Closing the Gender Gap in Science: New Evidence from urban China

Weili Ding
Queen's University and NYU-Shanghai
weili.ding@nyu.edu

Yipeng Tang
Capital Normal University
tangyi_pig@hotmail.com

Yongmei Hu*
Beijing Normal University and
National Innovation Center for Basic Education Quality Monitoring
huym0718@bnu.edu.cn

July 2019

Abstract

In this paper, we analyze recently collected data that conducts a unique assessment of high school student performance for over two thousand students from five Chinese provinces. Across three domains of scientific intelligence tested, we document heterogeneous gender gaps in academic performance. These differences generally arise due to differential productivity of inputs to the education production process and not differential levels of inputs. At many quantiles of the achievement distribution, girls perform better than boys when identifying scientific issues, whereas the converse holds on the portion of the assessment that measures whether one can apply scientific evidence. These differences may partially explain the subsequent gap in decision to major in specific STEM fields in college.

*We would like to thank Steven Lehrer for many helpful comments and suggestions on earlier drafts. We also wish to thank conference participants at the Canadian Economics Association Annual Meeting for additional comments. Ding wishes to thank the Project 985 visitor scholar funding for research support that faciliated conducting this project via visits to Beijing Normal University. We are responsible for all errors.

1 Introduction

For decades, the commonly accepted wisdom has been that girls are not as good as boys at either math or science. This has spawned an active academic literature that investigates whether or not this is true,[1] and if so, why. In a highly cited paper, Guiso et al. (2008) documents that teenage boys on average tend to outperform girls as measured by math test scores on the 2003 Program for International Student Assessment (PISA).[2] Additionally, these authors document substantial heterogeneity in gender performance gaps across industrialized countries and suggest that cultural explanations may be at play. The OECD further reported that in mathematics performance, boys outperformed girls in 54 out of the 65 countries that participated in the 2009 PISA. Results from China, where the 2009 PISA was only completed by teenagers in Shanghai, demonstrate that it is one of only 11 countries where on average, girls scored slightly higher, but the gap is not statistically significant.

In this paper, we extend this literature on gender gaps in math and science using recently

[1] To the best of our knowledge, this literature dates back to the 19th century. Romanes (1887) who was a student of Darwin, completed one of the very first studies in this area and concluded that gaps would begin in adolescence and persist, since mental abilities are secondary sex characteristics.

[2] More generally, the academic literature focuses more heavily on gender gaps in mathematics than science and posits two types of explanation: cultural and biological. Cultural factors include home and school inputs. For example, Dee (2007) and Carrell, Page, and West (2010) both find that the gender of the teacher matters for female performance on science and math. Fryer and Levitt (2010) suggests a role for stereotypes since parents are found to have lower math expectations for their daughters. Nollenberger et al. (2014) presents evidence that among immigrant populations the transmission of cultures related to gender equality from the country of ascent influences the size of gender gaps. However, using a different empirical strategy, Bharawadaj et al (2013) suggest that the evidence of the importance of culture as an explanatory factor in Guiso et al (2008) is sensitive to the inclusion of specific countries in the data being analyzed. Regarding biological explanations, Ceci et al (2009) conduct a review of studies in this area and note that much of the available evidence is mixed and often times contradictory. That said, Kucian et al, (2005) provides compelling evidence of functional differences in biological factors, since boys and girls are found to differ in which regions of their brain are activated when solving numerical problems.

collected data by the authors, who administered an examination based on the 2009 PISA to teenagers studying in five major urban cities across China other than Shanghai. This examination was influenced by the design of PISA but strictly focused on different domains of scientific intelligence and competence. Thus, we provide important information on children that were previously not examined in the literature. Second, by preparing questions on the assessment in separate and distinct domains of scientific intelligence, we follow Borghans et al. (2008), Cunha et al. (2010), among others, to account for the multidimensionality of a child's human capital.

The different dimensions being tested include the ability to i) simply formulate a research issue, ii) translate scientific knowledge, and iii) incorporate existing knowledge in new situations. These dimensions capture different skills that can prepare students for various careers. Intuitively, scientific literacy can help one make more informed choices that influence the direction of future scientific developments (e.g. stem cell research),[3] whereas advances in science would require more in depth knowledge of applying the procedures and practices associated with scientific enquiry; rather than just knowledge of the broad concepts and theories.

Our final extension to the literature is methodological and since there may be important policy relevant heterogeneity that is ignored if we focused solely on mean impacts, we use a recently developed microeconometric strategy to shed new light on the extent to which there is heterogeneity in these gender gaps in student performance. After all, Xie and Shauman (2003) have reported that gender gaps in the top 5 percent of the US high school math distribution have remained constant over a 20 year period.[4] Thus, our aim is to understand across the

[3]This is akin to mathematical competency being crucial to every citizen. After all, individuals' make decisions that require some forms of mathematical intelligence such as how to save for retirement or involving a distinction between fixed versus adjustable rate mortgages. Individuals with more advanced understanding may possess skills essential for certain careers including becoming an accountant, economist or actuary.

[4]Specifically, they find that a 2:1 ratio of boys to girls in the top 5 percent of the distribution. Further, Ellison and Swanson (2010) document that the gender gap in secondary-school math at high-achievement levels is present in every US high school, although the size of the gap varies between schools. Note, Bedard

distribution in multiple domains of scientific intelligence, whether gender differences in performance arise due to different inputs or differential productivity of these inputs. Our empirical approach first relaxes the need to impose common support as in Gevrek and Seiberlich (2014) when undertaking the test score gap decomposition and second can decompose the test score gap to three components rather than two, providing further insights.

From a policy perspective, there is substantial interest in reducing gender gaps in economic outcomes such as wages and decisions to major in STEM fields in college. Evidence surveyed in Altonji, Blom and Meghir (2012) concludes that there are strong links from high school math grades and mathematical ability to both subsequent field choice in college and early career outcomes. Exploring different dimensions of scientific intelligence may shed new light on the pathways by which the large gender differences arise from majoring in different fields of study within STEM. For example, Turner and Bowen (1999) use time series data to show that although there was a surge of women majoring in biology, there has been no change in the gender gap in either mathematical sciences or physics. Thus, understanding the potential sources behind the divergence in math and science test scores between boys and girls may provide guidance on the directions in which to reform education curriculums and in designing policies that aim to improve the future outcomes of young women.[5]

Our empirical analysis uncovers three major findings. First, there is significant difference in the gender gap across domains of scientific understanding. Girls outperform boys at every quantile of the unconditional distribution of identifying scientific issues. In contrast, above the 20th quantile, boys seem to outperform girls in the other competencies being tested that focus on deeper levels of understanding such as providing explanations and applying the scientific

and Cho (2010) review the literature that reports a gender gap in math scores across OECD countries

[5]Prior policy responses suggested role models where women would teach younger girls and often had an empirical basis being a case study or personal reflection. Carrington et al., (2008) formally test this hypothesis and find little support for those who advocate recruitment drives with role models. Naturally, if the differences are biological rather than cultural in origin, the policy responses would differ.

knowledge in new contexts. These differences in performance may partially explain why girls are more likely to major in biology relative to boys, whereas boys are more likely to choose college majors such as physics and engineering.

Second, the gender gap in almost every competency can be explained by differences in the productivity of inputs across genders. In general, inputs have a lower rate of productivity for girls than boys and this may suggest that there are gender differences in learning styles. Indeed, a growing body of evidence predominately from neuroscience that is surveyed in Miller and Halpern (2014) demonstrates gender differences in neural processing in different dimensions of mathematics and scientific knowledge among children. The authors conclude that the bulk of the evidence suggests that many of the differences in neural processing between genders emerge in part, due to educational policies and curricula that do not jointly maximize boy's and 'girl's cognitive potential in these subject areas.

Third, our results indicate there are significant differences in which factors influence achievement for boys and girls. This result is consistent with recent literature including Gong et al. (2018) who find gender differences in responses to teacher gender in China's middle schools. In particular, the returns to attending a key school significantly differs across many quantiles.[6] However, irrespective of the domain of scientific understanding being tested, little, if any, of the gender gap in student performance is due to differences in the levels of home and school inputs. Statistical tests do not provide any evidence of there being a composition

[6] A key school can be viewed as a summary indicator for several highly correlated school quality differences. These school quality differences that have been shown in Ding and Lehrer (2007) and Hoekstra et al. (2018) to be driven by teacher quality. Attending a key school has been repeatedly shown to significantly increase students' odds of attending a four-year college. Since decomposition analysis requires the estimation of contemporaneous education production functions by gender, the key school captures a combination of inputs received and these students attend courses with different teachers across subjects. Thus, this variable could also be viewed as capturing the cumulative effect of the full history of school inputs a child receives and without plausible exogenous variation in the contemporaneous inputs it would not be possible to separately identify an unbiased effect of that contemporaneous home or school input from historical inputs.

effect at any quantile in every domain being tested.

The above result may seem surprising given the large literature that documents gender differences in household investments in China. Yet, these results are consistent with growing evidence summarized in Zeng et al. (2014) that in contemporary urban China, partially as a result of the one child policy, gender gaps in upbringing, income and expectations have declined markedly and may have disappeared. This finding suggests that Chinese parents now recognize transformations in the labor market over the last two decades when optimizing their objectives by selecting child inputs. These transformations would include the rapidly increasing returns to schooling beyond secondary school that exceed most Western nations and the fact that on average, women in China now largely have the same educational opportunities as men. Thus, inputs and parental investments appearing more gender neutral in urban China should not be a surprise. Taken together, our findings suggest that policy discussions should shift from focusing on the potential consequences of improved understanding of sex preferences on input choice in urban China to understanding why gender differences in educational productivity emerge; and how can they be remedied.

This paper is organized as follows. In the next section, we describe the data that we collected in China and the structure of the assessment tool used to measure different domains of scientific competency. Section 3 outlines the econometric strategies used to undertake a distributional decomposition of the gender gap in student performance on elements within the designed assessment. Our empirical results are presented and discussed in Section 4. A concluding section summarizes our main findings and discusses directions for future research.

2 Data

To evaluate the impacts of youth science competitions on students' scientific learning outcome, the Institute of Economics of Education (IEE) at Beijing Normal University and the Children and Youth Science Center of the China Association for Science and Technology conducted a survey with direct student assessments across China.[7] China is a country with vast territory and enormous diversity. To ensure the representativeness of the survey, the survey was designed to be administered in the capital cities of five provinces respectively located in different regions of China.[8] Not only were these locations selected to reflect geographic diversity, but they are at different stages of economic development;[9] although none are among the booming coastal provinces that have driven much of China's recent growth.

Within each capital city, the project team first collected the list of all schools accredited by the local educational authority. Schools were subsequently sorted into two groups on the basis of their quality or effectiveness as determined by whether they were designated as a key or regular school. The project team randomly selected one key and one regular school in each capital city to participate in this study. Ding and Lehrer (2007) discuss how key schools and regular schools differ in terms of teacher quality, student graduation outcomes, peer quality and allow teaching styles to incorporate greater classroom autonomy. Given the budget for data collection, attention was paid to increase external validity by collecting data from few

[7] The Children and Youth Science Center of CAST is a non-profit organization comprised of scientists and engineers. This center both popularizes and promotes public knowledge of science as well as organizes numerous science contests within China.

[8] We account for geographic differences in our analyses since prefecture-level science associations in the provinces we surveyed in the eastern, central and west regions respectively spent in the 2007/8 academic year RMB1.61 million, 0.67 million and 0.59 million on adolescent science and technology activities. The uneven distribution of resources for science and technology education is also quite prevalent within provinces as well as across the 31 administrative regions in China.

