NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Alexey Bessudnov, Alexey Makarov

GENDER DIFFERENCES IN MATHEMATICAL PERFORMANCE AND THE SCHOOL CONTEXT: EVIDENCE FROM RUSSIA

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: EDUCATION
WP BRP 11/EDU/2013

Gender differences in mathematical performance and the school context: Evidence from Russia

Alexey Bessudnov[1] Alexey Makarov[2]

Abstract

Gender differences in mathematical performance have been long debated in psychology, economics, and sociology. We contribute to this literature by analyzing a large data set of high school graduates who in 2011 took a standardized mathematical test in Russia ($n = 738,456$). We find no substantial difference in mean test scores of boys and girls. However, boys have a greater variance of scores and are more numerous at the top of the distribution. We apply quantile regression to model the association between school characteristics and gender differences in test scores throughout the distribution. Male advantage in test scores, particularly at the top of the distribution, is concentrated in cities and in schools with the advanced curriculum. In ordinary high schools, especially in the countryside, gender differences in all parts of the distribution are very small. A separate analysis at the regional level confirms that male advantage in mean test scores is higher in more urbanized regions.

JEL classification: I21 (Analysis of Education), I24 (Education and Inequality).

Keywords: gender inequality, mathematical performance, school context.

[1]Department of Sociology, Centre for Advanced Studies, National Research University – Higher School of Economics. Email: abessudnov@hse.ru

[2]Department of Mathematics, National Research University – Higher School of Economics. Email: amakarov@hse.ru

Gender differences in school test scores in general and in mathematical performance in particular have over the past decades attracted considerable attention of scholars in several disciplines. A number of papers that explored this problem have been recently publushed in leading journals in psychology, economics, and sociology. The interest to this topic can be explained by several factors. Mathematical performance in school can affect career choice and is a good predictor of future earnings. Despite the rising rates of female participation in the labour market and in the educational system and the continuing trend for more gender equality in industrialized countries, there remains a significant underrepresentation of women in science, technology, engineering and mathematics (STEM) sectors. The study of gender differences in math performance in school and on standardized tests helps understand to what extent they are related to gender inequality in STEM and, more generally, to what extent they can contribute to gender inequality in the labour market.

Another reason for continuing interest in gender inequality in math performance is data availability. Cross-national studies of school children (such as PISA and TIMSS) allow researchers to look at gender differences in math in different national contexts and relate them to indicators of gender equality across countries. Extensive national data sets, mostly coming from the USA, help estimate gender differences in math more precisely.

In this working paper we contribute to this literature by analyzing a new large data set with the results of a standardized math test taken in 2011 by all high school graduates in Russia ($n = 738,456$). The size of the data set allows us not only to look at the gender differences in means and variances, but to reliably model differences throughout the distribution with the methods of relative distribution and quantile regression. After comparing the distributions for boys and girls in the whole data set, we model the association between school-level contextual factors (such as school type and location) and gender differences in the distribution of test scores. We also estimate the association between the size of a classroom and a school, gender ratios in classrooms and schools and the gender differences in test performance. In a separate analysis, we look at the determinants of the gender gap in math at the regional level.

The paper is structured as follows. Section 1 provides a review of recent studies of gender differences in mathematical performance, both at the national and cross-national levels. Section 2 describes our data and the statistical methods we use. Section 3.1 analyzes gender differences in test score distributions in our data set. Section 3.2 introduces the analysis at the regional level. Section 3.3 explores school- and classroom-level effects on the gender differences in test scores. Section 4 discusses the results substantively.

1 A literature review

There are several main themes in the literature that explores the gender gap in mathematics. First, scholars try to establish whether there are indeed differences in mathematical achievement of men and women. Second, if these differences exist, are they better explained by biological or social factors? Third, a separate literaure looks at the contextual (i.e. school- or neighbourhood-level) determinants of the gender differences in math.

The usual measure of gender differences in test scores is the effect size, d, that is the difference between mean values for men and women expressed in standard deviations. Positive values of d are evidence of male advantage and negative values show female advantage. Absolute values of d below 0.2 can be roughly considered small, between 0.2 and 0.5 medium and above 0.5 large.

