Economics of Education Review 29 (2010) 383–399

Contents lists available at ScienceDirect

Economics of Education Review

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

Are boys left behind? The evolution of the gender achievement gap in Beijing's middle schools

Fang Lai *

Humanities and Social Sciences, Steinhardt School of Culture, Education, and Human Development, New York University, 309W Kimball Hall, 246 Greene Street, New York, NY 10003, USA

ARTICLE INFO

Article history:
Received 12 February 2008
Accepted 30 July 2009

JEL classification:
I20
J11
J16

Keywords:
Gender achievement gap
Quartile regression
Missing data

ABSTRACT

Using one cohort of 7235 middle school students in Beijing, China, we examined the evolution of the gender achievement gap in middle school. Our study found a more significant female dominance than in U.S. studies: even though boys gradually caught up during middle school, especially in Math and Science, and the gender achievement gap decreased over the distribution of test scores, girls outperformed boys throughout primary and middle school and in each quartile of the performance distribution. As well, girls had a more positive school experience than boys, and boys had a higher dropout rate by the end of middle school. Despite significant gender differences in various important characteristics that have explained the gender achievement gap in the U.S., in our study, primary school test scores seemed to be the only significant source of the gender achievement gap at the end of middle school, indicating the importance of early intervention.

1. Introduction

The gender gap in academic achievement is an important issue to explore, not only because the gender achievement gap itself is an important aspect of educational inequality, but also because it is closely related to the gender wage gap in the labor market. Evidence has been found of significant links between academic achievement in school and success at a later educational stage, as well as future labor market status (e.g., Bedard & Ferrall, 2003; Christie & Shannon, 2001[1]; Rose, 2006). In particular, in China and many other Asian countries, Confucian ideology was distilled into their cultures and the merit-basis test system for career promotion embodied in this ideology has a history spanning thousands of years. As a result, although career promotion in China no longer depends solely on merit-basis exams, the test-oriented education system is still prevalent, and school performance is an essential criterion in the selection process for academic promotion, which, in turn, is a vital indicator of future job market placement. Therefore, the gender achievement gap is an important issue to examine under such context, and China is a case in point.

The gender achievement gap has been examined and explained in detail in the U.S. and many other Western countries, yet virtually no rigorous studies of the gender achievement gap have been conducted in China because of the limited availability of China's secondary education data. This paper is one of the first attempts to provide a statistical account of the gender achievement gap in China using a detailed data set of one cohort of 7235 middle school students in Beijing's Dongcheng District, and compare the Chinese evidence with data from the existing literature, mostly from Western countries such as the U.S. and Great Britain. It also explores various explanations for the gender gap.

* Tel.: +1 212 992 9371/998 5062; fax: +1 212 995 4832.
E-mail address: fang.lai@nyu.edu.

[1] This paper did not find a significant link between the gender gap in educational attainment and the gender wage gap, but it did find gender differences in the field of study of postsecondary graduates as an important determinant of the gender gap in earnings. Moreover, projections suggest that the ongoing improvement of women's attainment relative to men has decreased the gender gap in earnings and will continue to do so.

0272-7757/$ – see front matter
doi:10.1016/j.econedurev.2009.07.009

Many studies have shown significant gender achievement gaps, with boys generally outperforming girls in Math and Science and girls excelling at literacy subjects (e.g., AAUW, 1998; NCES, 2004; OECD, 2004). In addition, data have shown that the gender gaps in favor of girls at different educational levels increase over time (e.g., Card & Lemieux, 2000; Cho, 2007; Goldin, Katz, & Kuziemko, 2006).[2] Some recent studies have also demonstrated that girls receive better report cards in nearly all subjects, while boys demonstrate better performance on standardized tests such as the SAT and ACT, especially in Math and Science (e.g., NCES, 2004). Researchers have attempted to explain this pattern by suggesting, for example, that family background, school organization, and classroom factors such as classroom diversity, teacher qualification, and instructional practices might affect the gender gap (Tansel, 2002; Tinklin, 2003); that boys' mathematical and logical abilities develop more quickly than girls after a certain age; that girls strategically choose a less challenging curriculum (e.g., Pallas & Alexander, 1983); that girls' superior non-cognitive capabilities help them achieve better school report cards (e.g., Duckworth & Seligman, 2006; Jacob, 2002[3]; Tinklin, 2003); and that boys have superior abilities for dealing with standardized tests and complex problems (e.g., Bolger & Kellaghan, 1990). Parents' involvement in their child's study also differs among boys and girls, yet the influences of various forms of parental involvement on the gender achievement gap are ambiguous because of the complexity of parent–child interactions (e.g., Muller, 1998; Tinklin, 2003).[4] This present research intends to examine whether the gender achievement gap revealed by this Chinese data set is consistent with existing evidence in the U.S. and other developed Western countries. As well, it hopes to provide possible explanations for this observed gender gap.

The present results show that since graduation from elementary school, girls dominated boys in overall academic performance, and the within-school gender gap was larger during middle school than it was in primary school. Even though the gender achievement gap favoring girls decreased over time and over test score distribution, boys on all quartiles of this distribution demonstrated inferior overall performance and were not able to catch up with girls until the High School Entrance Exam (HSEE) at the end of middle school. In particular, significantly more boys dropped out of the regular public school system by the end of middle school and did not take the HSEE at all, which might partly explain the closing of the gender gap as measured by the HSEE. In general, the gender gaps disclosed in this study are closer to British research findings, which show girls' consistent dominance in educational achievement over time (e.g., Machin & McNally, 2005), than they are to the U.S. findings, which show that boys generally excel at Math and Science and girls excel at literacy subjects (e.g., NCES, 2004). We also find in the present data set that girls generally receive more parental care, have more cooperative attitudes toward schooling, and exert more effort in their studying. However, primary school performance and affiliation seem to be the only significant contributor to the gender achievement gap, and a substantial proportion of the gender achievement gap remains unexplained by the information in this data set. Therefore, intervention to decrease the gender gap should be implemented at an earlier stage than in middle school. Moreover, additional research with samples in different contexts and using more precise measurements is necessary, especially as the nature of the gender achievement gap in this study has raised concerns about the relatively disadvantaged status of boys in China's test-oriented education system.

The structure of this paper is as follows. Section 2 introduces the background of Chinese middle school education and data. Section 3 describes the evolution of the gender achievement gap over the middle school period, compares the gender achievement gap at the end of middle school using three measures of tests—the HSEE, the Middle School Graduation Exams (MSGE), and the fifth semester tests—and explains the discrepancy in gender gap estimates using these three measures. Section 4 explores possible sources of the gender achievement gap, and Section 5 concludes the paper.

2. Background and data

This study used census and administrative data on a cohort of 7235 students who entered public middle schools in Beijing's Dongcheng District in 1999 and graduated in 2002. The Dongcheng District spans 24.7 km^2 in the east central part of Beijing. As the second largest precinct in the old city section, it has the fourth highest population density among all districts and counties of Beijing. It also encompasses important Beijing sites such as the Forbidden City and Tian'anmen Square, local and national governmental institutes, and popular commercial areas. Its GDP per capita was 30,517.24 Yuan ($3690.114) in 2002. The residents come from diverse socioeconomic backgrounds, and many are employed in the district's two major industries: commerce and service. In sum, then, the Dongcheng District is a most important district of Beijing, and largely represents the metropolitan areas of China's developed regions in overall demographic and socioeconomic composition, education, and economy.

As of 1998, students in China enter school at the age of 6 and spend 6 years in primary school. Upon graduating from

[2] The first paper found that the gender gap favoring women in college attendance increased over the past three decades in the U.S; the second focused on the gender gap in high school performance and how it translated into the gender gap in college enrollment; the third focused on the narrowing and reversal of the gender gap in college attendance and graduation in the U.S. since the 1950s, and explored possible sources for this trend.

[3] Jacob (2002) found that nearly 90% of the gender gap among students attending higher education can be explained by higher non-cognitive skills and college premiums among women. Non-cognitive factors influence college enrollment even after controlling for high school achievement.