[9] These five capital cities are Shenyang city in Liaoning province, Lanzhou city in Gansu province, Wuhan city in Hubei province, Fuzhou city in Fujian province, and Chengdu city in Sichuan province. According to 2010 data from China National Bureau of Statistics, the per capita GDP of these five provinces were 5909, 5525, 4079, 3104 and 2368 US dollars respectively, and their rankings among all the 31 administrative regions across China were 9th, 10th, 13th, 23rd, 28th respectively.

randomly selected schools in five provinces as opposed to many schools in a single province.[10] Thus, within a province, we can make comparisons of some inputs, but across provinces the support of inputs also could change as does some features on how the school operates.

Each school principal was contacted and asked to provide a list of all classes offered at the grade 10 and 11 level and whether they would be considered regular or honors.[11] Each principal gave consent and provided the full class rosters for all classes offered at these levels. The project team next randomly selected one class within each school. In total, data was collected from all of the 2497 students that were registered in these 22 classes drawn from 11 schools in five capital cities in 2010.[12]

Each student in the sample completed a brief questionnaire to provide some demographics as well as information related to home inputs hypothesized to be associated with student's science learning.[13] The home and school inputs selected were based on evidence from the

[10] Within a province, comparisons of levels of inputs across schools can be made. However, across provinces, the support of inputs also could differ as well as certain features on how schools operate.

[11] Chinese law mandates nine years of free education, from primary school to the end of middle school and the largest drop in secondary education enrollment occurs prior to senior high school. For example, Yi et al. (2015) report that more than 51 percent of junior high school students in poor rural areas do not go on to high school. Further, Loyalka et al. (2017) report across China that between 4.2% and 7.4% of students who enrolled in academic senior high school dropped out before graduating as well as small gender disparities that likely arise since boys can more easily find higher wage employment. The evidence is suggestive that any gender differences in dropout behavior between grades 10 and 11 would not be concentrated in specific locations of the distribution. Further, since we do not have information on the same cohort in the prior academic year. That said, we cannot directly investigate differential gender disparity in dropout behavior between grades 10 and 11. However, multivariate regression do not find evidence supporting the notion of gender differences across grades, conditional on school fixed effects in our sample.

[12] In Wuhan, the project team collected data from 3 schools and as a result six classrooms.

[13] There were also questions related to individual participation in youth science and technology contests in China, such as the China Adolescents Science & Technology Innovation Contest (since 1982), Awarding Program for Future Scientists (since 2000), and China Science Olympiad of five subjects for middle school students. It is well recognized that these contests are contributing to identifying gifted youths who have special

economics of education literature.[14] Following the questionnaire, each student spent one-hour completing an assessment designed to measure domains of scientific literacy. As such, this data collection was named the Chinese Youth's Scientific Literacy Dataset (CYSLD).

The assessment tool follows how PISA measures scientific competency, which is considered to be the key aspect of scientific literacy.[15] Scientific literacy can be further broken down into three measurable cognitive capabilities, i) identifying scientific issues (ISI), ii) explaining phenomena scientifically (EPS), and iii) using scientific evidence (USE). The characteristics of these scales are detailed in Table 1, but essentially ISI can be viewed as the most simple level of learning since it just reflects basic scientific understanding, whereas USE reflects the highest level of learning since it involves a detailed understanding of a concept including scientific knowledge translation.

To develop an assessment related to the PISA framework of scientific competencies, CYSLD made some minor adjustment to reflect aspect of China's educational context. The adjustments to the assessment are fully detailed in Hu et al., (2012) and are minor in nature. The first adjustment is that each question on the assessment is developed by the Center for National Assessment of Education Quality, which is a national institute affiliated to China's Ministry of Education whose mission is to monitor and assess basic education quality in China. Second, the CYSLD expanded the contents of PISA's science test to five main areas - physical

interests in science study and will choose to start careers in science-related fields (Hu, et al., 2012)

[14] For example, Fryer and Levitt (2010) use the frequency of books in their investigation of gender differences in math achievement. Substantial research in survey design finds that ownership of household durable assets is easier to measure reliably than income or consumption expenditures and is regarded as a good indicator of long-term household wealth and other socio-economic household characteristics (see e.g. Balen et al. (2010)).

[15] That said, what is scientific capability and how to evaluate it has long been questioned and explored by educational researchers. As an influential and well-recognized international assessment program, PISA proposed the concept of scientific literacy and successfully elaborated this concept from four aspects - context, knowledge, attitude and competencies (OECD, 2007). Scientific competencies are often portrayed as being the most important component of the scientific literacy, and it mainly refers to the three measurable cognitive capabilities that we explore in our assessment.

science (physics and chemistry), life science, earth and space science, information technology and mathematics. Thus, the scientific competencies tested are broader than PISA. Finally, related to scoring individual questions, a single parametric Rasch model was used for scaling as opposed to PISA's practice of providing five plausible values. This alternative transformation does not change the ordinal nature of the test scores and is both more intuitive and easily replicated by other researchers, whereas the plausible values are ad-hoc.

2.1 Summary Statistics

Table 2 provides summary information on the student, family and school variables measured in the CYSLD. In total, 2436 students (97.56%) fully completed all aspects of the questionnaire and statistical tests suggest that there is no link between observables and the likelihood of completion. In this sample, there are 1064 girls and 1372 boys, a ratio of boys to girls is 1.29:1, which is consistent with China's demographic adolescent gender ratio in these provinces in 2010. The final column contains results of whether there are average differences in any of these observed variables between the genders. None of the school inputs, including class size and type of school, differ significantly at the 5% level between the genders.[16] Last, there is no significant gender difference in key school attendance.

Most of the children come from families that neither own a motor vehicle or a personal laptop. There is no gender difference in these assets including whether the child owns a cellphone. Many children claim to have access to a social web account as proxied by QQ, which is an instant messaging software service. Boys have a significantly higher rate of holding a QQ account and this may be the case since QQ also offers a variety of services, including online

[16] That said, analysis of our surveys conducted in the middle schools reveal that within cities, there is substantial heterogeneity in the instruments and equipment available in the physics, chemistry and biology labs. In many regular middle schools these materials are insufficiently provided, and our analysis further reveals that there are much fewer popular science magazines and books in the school libraries.

social games and platform of games and group.[17] Consistent with the literature on China's son preference, families with boys are more likely to be larger and this difference is driven strictly by the Western provinces.[18] On average, fewer families with girls claim to earn more than 5000 RMB (~830USD) a month and boys are significantly older by roughly 7 weeks.

Turning towards scientific competency, the bottom rows of Table 2 indicate that on average, boys and girls do not perform differently on the assessment. However, the females in the sample do perform significantly higher than the males on both the ISI and EPS subscales, whereas girls score slightly lower on the USE subscale than boys. The difference on the USE subscale is not statistically significant. Figure 1 presents Kernel density estimates of the distribution of scores by gender on each subscale and we observe in the bottom two panels that there is a slight rightward skew to the male distribution on the EPS and USE subscales. The converse pattern appears on the ISI scale and the female density curve is shifted to the right of that of males beyond the 70th percentile. On the assessment as a whole as well as each of the four subscales we observe that there is a larger mass of males in the lower percentiles.[19]

[17] Financial releases from Tencent the company that manufactures QQ indicate that there were 829 million active QQ accounts in late 2014. We should note that in the survey, the subjects not only provided details on their QQ account but also provided information on whether they or their parents had an email account and included the email address.

[18] Specifically, we conducted tests of differences in the proportion of families where the subject is the only child by gender across regions. In eastern China and central China two sided tests are unable to find significant differences (z=0.93 Pr(z>z*=0.35), and z=-0.50 Pr(z>z*=0.62) respectively. However in the Western provinces 58% of the boys are the only child in their family compared to 48.9% of the girls, which is statistically different z=3.34 Pr(z>z*=0.001) at the 1% level.

[19] A potential concern is that this may not reflect student knowledge and the distribution arises due to gender differences in response to the stakes of the examination. For example, Cai et al (2015) provide evidence from a field experiment conducted in a single county in the Fujian province that there are gender differences in the student effort based on the stakes of the examination. The results are interpreted as showing that compared to male students, females underperform on the high-stakes Gaokao, relative to their performance on the low-stakes mock examination held two months earlier. This is interpreted as being consistent with evidence that women appear to systematically underperform relative to men in competitive settings and women may

Finally, under each graph appears results from Kolmogorov-Smirnov tests of the equality of distribution between genders and indicates that there are significant differences in each test at the 5% level with the exception of using scientific knowledge which is significantly different at the 10% level. These differences are not being driven by extreme values which is important since the Kolmogorov–Smirnov test is well known to lack statistical power against differences in the tails of distributions.

The fact that these differences is not due to extreme values is important to stress since a controversial alternative hypothesis for gender differences in performance across quantiles is the greater male variability hypothesis of Ellis (1894). This hypothesis postulates that males are more variable than females in academic performance, and as such they are more likely to be overrepresented in the tails of a distribution.[20] Our econometric analysis outlined in the next section does not focus on testing this hypothesis and strives to improve our understanding

simply prefer to opt-out from competition (e.g. Bertrand, 2010, Niederle and Vesterlund, 2007) but it remains possible that women outperform men when competitive pressures are lower (e.g. Ors et al. (2013), Jurajda and Munich (2011) and Attali et al. (2011)). This possibility does present an alternative explanation for the increased mass of males in lower tails but not the shift to the right at higher percentiles on the USE and EPS subscale seen in the respective panels of Figure 1. However, it is extremely unlikely that this could explain the differing gender differences in performances across competencies since the assessment did not clearly inform the students of which competency each question evaluated.

[20]This hypothesis generates controversy since it has been used to explain dimensions of occupational segregation by suggesting that fewer women have the requisite ability for certain status jobs. While some studies (see e.g. Machin and Pekkarinen, 2008 and Laken, 2013, among others) do present evidence that supports the greater male variability hypothesis with specifics that differ according to whether mathematics or reading are being studied and tested, other studies contradict these findings and emphasize that cultural factors (see e.g. Makel et al., 2016, Hyde and Mertz, 2013) are responsible for the variability presented. More recently, O'Dea et al. (2018) compared gender differences in academic grades from over 1.6 million students and concluded that since the gender differences in both mean and variance of grades are smaller in STEM than non-STEM subjects, suggesting that greater variability in performance is insufficient to explain male over-representation in STEM careers.

of why performance gaps between the genders emerge at different points in the distribution. Our analysis focuses on the roles of school and home inputs as well as their differential returns in explaining performance differences across quantiles by gender.

3 Econometric methods

A growing literature in the economics of education analyzes the differences in outcomes between groups; males and females in our case.[21] Much of this literature undertakes a predominately descriptive decomposition analysis. Given the differential heterogeneity in performance documented in Figure 1, we believe that any decomposition of test scores between genders should be ideally conducted over the entire distribution. Thus, we utilize a semiparametric quantile regression technique developed in Chernozhukov, Fernandez-Val and Melly (2013) to undertake this decomposition across the distribution of test scores.[22]

Our aim is to understand the relative contribution of factors in explaining the differential gender gap and not to identify causal effects. To understand our empirical strategy, we assume that child i in school s has a education production function that is gender-specific and for females is given by

$$A_{is} = \beta_0^f + \beta_1^f X_{is} + \varepsilon_{is}^f \tag{1}$$

and the males' production process is expressed as

$$A_{is} = \beta_0^m + \beta_1^m X_{is} + \varepsilon_{is}^m \tag{2}$$

[21] Examples within the economics of education literature where test score gaps are decomposed between countries (Ammermuller, 2007), schools (Krieg and Storer, 2006) and ethnic groups (McEwan, 2004); among many others.