In an early meta-analysis Hyde et al. (1990) collected 100 studies published between 1963 and 1988 that reported 254 effect sizes based on the analysis of more than 3.1 mln people (mainly in the USA). Averaged over all samples, d was found to be 0.15. When the study was limited to the samples of the general population only, d decreased to -0.05, slightly favouring women. These are small values that indicate the absence of substantial gender differences in mean mathematical performance. However, it was found that while there were no gender differences in mathematical problem solving in elementary and middle school, the gender gap favouring men emerged in high school and college.

Hedges and Nowell (1995) reported the results of a secondary analysis of six large data sets collected in the USA between 1960 and 1992. They found slight advantage of men in mathematical achievement (d ranges from 0.03 to 0.26). In a more recent study, Lindberg et al. (2010) conducted a meta-analysis of 242 studies published between 1990 and 2007 that involved about 1.3 mln people. Overall, d was found to be 0.05. In a separate analysis of four large studies of American adolescents, they found d varying between -0.15 and 0.22.

Hyde et al. (2008) analyzed mathematical test scores of about 7 mln American school children from 10 states and found d to be less than 0.1 in all grades (also see Hyde and Mertz (2009)). Weighted d was 0.0065. Contrary to the results of a previous study (Hyde et al., 1990), there was no evidence of a gender gap in mathematical performance emerging in higher grades. This partly contradicts earlier findings of Leahey and Guo (2001) who with longitudinal data from the National Longitudinal Study of Youth and the National Educational Longitudinal Study 1979 documented a faster rate of acceleration in math (especially in geometry) for boys that results in a slight gender gap favouring boys by the 12th grade. Fryer and Levitt (2010) analyzed longitudinal data from the Early Childhood Longitudinal Study Kindergarten Cohort 2003 (n=20,000) and found that when children entered kindergarten they were equal in math and reading, but by the end of the 5th grade girls fell behind boys in mathematical test scores by 0.2 standard deviations.

Overall, recent research shows that currently in the USA boys and girls are either equal or very close in mean mathemetical performance. The evidence on whether some gender gap emerges in higher grades is contradictory, but in any case this gap is not large.

Another question is whether other characteristics of mathematical test score distributions differ by sex. A standard measure of variability of test scores is variance ratio, i.e. variance for men divided by variance for women. Variance ratios that are higher than one indicate greater male variability of test scores. Another way to measure variability is to look at the male/female ratio at different percentiles of the distribution.

It has long been argued that men show more variability than women in a number of characteristics, including mathematical performance. Indeed, most research conducted in the USA and cross-nationally produced variance ratios for mathematical test scores that were higher than one, confirming the hypothesis about greater male variability. In the meta-analysis by Hyde et al. (2008) variance ratios ranged from 1.1 to 1.2. Similar results were reported by Hedges and Nowell (1995). Lindberg et al. (2010) reported the average VR of 1.08. Machin and Pekkarinen (2008) analyzed cross-national data from PISA, TIMSS and PIRLS and documented greater variance for boys virtually in all countries (with minor exceptions), both for mathematics and reading. Ellison and Swanson (2010) looked at the gender composition of the participants of the American Mathematical Competitions, a selected sample of mathematically advanced American school children. While the proportions of boys and girls among all participants were not so different, the gender gap dramatically widened at the highest levels of achievement. At the 99.9 percentile the ratio of boys to girls was found to be 9 to 1, and all the top scorers were male.

Spelke (2005) presented some evidence against the hypothesis of greater male variability in math scores. However, she mostly discussed the SAT-M test and its potential gender bias. Later analyses by Hyde et al. (2008), Machin and Pekkarinen (2008) and Lindberg et al. (2010) left almost no doubt that boys indeed had more variable test scores than girls.

The central question of the studies of the gender gap in math achievement is whether the differences in the distributions of test scores have biological or social explanation. Extensive research demonstrated that there are some innate gender differences in the way human brain functions (Kimura, 2000), also see literature reviews in Penner (2008); Buchmann et al. (2008). These innate differences may well be related to gender differences in the distributions of mathematical test scores. For example,Geary (1996) provided an evolutionary explanation for greater male variability in mathematical performance.

