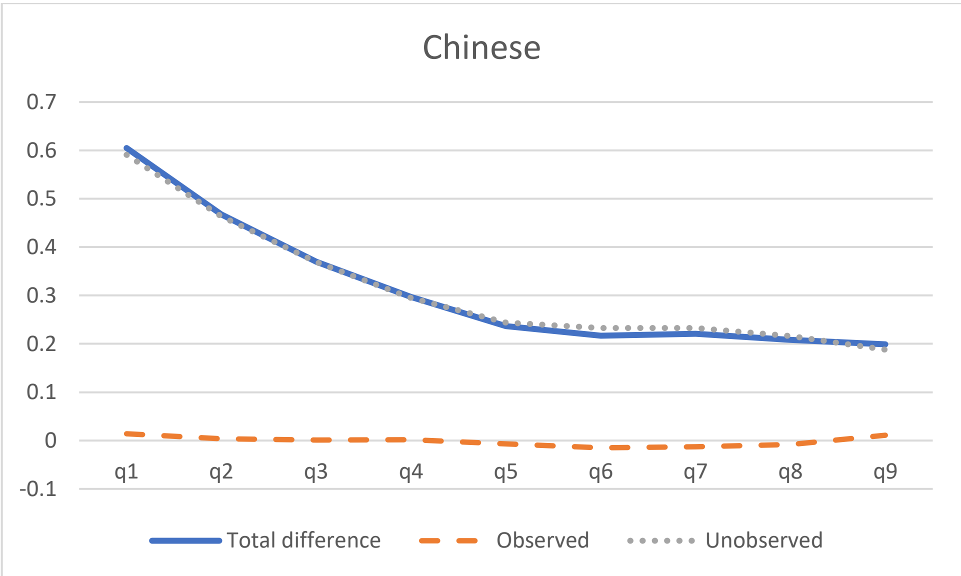
The quantile decomposition results suggest that boys at the bottom end of the conditional ‘ability’ distribution experience severe challenges in learning across all the major subjects, relative to girls. The reason for this finding could be gender differences in noncognitive ability, which harms boys from disadvantaged backgrounds most, or institutional factors such as opportunities for unskilled migrant labour, which discourages school education for rural boys.