[4] For example, Tinklin (2003) shows that parents pay more attention to high-achieving girls than high-achieving boys, and the reverse pattern is found for low-achieving students. She also finds that the gender gap among high attainers is positively related to the frequency of parents' school visits. Muller (1998) similarly found that the forms and focus of parental involvement differed between boys and girls, yet there was no evidence of differences between boys and girls in the relationship between involvement and performance.

primary school, they take an elementary school graduation examination administered by each individual elementary school in two major subjects, Chinese and Math. However, the test scores from the elementary school graduation examination do not factor into the students' middle school placements since they depend on randomized lotteries for school entry. Students then spend 3 years (six semesters) in public middle schools, which are mixed-sex schools. At the end of each semester, students take semester tests in all subjects learned during that semester. To graduate from middle school, students must pass the Middle School Graduation Examination (MSGE) in five major subjects: Chinese, Math, English, Physics, and Chemistry; the first three subjects are taught throughout the 3 years of middle school, while the last two are introduced in the third and fifth semesters, respectively. Graduation from middle school signifies the completion of China's 9-year compulsory education. In China, there are two tracks in the public high school system: regular high school and vocational high school. If students decide to proceed to regular high school after graduating from middle school, they are required to take the High School Entrance Exam (HSEE) in the same five subjects that are tested by the MSGE, though at a more difficult level in order to sort students across public high schools according to their aptitudes. Test scores on the HSEE are essential determinants of these regular high school placements. Students who do not intend to enter regular public high schools do not need to take the HSEE. In China, the public vocational high school system and the private or semi-private high school education system, though emerging and growing quickly, are still greatly under-developed. Thus, regular public high schools are by far the first choice of students who desire a well-rounded high school education rather than a vocational or private school education,[5] and students who seek quality high school education also take the HSEE. Moreover, students in regular high schools also have a much higher chance of promoting to higher education and have more promising future labor market prospects, compared to their counterparts in vocational high schools. In fact, passing the National College Entrance Exam would be nearly impossible without regular high school training, which also makes the HSEE essential in promoting students to higher education. However, only 50–60% of middle school graduates can be admitted by regular high schools. Most other students have to go to vocational high schools.

Semester tests, the MSGE, and the HSEE are uniform across the entire district. Grading of the semester tests and MSGE is administered by the individual schools. These exams are graded by the teachers who instruct the examinees, while the HSEE is graded by a centralized grading committee assigned by the District Education Bureau.

The data set used in this study combined two data sources on the students who entered the public middle schools in Beijing's Dongcheng District in 1999: census data collected by the District Education Bureau in 2002 and school administrative data. A total of 7004 out of 7235 students and their parents provided valid responses to the 2002 census of middle school students and parents. The census data contain detailed information on the students' individual characteristics and family backgrounds, as well as students' and parents' responses to questions about their experiences and attitudes toward schooling and education. The administrative records contain the students' middle school and primary school affiliations; more importantly, they contain the students' test scores from the elementary school graduation exams, all semester tests during middle school, the MSGE, and the HSEE. Because of administrative errors and technical difficulties in combining data from different sources, 5834 students had available test scores on the elementary school graduation exam in at least one subject; 6065 students had test scores for at least one semester; and 3852 students had MSGE scores. As for availability of HSEE scores, in addition to the aforementioned reasons, students who chose not to take the HSEE constituted another important reason for the lack of HSEE scores; thus, only 4586 HSEE scores were available. Moreover, as all middle schools busily prepare for the MSGE and HSEE during the sixth semester of middle school, most middle schools are not serious about administering the sixth semester tests or recording the scores. Therefore, the number of sixth semester test score records available is too small to provide any meaningful empirical evidence and are thus not included in the subsequent analysis. Problems caused by the limited availability of test score data will be considered in the subsequent analysis.

This data set contains unusually rich information on individual students, which facilitate the estimation and explanation of how gender achievement gaps have evolved over the years. Moreover, some empirical evidence in existing literature using U.S. data demonstrates that girls' strategic choice of easier curriculum compared with boys, especially in science subjects, might be a significant source of the gender achievement gap; because China implements uniform curricula and teaching plans for major subjects throughout all the stages of compulsory education, the use of Chinese data can rule out this possibility. Because the data set only contains information in a single cohort of the students in a single district in Beijing, the external validity of this study's conclusions might in fact be limited. However, this district is still worth examining because it is a most important district of Beijing and represents typical metropolitan areas in China's developed regions. Furthermore, given the extremely limited publication of test score data on Chinese students and the consequently small number of rigorous empirical studies on the gender achievement gap in China, evidence from the present study provides extremely valuable insights. In particular, in a country as large as China, issues of educational inequality, especially gender inequality, are distinctively different in well-developed metropolitan areas and less-developed regions; thus, it is worthwhile to examine different contexts separately and explore possible causes and solutions appropriate to each individual context. This study not only provides insights into the gender achievement gaps in

[5] As the vocational high school system was rather underdeveloped compared to the regular high school system in 2002, even students who desire serious vocational development would choose the regular high school track (they could enter a good vocational college after 3 years of regular high school education).

China's metropolitan area, but also takes an essential step toward stimulating future research in this area that can employ samples representing other subgroups of the population or even the general population.

3. The evolution of the gender gap

3.1. Estimation models

We basically estimated the gender achievement gap by regressing student academic performance in each semester on gender; thus, the magnitude and sign of the coefficient of the gender indicator imply the magnitude and direction of the gender performance gap. The changes in the gender gaps estimated over the 3 middle school years covered in this study depict the evolution of the gender gap during middle school.

Middle school academic performance was measured by three indicators: scores on semester tests, scores on the MSGE, and scores on the HSEE. For each test, the basic model is

$$y_{is} = \alpha + \beta \cdot female_i + \eta_s + \varepsilon_{is} \tag{1}$$

where y_{is} indicates the relevant test score of student i attending middle school s, which can be student test score in each semester in each subject as well as average score of all subjects for that semester. It can also be the test scores on the HSEE and MSGE, either in an individual subject or as the average score of the five major subjects. $female_i$ is the gender indicator equal to 1 if student i is a girl, and 0 if a boy. η_s are the school fixed effects to control for systematic difference in grading across schools,[6] and ε_{is} represent idiosyncratic error terms.[7] Thus, β is the estimate of the within-school gender achievement gap. This model was first applied for elementary school graduation tests, in which case η_s are the elementary school fixed effects and β measures the gender achievement gap within each elementary school; the model was then applied to each semester test on the five major subjects, the MSGE test scores in the five major subjects, and the HSEE scores in the same subjects. By comparing test scores across semesters, we can observe the evolution of raw gender achievement gaps in all subjects from elementary school graduation to the end of middle school.

In addition to the measure of the raw gender achievement gap, we sought to control for elementary school performance in order to examine the marginal gender gap evolution after students entered middle school. Thus, we implemented the following value-added model

$$y_{ies} = \alpha + \beta \cdot female_i + \theta \cdot E_{ie} + \gamma \cdot y_{0ies} + \delta \cdot y_{0ies} \cdot E_{ie} + \eta_s + \varepsilon_{ies} \tag{2}$$

where E_{ie} is the indicator of the elementary school attended by student i, which is equal to 1 if student i attended elementary school e, and 0 if not. $\theta \cdot E_{ie}$ captures the fixed effect of the elementary school on student performance measures. y_{0ies} is the elementary school graduation test score in the corresponding subject of student i who attended elementary school e; the interaction term $\delta \cdot y_{0ies} \cdot E_{ie}$ captures the systematic difference in the elementary school performance or the grading of elementary school graduation examinations, which affects the slope of the test score evolution from the graduation from elementary school to the time when the test on which the student i received score y_{ies} took place. Thus, the interpretation of β is different from Model 1: it estimates the incremental within-school gender achievement gap in the relevant semester compared with the gender gap at elementary school graduation. That is, the change in the gender achievement gap as estimated by Model 2 is solely attributable to the middle school experience.

To further capture a systematic difference in grading that is not fixed within the same middle school and thereby improve the precision of the gender gap estimates, we additionally controlled for the first semester test score and the interaction between the first semester test score and the middle school fixed effects. Thus, the model becomes[8]

$$y_{is} = \alpha + \beta \cdot female_i + \eta_s + \gamma \cdot y_{0is} + \delta \cdot y_{0is} \cdot \eta_s + \varepsilon_{is} \tag{3}$$

where y_{0is} is the first semester test score of student i who attended school s. The interaction term $\delta \cdot y_{0is} \cdot \eta_s$ captures the systematic difference in grading and performance that affects the slope of the test score evolution since the completion of the first semester. Again, this is a value-added model, and β estimates the additional within-school gender achievement gap in the relevant semester, compared to the first semester.

As existing studies show, the relative performance of boys and girls might differ over different academic performance levels, and the students' socioeconomic status may often have differential effects on their school performance as well. Thus, we adjusted these models to analyze the heterogeneity in gender achievement gaps. Quartile regressions were used to explore heterogeneity in the gender gap at different quartiles of the test score distributions; the heterogeneous effects of student socioeconomic status on gender achievement gaps were explored by adding the interactions of the gender indicator and the socioeconomic status variables to the basic form of the regressions.

3.2. Empirical results of the pattern of the gender gaps

Table 1 shows the estimated gender gap in overall achievement as well as for each individual subject around the beginning and end of middle school (or the beginning and end of the instruction of the relevant subjects). The row

[6] The school fixed effects also capture the common shocks at the school level, such as unobserved school characteristics.

[7] We also considered heteroscadasticity and the various serial correlations among the error terms such as clustering within classes, and obtained the robust variance estimators and the GLS classroom random-effects estimators, which were similar to the presented results.

[8] We also implemented the more common form of value-added models that used the immediate previous semester performance, instead of the first semester performance, as the baseline score to the current semester performance; the results were similar to the one used in this model, and will be briefly reported in the subsequent sections.

Table 1
Estimates of the gender achievement gaps over the years.