On the other hand, traditional gender stereotypes and norms that portray men as more capable for mathematics and science can definitely affect mathematical performance of boys and girls as well (see Buchmann et al., 2008, for a review of possible mechanisms). As mathematical performance cannot be measured at birth, it is a hard task to empirically separate innate and environmental reasons of any possible gender differences in test scores (also, some innate differences can express themselves at a later age and may only be present in certain social environments).

One of the approaches to this problem has been to study mathematical performance cross-nationally. It is unlikely that genetic differences in cognition between sexes vary by country. If we make this assumption then all non-random cross-national differences in d and variance ratios of test scores can be attributed to environmental rather than innate factors. Note that this approach does not help us show empirically whether there are any innate gender differences in performance, but only tests the presence of environmental effects.

The data that are used in cross-national analysis come from three large international studies of performance of school children: PISA, TIMSS and PIRLS. All studies conducted with these data sets universally indicated that there were large cross-national differences in the size and sometimes direction of the gender gaps in mathematics and reading. There is little doubt now that environmental factors that vary across countries do have an effect on gender differences in school performance.

What are the explanations for cross-national differences in the size and direction of the gender gap? The hypothesis that is often tested in the literature deals with gender stratification. It states that in the countries that achieved greater gender equality in society there is also greater equality between boys and girls in mathematical performance in school.

In an early test of the gender stratification hypothesis, Baker and Jones (1993) analyzed the data from SIMS (the predecessor of TIMSS) that had mathematical test scores of about 77,000 eight-graders from 19 countries. They found significant cross-national variation in the size and direction of the gender gap in mathematics. In most countries, boys showed higher average test scores than girls, but there were four countries in the sample where girls performed significantly better than boys (French Belgium, Finland, Hungary, Thailand). In support of the gender stratification hypothesis, the gender gap was found to be correlated with measures of female educational and occupational opportunities in adulthood. Note that Baker and Jones (1993) discussed how the size of the gender gap may be related to the male/female ratio in the sample that could be the evidence of differential self-selection to school. Indeed, in the French sample there were 77 boys for 100 girls, while in Nigeria the male-to-female ratio was 268 to 100.

Marks (2008) looked at the cross-national gender gaps in mathematics and reading with the data from PISA 2000 (PISA surveys 15-year old school children). Generally, boys outperformed girls in mathematics and girls outperformed boys in reading. As in other studies, significant cross-national variation in the size of the gender gaps was found. However, the explanation of this variation was inconclusive. Some characteristics of the school system were correlated with the gender gap in reading, but not in mathematics. The proportion of women in the work force, social inequality and public sector spending were found to be correlated with the gender differences in reading (but again, not in mathematics), although these correlations were weak and unstable.

Penner (2008) used the data on mathematical performance from TIMSS 1995 that included over 40 countries. TIMSS surveyed pupils in the final year of secondary school. In all countries boys did better in math than girls, but d varied from 0.05 (Hungary) to 0.63 (Netherlands). Penner looked not only at gender differences in mean performance, but also at other parts of test score distributions (applying quantile regression and logistic regression to predict the probabilities of being above a certain cut-point in the distribution). The gender gap in math was found to be smaller in countries with greater gender equality (measured in education and in the labour market, and also as adherence to traditional gender roles and as a status of women in society). The size of these effects varied in different parts of the distribution, and some of them (relative status of men and women) were present only at the top of the distribution.

Guiso et al. (2008) specifically tested the gender stratification hypothesis with the data from about 40 countries from PISA 2003. As in previous studies, girls on average did better than boys in reading, but the gender gap was reversed for math. In both cases, the size of the gap correlated with the indicators of gender equality, such as female economic activity rate and women's political empowerment.