Figure 2: Quantile Decomposition of Gender Gap in Exam and Cognitive Test Scores

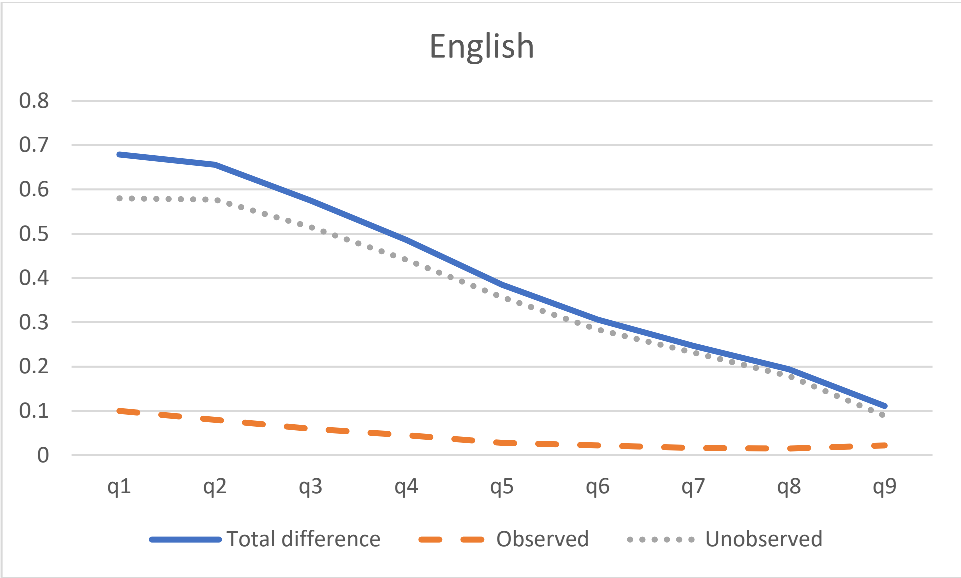
Panel A, Math scores



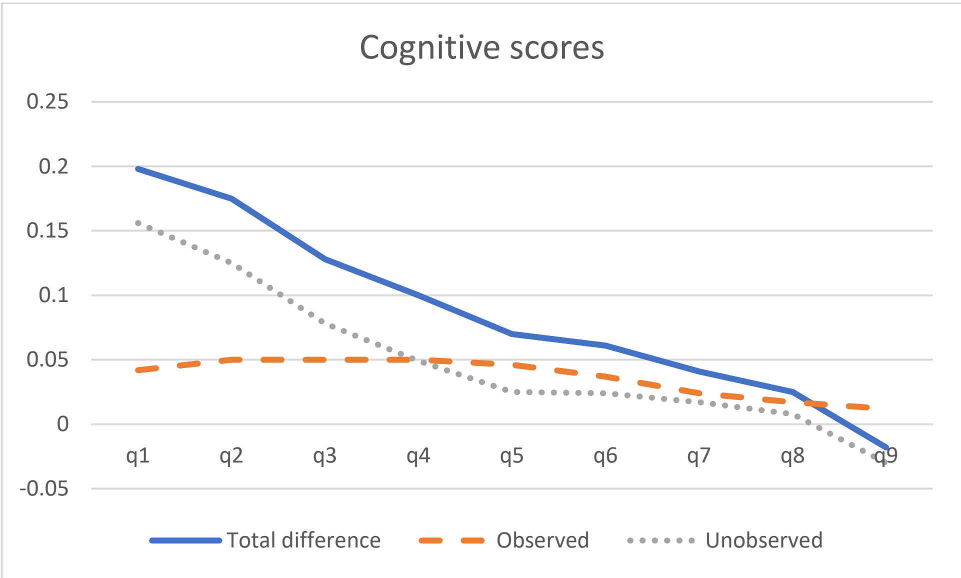
Panel B, Chinese scores



Penal C, English scores



Panel D, Cognitive scores



Note: The horizonal axis represents the quantiles of the conditional distribution of the standardised exam scores for various subjects and the cognitive test scores, where q1 indicates the bottom decile, q2 the second lowest decile, and so on.

7. Conclusions

Using a large nationally representative survey of students in grade 7 and grade 9 in China in 2013, we focus on the gender gap in exam scores in math, Chinese, and English, the three core academic subjects at the final stage of compulsory education. Compared with previous papers on gender differences, we focus on the heterogeneity in the gender gap between boys and girls and highlight the role of noncognitive skills. We take advantage of the rich information in the characteristics of the student, the parent, and the school attended and control for school fixed-effects and standardised cognitive ability test scores. More importantly, this study benefits from randomised student assignment across classes within schools.

Although boys and girls have similar cognitive scores, girls outperform boys in exam scores across all subjects except for math, even conditional on *hukou* status, school selection, streaming, and cognitive ability. The gender gap in favour of girls is the smallest in math and largest in Chinese, with English in between. These gaps are large relative to the PISA evidence from other countries. The decomposition results indicate that for Chinese and English, virtually all of the gender achievement gap in favour of girls across quantiles is explained by the unexplained 'girl premium'. Moreover, for all subjects, the unexplained 'girl premium' diminishes further up the conditional exam score distribution and vanishes at the top decile in math. We observe a larger gender gap in rural schools than in urban schools. We also observe that girls generally perform better when living with both their parents, and this finding is more pronounced for girls with a rural *hukou*. Notably, the gender gap remains large and persistent even with a rich set of variables, including a cognitive score.

Our findings imply that the gender gap is highly heterogeneous across socioeconomic groups and that noncognitive skills may play a predominate role in explaining the remaining gender difference in academic achievements, after accounting for school selection, streaming, and cognitive ability. From a psychological point of view on the gender difference, girls are more self-disciplined than boys are, and this might be particularly salient at the bottom decile in the conditional exam score distribution (Duckworth and Seligman, 2006). At the top decile, boys may share more similarities with girls in noncognitive skills.

Although the international literature based on PISA studies have emphasised the 'cultural factors' in explaining the gender attainment gap, our evidence suggests that to the extent that traditional gender role attitudes still exist in China, they are relatively small and more than offset by other factors. The main explanation of a large persistent gap favouring girls in China is the educational system, which strongly focuses on general academic education at the expense of vocational education, which presumably disadvantages boys at the lower tail of the talent distribution who ultimately pursue unskilled jobs. The research has also established that returns to senior high school are very low in China because its educational system focuses on preparing students for the highly competitive university entrance exam and offers low value for those who do not enrol in Higher Education(see Li et al. 2012, Awaworyi and Mishra 2018 and Wu et al. 2019). Therefore, it could be rational for rural students, especially boys who have relatively good opportunities to work as a low-skilled migrant worker, to leave school as soon as (or even drop out before) they complete compulsory education at the end of grade 9. According to Zhao *et al*. (2017), 'In 1990, 2000 and 2010, the transition rates from junior to Senior High School in cities were 40.41%, 66.71% and 88.11% respectively, while in the same years, the transition rates in towns and rural areas were 18.96%. 22.06% and 38.36%, respectively'. Other school-related factors might include any teaching methods and teacher attitudes that might unintentionally favour girls over boys. For instance, the prevalent teacher-led teaching mode used in relatively big classes which tends to focus on classroom discipline, memorisation, and recitation might have a disproportionately detrimental effect on boys at the lower tail of the talent distribution.

More broadly, the *hukou* system and the consequent urban–rural gap in socioeconomic conditions could also play a role in explaining the education gender gap. The large significant additional girl premium in English for students with rural *hukou* is especially intriguing.

China declared that 93.4% of students completed their 9-year compulsory education and a nearly 100% enrolment rate in 2016, three decades after its introduction (Ministry of Education, 2017). As overall educational attainment is rapidly improving, more attention and effort should be devoted to ameliorating the enormous and often increasing inequality in its education system, along various dimensions including the gender and urban–rural divide. More research is necessary to provide the evidence base required for more precisely targeted policies in this field.