[22] An additional advantage of using this methodology and moving beyond carrying out a decomposition at the mean is that as we outline in this section, we are estimating a more flexible model for heterogeneous data at various points of the conditional achievement distribution. Last we note that this decomposition builds off results in Melly (2005) and it is also considered to be more efficient than the semiparametric decomposition developed in Machado and Mata (2005).

where X is a matrix of observed individual and school level inputs. Similar to the Blinder-Oaxaca decomposition taken at the mean, the parameters are not restricted to be equal between the genders. For each quantile τ, we first estimate these two equations using the conditional quantile regression estimator providing us with estimates of $\beta_1^m(\tau)$ and $\beta_1^f(\tau)$.[23]

The distributional decomposition of Chernozhukov, Fernandez-Val and Melly (2013) extends the framework developed in Juhn et al. (1993) that provides a tool to describe three components that may cause the density of achievement to change between groups; measured returns to inputs, measured quantities of inputs and residuals. The underlying idea is to construct a counterfactual unconditional achievement distribution for girls and insights can be gleaned by comparing it to the actual female achievement distribution. To carry out this decomposition an overlapping support assumption on both observed and unobserved factors is maintained.[24] From the perspective of an applied econometrician, a particularly attractive feature of this method is that it does not assume rank-preservation, so that girls do not need to have the same rank in the counterfactual distribution as the actual distribution. Further, one does not need to assume homoskedasticity with the Chernozhukov et al. (2013) decomposition.

To provide some intuition for why a counterfactual distribution is constructed, we define F_{A^m} and F_{A^f} as the respective distribution functions of test scores by gender. Just as a comparison of means between groups can only can test for statistical differences, comparing the τ^{th} quantile of F_{A^m} and F_{A^f} would not permit any additional economic insights.[25] We

[23]The quantile regression estimator is more flexible than an OLS estimator since it allows the impact of each covariate to vary over the distribution of unobserved factors (i.e., ability) that affect achievement on each portion of the assessment.

[24]Ex ante we do not see any reason why this assumption would not hold between genders in China or elsewhere. Note that, whereas ordinary least square regressions have the property that the mean of the dependent variable and the mean of the explanatory variables are on the regression line, thereby making the Blinder-Oaxaca decomposition of the dependent variable fairly straightforward; quantile regression estimators do not share this property.

[25]The main technical challenge when extending a decomposition procedure from simple means to quantiles

define $\widehat{q}_\tau(\widehat{\beta}_1^f, X^f)$ and $\widehat{q}_\tau(\widehat{\beta}_1^m, X^m)$ to be the unconditional achievement distributions for each gender at quantile τ, where $\widehat{q}_\tau$ is a natural estimator of the τ^{th} quantile of the unconditional achievement distribution. Suppose that we both add and subtract a term to $\widehat{q}_\tau(\widehat{\beta}_1^m, X^m)$ $-$ $\widehat{q}_\tau(\widehat{\beta}_1^f, X^f)$. This term $\widehat{q}_\tau(\widehat{\beta}_1^f, X^m)$ reflects this unconditional counterfactual distribution for boys is designed to represent the achievement of males if they had the same production function parameters ($\widehat{\beta}_1^f$) as females. Doing so generates

$$\widehat{[q_\tau}(\widehat{\beta}_1^m, X^m) - \widehat{q}_\tau(\widehat{\beta}_1^f, X^m)] + [\widehat{q}_\tau(\widehat{\beta}_1^f, X^m) - \widehat{q}_\tau(\widehat{\beta}_1^f, X^f)] \tag{3}$$

The terms represented in the squared parentheses of equation (3) yield specific insights. The first term represents the portion of the gender gap attributable to group differences in the coefficients including the intercept. In our application, the difference in the intercept is quite small so we will use the terminology differential productivity of inputs rather than refer to this as an unexplained component which is occasionally done in the labor economics literature. The second term, captures the contribution to the gender gap attributable to group differences in the distribution of characteristics at quantile τ. However, since equation (3) does not consider there to be differences in the residuals, an additional step is required.

Formally, to add this next step and undertake this decomposition, we define the conditional achievement distribution for girls as

$$F_{A^f|X}(A^f|X = x) = \int_0^1 1\{F^{-1}_{A^f|X}(\tau|X = x) \preceq q_\tau\} d\tau \tag{4}$$

where τ is the τ^{th} quantile of the unconditional achievement distribution functions. The counterfactual unconditional achievement distribution for boys is generated by integrating the conditional achievement of girls over the marginal distribution of covariates among boys

arises since unconditional quantiles differ from the integral of the conditional quantiles. To solve this problem, we construct a counterfactual distribution as outlined in this section.

and is expressed as

$$F^c_{A^m}(q_\tau) = \int \left[\int_0^1 1\{F^{-1}_{A^f|X}(\tau|X = x) \preceq q_\tau\} d\tau \right] dF_{A^m}(x). \tag{5}$$

In practice, this unconditional counterfactual distribution is recovered by replacing $F^{-1}_{A^f|X}(\tau|X = x) \preceq q_\tau$ with consistent estimates of $\widehat{\beta}^f_1(\tau)X_{is}$ that are obtained from the conditional quantile regressions and then simply inverting the distribution function $F^c_{A^m}$. That is, estimates of $\widehat{\beta}^m_1(\tau)$ and $\widehat{\beta}^f_1(\tau)$ are used in place of the conditional quantile regression function in equation (4). Last, by inverting the unconditional counterfactual distribution function $F^c_{A^m}(q_\tau)$, the unconditional quantiles of interest are recovered which can then be compared to quantiles of the true unconditional distribution function for boys ($F_{A^m}(q_\tau)$) to measure the gender gap.

To undertake the decomposition, we first add and subtracting the term $\widehat{q}_\tau(\widehat{\beta}^f_1, X^m)$ reflecting the counterfactual distribution from $\widehat{q}_\tau(\widehat{\beta}^m_1, X^m) - \widehat{q}_\tau(\widehat{\beta}^f_1, X^f)$ as shown in equation (3). To next separate the effects of the coefficients from the effects of the residuals, a $(NX1)$ vector β^* is defined, with its n^{th} component

$$\beta^*_1(\tau_n) = \widehat{\beta}^m_1(0.5) - \widehat{\beta}^f_1(0.5) + \widehat{\beta}^f_1(\tau_n) \tag{6}$$

is constructed. In the above equation, $\widehat{\beta}^f_1(0.5)$ and $\widehat{\beta}^m_1(0.5)$ are the estimated coefficient vectors from a median regression for each gender is defined. By doing so and both adding and subtracting $\widehat{q}_\tau(\widehat{\beta}^*_1, X^m)$ to equation (3) allows the overall decomposition to be reexpressed as

$$[\widehat{q}_\tau(\widehat{\beta}^*_1, X^m) - \widehat{q}_\tau(\widehat{\beta}^f_1, X^m)] + [\widehat{q}_\tau(\widehat{\beta}^f_1, X^m) - \widehat{q}_\tau(\widehat{\beta}^f_1, X^f)] + [\widehat{q}_\tau(\widehat{\beta}^m_1, X^m) - \widehat{q}_\tau(\widehat{\beta}^*_1, X^m)]. \tag{7}$$

The first component of equation (7) indicates the part of the achievement gap which is due to differences in coefficients of each explanatory variable. Using the education production framework, this portion can be interpreted as differences in the productivity of inputs between the genders. The second set of terms in square parentheses indicates which part of the gap between the groups is due to differences in explanatory variables. That is, to what extent do the differences arise from differences in the types of inputs boys and girls receive.[26] The remaining term in equation (7) is calculated as the difference between the estimated coefficient vector and the median coefficient vector. This term measures within group inequality and reflects the estimated dispersion of student performance that is exclusively attributable to residual inequality.

Standard errors for each of these components can be obtained by the bootstrap allowing us to test functional hypotheses about counterfactual effects including whether there is any significant effect; $H_0 : \beta(\tau) = 0; \forall\tau \in \tau$; versus $H_A : \beta(\tau) \neq 0$; for some $\tau \in \tau$.[27] Further, we formally test for stochastic dominance as well as production function effect heterogeneity; $H_0 : \beta(\tau) = \beta(0.5); \forall\tau \in \tau$; versus $H_A : \beta(\tau) \neq \beta(0.5)$ for some $\tau \in \tau$.[28] Tests can also be carried out on portions of the decomposition in equation (7) to shed further light on the extent of the heterogeneity and its potential sources.

Based on the Monte Carlo evidence in Chernozhukov et al. (2013), plug-in estimates of the gender-specific education production functions using a quantile regression estimator are used as opposed to a distribution regression estimator in equation (4). After all, the test score measures utilized in this paper are continuous variables since they are calculated based on one-parameter Rasch processes and Chernozhukov et al. (2013) shows that quantile regression performs better than distribution regression if the unconditional test score distributions (presented in Figure 1) do not involve mass points due to rounding.

[26] A large literature documents differences in education inputs across genders in both the developed world (e.g. see Lundberg and Rose, 2002 and Baker and Milligan, 2013, among others) as well as in China (see Song, 2001, Knight et al, 2010, Ding and Zhang, 2014, among others). It is beyond the scope of this paper to explain why these differences persist but mechanisms based on tastes as well as differences in investments due to statistical discrimination, have been posited in the literature.

[27] We note that this procedure involves exchangeable bootstrap distributions and would be time consuming with larger datasets than we utilize in this paper. In those situations, Melly (2005) does provide formulas to compute the analytic standard errors. In our implementation we consider 200 repetitions.

[28] The tests are focused on the full distribution and as such, do not require corrections for multiple testing.

4 Results

The four panels of Figure 2 present the results of decomposition analyses for the full assessment as well on each subscale. In each plot, we represent for that specific assessment the estimated unconditional gender performance score gap as a solid line, the coefficients effect (the differential productivity) with a medium-dashed line, the characteristics effect with a short-dashed line, and the residuals with long-dashed line. Notice that the total performance gender gap (solid line) is the sum of the remaining (dashed) effects.

Exploring the solid line in each competency being tested, we generally observe that in the lowest quantiles, girls outperform boys. In both the USE and EPS test, boys outperform girls at most quantiles, whereas girls outperform boys on the ISI test at nearly every quantile. The shape of the overall score gap most closely mirrors the shape of the score on the EPS scale, but is not marked as such.

Once we examine the components of the decomposition, we see the first surprising finding, in that irrespective of the competency being tested, the characteristics effects lie on top of 0 and does not display any heterogeneity. In fact, bootstrap tests on the quantile process are unable to reject either the Null of no effect at all quantiles or the Null that there is heterogeneity in the effect of characteristics at any quantile for all four test score outcomes being investigated. This result suggests that in urban China there is no significant differences in this set of explanatory variables between genders anywhere in the unconditional distribution.

The second major result is that, on the total score, the test score gap arises due to effects in the component of the decomposition related to the differential productivity which generally exceed the component related to residual heterogeneity at the lowest quantiles. Above the 50th percentile, the gap solely arises due to these differences in productivity. These results indicate that in the lower parts of the distribution there are gender different productivity relationships between inputs and achievement. In general, boys receive larger and more positive returns to the inputs, which is why the effects of residual heterogeneity is negative. Recall, the hypothesis on gender differences in student effort, if true, would suggest that this underestimates the productivity gap. Since we condition on composition effects, the effects of residual heterogeneity suggest that there is more dispersion in male test scores that slowly decreases at higher percentiles.