Fryer and Levitt (2010) tested the gender stratification hypothesis with both PISA 2003 and TIMSS 2003, comparing the results from two data sets. There was indeed correlation between the size of the gender gap in math and measures of gender equality in PISA 2003. However, this correlation disappeared in TIMSS 2003, mainly because of the presence of a larger number of Muslim Middle Eastern countries in the sample. Somewhat surprisingly, adherence to traditional gender roles in Muslim countries did not preclude greater equalization of mathematical performance of boys and girls. Fryer and Levitt claimed that this may be due to single sex schooling in Muslim countries that may foster girls' achievement in math.

Else-Quest et al. (2010) also performed an analysis of both PISA 2003 and TIMSS 2003 and found that the average effect sizes in both data sets were less than 0.15, indicating little difference in mathematical achievement between male and female pupils. However, the range of d across countries was found to be from -0.42 to 0.4. Else-Quest et al. (2010) looked at the correlations between the country-level gender gaps in math and various measures of gender equality. Some of these correlations proved statistically significant. In TIMSS, gender differences in mathematical performance were associated with gender ratios in school enrollment. In PISA, statistically significant predictors of the gender gap were different: women's shares of research positions and parliamentary seats and gender differences in economic activity rates.

In a recent study, Kane and Mertz (2012) looked at the data from TIMSS 2003 and 2007 (and, in some analyses, PISA 2003 and 2009). In both TIMSS data sets there were no statistically significant gender differences in mean mathematical performance either among fourth or eighth graders, except for eighth graders in 2007 when girls slightly outperformed boys. The variance ratios almost in all countries were found greater than one, suggesting greater male variability. In the TIMSS 2007 data set for eighth-graders variance ratio correlated negatively with the effect size d, so that in the countries with the largest variance ratios girls on average performed better than boys. In most data sets that were analyzed there was no significant correlation between the gender gap in math and the Global Gender Inequality Index (and, in fact, in some data sets the correlation was in the opposite to the theoretically predicted direction, so that in more gender equal countries boys outperformed girls in math to a larger degree). Kane and Mertz also tested the single sex schooling hypothesis and did not find evidence to support it.

Taken together, cross-national research of the gender gap in math shows the following picture. In older data sets from the 1990s, boys outperformed girls in most or even all countries (Penner, 2008; Marks, 2008; Baker and Jones, 1993). In more recent data sets, the evidence is more contradictory. PISA 2003 shows some average male advantage (Guiso et al., 2008; Fryer and Levitt, 2010), although it is small (Else-Quest et al., 2010). In TIMSS 2003 and 2007 the average gender gap in math was virtually nonexistent (Else-Quest et al., 2010; Kane and Mertz, 2012). In most countries, male-to-female variance ratios were greater than one, although there were some exceptions. In all data sets, there was significant cross-national variation in the size and direction of the gender gap. The attempts to explain this variation with some country-level predictors have so far been inconclusive. In some data sets there was correlation between the gender gap and various measures of societal gender equality. However, these correlations were absent in other data sets and in any case they cannot prove the causal effect because of unobserved confounders.

In an attempt to test the gender stratification hypothesis within one country, Pope and Sydnor (2010) looked at the gender gap in math across the US states, using data on white students from the National Assessment of Educational Progress. There was variation in the size of the gender gap across states, and the gender gap in math and science correlated with gender attitudes. In the states with more traditional attitudes to gender roles the gap favoured boys to a larger degree. While this correlation at the state level is informative, it cannot be taken as evidence of a direct causal link between societal gender roles and attitudes and the size of the gender gap in math (for the same reason of unobserved heterogeneity). Interestingly, Pope and Sydnor also found negative correlation between gender gaps in math and science, on one hand, and the gender gap in reading, on the other hand. In the states where boys to a larger extent outperformed girls in math and science, the reverse gender gap in reading was also larger.

While the gender gap in math varies cross-nationally, it is also different across socio-economic contexts. With the data from the Beginning School Study conducted in Baltimore, Entwisle et al. (1994) showed that while mean performance of boys and girls was similar, the highest scoring boys from more affluent neighbourhoods outperformed the highest performing girls, while on the other end of the distribution the disadvantaged boys did worse compared to girls. This is consistent with the finding about greater male variability. Entwisle et al. explained the emergence of this variability by greater susceptibility of boys to the effects of neighbourhood resources. Boys that came from socially disadvantaged neighbourhoods were likely to fall behind in their studies, while the effect on girls was less negative.