		Overall		Math		Chinese		English		Physics		Chemistry		Number of observations	Control		
		Estimate	*p*-Value	Estimate	*p*-Value	Estimate	*p*-Value	Estimate	*p*-Value	Estimate	*p*-Value	Estimate	*p*-Value		School fixed effects	Primary school test score	First semester test score
Dependent variable: the test score of each subject below																	
HSEE	(1)	0.17***	[0.00]	−0.04	[0.15]	0.52***	[0.00]	0.39***	[0.00]	−0.12***	[0.00]	0.00	[0.87]	4477	Y		
	(2)	0.06*	[0.05]	−0.08**	[0.02]	0.41***	[0.00]	0.10***	[0.00]	−0.22***	[0.00]	−0.10***	[0.00]	3491	Y	Y	
	(3)	−0.11***	[0.00]	−0.14***	[0.00]	0.26***	[0.00]	0.26***	[0.00]	−0.22***	[0.00]	−0.05**	[0.04]	3368	Y		Y
MSGE	(1)	0.28***	[0.00]	0.18***	[0.00]	0.48***	[0.00]	0.41***	[0.00]	0.02	[0.0 9]	0.05*	[0.09]	3767	Y		
	(2)	0.16***	[0.00]	0.15***	[0.00]	0.31***	[0.00]	0.30***	[0.00]	−0.08**	[0.70]	0.01	[0.70]	3090	Y	Y	
	(3)	0.02	[0.38]	0.10***	[0.81]	0.18***	[0.81]	0.15***	[0.81]	−0.05*	[0.09]	−0.01	[0.81]	3110	Y		Y
Fifth semester	(1)	0.27***	[0.00]	0.15***	[0.00]	0.53***	[0.00]	0.15***	[0.00]	0.15**	[0.0 8]	0.15***	[0.00]	5060	Y		
	(2)	0.15***	[0.00]	0.12***	[0.00]	0.33***	[0.00]	0.25***	[0.00]	−0.05**	[0.04]	0.04	[0.14]	4182	Y	Y	
	(3)	−0.06***	[0.00]	0.03	[0.14]	0.17***	[0.00]	0.02	[0.21]	−0.17***	[0.00]			4739	Y		Y
3rd semester	(1)									0.15***	[0.00]			4999	Y		
	(2)									0.00	[0.87]			4200	Y	Y	
First semester	(1)	0.38***	[0.00]	0.15***	[0.00]	0.52***	[0.00]	0.41***	[0.00]					5331	Y		
	(2)	0.25***	[0.00]	0.12***	[0.00]	0.31***	[0.00]	0.30***	[0.00]					4448	Y	Y	
Elementary	(1)	0.19***	[0.00]	0.05*	[0.06]	0.36***	[0.00]							5629	Y		

Each row reports the estimated gender achievement gap in the subject indicated by the corresponding column title on the test indicated by the corresponding row title, controlling for the variables indicated in the "control" panel. When controlling for the elementary graduation test scores in the corresponding subjects, the middle school fixed effects, elementary school fixed effects, and the interactions between the elementary school dummies and the elementary school graduation test scores are also included. When controlling for the first semester test scores of the corresponding subjects, both the middle school fixed effects and the interactions between the middle school dummies and the first semester test scores are also included. Physics is first introduced in the third semester, and Chemistry is first introduced in the fifth semester. Thus, the test scores of the semester in which the subjects were first introduced were controlled in Model 3.

* Significant at 10%.

** Significant at 5%.

*** Significant at 1%.

titles indicate the relevant tests, and the column titles indicate the relevant statistics as well as variables controlled in each regression, as presented by the corresponding row. Row 1 of each test reports the estimated gender gap from Model 1, while rows 2 and 3 report the estimated gender gap using the value-added Models 2 and 3, respectively. All test score measures are normalized so that the coefficients indicate units of standard deviations from the school mean test scores.

The first column shows that once students graduate from elementary school, the within-elementary school gender gap is 0.19 standard deviations in favor of girls, which is significant at the 0.01 level. Throughout the middle school period and within each middle school, girls have significantly higher average performance across all subjects than boys, and the gender gap is larger here than when they graduated from elementary school, as shown by the significant positive coefficients in rows 1 and 2 for all middle school tests. However, there seems to be no significant increase of the within-school gender gap in overall performance after the first semester of middle school; instead, there is a significant decrease, compared with the first semester, in this gender gap by the end of middle school (significantly negative coefficients of row 3 for the fifth semester test and HSEE test). We also implemented a more common form of value-added model using the current semester score as the dependent variable and the previous semester performance, instead of the first semester performance, as the baseline score (the results are not reported in the table). We found no significant change in the gender achievement gap favoring girls from one semester to the next, except for a weakly significant decrease in the fifth semester. Additionally controlling for students and parents' demographic and socioeconomic status did not significantly change the results.

The subsequent columns show the estimated within-school gender achievement gaps of the three subjects (Math, Chinese, English) taught throughout the 3 years of middle school, as well as two science subjects (Physics, Chemistry) introduced in the third and fifth semester, respectively. As English, Chemistry, and Physics were not among the test subjects required for elementary school graduation, there were no test scores in these subjects at the elementary level, and row 2 for all tests in these three subjects reports estimates from the value-added Model 2 controlling for overall elementary school performance as a proxy of academic background upon entering middle school. The coefficients of the female indicator show that for English and Chinese, the girls' performance was significantly better than the boys'; moreover, the significant positive coefficients of rows 2 and 3 for each test indicate that the gender gaps increased, compared with the elementary school and the first semester of middle school. For Math, however, even though the girls performed better than the boys over the middle school period until the administration of the HSEE, and for most semesters, the gender gap favoring girls was larger during middle school than when they graduated from elementary school, the gender gap in the fifth semester was not significantly different from that in the first semester. Moreover, the gender gap in Math measured by the HSEE was insignificantly negative and significantly smaller than that measured by the first semester Math test.

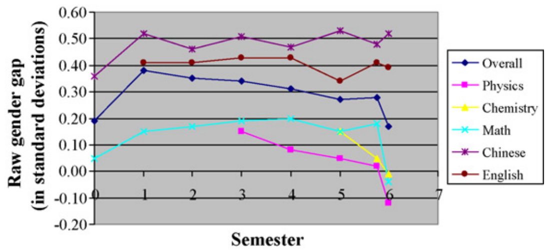

Fig. 1. Raw gender achievement gaps in all subjects during middle school.

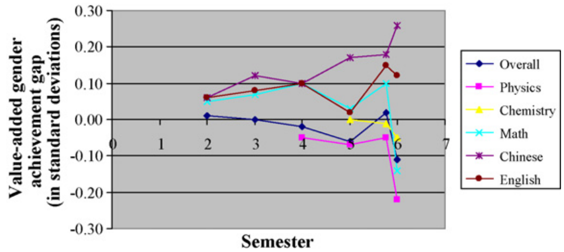

Fig. 2. Value-added gender achievement gaps in all subjects during middle school compared to the first semester.

This pattern of decreased gender gap is more evident for Physics and Chemistry. Physics is introduced in the third semester of middle school, and rows 2 and 3 for all Physics tests show that the gender gap decreased, compared with both the gender gap in baseline academic background proxied by average test scores on elementary graduation tests as well as the gender gap measured by the third semester Physics score (i.e., the first available Physics score), respectively. The raw gender gap in Physics was even reversed on the HSEE, with the boys' performance on average exceeding the girls' by 0.12 standard deviations. Chemistry is introduced only in the fifth semester; we observed that the gender gap in Chemistry performance favoring girls decreased rapidly over the year to an insignificant raw gender gap, as measured by HSEE performance.

Figs. 1 and 2 provide an intuitive overview of the trends of gender gap evolution in all subjects, as well as of overall performance over the 3 years of middle school. The y-axis indicates the estimated gender gap in units of standard deviation from the mean test score, and the x-axis indicates the semester index. Fig. 1 presents the raw gender gaps, and Fig. 2 presents the value-added gender gaps compared with the first semester performance from Model 3. According to the actual length of the interims[9] among the fifth semester tests, the MSGE, and the HSEE, we indexed the time of the MSGE as 5.75 and the HSEE as 6. In general, the gender gap in academic achievement increased compared with the elementary school (indexed by 0 on the x-axis). Within the 3 years of middle school, the raw gender gap, though decreas-

[9] The sixth semester starts in mid-February; the MSGE is usually taken in early May and the HSEE in mid-June (although a typical semester usually ends in mid-July).

ing over time in overall performance, favored the girls most often, with a reversal on the HSEE at the end in Math and Physics—and this was especially significant in Physics. The gender gap was also significantly larger in the two literacy subjects, Chinese and English, than in the three science subjects. The trend in Fig. 2 shows that the gender gap favoring girls steadily increased in Chinese; the gender gaps in both Math and English also steadily increased until the trend reversed in the fifth semester, and the gaps increased again on the MSGE before a final drop on the HSEE. In Chemistry and Physics, the gender gaps steadily dropped until a comeback was evidenced on the MSGE, followed by a final sharp reversal favoring the boys on the HSEE. The gender gaps in the overall score more closely followed the patterns of the gender gaps in the science subjects than they did in the literacy subjects, yet much less dramatic than the gender gaps in the science subjects. We also used the immediate previous semester test scores instead of the first semester scores in each subject as covariate (baseline score) to capture the change in gender gap semester by semester. The patterns found were consistent with those described in Figs. 1 and 2.