The third finding is that this pattern of offsetting effects of residual inequality from the effects of coefficients has its origins on the USE and EPS subscales. On the EPS subscale, these patterns solely appear in the bottom 15 percent of quantiles, whereas this appears throughout the distribution on the USE subscale. Bootstrap tests of whether the effects of coefficients are constant across all quantiles are rejected at the 2% level or lower on each of these two scales and Kolmogorov-Smirnov tests reject the hypothesis of no stochastic dominance of boys on coefficients;[29] thereby providing convincing evidence that not just the education production parameters differ, but those for boys are significantly larger across the distribution.

On the ISI portion of the test, girls outperform boys at every percentile. However, this difference is generally neither driven by differences in characteristics nor the effect of coefficients. Surprisingly, we observe that above the 20th percentile, boys have larger and more positive productivity parameters in their education production function than girls. Further and in contrast to the other two scales, only above the 50th percentile on the ISI subscale do we see the component associated with residual inequality being positive. This indicates that there is much less within gender heterogeneity at the lowest quantiles in the ISI scale. This result is somewhat surprising since Figure 1 suggested that there was a longer tail in the ISI scale on the lower end. Last, only at the very bottom of the distribution can one argue that the gains in female performance on ISI are due to their higher productivity parameters.

Taken together, these results provide evidence that gender differences in scientific performance differ based on the domain being tested. Irrespective of the domain being tested, the effect of coefficients suggests that boys have larger productivity parameters than girls at many

[29] The p-value for USE is 0.02, p-vale for EPS is 0.00 and the result also holds if we use Cramer-Von-Mises tests that simulations suggest have better power than Kolmogorov-Smirnov tests.

quantiles. Surprisingly, none of these gender differences in performance can be attributed to differences in characteristics in urban China. This suggests that more attention needs to be paid towards understanding the gender differential productivity process in both scientific learning across domains tested.

For robustness, Appendix Table 1 presents results of Oaxaca-Blinder decomposition of the average test score gap in each domain tested. These results continue to demonstrate that the lion's share of the average gap is explained by differences in coefficients and the remaining amount explained by differences in the coefficients is quite small. However, the advantage of carrying out the decomposition over the distribution, as opposed to just looking at average differences, is that we observed from the former that the difference in the effect of the coefficients is particularly pronounced and negative only at the lower percentiles. Detailed decomposition results reveal this to be the portion of the population who contributes to the prevalent gender gap and suggests that any policy responses should target these individuals.

In Table 3, we present quantile regression estimates of equations (1) and (2) that were estimated in the first step of constructing the decomposition to shed further light on the two sequences of quantile coefficients that contribute to the decomposition.[30]. To understand the differential productivity requires one to understand how the returns to different inputs affect the performance of boys ($\widehat{\beta}_1^m(\tau)$) and girls ($\widehat{\beta}_1^f(\tau)$). After all, these consistent estimates are used to replace $F^{-1}_{A^f|X}(\tau|X=x) \preceq q_\tau$ in equation (4).

To validate the appropriateness of conditional quantile regression in our application, we additionally conduct the bootstrap test developed in Rothe and Wied (2012) of the entire conditional distribution of achievement given observable inputs. We are unable to reject the

[30] To be explicit, these effects are on the conditional outcome distribution and as Ding and Lehrer (2011) discuss differ from those obtained by the unconditional quantile regression estimator of Firpo et al. (2009). Estimates of how changes in covariates influence the quantiles of the unconditional distribution of an outcome variable are available from the authors upon request and are generally quite similar in sign, statistical significance and magnitude.

Null and obtain a p-value of the test of the equation for girls and boys respectively of 0.62 and 0.11. This result reinforces that the education production process can be reasonably described by a linear quantile regression model.

The evidence in Table 3 shows that the magnitude and statististical significance of elements of $\widehat{\beta}_1^m(\tau)$ and $\widehat{\beta}_1^f(\tau)$ differ. For example when explaining the total score, family income and household computers both are found to significantly influence boys' performance, whereas proxies for other non-educational inputs such as the family owning a car or having a QQ account only influence girls performance in a statistically significant manner. Attending a key school does influence both gender, but the magnitude of the effect is approximately 30-50% larger for boys than girls at most quantiles. The larger effect of attending a key school which employ different teaching resources also appears at most quantiles on the EPS and ISI scales. Household location is another explanatory variable that is found to play a large role for boys that relatively improved performance on the USE scale; where the effect of living in an Eastern province relative to a Western province has nearly double the effect at most quantiles, as compared to girls. Interestingly, household income does not significantly affect girls' performance on any competency being tested at any quantile. Although statistically significant, we find that reductions in class size lead to very small gains in performance for both genders. Taken with the large effects for key schools, this suggests that other school inputs, particularly teacher quality, is quite important.[31] As a whole, these results provide some guidance as potential avenues that can reduce the gender gap in each scientific competency.

5 Conclusion

Explanations for gender differences in measures of student scientific and mathematics achievement can be generally divided into three categories, cultural, attitudinal, and educational.

[31] Ding and Lehrer (2007) and Hoekstra et al. (2018) present evidence of the various dimensions that school quality differs in urban China and find that teacher quality is quite important in boosting student achievement.

Within the economics of education, much cross-country research has suggested cultural factors play a leading role and an emerging literature indicates the importance of attitudinal and other non-cognitive skills; thereby reducing attention to educational explanations. However, using recent data collected by the authors we provide new evidence that in urban China much of the existence of heterogeneous gender gaps in student scientific capability in China can be accounted for by differential productivity of inputs.

From a policy perspective, educational explanations for the gender gap are easier to address than cultural or attitudinal factors since many school-related parameters including teaching and assessment methods traditionally used in math an science classes likely can be altered. For instance evidence in Hoffmann (2002) and Zohar and Bronshtein (2005) suggest that within physics classes, the notion of what it means to understand physics, and physics curricula are heavily biased towards boys' interests, knowledge, and abilities. Similar to results on differential effects of teacher gender on student performance by student gender, questioning if there is an optimal assignment of students to teachers, our results question how course time should be allocated by gender. Changing teaching methods and how curricula are developed can be low cost, so may offer a decent cost to benefit ratio. To inform this policy option future work is needed that could examine quasi-random variation in the weights paid to the different competencies in both teaching time within courses and on examinations to see if it does change student outcomes by gender.

An enormous body of literature has documented sex preferences and discriminatory investments in China. However, more recent analyses from urban China demonstrate a levelling of the playing field where urban girls grow up with nearly identical opportunities as their male counterparts. Despite these positive developments, our results of gender differences in educational productivity can potentially inform discussions surrounding other topics where gender differences remain prevalent such as those related to the continued under-representation of women in scientific fields that are mathematically intensive in colleges, graduate programs and the labor force. Our result of gender differences across domain being tested allow one to reconsider to what extent is it truly a puzzle that boys and girls are not equally likely to plan a career that involves mathematics or science. For example, if learning styles do indeed differ and girls make decisions related to college major based on self-beliefs of their likely success, it would not be a surprise that they eschewed majors that stressed competencies where they are relatively weak.

Taken together, our empirical results suggest that the policy discussion should move beyond arguments related to preferences or discrimination and instead focus more heavily on understanding how current environments shape our developing brain and learning style. Indeed, we observe that girls do outperform boys in certain areas of scientific competency and we propose that changes in curricula and teaching practices may prove to be effective in reducing the remaining gender gaps. Issues related to gender equality have long been neglected in curriculum reforms by committees responsible for China's basic education.

It is worth stressing that our results only provide a snapshot of the gender gap in scientific learning among high school students in five Chinese provinces in 2010. It would be interesting to extend this analysis to other domains such as mathematical intelligence as well as exploring the evolution of these gender gaps. Since the landscape in which Chinese children grow up continues to rapidly and heterogeneously evolve across regions in China, it would not be surprising that the relative roles of the components in the gender gap shifted.

Our main finding of gender differences within finer domains of scientific competencies is consistent with a larger body of interdisciplinary research. These results are suggestive that future work can construct research designs to evaluate whether targeted teaching techniques can successfully address gender difference in learning styles. We believe developing effective policy in this area requires incorporating findings from several scientific fields. This would not only hold potential for subsequent basic education reforms to truly balance human development between genders, but would potentially promote subsequent economic growth by increasing the levels of human capital for subsequent generations.

References

[1] Altonji, J. G., E. Blom and C. Meghir (2012), "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers," *Annual Review of Economics*, 4(1), 185–223.

[2] Ammermuller, A. (2007), "PISA: What makes the difference? Explaining the gap in test scores between Finland and Germany," *Empirical Economics*, 33(2), 263- 287.

[3] Attali, Y., Z. Neeman and A. Schlosse (2011), "Rise to the Challenge or Not Give a Damn: Differential Performance in High vs. low-stakes Tests." *IZA Discussion Paper No. 5693.*

[4] Baker, M. and K. Milligan (2016), "Boy-Girl Differences in Parental Time Investments: Evidence from Three Countries", *Journal of Human Capital*, 10(4), 399-441.

[5] Balen J., D. P. McManus, Y. S. Li, Z. Y. Zhao, L. P. Yuan, J. Utzinger, G. M. Williams, Y. Li, M-Y. Ren, Z-C. Liu, J. Zhou and G. Raso (2010), "Comparison of two approaches for measuring household wealth via an asset-based index in rural and peri-urban settings of Hunan province, China," *Emerging Themes in Epidemiology*, 7(1), doi: 10.1186/1742-7622-7-7.

[6] Bedard, K. and I. Cho (2010), "Early gender test score gaps across OECD countries," *Economics of Education Review*, 29(3), 348–363.

[7] Bertrand, M. (2010), "New Perspectives on Gender," in *Handbook of Labor Economics*, O. Ashenfelter and D. Card eds., Vol 4B: 1543-1590.

[8] Bharawadaj, P., G. de Giorgi, D. Hansen and C. Neilson (2016), "The gender gap in mathematics: Evidence from low- and middle-income countries," *Economic Development and Cultural Change,* 65(1) 141-166.

[9] Borghans, L., H. Meijers and B. ter Weel (2008), "The role of noncognitive skills in explaining cognitive test scores", *Economic Inquiry*, 46(1), 2–12.

[10] Cai, X., Y. Lu, J. Pan and S. Zhong (2017), "Gender Gap Under Pressure: Evidence from China's National College Entrance Examination", *forthcoming in The Review of Economics and Statistics.*

[11] Carrell, S., M. Page, and J. West (2010): "Sex and science: How professor gender perpetuates the gender gap," *The Quarterly Journal of Economics*, 125(3), 1101–1144.

[12] Carrington, B., P. Tymms and C. Merrell (2008), "Role Models, School Improvement and the "Gender Gap"–Do Men Bring out the Best in Boys and Women the Best in Girls?," *British Educational Research Journal*, 34(3), 315 - 327.

[13] Ceci, S., W. Williams, and S. Barnett (2009), "Women's underrepresentation in science: Sociocultural and biological considerations," *Psychological Bulletin*, 135(2), 218- 261.

[14] Chernozhukov, V., I. Fernandez-Val and B. Melly (2013), "Inference on Counterfactual Distributions," *Econometrica*, 81(6), 2205–2268.

[15] Cunha, F., J. J. Heckman and S. M. Schennach (2010), "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 78(3), 883–931.

[16] Dee, T. (2007), "Teachers and the gender gaps in student achievement," *Journal of Human Resources*, 42(3), 528–554.

[17] Ding, W. and Y. Zhang (2014), "When a son is born: The impact of fertility patterns on family finance in rural China," *China Economic Review,* 30, 192–208.

[18] Ding, W. and S. F. Lehrer (2011), "Experimental Estimates of the Impacts of Class Size on Test Scores: Robustness and Heterogeneity," *Education Economics*, 19(3), 29-252.