In a later and more methodologically rigorous quasi-experimental study, Legewie and DiPrete (2012) found a similar effect with the data on socio-economic composition and reading test scores of pupils in some elementary and upper-secondary schools in Berlin. Generally, girls did better in reading than boys. However, in schools with pupils of higher socio-economic status (SES) and better average performance, the gender gap favouring girls was considerably smaller. Legewie and DiPrete explained this effect by the differential impact of school environment on boys and girls. Schools with lower average SES were characterized by a strong conception of masculinity in male peer culture that promoted anti-school attitudes. This anti-school male culture was less frequent in schools with higher SES. Girls did not have the same anti-school peer culture, and for them the effect of school environment on performance was weaker.

2 Data and methods

In 2009 Russia introduced compulsory universal state examinations (USE, known in Russia by the acronym EGE) that high school pupils must take in order to graduate. Russian and mathematics are compulsory USE subjects, and more subjects are required to be able to apply to most universities. Universities have to accept USE as entrance examinations. The centrally administered USE replaced an old system of separate school and universities exams, conducted locally.

The Russian school system has several tracks. First, there are several types of high schools that differ in terms of quality of education and socio-economic status of pupils. The majority of schools do not have any special status. Some schools, however, offer more advanced training in one or several subjects (such as mathematics, languages, etc.). Another type of schools, *lycees* and *gymnasiums* usually have more advanced curricula in several subjects, better funding and are generally considered to be educational institutions of better quality. On the other hand, less academically successful pupils often go to evening schools. Apart from these main types, there is a smaller number of schools for children with disabilities, military schools, etc.

School education up to the 9th grade is compulsory. After that pupils have a choice and can stay in school for another two years or leave and go to a vocational school. Staying in school for 10th and 11th grades is the academic track that usually leads to applying to a university. Vocational schools train for routine manual occupations (drivers, industrial workers, etc.) that usually takes two years, or service and lower professional occupations (nurses, primary school teachers, hospitality workers, etc.) that takes four years. After completing a vocational school graduates can change their track and apply to universities. USE is mainly taken by the 11th grade graduates, although a small number of students in vocational schools also take the exams in order to be able to continue their education in universities.

For the analysis in this paper we use the data set with the results of the USE in mathematics conducted in 2011. The data set contains valid records for the whole population of test takers (738,456 people). Most of them are 11th grade high schools pupils. Note that this is only part of the respective age cohort, as most pupils who left school after 9th grade and took the vocational track did not take USE. The number of pupils who finished 9th grade in 2009 was 1,178,500 (Rosstat, 2012). This is the same cohort of pupils who took USE in 2011, so the proportion of USE takers among the 9th grade graduates was roughly 0.63.

The USE mathematical test in 2011 consisted of two parts. The first part had twelve problems; each correct answer was graded one point. The second part had six more complicated problems in the ascending order of difficulty, and students were required to present full solutions. Problems in this part could be graded 2, 3 or 4 points. The exam covered the Russian high school curriculum for algebra and geometry. The maximum test score was 30. In order to standardize the scale for all subjects, the original score was rescaled by USE organizers to the 0 to 100 scale. In this paper, we use the original 30-point scale.

Apart from data on test scores and sex, the data set contains identifying variables for region, school and class, information on school location (urban vs. countryside) and school type. Descriptive statistics for all the variables is given in Table 3.

We measure gender differences in the USE math test using Cohen's d that is the difference in means in two groups divided by pooled within-group standard deviation.

$$d = \frac{\overline{Y}_b - \overline{Y}_g}{S_{pooled}}$$

$$S_{pooled} = \sqrt{\frac{(n_b - 1)S_b^2 + (n_g - 1)S_g^2}{n_b + n_g - 2}}$$

where $\overline{Y}_b$ and $\overline{Y}_g$ are mean scores for boys and girls, S_b^2 and S_g^2 are variances for boys and girls, and n_b and n_b are the number of boys and girls in the sample.