This interesting dramatic change in the gender gap during the last semester cannot be explained by trend alone. It indicates discrepancies in gender gaps as measured by the three tests, with the MSGE demonstrating the largest gender gap favoring girls, followed by the fifth semester test, and finally the HSEE. The HSEE is the essential criterion for determining (1) whether a student can continue on to high school and (2) into which high school the student will be placed. Thus, the gender gaps measured by the HSEE are meaningful only for students who want to continue on to high school. Conditional on the decision to proceed to high school, the girls in this study continued to dominate in their overall performance and specifically in two literacy subjects, while losing ground to boys in Math (no statistical significance) and in Physics (significantly at the 0.01 level). The implications of these discrepancies in the gender gap on different tests will be discussed in later sections.

To further explore the heterogeneity in gender achievement gaps over the distribution of student performance, we estimated the raw and value-added gender gaps for different quartiles of the corresponding test score distribution semi-parametrically with bootstrapped standard errors. For all semesters, even though girls outperformed boys in all quartiles, the raw gender gap significantly decreased over the distribution of the overall performance in all tests throughout the 3 years of middle school. Compared to the first semester, the value-added gender gap is, in general, not significantly different across quartiles.

The decrease in the gender achievement gaps over the distribution of the test scores is consistent with an important finding in the literature (e.g., Entwisle, Alexander, & Olson, 1994; Feingold, 1992): boys might generally have a larger variance in performance than girls. To compare the within-school performance variation[10] among boys with

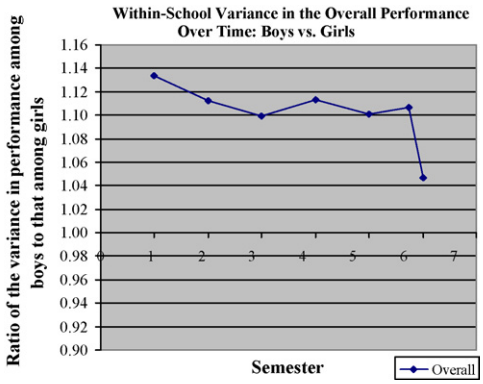

Fig. 3. Within-school variance in the overall performance over time: boys vs. girls.

that among girls over time, we used the residuals of the regression of the test scores on middle school dummies for each test over the 3 years, calculated the variance of these residuals[11] among boys and girls, respectively, and plotted the ratio of this variance among boys to that among girls in Fig. 3. The ratio was stable over the middle school period, ranging from 1.10 to 1.14, until it suddenly dropped to approximately 1.04, as measured by the HSEE score. Thus, the decrease in the gender achievement gaps over the test score distribution might reflect larger variance in boys' performance than girls'.

We then explored the heterogeneity in gender achievement gaps across students' socioeconomic status, as the literature suggests that socioeconomic status might have different effects on the academic performance of boys and girls. For example, Tansel (2002) showed that the effects of parents' income and education had larger effects on the educational attainments of Turkish girls than boys. Other studies showed that boys from disadvantaged family backgrounds might have more behavioral problems and lower performance than girls from similar backgrounds; and that parents of higher socioeconomic status might be relatively gender-neutral in their willingness to provide for their child's education, while parents of lower socioeconomic might favor boys. To see whether parental socioeconomic status had differential effects on boys' and girls' performance, we added interactions between the female indicator and parental socioeconomic status variables such as "parental income higher than the mean level" and "parents' average years of education" to Models 1, 2, and 3, controlling for the main effects of the socioeconomic variables. We did not find consistent evidence that parents' education or income had differential effects on the performance of boys and girls (not reported in the table).

[10] We are interested in within-school variation as different schools might have different grading policies for the semester tests and the MSGE. Therefore, we need to control for the school fixed effects and use the residuals to calculate the variance in the performance of boys and girls, respectively.

[11] Correlation within classroom was considered in the residual estimation; when we also calculated the GLS random-effects estimators, the results were similar.

3.3. Missing test scores and discrepancy in end-of-middle-school gender achievement gaps measured by different tests

The gender gaps at the end of middle school education, as measured by the scores of three tests (HSEE, MSGE, and fifth semester test), differed significantly, even though those tests were all taken within one semester. Several reasons might have driven the discrepancies in gender achievement measurements using these three tests. First, as many studies show, the advantage that girls have over boys in performance decreases as question difficulty increases. Thus, the larger gender gap caused by the MSGE and not the other two tests might result from the overly simple questions found on the MSGE—arguably the easiest test of the three. However, the complexity of the semester tests and the HSEE should be comparable because the semester tests usually follow the HSEE to familiarize students with the latter. Second, according to current research, even though the single or multiple-choice questions that favor boys constitute a major part of the fifth semester test, the MSGE, and the HSEE, the fact that teachers instructing the examinees also grade the first two tests might detract from the standardized nature of these tests. One reason for this may be that the teachers' (graders') first-hand impressions of the students might, even unconsciously, affect their grading to favor girls, particularly via the test's open-ended questions. It is difficult to identify the effects of these two key sources without an experimental setting. Thus, even though we will address them in a later analysis, this paper does not focus rigorously on identifying the influences from these two sources. Nevertheless, the patterns indicated in this paper warrant further research along these lines to arrive at a more precise explanation of the effects of test form and complexity as well as of teachers on the gender gap using different test measures. The implications of such findings would be most relevant for educational equity and quality.

Finally, because only students who desired to enter high school took the HSEE in 2002, and many student records of the MSGE and semester test scores were missing in the current data set, this lack of data might have resulted in biased estimates. This might have also driven the discrepancies in gender achievement gaps, as measured by HSEE, MSGE, and fifth semester scores. This potential reason will be the focus of this section.

We propose three strategies to explore this reason for the discrepancies in gender gaps measured by the different tests above. First, we examine whether the students in the whole sample and those with the relevant available test scores had significantly different characteristics. We may then examine how the difference between the sub-samples and the whole sample might have biased the estimated gender achievement gap. Second, we may use common sub-samples with both available HSEE and MSGE scores, or with both available HSEE and semester test scores, or with all three available test scores combined, and examine whether the gender gaps measured by these different tests were similar when using these common sub-samples. Third, we may impute the missing data using a censored model with multiple imputation, which is an appropriate model for data that is not "missing at random" (MAR). To some extent, the results using the imputed data will correct the missing data bias in the gender gap estimation.

3.3.1. Comparison of characteristics between sub-samples and the whole sample

Table 2 shows summary statistics of important individual and parental characteristics for students in the whole sample and various sub-samples with different available test scores. This table shows that students with available fifth semester scores were most similar to the whole sample. As for the sample with no missing MSGE scores, students in the sub-sample had lower socioeconomic status than those in the whole sample, yet the available semester test scores were slightly higher for students with no missing MSGE than for the whole sample. As for the sample with no missing HSEE scores, students in this sub-sample had significantly superior family background and semester academic performance than students in the whole sample. In particular, their mean semester test scores exceeded the mean of the whole sample by 0.35 standard deviations. Thus, we can conclude that students with no missing MSGE and/or HSEE scores belong to the higher portion of the distribution of academic performance. As thus far observed in this study and supported by current studies, the gender achievement gap favoring girls decreased over the distribution of performance; thus, we may conclude that the missing test scores might have contributed to an underestimation of the gender gap, if it, in fact, even had an impact on this estimation.

Furthermore, 53% of the students with available HSEE scores were girls, indicating a 6% gender gap in HSEE score availability, which is a significant difference at the 0.01 level. If the HSEE scores were missing because of administrative errors or difficulty in combining data, there should not be a significant difference in cases with missing test scores between boys and girls; thus, this difference indicates that boys might be less likely to take the HSEE required for regular high school admission. As the vocational high schools and private high schools are rather underdeveloped, schools and parents usually encourage students to take the HSEE if they stand even a slight chance of passing it to enter regular high schools; this is especially true in Beijing, where a greater number of high school seats are available than in most other parts of the country. Nonetheless, only 50–60% of middle school graduates are able to be accommodated by regular high schools, and thus many students still need to decide whether to take the HSEE or pursue admission to vocational high schools. Semester test scores are the primary tool students use for this decision. Because admission to vocational high schools takes place before the HSEE, students admitted to the vocational high schools do not take the HSEE. Thus, the abrupt closure of the gender gap as measured by HSEE scores might be driven by the absence of the HSEE among boys who were discouraged by their relatively poor semester test performance instead of actual improvement in boys' academic performance. When we ran a logit regression of indicator of missing HSEE scores on gender and school dummies, the gender coefficient was positively significant. It became insignificant after we included no other

Table 2
Summary statistics of the whole sample and sub-samples with the corresponding test scores available.