[19] Ding, W. and S. F. Lehrer (2007), "Do peers affect student achievement in China secondary schools?", *Review of Economics and Statistics,* 89(2), 300–312.

[20] Ellis, H. (1894), *Man and woman: A study of human secondary sexual characters.* London: Walter Scott; New York: Scribner's.

[21] Ellison, G. and A. Swanson (2010), "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions," *Journal of Economic Perspectives*, 24(2), 109–128.

[22] Fryer, R. and S. Levitt (2010), "An empirical analysis of the gender gap in mathematics," *American Economic Journal: Applied Economics*, 2(2), 210- 240.

[23] Gevrek Z. E., and R. R. Seiberlich (2014), "Semiparametric decomposition of the gender achievement gap: an application for Turkey," *Labour Economics*, 31(1), 27–44.

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

[25] Guiso, L., F. Monte, P. Sapienza and L. Zingales (2008), "Culture, Gender, and Math," *Science,* 320, 1164–1165.

[26] Hoekstra, M., P. Mouganie and Y. Wang (2018) "Peer Quality and the Academic Benefits to Attending Better Schools," *Journal of Labor Economics,* 36(4), 841-884.

[27] Hu, Y.M., Li, D.H., and H. P Xue (2012), "*Evaluation of Chinese Adolescent Science and Technology Contests and International Comparative Study*," Beijing: Beijing Normal University Press.

[28] Hyde, J. S. and J. E. Mertz (2009), "Gender, culture, and mathematics performance," *Proceedings of the National Academy of Sciences*, 106(22), 8801-8807.

[29] Jurajda, S. and D. Munich (2011), "Gender gap in performance under competitive pressure: Admission to Czech Universities", *American Economic Review*, 101(3), 514–518.

[30] Knight, J., L. Shi and D. Quheng (2010), "Son preference and household income in rural China," *Journal of Development Studies*, 46(10), 1786–1805.

[31] Krieg, J. and P . Storer (2006), "How much do students matter? Applying the Oaxaca decomposition to explain determinants of adequate yearly progress," *Contemporary Economic Policy*, 24(4), 563- 581.

[32] Kucian, K., T . Loenneker, T. Dietrich, E. Martin and M. von Aster (2005), "Gender differences in brain activation patterns during mental rotation and number related cognitive tasks," *Psychology Science*, 47(1), 112- 131.

[33] Lakin, J. M. (2013), "Sex differences in reasoning abilities: Surprising evidence that male-female ratios in the tails of the quantitative reasoning distribution have increased," *Intelligence*, 41, 263-274.

[34] Loyalka, P., C. James, J. Wei, J. Natalie and R. Joel (2017), "Inequalities in the pathway to college in China: When do students from poor areas fall behind?," *The China Quarterly*, 229, 172–194.

[35] Lundberg, S., and E. Rose (2002), "The effects of sons and daughters on men's labor supply and wages," *Review of Economics and Statistics*, 84, 251–268.

[36] Machin, S. and T. Pekkarinen (2008), "Global sex differences in test score variability," *Science*, 322(5906), 1331-1332.

[37] Makel, M.C., J. Wai, K. Peairs and M. Putallaz (2016), "Sex differences in the right tail of cognitive abilities: An update and cross cultural extension," *Intelligence*, 59, 8-15.

[38] McEwan, P. J. (2004), "The Indigenous Test Score Gap in Bolivia and Chile," *Economic Development and Cultural Change*, 53(1), 157- 90.

[39] Melly, B. (2005), "Decomposition of differences in distribution using quantile regression," *Labour Economics*, 12(4), 577-590.

[40] Melly, B. (2006), *Applied quantile regression* (Doctoral dissertation, Dissertation no. 3255. University of St. Gallen).

[41] Miller D. I. and D. F. Halpern (2014), "The new science of cognitive sex differences," *Trends in Cognitive Sciences*, 18(1), 37–45.

[42] Niederle M. and L. Vesterlund (2007), "Do Women Shy Away from Competition? Do Men Compete Too Much?," *The Quarterly Journal of Economics*, 122(3), 1067-1101.

[43] Nollenberger, N., N. Rodríguez-Planas and A. Sevilla (2016), "The Math Gender Gap: The Role of Culture," *American Economic Review*, 106(5), 257-261.

[44] O'Dea, R.E., M. Lagisz, M. D. Jennions and S. Nakagawa (2018), "Gender Differences in Individual Variation in Academic Grades Fail to Fit Expected Patterns for STEM," *Nature Communications,* 9, 3777.

[45] OECD. (2007). PISA 2006 Science Competencies for Tomorrow's World, PISA. OECD Publishing.

[46] OECD. (2006). Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006, PISA, OECD Publishing.

[47] OECD. (2009). PISA Data Analysis Manual SPSS® 2nd Edition, OECD Publishing.

[48] Ors, E., F. Palomino and E. Peyrache (2013) "Performance Gender Gap: Does Competition Matter?", *Journal of Labor Economics*, 31(3), 443-499.

[49] Romanes, G. J. (1887), "Mental differences between men and women," reprinted in *The Education Papers. Women's Quest for Equality in Britain 1850–1912,* Spender, D. (ed.) 1982, Routledge Press, 10-31.

[50] Rothe, C. and D. Wied (2012), "Misspecification Testing in a Class of Conditional Distributional Models," *Journal of the American Statistical Association*, 108, 314–324.

[51] Song, L. (2001). Gender effects on household resource allocation in rural China. In C. Riskin, Z. Renwei,& L. Shi (Eds.), *China's retreat fromequality: income distribution and economic transition* (pp. 276–302). New York: M.E.

[52] Turner, S. E. and W. G. Bowen (1999), "Choice of Major: The Changing (unchanging) Gender Gap." *Industrial and Labor Relations Review,* 52(2), 289–313.

[53] Xie, Y. and K. A. Shauman (2003), *Women in science: Career processes and outcomes*, Harvard University Press.

[54] Yi, H. M., Y. Q. Song, Ch. F. Liu, X. T. Huang, L. X. Zhang, Y. L. Bai and Y. J. Shi (2015), "Giving kids a head start: The impact of early commitment of financial aid on poor seventh grade students in rural China," *Journal of Development Economics*, 113(3), 1–15.

[55] Yip, D.Y., M. M. Chiu and E. S. C. Ho (2004), "Hong Kong Student Achievement in OECD-PISA Study: Gender Differences in Science Content, Literacy Skills, and Test Item Formats," *International Journal of Science and Mathematics Education*, 2,91-106.

[56] Zeng, J., X. Pang, L. Zhang, A. Medina and S. Rozelle (2014) "Gender Inequality in Education in China: A Meta–Regression Analysis."*Contemporary Economic Policy,* 32(2), 474–491.

[57] Zhang, Y. (2013), "Does private tutoring improve students' National College Entrance Exam performance? — A case study from Jinan, China," *Economics of Education Review*, 32(1), 1–28.

Table 1: Details on the Three Dimensions and Connotations of Scientific Competencies Defined by the OECD

Larger Groupings	Dimensions within the Grouping
Identifying scientific issues (ISI)	Recognizing issues that it is possible to investigate scientifically
	Identifying keywords to search for scientific information
	Recognizing the key features of a scientific investigation
Explaining phenomena scientifically (EPS)	Applying knowledge of science in a given situation
	Describing or interpreting phenomena scientifically and predicting changes
	Identifying appropriate descriptions, explanations, and predictions
Using scientific evidence (USE)	Interpreting scientific evidence and making and communicating conclusions
	Identifying the assumptions, evidence and reasoning behind conclusions
	Reflecting on the societal implications of science and technological developments

Note: This table is a direct reproduction of material provided on OECD (2006), Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006, PISA, OECD Publishing, p.29.

Table 2: Summarizing the Data by Gender

Variable	Girls	Boys	Test of Equality of Means
Family owns a Personal Car	0.272 (0.439)	0.276 (0.442)	-0.231 [0.817]
Family owns a Personal Laptop	0.197 (0.393)	0.195 (0.392)	0.118 [0.906]
Child Age	17.171 (12.620)	17.368 (13.251)	4.982*** [0.000]
Child has a Sibling	0.400 (0.497)	0.443 (0.490)	2.119** [0.034]
Child owns a Cell Phone	0.555 (0.500)	0.515 (0.497)	-1.970** [0.049]
Child has a QQ Account	0.663 (0.473)	0.733 (0.442)	3.742*** [0.0002]
Child's Family is an Ethnic minority	0.029 (0.178)	0.033 (0.168)	0.521 [0.602]
Household income >5000RMB a month	0.348 (0.447)	0.332 (0.448)	0.864 [0.388]
Key School	0.592 (0.492)	0.628 (0.484)	-1.804* [0.071]
Class Size	55.575 (0.341)	55.951 (0.406)	-0.714 [0.476]
Child Resides in a Eastern Province	0.287 (0.453)	0.253 (0.435)	1.889* [0.059]
Child Resides in Central China	0.192 (0.394)	0.179 (0.383)	0.870 [0.384]
Child Resides in a Western Province	0.520 (0.500)	0.569 (0.496)	-2.368** [0.018]
Total Score on the Assessment	59.357 (8.641)	58.939 (8.415)	1.197 [0.231]
Score on ISI subscale	52.112 (13.527)	51.664 (13.013)	1.902* [0.057]
Score on EPS subscale	64.384 (10.326)	63.517 (10.046)	2.080** [0.038]
Score on USE subscale	60.466 (15.198)	61.660 (15.548)	0.856 [0.409]
Observations	1372	1064	

Note: The last column reports the results of a difference in the means in the respective row variable between the genders. T-tests are conducted for test scores and a test of the difference in proportions is conducted for the remaining variables. Standard errors are reported in ( ) brackets and p-values for the associated test are reported in [ ] brackets. ***significant at 1% level, **5% level, *10% level.