The difference in variability of test scores is given by variance ratio (VR).

$$VR = \frac{S_b^2}{S_g^2}$$

To compare the distributions of test scores for boys and girls beyond the differences in means and variances we employ the concept of relative distribution as suggested in Handcock and Morris (1999). We provide a brief description of this method in section 3.1.

The analysis at the aggregate level (section 3.2) is based on standard OLS regression. The analysis at the individual level applies OLS regression and quantile regression (Hao and Naiman, 2007; Handcock and Morris, 1999) to model the effects of predictors not only on the mean, but the entire USE test score distribution. Details are given in section 3.3. Quantile regression was estimated with the R package `quantreg` (Koenker, 2013). The relative distribution was estimated with the `reldist` package (Handcock, 2013).

3 Results

3.1 Characteristics of the test score distribution

Table 1 presents some descriptive statistics, the effect size and VR for the whole data set of USE test takers. The effect size, Cohen's d, is 0.05 that indicates that there is virtually no difference in mean test scores of boys and girls. The variance ratio (VR) is 1.12 showing somewhat greater variability of test scores for boys. Both findings are consistent with results of previous studies of the gender differences in mathematical achievement.

Figure 1 plots test score distributions for boys and girls. To make a comparison of two distributions beyond the differences in means and variances we employ the concept of a relative distribution suggested in Handcock and

Table 1: Statistics for the whole sample

mean test score (boys)	10.15
mean test score (girls)	9.91
n (boys)	330111
n (girls)	408345
ratio (b/g)	0.81
Cohen's d	0.05
variance ratio	1.12

Figure 1: Test score distributions for boys and girls

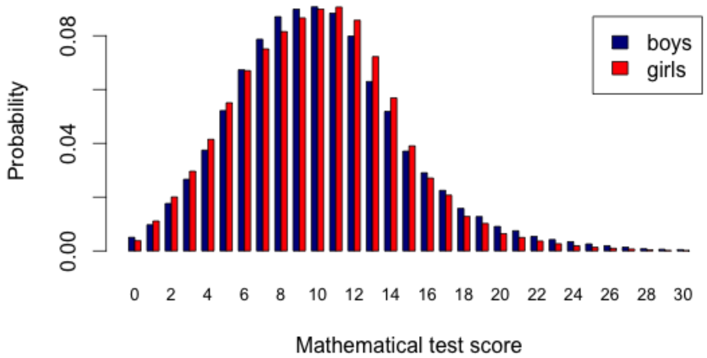
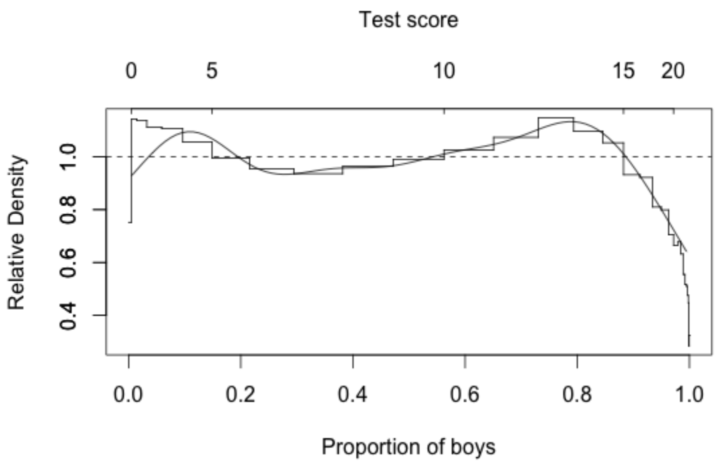


Figure 2: The relative distribution of mathematical test scores. Boys are the reference group



Morris (1999). Figure 2 shows the probability density function (pdf) for the relative distribution.