References

Attanasio, O. Blundell, R. Conti, G. and Mason, G. (2020). Inequality in socio-emotional skills: A cross-cohort comparison. *Journal of Public Economics*, forthcoming.

Awaworyi, S. and Mishra, V. (2018) Returns to education in China: A meta-analysis. *Applied Economics* 50(54), 5903-5919.

Barro, R. and J.W. Lee (2013). A New Data Set of Educational Attainment in the World, 1950-2010, *Journal of Development Economics,* Vol. 104, pp. 184-198.

Bertocchi, G. Bozzano, M. (2019) Gender Gaps in Education. *IZA Discussion Paper* No. 12724.

DiNardo, J., Fortin, N.M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica* 64(5): 1001-1044.

Chen, Y., and S. Feng (2013). Access to public schools and the education of migrant children in China." *China Economic Review* 26, 75–88.

Chernozhukov, V., Fernandez-Val, I., and Melly, B. (2013). Inference on counterfactual distribution. *Econometrica*, Vol. 81(6).

Giles, J. and Huang, Y. (2020). Migration and human capital accumulation in China: Migration may generate detrimental long-term impacts by widening the urban–rural educational gap. *IZA World of Labor* 2020: 476.

Duan, W., Guang, Y. and Bu, H. (2018). The Effect of Parental Involvement and Socioeconomic Status on Junior School Students' Academic Achievement and School Behavior in China. *Frontiers in Psychology* 9: 952.

Duckworth, A.L., and Seligman, M.E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Education Psychology*, 98(1), 198:208.

Fortin, N. M., Oreopoulos, P. and Phipps, S. (2014). Leaving boys behind: Gender disparities in high academic achievement. *Journal of Human Resources*, Vol. 50 (3), 549-578.

Golsteyn, B. H.H. and Schils, T. (2014). Gender gaps in primary school achievement: A decomposition into endowments and returns in IQ and non-cognitive factors. *Economics of Education Review*, Vol. 41, 176-187.

Gong, J., Lu, Y. and Song, H. (2018). The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes, *Journal of Labor Economics* 36(3): 743-778.

Hanushek, E.A. (1979). Education Production Functions. In Steve Bradley and Colin Green (ed.) Economics of Education, 2nd Edition, pp 161-170. Academic Press, London.

Haspido, L., and Moral-Benito, E. (2014) The public sector wage premium in Spain: Evidence from longitudinal administrative data. IZA Discussion Paper, No. 8315.

Hermann, Z., Kopasz, M. (2019). Educational policies and the gender gap in test scores: a cross-country analysis, *Research Papers in Education*, DOI: 10.1080/02671522.2019.1678065.

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal* 8(4): 453-479.

Juhn, C., Murphy, K.M., Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy* 101(3), 410–442.

Li, H., Liu, P.W. and Zhang, J. (2012). Estimating Returns to Education Using Twins in Urban China. *Journal of Development Economics* 97: 494–504.

Liu, R. and Chiang, YL (2019) Who is more motivated to learn? The roles of family background and teacher-student interaction in motivating student learning. *Journal of Chinese Sociology* 6:6.

Machado, J, A.F., and Mata, J. (2005). Counterfactual decomposition of changes in wage distribution using quantile regression. *Journal of Applied Econometrics*, Vol. 20: 445-465.

Matthews, J.S., Cameron, C.E., and Morrison, F.J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Education Psychology*, Vol. 101(3), 689:704.

Munir, F. and Winter-Ebmer, R. (2018) Decomposing intergenerational gender test score differences. *Journal for Labour Market Research* 52:12.

OECD (2015a), *Education at a Glance 2015: OECD Indicators*, OECD Publishing, Paris.

OECD (2015b), Emerging gender gaps in education, in *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence,* OECD Publishing, Paris.

OECD (2020), Social and emotional skills: Well-being, connectedness and success. OECD Publishing, Paris.

San Román, AG and Rica, S. (2020) Gender gaps in pisa test scores: the impact of social norms and the mother's transmission of role attitudes. *IZA DP* No. 6338.

Xu, D., Li, Q. (2018). Gender achievement gaps among Chinese middle school students and the role of teachers' gender. *Economics of Education Review* 67, 82-93/

Wang, X., Luo, R., Zhang, L. and Rozelle, S. (2017a). The education gap of China's migrant children and rural counterparts. *Journal of Development Studies* 53(11), 1865-1881.

Wang, X., Bai, Y., Zhang, L. and Rozelle, S. (2017b). Migration, schooling choice and student outcomes in China. *Population and Development Review* 43, 625-643.

Wang, H., Cheng, Z. and Smyth, R. (2018) Do migrant students affect local students' academic achievements in urban China? *Economics of Education Review* 63, 64-77/

Wu, J., Wei, X., Zhang, H. and Zhou, X. (2019) Elite schools, magnet classes, and academic performances: Regression-discontinuity evidence from China. *China Economic Review* 55, 143-167.

Zhao, G., Ye, J., Li, Z. and Xue, S. (2017) How and why do Chinese urban students outperform their rural counterparts? *China Economic Review* 45, 103-123.