Table 3: Conditional Quantile Regression Estimates of the Education Production Function by Gender

Quantile	5	10	20	30	40	50	60	70	80	90	95
Girls Total Score N=1063											
Family owns a Personal Car	-0.33 [2.11]	-1.52 [1.35]	-0.88 [1.08]	-1.23 [0.86]	-2.15*** [0.80]	-1.61** [0.66]	-1.17* [0.64]	-1.45** [0.60]	-1.13* [0.67]	-0.33 [0.78]	0.45 [0.93]
Family owns a Personal Laptop	2.27 [2.02]	1.75 [1.30]	-0.17 [1.04]	-1.13 [0.82]	-1.23 [0.77]	-0.59 [0.63]	-0.75 [0.61]	-0.30 [0.58]	-0.13 [0.65]	0.13 [0.74]	-0.61 [0.89]
Household income >5000RMB a month	-1.63 [2.10]	-0.45 [1.35]	0.58 [1.08]	0.32 [0.86]	1.07 [0.80]	0.49 [0.66]	0.22 [0.63]	-0.06 [0.60]	-0.36 [0.67]	-0.08 [0.77]	0.18 [0.92]
Key School	1.06 [1.84]	3.66*** [1.18]	4.49*** [0.94]	5.49*** [0.75]	5.68*** [0.70]	5.79*** [0.58]	5.57*** [0.56]	5.86*** [0.53]	5.08*** [0.59]	4.73*** [0.68]	4.55*** [0.81]
Child Resides in a Eastern Province	5.79*** [2.03]	5.31*** [1.30]	5.50*** [1.04]	4.93*** [0.83]	4.36*** [0.77]	4.33*** [0.64]	4.01*** [0.61]	3.72*** [0.58]	3.66*** [0.65]	2.33*** [0.75]	2.94*** [0.89]
Child Resides in Central China	2.10 [2.31]	0.45 [1.48]	1.76 [1.18]	1.39 [0.94]	1.12 [0.87]	1.51** [0.73]	1.30* [0.70]	1.45** [0.66]	1.17 [0.74]	1.09 [0.85]	1.73* [1.01]
Child has a QQ Account	2.09 [1.79]	2.86** [1.15]	1.74* [0.92]	1.70** [0.73]	1.87*** [0.68]	1.36** [0.56]	1.18** [0.54]	1.21** [0.51]	0.90 [0.57]	0.50 [0.66]	0.79 [0.79]
Child has a Sibling	-1.32 [1.72]	-1.37 [1.11]	-0.82 [0.88]	-0.50 [0.70]	-0.81 [0.65]	-0.62 [0.54]	-0.74 [0.52]	-0.07 [0.49]	-0.80 [0.55]	-0.78 [0.63]	-0.81 [0.76]
Class Size	-0.16** [0.07]	-0.08* [0.05]	-0.10*** [0.04]	-0.11*** [0.03]	-0.11*** [0.03]	-0.10*** [0.02]	-0.10*** [0.02]	-0.10*** [0.02]	-0.09*** [0.02]	-0.11*** [0.03]	-0.08** [0.03]
Child owns a Cell Phone	-2.84* [1.60]	-1.89* [1.03]	-1.87** [0.82]	-1.66** [0.65]	-1.55** [0.61]	-1.27** [0.50]	-1.16** [0.48]	-0.93** [0.46]	-1.39*** [0.51]	-1.13* [0.59]	-0.95 [0.70]
Child Age	-13.80 [13.25]	-7.64 [8.51]	-4.50 [6.80]	-2.46 [5.41]	-0.87 [5.03]	1.03 [4.17]	3.19 [4.01]	4.43 [3.79]	4.66 [4.24]	7.28 [4.89]	8.56 [5.83]
Child Age Squared/100	37.57 [38.41]	18.86 [24.66]	9.97 [19.72]	4.75 [15.69]	-0.32 [14.57]	-6.18 [12.08]	-12.83 [11.62]	-16.36 [11.00]	-16.56 [12.30]	-23.42* [14.17]	-27.93* [16.89]
Boys Total Score N=1369											
Family owns a Personal Car	-1.09 [1.11]	-1.65 [1.28]	-0.26 [0.97]	-0.57 [0.82]	-0.04 [0.69]	-0.11 [0.75]	-0.01 [0.71]	-0.05 [0.72]	0.33 [0.64]	0.73 [0.97]	0.34 [1.32]
Family owns a Personal Laptop	-1.85 [1.13]	-1.11 [1.30]	-0.06 [0.98]	-0.73 [0.83]	-1.37** [0.70]	-1.25 [0.76]	-1.29* [0.72]	-1.61** [0.73]	-2.02*** [0.65]	-1.55 [0.98]	0.05 [1.34]
Household income >5000RMB a month	2.04* [1.09]	1.30 [1.26]	1.16 [0.95]	1.43* [0.81]	1.65** [0.68]	1.40* [0.74]	1.40** [0.70]	1.70** [0.71]	1.47** [0.63]	1.57* [0.95]	-0.12 [1.30]
Key School	7.11*** [1.05]	7.71*** [1.22]	7.06*** [0.92]	7.26*** [0.78]	7.18*** [0.65]	6.40*** [0.71]	5.88*** [0.67]	5.26*** [0.68]	6.22*** [0.61]	6.20*** [0.92]	7.00*** [1.25]
Child Resides in a Eastern Province	7.48*** [1.08]	7.23*** [1.24]	6.23*** [0.94]	6.23*** [0.80]	5.92*** [0.67]	5.99*** [0.73]	5.42*** [0.69]	5.22*** [0.70]	5.06*** [0.62]	4.92*** [0.94]	5.38*** [1.28]
Child Resides in Central China	0.74 [1.27]	-0.46 [1.46]	-0.10 [1.10]	-0.57 [0.93]	0.59 [0.79]	1.26 [0.86]	1.18 [0.81]	1.65** [0.82]	1.57** [0.73]	1.75 [1.10]	2.01 [1.50]
Child has a QQ Account	-0.38 [0.99]	-0.49 [1.14]	0.78 [0.86]	0.06 [0.73]	-0.56 [0.61]	-0.97 [0.67]	-0.73 [0.63]	0.24 [0.64]	-0.03 [0.57]	-0.26 [0.86]	-0.18 [1.17]

Child has a Sibling	0.47 [0.98]	0.18 [1.13]	-1.59* [0.86]	-1.61** [0.72]	-2.40*** [0.61]	-3.03*** [0.66]	-2.46*** [0.62]	-2.57*** [0.63]	-1.78*** [0.56]	-2.01** [0.85]	-2.03* [1.16]
Class Size	-0.09** [0.04]	-0.06 [0.04]	-0.04 [0.03]	-0.06** [0.03]	-0.06** [0.02]	-0.04 [0.03]	-0.04* [0.02]	-0.03 [0.02]	-0.02 [0.02]	-0.04 [0.03]	-0.07 [0.05]
Child owns a Cell Phone	-3.14*** [0.92]	-2.04* [1.06]	-1.72** [0.80]	-1.06 [0.68]	-0.62 [0.57]	-0.27 [0.62]	0.00 [0.59]	-0.71 [0.60]	-0.17 [0.53]	-1.12 [0.80]	-1.13 [1.09]
Child Age	6.64 [11.79]	14.24 [13.58]	12.95 [10.27]	10.20 [8.69]	1.82 [7.30]	1.29 [7.96]	6.03 [7.49]	0.93 [7.60]	-5.39 [6.77]	-2.12 [10.24]	5.10 [13.97]
Child Age Squared/100	-21.41 [34.21]	-43.66 [39.42]	-39.25 [29.80]	-31.15 [25.22]	-6.88 [21.19]	-5.26 [23.11]	-18.73 [21.74]	-4.20 [22.07]	13.56 [19.66]	5.00 [29.73]	-16.26 [40.54]
Quantile	5	10	20	30	40	50	60	70	80	90	95
Girls ISI Score N=1063											
Family owns a Personal Car	1.04 [3.33]	0.02 [1.98]	-0.78 [1.63]	-0.40 [1.47]	-1.49 [1.51]	-0.72 [1.34]	-0.04 [1.30]	-0.27 [1.39]	-2.05 [1.52]	-0.30 [1.53]	-1.38 [2.23]
Family owns a Personal Laptop	-2.25 [3.19]	0.64 [1.89]	0.42 [1.56]	-2.34* [1.41]	-2.11 [1.44]	-1.48 [1.28]	-1.44 [1.24]	-0.89 [1.33]	0.15 [1.46]	1.23 [1.47]	0.67 [2.13]
Household income >5000RMB a month	-0.67 [3.31]	-0.78 [1.97]	-1.74 [1.62]	-0.16 [1.47]	-0.31 [1.50]	-0.89 [1.33]	-1.29 [1.29]	-1.69 [1.38]	-0.03 [1.52]	0.64 [1.53]	-0.23 [2.22]
Key School	5.82** [2.91]	6.44*** [1.73]	7.61*** [1.42]	7.63*** [1.29]	6.87*** [1.32]	7.47*** [1.17]	7.41*** [1.13]	6.82*** [1.21]	5.86*** [1.33]	4.60*** [1.34]	3.03 [1.95]
Child Resides in a Eastern Province	9.84*** [3.21]	8.03*** [1.90]	8.36*** [1.57]	7.53*** [1.42]	7.72*** [1.45]	5.74*** [1.29]	5.53*** [1.25]	4.49*** [1.34]	4.12*** [1.47]	4.49*** [1.48]	3.95* [2.15]
Child Resides in Central China	-4.70 [3.64]	-0.06 [2.16]	0.71 [1.78]	0.26 [1.61]	2.99* [1.65]	1.74 [1.46]	2.10 [1.42]	0.74 [1.52]	1.52 [1.67]	2.78* [1.68]	2.28 [2.44]
Child has a QQ Account	7.89*** [2.82]	3.86** [1.68]	1.87 [1.38]	1.81 [1.25]	1.10 [1.28]	1.19 [1.13]	1.51 [1.10]	1.20 [1.18]	1.02 [1.29]	1.53 [1.30]	0.23 [1.89]
Child has a Sibling	-1.08 [2.72]	-0.58 [1.61]	-0.71 [1.33]	-1.47 [1.20]	-1.99 [1.23]	-0.43 [1.09]	-0.82 [1.06]	-1.63 [1.13]	-4.10*** [1.24]	-4.07*** [1.25]	-3.54* [1.82]
Class Size	-0.19* [0.11]	-0.16** [0.07]	-0.15*** [0.06]	-0.14*** [0.05]	-0.12** [0.05]	-0.15*** [0.05]	-0.11** [0.04]	-0.12*** [0.05]	-0.09* [0.05]	-0.06 [0.05]	-0.15** [0.08]
Child owns a Cell Phone	-4.83* [2.53]	-3.89*** [1.50]	-2.69** [1.24]	-1.56 [1.12]	-1.93* [1.15]	-1.72* [1.02]	-1.53 [0.99]	-1.50 [1.06]	-1.62 [1.16]	-0.37 [1.17]	0.08 [1.69]
Child Age	-18.67 [20.92]	-15.40 [12.43]	-9.76 [10.24]	-6.10 [9.26]	-6.32 [9.47]	-12.10 [8.41]	-7.70 [8.16]	-4.78 [8.74]	-8.11 [9.57]	-3.64 [9.64]	-17.03 [14.01]
Child Age Squared/100	48.25 [60.63]	40.95 [36.03]	25.11 [29.67]	14.05 [26.84]	14.70 [27.46]	30.76 [24.37]	18.47 [23.66]	10.83 [25.32]	21.18 [27.73]	8.76 [27.93]	46.25 [40.60]
Boys ISI Score N=1369											
Family owns a Personal Car	2.78 [3.47]	-0.70 [2.54]	-1.17 [1.67]	-0.91 [1.49]	-1.97 [1.55]	-0.55 [1.59]	-0.22 [1.64]	1.04 [1.59]	0.95 [1.34]	-2.31 [1.95]	-1.66 [2.15]
Family owns a Personal Laptop	4.00 [3.51]	2.86 [2.57]	0.20 [1.69]	-0.06 [1.51]	0.83 [1.56]	-1.16 [1.60]	-2.54 [1.66]	-4.14*** [1.60]	-2.55* [1.35]	-2.79 [1.98]	-3.13 [2.17]