Sample	Parents' income, RMB	Parents' education, years	Elementary graduation test score, standardized	First semester score, standardized	Mean semester scores, standardized	Ideal education degree (student's) (1–7)	Ideal education degree (parents') (1–7)	Father's SES (1–5)	Mother's SES (1–5)	Like school (1–5)	Gender: 0—male, 1—female
Whole sample											
Mean	1546	13.1	0.00	0.00	0.00	4.8	4.4	2.9	2.6	3.4	0.50
# of obs	6543	6798	5764	5433	5957	6820	6728	6471	6533	6839	7096
With HSEE											
Mean	1645	13.4	0.29	0.31	0.35	5.1	4.7	3.0	2.7	3.5	0.53
# of obs	4216	4353	3617	3429	3762	4360	4321	4180	4223	4375	4477
With the fifth semester score											
Mean	1503	13.0	0.00	0.00	0.00	4.8	4.4	2.8	2.6	3.4	0.50
# of obs	4749	4923	4412	4847	5168	4939	4878	4693	4719	4957	5159
With MSGE											
Mean	1474	13.0	0.07	0.08	0.09	4.9	4.5	2.9	2.6	3.4	0.51
# of obs	3490	3647	3236	3198	3328	3642	3589	3465	3506	3657	3777
With both HSEE and the fifth semester score											
Mean	1614	13.3	0.29	0.31	0.36	5.1	4.7	3.0	2.7	3.5	0.52
# of obs	3059	3157	2762	3059	3257	3161	3135	3044	3065	3173	3257
With both HSEE and MSGE											
Mean	1510	13.1	0.29	0.34	0.39	5.1	4.7	2.9	2.6	3.5	0.53
# of obs	2376	2465	2197	2115	2195	2462	2438	2355	2393	2471	2526
With HSEE, MSGE, and the fifth semester score											
Mean	1484	13.0	0.28	0.33	0.38	5.1	4.6	2.9	2.6	3.5	0.53
# of obs	1905	1979	1784	1982	2033	1976	1957	1900	1929	1983	2033

Note: Please refer to Appendix A for a detailed description of the variables in this table.

individual and family background characteristics except the semester test scores as controls, which is consistent with the above rationale. The sudden drop in the ratio of within-school variance in performance among boys to that among girls on the HSEE (Fig. 3) might also represent the fact that boys disproportionately fell into the bottom of the performance distribution in semester tests and did not take the HSEE, thus decreasing the variance in HSEE performance among boys. This raises concerns over the status of the boys' education once they pass the compulsory stage and graduate from middle school.

3.3.2. Estimating gender achievement gaps at the end of the middle school education using common sub-samples

In this section, we estimated the gender gaps using common sub-samples to examine how much the discrepancy in gender gaps at the end of middle school education using different test score measures was driven by sample selection due to missing test scores. As Table 2 shows, students in the whole sample and most of the sub-samples were significantly different. Therefore, we did not expect to obtain unbiased estimates by using common sub-samples; our intention instead was to examine whether the discrepancy among the three test score measures dropped by using common sub-samples.

Table 3 presents this set of results. Each row panel title in Table 3 indicates the test score measure used for the gender gap estimation in that panel. The column panel title indicates the sub-sample used for that panel. Each sub-column represents the raw gender gap estimate by an individual regression controlling for variables indicated in the "control" panel at the bottom, and using the corresponding sub-sample (column panel title) and test scores (row panel title).

The discrepancies narrowed significantly when using common sub-samples, suggesting that the discrepancies among the gender gaps at the end of middle school education using different test measures are, to a great extent, driven by sample selection. Nonetheless, when using the sub-sample with all three available test scores, we find that the gender gap estimated by using the MSGE is the largest, followed by the gender gap using the fifth semester score, and then by using the HSEE. As mentioned earlier, this might have resulted from the MSGE's low level of difficulty compared with the other two tests, as well as from other reasons that cannot be confirmed using data available in this study, such as the grading bias, particularly because the MSGE and semester tests were graded by teachers who instructed the students.

3.3.3. Correction for sample selection bias in the gender gap estimation using the HSEE

In this section, we discuss using the pattern of missing HSEE scores to conduct data imputation, and thus correct for sample selection bias in estimating the gender gap using the HSEE scores. We first assume that within each school, missing HSEE scores due to administrative errors and difficulties in combining data is random. Thus, this portion of missing data can be imputed by multiple imputation methods for data missing at random (MAR), as reviewed in Allison (2002). The procedure is

1. Obtain the starting values for the parameter estimates by listwise deletion estimation.
2. Use the current values of the parameters to generate predicted values for all the missing test scores that are assumed to be MAR.
3. To each predicted value, add a random draw from the residual distribution for the test score distribution.
4. Using the imputed data set with both observed and imputed values, recalculate the parameter estimates.
5. Based on newly calculated means and covariance matrix, make a random draw from the posterior distribution of the parameter estimates to obtain the parameter estimates of this iteration.
6. Go back to step 2 and continue the iterations until convergence is achieved.
7. Sequential chains of data augmentation: Use the parameter estimates from the last iteration as starting values, and follow steps 2–5 for another chains of iterations, with the number of iterations equal to the number of iterations to reach the convergence in step 6 divided by the number of desired imputed "complete data sets." The imputed "complete data set" in the last iteration is used as the first data set for the final-stage parameter estimation.
8. Starting with the current values, repeat step 7 until the desired number ($M=5$ in our case) of imputed "complete data sets" are generated; calculate the parameter estimates of each data set.

The final parameter estimates are the means of the parameter estimates from all imputed "complete data sets." The standard errors of the parameter estimates are calculated following the formula by Rubin (1987)[12] below:

$$S(\bar{\beta}) = \sqrt{\frac{1}{M}\sum_{k} s(\beta_k)^2 + \left(1+\frac{1}{M}\right)\left(\frac{1}{M-1}\right)\sum_{k}(\beta_k-\bar{\beta})^2} \tag{4}$$

where β_k indicates the estimates of β using the kth imputed data set of the total M imputed data sets. $S(\bar{\beta})$ is the standard error of the estimated β, i.e., $\bar{\beta}=\sum_K \beta_k$. Multiple imputations using this method will not underestimate the standard deviation of the estimates as traditional methods of data imputation do for the following reasons. Instead of running the imputation model and just adding a random error to the predicted values, this method employs a data augmentation process, in which the parameter estimates for each iteration are randomly drawn from the posterior distribution of the parameter estimates in that iteration; with the sequential chain approach, convergence to the true posterior distribution is more likely to be achieved. Moreover, this imputation method uses more than one imputed data set (here, five has proven to be an appropriate number of data sets that need to be imputed) instead of taking an arbitrary imputed data set as given. Therefore, the relevant test statistics are more reliable using this method. The regression model for the imputation itera-

[12] Rubin (1987) and some other relevant works also provide the formula for the corresponding degrees of freedom.

Table 3
Estimates of the end-of-middle-school gender gaps in the overall scores using sub-samples with the relevant test scores available.

Comparable samples	(1) With HSEE and MSGE			(2) With HSEE and the fifth semester score			(3) With HSEE, MSGE and the fifth semester score		
(1) Gender gap estimates using HSEE									
Female	0.14***	0.04	−0.11***	0.15***	0.04	−0.12***	0.13***	0.03	−0.11***
p-Value	[0.00]	[0.29]	[0.00]	[0.00]	[0.25]	[0.00]	[0.00]	[0.57]	[0.00]
(2) Gender gap estimates using MSGE									
Female	0.17***	0.09***	−0.02				0.17***	0.07**	−0.02
p-Value	[0.00]	[0.01]	[0.31]				[0.00]	[0.04]	[0.36]
(3) Gender gap estimates using the fifth semester score									
Female				0.18***	0.10***	−0.06***	0.15***	0.08**	−0.06***
p-Value				[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]
# of obs	2526	2112	2115	3257	2677	3059	2033	1727	1982
Control									
School fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Primary school score		Y			Y			Y	
First semester score			Y			Y			Y

p values in brackets. Each sub-column reports gender achievement gaps estimated by regression using as dependent variable the scores of tests indicated by the corresponding row panel title, controlling for variables listed in the “control” panel in the same column. The column panel titles indicate the sub-sample used: (1) sub-sample of students with both the HSEE and MSGE scores available; (2) sub-sample of students with both the HSEE and the fifth semester average test scores available; (3) sub-sample of students with the HSEE, MSGE and the fifth semester scores available. When controlling for the elementary graduation test scores in the corresponding subjects, the middle school fixed effects, elementary school fixed effects, and the interactions between the elementary school dummies and the elementary school graduation test scores are also included. When controlling for the first semester test scores of the corresponding subjects, both the middle school fixed effects and the interactions between the middle school dummies and the first semester test scores are also included.

* Significant at 10%.

** Significant at 5%.

*** Significant at 1%.

tions used the students’ average semester test scores and their squared term, middle school fixed effects, students’ gender, elementary graduation test scores, and students’ parents’ income, education, and ages to impute the HSEE scores that were assumed to be missing at random. We also included other variables in the imputation models and obtained similar results.