A relative distribution requires a reference group (in our case, boys) and a comparison group (girls). It shows what would be the rank of an observation taken from the comparison group in the reference group. For example, the relative distribution shows where in the male distribution would be a girl with the test score, say, 10. The pdf of the relative distribution is simply the ratio of pdfs for comparison and reference groups. So figure 2 shows the ratio of the heights of blue and red bars from figure 1, plotted against the proportion of boys in the test score distribution.

If two distributions were identical the relative pdf would simply equal one throughout the reference distribution. However, we see on the plot that in some parts of the distribution relative density is clearly different from one. Girls are more likely than boys to be in parts of the distribution where relative density is greater than one and less likely to be in the parts of the distribution where it is less than one. The plot shows that there is a higher chance for girls to score under 5 points and from 11 to 15 points. The probability of scoring in the range between 5 and 11 points is only very slightly higher for boys. However, boys have a considerably higher probability of being in the top part of the distribution (starting from 16 points) and this male advantage increases closer to the very top of the distribution. Note that only 3% of test takers get scores of 20 points and higher where male advantage is most evident.

As shown in Table 1, the ratio of the number of boys to the number of girls in the data set is 0.81. As the proportions of both sexes in the whole population in this age group are about equal, this indicates different rates of transition from the 9th to 10th grade for boys and girls. If we assume that selection to 10th grade is at least partly based on academic ability, then, given that a larger proportion of boys leave school after 9th grade, the estimate of d in our sample is likely to be biased against girls. Given our data, it is impossible to estimate the strength of this bias. However, it is unlikely to affect the top of the test score distribution as best students stay in the academic track and do not go to vocational schools.

3.2 Determinants of the gender differences in math at the regional level

Figure 3 that plots d on the map of Russia shows that it varies from -0.17 to 0.16. Note that the map shows clear geographic patterns in the distribution of d. There are several regions where girls on average performed better than boys at the test. These are most ethnic republics in the North Caucasus, some ethnic republics in the Volga region (Chuvashia, Mari El and also neighbouring Mordovia and ethnically Russian Tambov and Penza *oblasti*), several regions in Southern Siberia, mostly ethnic republics (Altay, Buryatiya, Tuva and also Altay kray and Zabaykalsky kray). On the other hand, many regions where boys have the greatest advantage over girls are concentrated in the north: Murmansk, Vologda, Novgorod, Leningrad *oblasti*, and the Komi Republic.

Figure 3: The effect size d across Russian regions

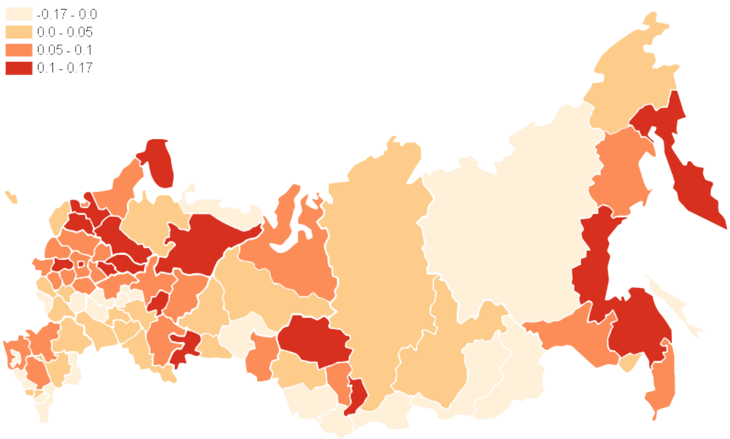


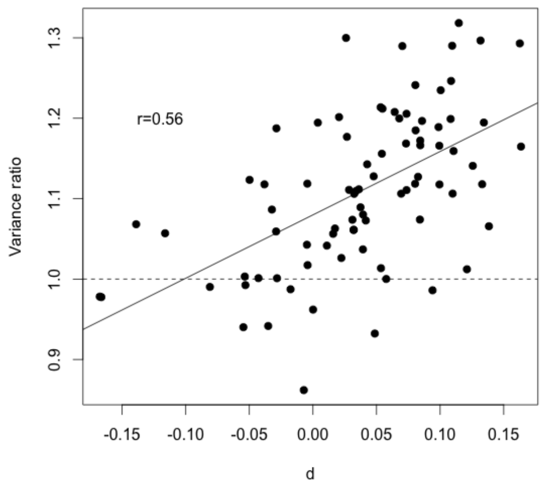
Figure 4 presents a scatter plot of d and VR at the regional level. In most regions VR is greater than one; however, there are some exceptions. Note positive correlation between d and VR ($r = 0.56$). This is opposite to the results reported with the TIMSS data by Kane and Mertz (2012) who found strong negative correlation between d and VR at the country level.