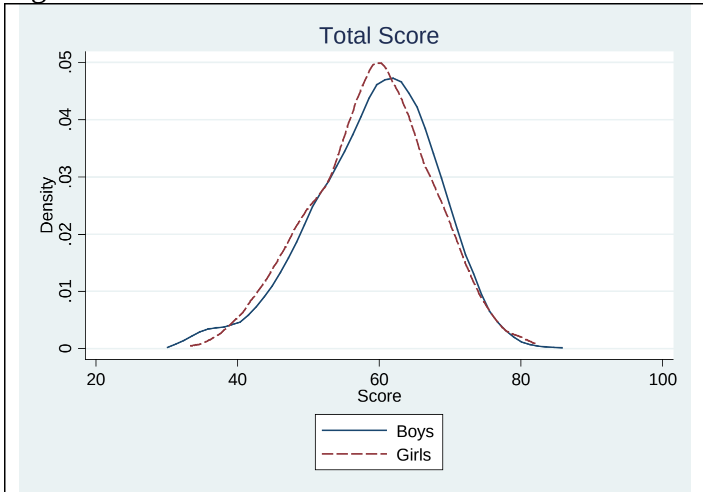
Household income >5000RMB a month	-3.95 [3.41]	1.44 [2.49]	2.27 [1.64]	3.39** [1.46]	4.35*** [1.52]	3.96** [1.56]	4.78*** [1.61]	5.26*** [1.56]	3.93*** [1.31]	4.15** [1.92]	1.52 [2.11]
Key School	10.51*** [3.29]	10.50*** [2.40]	9.67*** [1.58]	9.66*** [1.41]	8.76*** [1.46]	8.95*** [1.50]	8.03*** [1.55]	6.92*** [1.50]	7.93*** [1.27]	8.85*** [1.85]	8.13*** [2.03]
Child Resides in a Eastern Province	8.58** [3.36]	7.83*** [2.46]	5.79*** [1.62]	5.81*** [1.44]	4.98*** [1.50]	5.46*** [1.54]	6.26*** [1.59]	5.87*** [1.54]	6.30*** [1.30]	5.32*** [1.89]	4.40** [2.08]
Child Resides in Central China	-6.06 [3.95]	-1.41 [2.89]	0.32 [1.90]	0.02 [1.69]	-1.38 [1.76]	-1.02 [1.80]	-1.01 [1.87]	1.53 [1.80]	1.64 [1.52]	1.88 [2.22]	1.25 [2.45]
Child has a QQ Account	-3.05 [3.08]	0.63 [2.26]	-0.35 [1.48]	0.06 [1.32]	1.14 [1.37]	1.33 [1.41]	0.06 [1.46]	-0.52 [1.41]	-0.81 [1.19]	-1.04 [1.73]	-2.91 [1.91]
Child has a Sibling	-0.33 [3.06]	-2.38 [2.24]	-3.24** [1.47]	-1.61 [1.31]	-1.92 [1.36]	-2.08 [1.40]	-2.76* [1.45]	-2.51* [1.40]	-1.74 [1.18]	-1.89 [1.72]	-1.66 [1.90]
Class Size	-0.17 [0.12]	-0.13 [0.09]	-0.12** [0.06]	-0.12** [0.05]	-0.12** [0.05]	-0.14** [0.05]	-0.15*** [0.06]	-0.14*** [0.05]	-0.12** [0.05]	-0.09 [0.07]	-0.06 [0.07]
Child owns a Cell Phone	0.06 [2.88]	-0.77 [2.10]	-0.37 [1.38]	-0.69 [1.23]	-0.93 [1.28]	-2.06 [1.31]	-1.87 [1.36]	-1.87 [1.32]	-0.45 [1.11]	1.19 [1.62]	2.05 [1.78]
Child Age	7.81 [36.72]	1.32 [26.87]	15.34 [17.66]	13.87 [15.77]	7.75 [16.35]	9.27 [16.78]	15.93 [17.37]	-1.12 [16.79]	17.45 [14.16]	21.80 [20.66]	24.23 [22.74]
Child Age Squared/100	-17.90 [106.58]	-1.92 [77.99]	-43.27 [51.24]	-41.65 [45.76]	-22.35 [47.45]	-27.71 [48.69]	-45.98 [50.42]	4.10 [48.72]	-47.41 [41.09]	-61.20 [59.96]	-70.47 [66.01]
Quantile	5	10	20	30	40	50	60	70	80	90	95
Girls EPS Score N=1063											
Family owns a Personal Car	-1.85 [3.74]	-0.49 [1.83]	-1.59 [1.19]	-2.72*** [0.88]	-2.28** [0.90]	-2.11*** [0.80]	-1.75** [0.83]	-2.01** [0.84]	-1.61** [0.70]	-1.22 [0.96]	-0.14 [1.18]
Family owns a Personal Laptop	1.42 [3.58]	-1.33 [1.76]	-2.02* [1.14]	0.17 [0.84]	0.07 [0.86]	-0.82 [0.76]	-1.00 [0.79]	0.45 [0.80]	0.53 [0.67]	0.05 [0.92]	0.09 [1.13]
Household income >5000RMB a month	0.20 [3.72]	-0.54 [1.83]	1.49 [1.19]	1.52* [0.88]	1.12 [0.89]	0.74 [0.79]	0.73 [0.82]	0.35 [0.83]	-0.09 [0.70]	-0.07 [0.96]	-0.71 [1.18]
Key School	3.93 [3.26]	5.60*** [1.60]	6.14*** [1.04]	6.15*** [0.77]	5.71*** [0.78]	5.71*** [0.70]	5.41*** [0.72]	5.55*** [0.73]	5.28*** [0.61]	5.17*** [0.84]	5.66*** [1.03]
Child Resides in a Eastern Province	7.46** [3.60]	6.09*** [1.77]	4.95*** [1.15]	4.59*** [0.85]	4.36*** [0.87]	4.67*** [0.77]	4.20*** [0.80]	3.45*** [0.81]	3.41*** [0.67]	2.41*** [0.93]	2.79** [1.14]
Child Resides in Central China	0.42 [4.09]	0.59 [2.01]	0.30 [1.30]	-0.25 [0.96]	0.56 [0.98]	0.81 [0.87]	1.01 [0.91]	0.07 [0.92]	0.20 [0.76]	-0.21 [1.06]	0.32 [1.29]
Child has a QQ Account	4.14 [3.17]	1.73 [1.55]	2.11** [1.01]	0.97 [0.75]	0.90 [0.76]	1.46** [0.67]	1.34* [0.70]	1.91*** [0.71]	1.37** [0.59]	1.25 [0.82]	1.30 [1.00]
Child has a Sibling	0.32 [3.05]	-0.61 [1.50]	-0.29 [0.97]	-0.58 [0.72]	-0.47 [0.73]	0.18 [0.65]	0.17 [0.68]	0.05 [0.68]	-0.36 [0.57]	-1.74** [0.79]	-0.37 [0.97]
Class Size	-0.21* [0.13]	-0.14** [0.06]	-0.12*** [0.04]	-0.12*** [0.03]	-0.10*** [0.03]	-0.10*** [0.03]	-0.08*** [0.03]	-0.09*** [0.03]	-0.06*** [0.02]	-0.08** [0.03]	-0.06 [0.04]
Child owns a Cell Phone	-3.67 [2.84]	-0.72 [1.39]	-1.56* [0.91]	-0.83 [0.67]	-0.62 [0.68]	-0.80 [0.61]	-0.70 [0.63]	-0.85 [0.64]	-0.70 [0.53]	-0.58 [0.73]	-1.50* [0.90]

Child Age	-7.88 [23.51]	-2.20 [11.52]	-9.14 [7.49]	-4.21 [5.54]	-4.87 [5.65]	-0.41 [5.01]	-0.07 [5.20]	3.76 [5.26]	6.03 [4.39]	8.08 [6.06]	3.21 [7.44]
Child Age Squared/100	17.48 [68.13]	1.43 [33.40]	24.15 [21.71]	9.57 [16.06]	11.01 [16.37]	-2.20 [14.51]	-3.55 [15.08]	-15.37 [15.24]	-21.94* [12.73]	-27.86 [17.57]	-14.60 [21.55]
Boys EPS Score N=1369											
Family owns a Personal Car	1.41 [1.69]	-0.78 [1.15]	-0.34 [1.20]	-1.31 [0.97]	-1.20 [0.94]	0.08 [0.93]	0.64 [0.77]	0.42 [0.84]	0.18 [1.12]	0.17 [1.18]	1.19 [1.41]
Family owns a Personal Laptop	-0.24 [1.71]	-0.40 [1.17]	-0.84 [1.22]	-0.74 [0.98]	-1.04 [0.95]	-1.74* [0.94]	-1.57** [0.78]	-1.24 [0.85]	-1.02 [1.14]	-0.41 [1.19]	2.37* [1.43]
Household income >5000RMB a month	-0.07 [1.66]	1.37 [1.13]	1.13 [1.18]	1.41 [0.95]	0.82 [0.93]	0.42 [0.91]	0.46 [0.76]	0.13 [0.82]	0.66 [1.10]	1.67 [1.16]	0.32 [1.39]
Key School	6.77*** [1.60]	7.12*** [1.09]	8.19*** [1.14]	8.23*** [0.92]	7.88*** [0.89]	7.35*** [0.88]	7.40*** [0.73]	7.05*** [0.79]	6.94*** [1.06]	6.52*** [1.11]	6.71*** [1.34]
Child Resides in a Eastern Province	8.25*** [1.64]	8.03*** [1.12]	5.98*** [1.17]	6.13*** [0.94]	6.84*** [0.91]	6.46*** [0.90]	6.38*** [0.75]	5.83*** [0.81]	5.25*** [1.09]	5.39*** [1.14]	5.14*** [1.37]
Child Resides in Central China	2.39 [1.92]	1.21 [1.31]	-1.14 [1.37]	-1.18 [1.10]	0.01 [1.07]	0.73 [1.06]	1.61* [0.88]	1.28 [0.95]	1.21 [1.28]	1.99 [1.34]	2.77* [1.61]
Child has a QQ Account	-1.14 [1.50]	-0.98 [1.02]	0.29 [1.07]	-0.01 [0.86]	0.03 [0.84]	-0.51 [0.83]	-1.01 [0.69]	-1.23* [0.74]	-0.65 [1.00]	-0.67 [1.04]	-0.92 [1.26]
Child has a Sibling	-1.05 [1.49]	-1.99* [1.02]	-1.89* [1.06]	-2.52*** [0.85]	-2.56*** [0.83]	-2.75*** [0.82]	-2.77*** [0.68]	-2.79*** [0.74]	-2.11** [0.99]	-3.29*** [1.04]	-3.43*** [1.25]
Class Size	-0.08 [0.06]	-0.01 [0.04]	-0.04 [0.04]	-0.06* [0.03]	-0.06* [0.03]	-0.06* [0.03]	-0.04 [0.03]	-0.06* [0.03]	-0.03 [0.04]	0.03 [0.04]	0.05 [0.05]
Child owns a Cell Phone	-1.04 [1.40]	-1.24 [0.96]	-1.69* [1.00]	-2.18*** [0.80]	-1.60** [0.78]	-0.40 [0.77]	-0.39 [0.64]	-0.25 [0.69]	-0.45 [0.93]	-0.51 [0.98]	-0.83 [1.17]
Child Age	3.31 [17.90]	9.49 [12.20]	14.65 [12.72]	12.81 [10.23]	1.03 [9.97]	1.37 [9.86]	8.37 [8.19]	10.60 [8.84]	7.36 [11.89]	-11.39 [12.45]	-7.06 [14.96]
Child Age Squared/100	-12.13 [51.95]	-30.86 [35.42]	-45.86 [36.93]	-39.07 [29.68]	-5.29 [28.95]	-6.70 [28.61]	-26.62 [23.78]	-32.33 [25.67]	-22.08 [34.52]	33.09 [36.13]	20.85 [43.42]
Quantile	5	10	20	30	40	50	60	70	80	90	95
Girls USE Score N=1063											
Family owns a Personal Car	-4.00* [2.20]	-4.09** [2.01]	-1.64 [1.82]	-1.41 [1.49]	-1.34 [1.27]	-1.26 [1.00]	0.00 [1.00]	-0.42 [0.89]	-1.06 [1.09]	-0.77 [1.39]	-1.49 [1.43]
Family owns a Personal Laptop	3.82* [2.11]	2.70 [1.93]	-0.72 [1.75]	-1.43 [1.43]	-1.46 [1.22]	-1.75* [0.96]	-2.26** [0.96]	-1.07 [0.85]	-1.45 [1.04]	-1.54 [1.33]	-1.99 [1.37]
Household income >5000RMB a month	2.13 [2.19]	1.55 [2.01]	0.80 [1.82]	-0.60 [1.49]	-0.90 [1.27]	-0.86 [1.00]	-1.10 [1.00]	0.21 [0.89]	0.29 [1.08]	0.97 [1.39]	0.40 [1.42]
Key School	3.59* [1.92]	5.20*** [1.76]	7.34*** [1.59]	9.04*** [1.31]	8.86*** [1.11]	8.69*** [0.88]	8.36*** [0.87]	7.64*** [0.78]	6.86*** [0.95]	7.35*** [1.21]	6.67*** [1.25]
Child Resides in a Eastern Province	6.72*** [2.12]	9.49*** [1.94]	8.83*** [1.76]	8.12*** [1.44]	6.50*** [1.23]	6.14*** [0.97]	4.96*** [0.96]	5.12*** [0.86]	4.29*** [1.05]	4.52*** [1.34]	3.01** [1.37]