When an HSEE score was missing because a student did not take the test, two scenarios can be established. First, some students had such poor school performance that they were discouraged from taking the HSEE; second, some students had such outstanding performance that they were pre-recruited by their desired schools and thus did not need to take the HSEE. In neither case were the HSEE scores missing in a random manner, and the former scenario should be considered the prevalent one. Considering this, then, we used a two-sided censored regression to correct for sample selection bias. These censored models were used because we lacked valid instruments that were related to HSEE attendance, yet not directly related to the HSEE scores in order to implement a two-stage estimation that would first estimate the propensity of taking the HSEE and then estimate the gender gap controlling for that propensity. We made the following assumptions about the censoring. First, students with missing HSEE scores whose average semester score was lower than 65 points[13] were assumed to follow the first scenario described above. Thus, their test scores were censored at 389—the official passing threshold for high school admission. Second, students with missing HSEE scores whose average semester score was higher than or equal to 85 points[14] were assumed to follow the second scenario. Thus, their test scores were censored at 455—the official threshold for admission to a top-tier high school. Consequently, we set up the joint log-likelihood function of the left-censored, uncensored (including both the available HSEE scores and the imputed values of the scores that were assumed to be MAR), and right-censored data, and used the maximum likelihood estimation to estimate the relevant parameters. Given the arbitrary nature of choosing the relevant thresholds, the relevant estimates might not be precise, but they most likely indicate the bias of the estimation by using the limited sample. The estimated within-school raw gender gap in the overall HSEE score was 0.19 standard deviations, which was significantly positive at the 0.01 level and higher than the estimates using any common samples as well as the sample with available HSEE scores, as was expected. We also applied other reasonable thresholds to determine the censoring of the observations with missing HSEE scores, and obtained similar results or slightly larger gender gaps.

[13] We used 65 since 60 is the passing grade; 65 points is a low enough score to discourage HSEE attendance.

[14] We used 85 since it is the widely-acknowledged excellence grade.

4. Explaining the gender achievement gap

4.1. Difference in important mediators for gender gap between boys and girls

In this section, we explore the sources of the gender achievement gap. First, the sample selection by primary school dropout is unlikely to drive the gender gap in Beijing. It is true that despite the law of compulsory education, girls are more likely to drop out of primary school than boys, especially in the rural area, which might cause sample selection bias for the gender gap estimate during middle school. However, the law of compulsory education is strictly enforced in Beijing, and there are no dropouts in primary school in Beijing.

Second, the gender gaps might be driven by some important gender differences in individual and family characteristics and experiences that were relevant to academic performance. To explore this possibility, we included some of these characteristics as mediators for gender gap in the regression Model 1 on average HSEE score, MSGE score, and fifth semester score. Then, we dropped the whole set of mediators in a specific aspect, examining how this would affect the gender gap estimate and the *R*-square of the model. By eliminating that set of mediators from the complete model including all mediators, the drop in *R*-square demonstrates the marginal contribution of the excluded characteristics to academic performance after controlling for other important characteristics. The joint significance of the excluded mediators can be tested using the *F*-test.

We examined six important aspects: demographic and socioeconomic characteristics (student age, parents' income, education, Communist Party membership, and index of socioeconomic status); parental care (index of parental care, lack of attention in making school choice in 1999, ideal final education degree desired for child, attendance at parents' meetings, and knowledge of child's teachers); non-cognitive skills or attitudes towards studying (students' ideal final education degree, average interest in all major subjects, and average hours spent on all major subjects before the last semester); whether students fit into the school atmosphere (relationship with schoolmates and teachers, whether they like school or not, and how much responsibility they take as members of the student body[15]); students' cognitive skills measured by how difficult they consider studying to be; and students' academic backgrounds (elementary school test scores, elementary school attended, interaction between school performance, and elementary school attended).

Table 4 shows the summary statistics and the within-school difference between boys and girls in the above variables.[16] These differences were measured by the coefficients of the gender indicator in the model, regressing the relevant characteristics on gender indicator and controlling for school fixed effects. We measured the within-school gender differences in these characteristics to be consistent with the within-school gender gap measures. First, contrary to what was observed in studies using sample from rural area in China, girls in our sample registered significantly earlier than boys; and no parental socioeconomic characteristics between boys and girls were significant at the 0.05 level. Second, significantly more parental care was offered to girls than to boys, although the differences were small in magnitude. Parents of girls lavished significantly more care in choosing a middle school during the middle school application, and had better knowledge of teachers.[17] Empirical evidence shows that parental involvement is strongly associated with student performance (e.g., Muller, 1998; Tinklin, 2003). Thus, the greater parental support received by girls might be one reason driving the gender gap.

In addition to better elementary school performance as compared with boys, and consistent with the evidence from existing literature, girls seemed to have a more cooperative attitude toward schooling and more favorable schooling experiences. Jacob (2002) reports that girls possessed superior non-cognitive skills to boys, and that the gender gap favoring girls in enrollment in U.S. colleges dropped by 40% after controlling for non-cognitive skills. Tinklin (2003) reports that the most significant explanation for the gender differences in high academic attainment in Scottish schools is that girls take school more seriously than boys. In our sample, girls reported having better relationships with schoolmates, and liked school significantly more than boys did. Specifically as related to studying, girls were less likely to report that studying was a tough burden and placed greater emphasis on studying. For example, girls spent significantly more time studying than boys did. Interestingly, even though girls were significantly more interested in Chinese and English than boys, and were less interested in the three science subjects, they still spent more time on the latter compared with boys and more strongly believed in improving their abilities through personal effort. Finally, girls assumed far more responsibilities on the school student committees than boys did in both elementary and middle schools.

These findings demonstrate that girls seemed to possess superior non-cognitive skills; more readily complied with the centralized curriculum, teaching plan, and test-oriented atmosphere; and thus fit more easily into the current educational system than boys did. Interesting questions arising from these findings are whether these superior "non-cognitive skills" are desirable for improving the quality of education for all students and whether the system, and not the boys' attitudes, requires change. The

[15] We used the highest position held in the student committees during elementary school as a proxy for students' position on middle school committees, because the latter might be endogenous to their test scores. For the other attitude measures, we tried both retrospective measures (for primary school or beginning of middle school) to avoid the endogeneity and current measures; general results were similar.

[16] We restricted the sample to observations with complete information on the mediators of the gender gaps listed in last paragraph. We also examined gender gaps using the whole sample as well as three additional samples which further restricted the sample used in Table 4 to observations with complete information on the students' test scores on the HSEE, MSGE, and fifth semester test, respectively. Most results were similar to what we present here.

[17] The differences in parental care measures were even more significant when using the whole sample.

Table 4
Gender difference in important individual and family characteristics and schooling experiences (with complete data of the mediators).

	Student's age, years	Parents' education, years	Parents' income, log	Father's SES (1–5)	Mother's SES (1–5)	Parents' years in Beijing, years	Father's Party membership (0/1)	Mother's Party membership (0/1)
Parental socioeconomic status								
Within-school difference	–0.06***	0.00	95.75*	–0.07*	0.00	0.16	–0.02	0.00
Mean	15.80	12.94	1481.15	2.84	2.57	39.39	0.27	0.16
Observations	3808	3808	3808	3689	3713	3718	3739	3722

	Evaluation of the importance of school (1–5)	Lack of attention to school choice (1–5)	Parental care index (1–5)	Ideal educational degree for the child (1–6)	Knowing about the head teacher (0/1)	Knowing about the major teachers (0/1)	Parental meeting attendance (0/1)
Parental care and attitudes toward child's education							
Within-school difference	0.02	–0.03*	–0.01	0.05	0.05***	0.05***	0.01
Mean	4.50	1.22	3.95	4.41	0.71	0.53	0.92
Observations	3717	3808	3808	3808	3808	3808	3808

	Relationship w/schoolmates (1–5)	Relationship w/teachers (1–5)	Like school (1–5)	Position in the student body (1–5)	Position in the student body during primary school (1–5)	Belief in efforts (1–5)	Importance of study (1–5)	Ideal degree of education (1–6)
Students attitudes, non-cognitive skills, and schooling experience								
Within-school difference	0.07***	0.02	0.07**	0.22***	0.29***	0.10***	0.15***	0.04
Mean	4.11	3.85	3.40	2.07	2.37	4.15	3.51	4.81
Observations	3808	3808	3808	3804	3808	3779	3808	3808

	Difficulty in study (1–5)	Complaining about lecture progress (1–5)	Average study time on five important study activities: remedy session, advanced session, office hour, homework, and other voluntary time spent on study, h/week					Overall study time—the last semester, h/week	Overall study time before the last semester, h/week
			Chinese	Math	English	Physics	Chemistry		
Within-school difference	–0.07**	–0.02	–0.01	0.07**	0.07**	0.10***	0.06*	0.06*	0.06***
Mean	2.49	2.98	1.60	1.90	1.78	1.74	1.70	1.74	1.13
Observations	3808	3792	3806	3808	3808	3806	3807	3808	3808

	Interest in study in all subjects (1–5)				
	Chinese	Math	English	Physics	Chemistry
Within-school difference	0.20***	–0.16***	0.33***	–0.44***	–0.19***
Mean	3.38	3.53	3.40	3.50	3.60
Observations	3808	3805	3807	3804	3807

The gender differences are measured by regressing the characteristics on the female indicator, controlling for middle school fixed effects. Ideal degree of education: 1—middle school degree; 2—high school/vocational high school degree; 3—professional college degree; 4—university degree; 5—master degree; 6—doctoral degree.