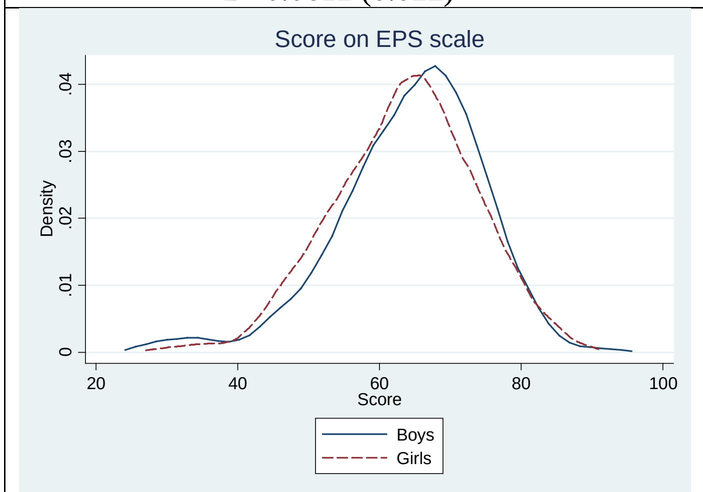
Child Resides in Central China	3.56	2.35	2.22	2.88*	1.08	1.30	1.24	2.40**	1.85	2.10	1.80
	[2.41]	[2.20]	[2.00]	[1.64]	[1.40]	[1.10]	[1.09]	[0.98]	[1.19]	[1.52]	[1.56]
Child has a QQ Account	0.88	1.27	2.45	2.84**	1.22	1.33	1.50*	0.60	1.64*	0.98	-0.47
	[1.87]	[1.71]	[1.55]	[1.27]	[1.08]	[0.85]	[0.85]	[0.76]	[0.92]	[1.18]	[1.21]
Child has a Sibling	-0.09	-2.06	-2.36	-1.97	-0.80	-1.03	-0.78	-0.71	-1.40	-0.54	-1.38
	[1.80]	[1.65]	[1.49]	[1.22]	[1.04]	[0.82]	[0.82]	[0.73]	[0.89]	[1.14]	[1.16]
Class Size	-0.16**	-0.16**	-0.17***	-0.18***	-0.15***	-0.12***	-0.10***	-0.12***	-0.18***	-0.15***	-0.22***
	[0.07]	[0.07]	[0.06]	[0.05]	[0.04]	[0.03]	[0.03]	[0.03]	[0.04]	[0.05]	[0.05]
Child owns a Cell Phone	-1.87	-4.69***	-4.02***	-3.72***	-2.35**	-1.72**	-2.01***	-1.89***	-2.30***	-1.47	-1.37
	[1.67]	[1.53]	[1.39]	[1.14]	[0.97]	[0.76]	[0.76]	[0.68]	[0.83]	[1.06]	[1.08]
Child Age	-14.82	-6.86	-1.37	2.37	4.82	6.09	8.15	10.69*	10.41	12.68	18.70**
	[13.84]	[12.67]	[11.47]	[9.40]	[8.02]	[6.30]	[6.29]	[5.61]	[6.84]	[8.75]	[8.97]
Child Age Squared/100	38.83	14.19	-1.68	-11.00	-18.33	-21.38	-27.61	-35.08**	-32.69*	-39.95	-58.00**
	[40.12]	[36.71]	[33.24]	[27.24]	[23.23]	[18.26]	[18.22]	[16.26]	[19.83]	[25.35]	[25.99]
Boys USE Score N=1369											
Family owns a Personal Car	-4.30**	-1.05	-1.12	-0.19	0.58	0.53	-0.05	0.27	-0.42	0.09	0.33
	[2.14]	[2.39]	[1.52]	[1.45]	[1.23]	[1.20]	[1.14]	[1.01]	[1.27]	[1.43]	[1.62]
Family owns a Personal Laptop	-2.75	-2.95	-1.71	-1.96	-2.69**	-2.52**	-2.97**	-3.03***	-1.91	-0.93	-3.01*
	[2.16]	[2.42]	[1.54]	[1.46]	[1.24]	[1.22]	[1.15]	[1.02]	[1.29]	[1.44]	[1.64]
Household income >5000RMB a month	0.92	-0.15	2.52*	1.44	2.42**	1.98*	3.04***	3.41***	2.69**	0.35	0.78
	[2.10]	[2.35]	[1.49]	[1.42]	[1.21]	[1.18]	[1.12]	[0.99]	[1.25]	[1.40]	[1.59]
Key School	11.72***	13.09***	11.47***	10.63***	8.89***	7.96***	7.42***	7.57***	7.36***	7.95***	7.13***
	[2.02]	[2.27]	[1.44]	[1.37]	[1.16]	[1.14]	[1.08]	[0.95]	[1.21]	[1.35]	[1.53]
Child Resides in a Eastern Province	12.01***	12.77***	10.32***	7.88***	6.84***	7.36***	6.36***	6.02***	5.83***	8.94***	9.56***
	[2.07]	[2.32]	[1.47]	[1.40]	[1.19]	[1.17]	[1.10]	[0.98]	[1.24]	[1.38]	[1.57]
Child Resides in Central China	-1.83	-1.18	-1.83	-1.90	-0.56	0.16	0.47	0.55	1.95	2.24	3.82**
	[2.43]	[2.72]	[1.73]	[1.65]	[1.40]	[1.37]	[1.29]	[1.15]	[1.45]	[1.62]	[1.84]
Child has a QQ Account	1.54	-0.51	-0.28	0.54	0.04	-0.66	-0.99	-1.49*	-0.38	-0.30	1.08
	[1.90]	[2.13]	[1.35]	[1.29]	[1.09]	[1.07]	[1.01]	[0.89]	[1.13]	[1.27]	[1.44]
Child has a Sibling	2.22	0.20	-1.04	-2.03	-2.65**	-3.21***	-2.73***	-2.47***	-2.38**	-1.33	-2.65*
	[1.89]	[2.11]	[1.34]	[1.28]	[1.08]	[1.06]	[1.00]	[0.89]	[1.12]	[1.26]	[1.43]
Class Size	-0.12*	-0.15*	-0.07	-0.08*	-0.04	-0.05	-0.05	-0.08**	-0.06	-0.10**	-0.03
	[0.07]	[0.08]	[0.05]	[0.05]	[0.04]	[0.04]	[0.04]	[0.03]	[0.04]	[0.05]	[0.06]
Child owns a Cell Phone	-4.63***	-1.80	-1.31	-1.15	-0.88	-0.67	-0.09	-0.91	-1.33	-1.38	-2.98**
	[1.77]	[1.98]	[1.26]	[1.20]	[1.02]	[1.00]	[0.94]	[0.83]	[1.06]	[1.18]	[1.34]
Child Age	-3.53	6.87	13.30	11.59	3.23	-11.99	-18.61	-15.54	-19.87	-21.54	3.58
	[22.60]	[25.32]	[16.08]	[15.31]	[12.99]	[12.72]	[12.04]	[10.66]	[13.49]	[15.11]	[17.11]
Child Age Squared/100	7.89	-23.85	-41.92	-36.94	-11.55	33.04	51.76	44.14	56.54	57.39	-13.16
	[65.61]	[73.48]	[46.69]	[44.44]	[37.70]	[36.92]	[34.94]	[30.94]	[39.15]	[43.85]	[49.67]

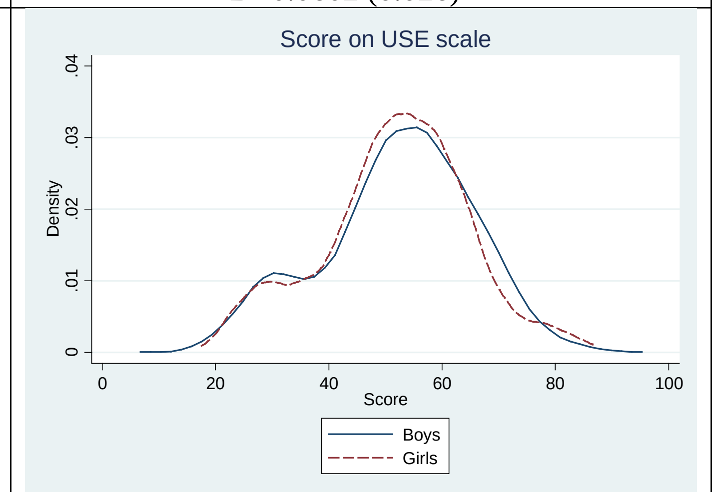
Note: Each equation also includes the full list of covariates summarized in Table 2 including ethnic minority and a constant. Standard errors in parentheses. . ***significant at 1% level, **5% level, *10% level.

Figure 1: Distribution of Student Performance on Components of the Examination by Gender



D=0.0612 (0.022)**	D=0.0602 (0.026)**





D=0.0764 (0.002)***	D=0.0502 (0.098)*

Note: The sample includes all students who wrote the examination. Below each graph is results from a Kolmogorov–Smirnov tests of the equality of distributions between the genders. D indicates the maximum of the tests statistics associated with the directional hypotheses associated with the test and the exact p-value is presented in parentheses. ***significant at 1% level, **5% level, *10% level.

Figure 2: Decomposing the Gender Gap in Student Performance on the Full Examination and Each Component

Note: Each panel decomposes the gap in the respective score presented as a thick solid line into components as described in equation (7).

Appendix Table 1: Oaxaca-Blinder Estimates of the Gender Gap in Student Performance on the Full Examination and Each Component

Outcome	Total Score	ISI Score	EPS Score	USE Score
Overall Difference at the Means $\overline{A}^{m} - \overline{A}^{f}$	0.418 (0.349)	-1.193* (0.631)	0.867** (0.416)	0.449 (0.542)
Components of the Decomposition				
Difference in Coefficients $(\hat{\beta}^{m} - \hat{\beta}^{f})\, \overline{X}^{f}$	0.359 (0.308)	-1.258** (0.595)	0.816** (0.381)	0.379 (0.491)
Difference in Characteristics $(\overline{X}^{m} - \overline{X}^{f})\, \hat{\beta}^{m}$	0.014 (0.193)	-0.035 (0.259)	-0.002 (0.208)	0.011 (0.261)
Interaction $\{E(X^{m})-E(X^{f})\}'(\hat{\beta}^{m} - \hat{\beta}^{f})$	0.044 (0.076)	0.098 (0.131)	0.053 (0.090)	0.059 (0.104)

Note: For each component of the decomposition, clustered standard errors at the school level are contained in parentheses. ***significant at 1% level, **5% level, *10% level.