* Significant at 10%.

** Significant at 5%.

*** Significant at 1%.

Table 5
Estimates of the gender achievement gap controlling for the mediators.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Dependent variable: the HSEE average test score													
Female	0.12***	0.05	0.18***	0.18***	0.18***	0.15***	0.08**	0.07**	0.14***	0.06*	0.05	0.09**	0.08**
p-Value	[0.00]	[0.11]	[0.00]	[0.00]	[0.00]	[0.05]	[0.04]	[0.04]	[0.00]	[0.10]	[0.13]	[0.01]	[0.04]
# of observations	4361	3466	4272	3535	3656	4367	2503	2508	2905	2555	2932	2840	2512
R-squared	0.07	0.25	0.1	0.11	0.02	0.07	0.34	0.33	0.18	0.33	0.31	0.32	0.33
p-Value of the joint *F*-test of the marginal contribution of the set of mediators excluded in the column								0.04	0.00	0.00	0.00	0.09	0.00
Dependent variable: the MSGE average test score													
Female	0.17***	0.15***	0.25***	0.22***	0.26***	0.24***	0.14***	0.15***	0.17***	0.11***	0.15***	0.14***	0.13***
p-Value	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
# of observations	3642	3016	3581	2944	2954	3644	2024	2030	2337	2065	2363	2475	2033
R-squared	0.18	0.39	0.28	0.26	0.06	0.16	0.5	0.49	0.37	0.48	0.46	0.51	0.49
p-Value of the joint *F*-test of the marginal contribution of the set of mediators excluded in the column								0.00	0.00	0.00	0.00	0.21	0.00
Dependent variable: the fifth semester average test score													
Female	0.14***	0.13***	0.19***	0.19***	0.25***	0.22***	0.12***	0.13***	0.14***	0.10***	0.12***	0.10***	0.11***
p-Value	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
# of observations	4944	4161	4842	3996	4138	4945	2886	2891	3286	2951	3363	3406	2900
R-squared	0.16	0.42	0.27	0.25	0.08	0.15	0.59	0.59	0.44	0.56	0.57	0.5	0.59
p-Value of the joint *F*-test of the marginal contribution of the set of mediators excluded in the column								0.03	0.00	0.00	0.00	0.06	0.00
Mediators													
Fitting in the school	Y	–	–	–	–	–	Y	–	Y	Y	Y	Y	Y
Cognitive skills and academic background	–	Y	–	–	–	–	Y	Y	–	Y	Y	Y	Y
Attitudes to study	–	–	Y	–	–	–	Y	Y	Y	–	Y	Y	Y
Parental care	–	–	–	Y	–	–	Y	Y	Y	Y	–	Y	Y
Demographic and parents' socioeconomic status	–	–	–	–	Y	–	Y	Y	Y	Y	Y	–	Y
Cognitive skills	–	–	–	–	–	Y	Y	Y	Y	Y	Y	Y	–

p values in brackets. Each column indicates a regression of the dependent variable on the female indicator, controlling for the mediators included in that column and school fixed effects.

* Significant at 10%.

** Significant at 5%.

*** Significant at 1%.

boys' negative attitudes might simply be reflections of a defective test-oriented education system. Such a discussion is beyond the scope of this paper, but deserves in-depth exploration in future research.

The higher levels of parental care and superior non-cognitive skills possessed by girls might also reflect the gender gap in returns to education in China. Studies have shown that returns to education for females exceed that for males (e.g., Jamison & Van der Gaag, 1987; Johnson & Chow, 1997).[18] Moreover, gender occupational segregation, which is an important source of the wage gender gap, decreases as educational level increases (e.g., Wu & Wu, 2007), and the gender wage gap is smaller at the upper part of the wage distribution (e.g., Millimet & Wang, 2006). All these might provide more incentives for girls than for boys to improve their school performance, and for parents to pay more attention to girls' education. Moreover, prevalent gender discrimination against women in the labor market might also compel girls and their parents to take girls' schooling more seriously in order to compensate for this future disadvantage. As a result, compared with girls, boys on average have inferior non-cognitive skills and tougher school experiences, and fall behind girls in academic achievement, thereby directly affecting the relative status of boys and girls in high school admission.

4.2. Explaining gender achievement gaps via mediators

Table 5 shows the results of the contribution of each set of mediators to the gender achievement gap by the end of middle school. The joint *F*-tests in columns 8–13 show that all sets of mediators are significant predictors of student performance, at least at the 0.05 level, except for parental socioeconomic status which is only significant at the 0.1 level in the regression using fifth semester and HSEE scores. However, none of the exclusions of the mediators from the complete model (column 7) significantly increased the gender gap estimates in column 7, except for "student cognitive skills and academic background." This indicates that the major contributor of the gender gap at the end of middle school is actually the gender achievement gap before students entered middle school. This highlights the importance of early intervention to decrease the gender achievement gap.

Moreover, even though the gender gap estimates in column 7 significantly decreased in magnitude compared with those using Model 1 without controls, they were still significantly positive at the 0.05 level. Thus, some factors may not have been fully captured by the rich set of observed characteristics in this study.[19] It is also possible that current measurements of the characteristics are not accurate enough for the analysis. These results are unlikely to be driven by the endogeneity of the predictors because endogeneity bias should have led to the opposite results. The findings here call for additional research using more precise measurements of both students' and teachers' attitudes and collecting more information than was available in this data set. In order to examine whether these results were driven by the sample selection due to missing data, we also restricted the analysis to common samples, such as the sample with the scores of all three tests and all independent variables available. The patterns are in general robust to sample selection.

5. Conclusion

This research included a cohort of 7235 students who entered public middle schools in Beijing's Dongcheng District to examine the evolution of the gender achievement gap. Results showed that, compared with boys, girls have higher test scores in both Chinese and Math upon elementary school graduation, and the within-school gender gap becomes larger during middle school than in elementary school. Even though the gender achievement gap favoring girls decreased during the middle school, boys demonstrated significantly inferior performance in all semester tests, the MSGE, and the HSEE, and only significantly outperformed girls on the HSEE in Physics. Even though the gender gap favoring girls decreased over the distribution of the overall score, indicating a larger variance in performance among boys than girls, girls outperformed boys in every quartile of the overall score distribution throughout middle school. In particular, significantly more boys dropped out of the regular public school system by the end of middle school and did not take the HSEE at all, driving a spurious sharp closure of the gender gap on the HSEE. To some degree, these findings resemble current research on the U.S. gender gap, yet they more closely parallel U.K. findings which also feature in more obvious female dominance in academic performance, even in Math.

For various reasons such as gender difference in returns to education, parents were more attentive to girls' education. As existing literature shows, girls also seemed to have higher non-cognitive skills than boys; complied more with the test-oriented education system; and had more positive schooling experiences. Thus, the results of this study suggest that adjusting the current educational system to develop the full academic potential of both genders is a priority issue concerning a fair and equitable education for both genders.

Importantly, even though both students' academic background and other factors, such as parental care and students' non-cognitive skills are strongly related to student performance, students' academic background is the only significant contributor to gender gap. Therefore, efforts should be made to decrease the gender achievement gap at the end of compulsory education, even before students enter middle school. Moreover, the residual significance of the gender achievement gap favoring girls after controlling for all the mediators suggests that additional factors not measured by the information in this data set were also quite likely to drive the gender gap. For exam-

[18] Relevant research using samples of other countries also showed significant gender difference in the return to schooling, and that this difference varied at different levels of education (e.g., Daoud, 2005; Duraisamy, 2002; Hawley, 2004; Nam, 1996; Toumanoff, 2005).

[19] We also applied the Blinder-Oaxaca Decomposition of the gender gap in test scores, which also showed that the differential effects of the observed characteristics on performance between boys and girls were not likely to drive the gender gap favoring girls, and that unobserved characteristics might be the main source of the gender gap.

ple, the non-anonymous nature of fifth semester and MSGE exams might favor girls since their cooperative attitudes usually make a stronger impression on teachers (the exam graders). On the other hand, teachers might unconsciously equalize the performance between boys and girls by favoring each gender in their weak subjects: as for the HSEE which was not graded by the teachers, compared with gender gaps in the tests graded by teachers, we observe larger incremental gender gaps favoring boys in Math and Physics and girls in Chinese (see Fig. 2). It is difficult to test these hypotheses without a more precise experimental design and more rigorous measurements of the latent variables. All of these possibilities require further research.

The data set used in our analysis contained only one cohort in one central administrative district in Beijing, and only their middle school performance records were available. This limited the external validity of the study's empirical results. Further studies of greater breadth and depth are necessary to assess the external validity of these results. Nonetheless, even though the Dongcheng District sample does not represent the whole Chinese population, especially in the economically and socially less-developed regions and rural areas of the country, this sample is still important to examine for several reasons. First, given the lack of rigorous empirical evidence on the gender achievement gap in China, this study provides valuable evidence for the evolution of the gender achievement gap during middle school. In particular, given the importance of the Dongcheng District to China's capital city of Beijing, issues of educational inequalities in this district are worth exploring. Moreover, as this district largely represents metropolitan areas in China's developed regions, the research findings will be instructive for creating new lenses with which to view China's fast economic growth and urbanization.

Finally, because of the radical economic and social transitions that have occurred in a country as large as China, rural/less-developed and urban/more-developed areas have pursued drastically different approaches to economic and human development since liberation and thus face unique challenges in education as well. For example, in rural areas where the One Child Policy was never strictly implemented and educational resources were scarce, one would expect to find a gender achievement gap favoring boys and pervasive underlying issues based on a traditional bias toward boys' education to the neglect of girls'. By contrast, in well-developed metropolitan areas where most children are the "only child," the lifestyle is less traditional, and educational resources are more abundant, one would expect the opposite pattern in the gender gap related to defects in the educational system itself. The present study offered such evidence. Therefore, to identify relevant problems and seek effective solutions, it is important to examine educational inequalities in different contexts and devise solutions specific to those contexts, instead of examining the whole nation for general patterns and a "one-size-fits-all" remedy. This study conducted in Beijing's Dongcheng District raises concerns about the relatively inferior academic performance of boys that might be related to China's test-oriented educational systems. As a possible impending epidemic in China's metropolitan area schools, this issue deserves closer examination using broader samples, more diverse settings, and different developmental stages of students.

Appendix A.

Description of some important variables in Tables 2 and 4.

1. Parents' socioeconomic status (SES): This variable is coded based on parents' self-reported profession and their rank in that profession. There are five tiers for their rank: 1—low; 2—low-medium; 3—medium; 4—medium to high; 5—high.
2. Parents' Party membership: 0—not a Party member; 1—Party member.
3. Evaluation of the importance of school: Parents' response to a survey question asking them to recall their opinion of the importance of school in their child's development 3 years ago, when they were about to make the middle school choice decision: 1—not at all important; 2—somewhat unimportant; 3—maybe important, maybe not; 4—somewhat important; 5—very important.
4. Lack of attention to school choice: Parents' self-reported level of attention to their school choices in 1999: 1—paid very high attention; 2—paid high attention; 3—paid some attention; 4—did not pay much attention; 5—no attention at all.
5. Parental care index: The average of parents' responses to a series of questions about their care of their child: (1) their attention to their child's study; (2) their attention to their child's hobbies; (3) their attention to their child's friends; (4) their attention to their child's everyday plans. Response to each question is on a scale from 1 to 5: 1—never pay attention to it; 2—rarely pay attention to it; 3—sometimes pay attention to it; 4—often pay attention to it; 5—always pay attention to it.
6. Ideal educational degree for the child: Parents' ideal of their child's final education degree: 1—middle school degree; 2—high school or vocational high school degree; 3—professional college degree; 4—university degree; 5—master's degree; 6—doctoral degree.
7. Relationship with schoolmates (1–5): 1—very bad; 2—not good; 3—just so-so; 4—good; 5—very good.
8. Relationship with teachers (1–5): 1—very bad; 2—not good; 3—just so-so; 4—good; 5—very good.
9. Whether the student likes school or not: 1—not at all; 2—somewhat dislike; 3—just so-so; 4—somewhat like; 5—like very much.
10. Position in the student body (1–5): 1—none; 2—class level (lower rank); 3—class level (higher rank); 4—grade level; 5—school level. This coding applies to positions in both middle school and primary school.
11. Belief in efforts (1–5): Students' self-reported opinion of whether they believe that ability could be improved by efforts: 1—strongly disagree; 2—somewhat disagree; 3—neither disagree nor agree; 4—somewhat agree; 5—strongly agree.
12. Importance of studying (1–5): Students' self-reported opinion of whether they consider studying is the most

important thing in their life: 1—strongly disagree; 2—somewhat disagree; 3—neither disagree nor agree; 4—somewhat agree; 5—strongly agree.

13. Ideal degree of education: Students' self-reported ideal of final education degree: 1—middle school degree; 2—high school or vocational high school degree; 3—professional college degree; 4—university degree; 5—master's degree; 6—doctoral degree.
14. Difficulty in studying (1–5): Students' response to the statement "Studying is too hard for me": 1—strongly disagree; 2—somewhat disagree; 3—neither disagree nor agree; 4—somewhat agree; 5—strongly agree.
15. Complaining about lecture progress: Students' response to the statement "The pace of the lecture does not fit me": 1—strongly disagree; 2—somewhat disagree; 3—neither disagree nor agree; 4—somewhat agree; 5—strongly agree.

References

Allison, P. (2002). *Missing data*. Thousand Oaks, CA: Sage Publications.

American Association of University Women. (1998). *Gender gaps: Where schools still fail our children*. Washington, DC: American Association of University Women Educational Foundation.

Bedard, K., & Ferrall, C. (2003). Wage and test score dispersion: Some international evidence. *Economics of Education Review*, *22*(1), 31–43.

Bolger, N., & Kellaghan, T. (1990, Summer). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement*, *27*(2), 165–174.

Card, D., & Lemieux, T. (2000). *Dropout and enrollment trends in the post-war period: What went wrong in the late 1970s?* NBER working paper 7658.

Cho, D. (2007). The role of high school performance in explaining women's rising college enrollment. *Economics of Education Review*, *26*(4), 450–462.

Christie, P., & Shannon, M. (2001). Educational attainment and the gender wage gap: Evidence from the 1986 and 1991 Canadian censuses. *Economics of Education Review*, *20*(2), 165–180.

Daoud, Yousef. (2005). Gender gap in returns to schooling in Palestine. *Economics of Education Review*, *24*(6), 633–649.

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

Duraisamy, P. (2002). Changes in returns to education in India, 1983–94: By gender, age-cohort and location. *Economics of Education Review*, *21*(2002), 609–622.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1994). The gender gap in math: Its possible origins in neighborhood effects. *American Sociological Review*, *59*(6), 822–838.

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research*, *62*(1), 61–84.

Goldin, C., Katz, L. F., & Kuziemko, I. (2006). The homecoming of American college women: The reversal of the college gender gap. *Journal of Economic Perspectives*, *20*(4), 133–156.

Hawley, J. D. (2004). Changing returns to education in times of prosperity and crisis, Thailand 1985–1998. *Economics of Education Review*, *23*, 273–286.

Jacob, B. A. (2002). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review*, *21*(6), 589–598.

Jamison, D. T, & van der Gaag, J. (1987). Education and earnings in the People's Republic of China. *Economics of Education Review*, *6*, 161–166.

Johnson, E. N., & Chow, G. C. (1997). Rates of return to schooling in China. *Pacific Economic Review*, *2*, 101–113.

Machin, S., & McNally, S. (2005). Gender and student achievement in English schools. *Oxford Review of Economic Policy*, *21*(3), 357–372.

Millimet, D. & Wang, L. (2006). A Distributional Analysis of the Gender Earnings Gap in Urban China. *Contributions to Economic Analysis & Policy*, *Berkeley Electronic Press*, *5* (1), 1461–1461.

Muller, C. (1998, October). Gender differences in parental involvement and adolescents' mathematics achievement. *Sociology of Education*, *71*(4), 336–356.

Nam, Y. S. (1996). Schooling and changes in earnings differentials by gender in South Korea, 1976–1991. *Economics of Education Review*, *15*(3), 245–258.

National Center for Education Statistics. (2004). *The condition of education 2004 in brief*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Organization for Economic Cooperation and Development (OECD). (2004). *Learning for tomorrow's world*. Paris: PISA. Retrieved from http://www.pisa.oecd.org.

Pallas, A. M., & Alexander, K. L. (1983). Sex differences in quantitative SAT performance: New evidence on the differential coursework hypothesis. *American Educational Research Journal*, *20*(2), 165–182.

Rose, H. (2006). Do gains in test scores explain labor market outcomes? *Economics of Education Review*, *26*(4), 430–446.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Tansel, A. (2002). Determinants of school attainment of boys and girls in Turkey: Individual, household and community factors. *Economics of Education Review*, *21*, 455–470.

Tinklin, T. (2003). Gender differences and high attainment. *British Educational Research Journal*, *29*(3), 302–332.

Toumanoff, P. (2005). The effects of gender on salary-at-hire in the academic labor market. *Economics of Education Review*, *24*, 179–188.

Wu, Y. X., & Wu, X. G. (2007, September). Gender occupational segregation in China's non-agricultural sector: 1982–2000, manuscript.