To establish the determinants of regional differences in the gender gap in math, we regress d and VR on two predictors: the ratio of the number of boys taking the USE to the number of girls and the proportion of pupils from schools located in the cities among the USE takers. The results are presented in Table 2.

The ratio of the number of boys to the number of girls reflects differential rates of transition of boys and girls to 10th grade. The population in this age cohort has about an equal number of boys and girls. However, boys are more likely to choose the vocational track and less likely to stay in school after 9th grade and take USE. Hence, there are more girls than boys in the USE data. However, the exact ratio varies across the regions, and in some regions there are actually more boys among test takers. For instance, in Yaroslavl and Novgorod *oblasti* the boys-to-girls ratio is 0.7, while in Dagestan it is 1 and in Chechnya 1.2.

As the results presented in Table 2 indicate, the more boys stay in school after the 9th grade, the lower is the mean test score for boys in the region and the smaller is the size of male advantage in math. In fact, in the regions with the highest boys-to-girls ratios, girls often performed on average better

Figure 4: d and variance ratio at the regional level



	d	VR
(Intercept)	0.04***	1.11***
	(0.01)	(0.01)
b/g ratio (st.)	−0.02***	0.00
	(0.01)	(0.01)
proportion urban (st.)	0.03***	0.06***
	(0.01)	(0.01)
Adj. R^2	0.47	0.31
n	83	83

***$p < 0.001$; standard errors are in parentheses
The predictors were standardized with mean of zero and standard deviation of one

Table 2: Region-level predictors of d and VR

than boys on the test. As selection to 10th grade is driven by academic ability, it is not surprising that in the regions where more boys of lower academic ability stay in school, mean performance of boys is weaker.

Note that the boys-to-girls ratio is not associated with regional differences in VR. Indeed, higher variance of male scores is mainly explained by higher proportions of boys at the top of the test score distribution. These are the most academically successful pupils who are not at risk of leaving school after the 9th grade.

Controlling for the sex ratio, the regional proportion of pupils from urban schools is another statistically significant predictor of d. In more urbanized regions there is a higher chance for boys to do better in math than girls. One standard deviation change in the variable that measures the proportion of urban test takers is associated with the change of d by 0.03. This is a stronger effect than the effect of sex ratio.

The descriptive analysis of the distribution of d across regions indicates that d tends to be smaller in ethnic republics. There is indeed positive correlation between d and the proportion of ethnically Russian population in the region (according to the 2010 census). However, as ethnic republics tend to be less urbanized, this correlation disappears after controlling for the proportion of pupils from urban schools.

3.3 School and classroom context and the gender gap in math

In this section we model the association between the gender differences in math and a number of characteristics measured at the school and classroom levels. In particular, we look at how the gender gap in math depends on the type of school, its location, classroom and school size, and the proportion of boys in school and in classroom.

As shown in section 3.1, mean performance of boys and girls is about equal, but there are other significant distributional differences. In particular, boys are more likely than girls to be among the top achievers (hence the higher variance for boys). To model not only conditional mean test score, but also differences at various quantiles of the distrbution we apply quantile regression (Hao and Naiman, 2007).

We model the USE math test score as a function of sex, school type, school location, school size, classroom size, proportions of boys in school and in classroom and the interaction effects between sex and all other predictors. All quantitative predictors have been standardized with mean zero and standard deviation one. We excluded from the analytical sample pupils from schools for children with disabilities and other non-standard types of schools, and also schools and classrooms with only one pupil, classrooms with more than 40 pupils and schools with more than 200 pupils taking the test (these are likely to be coding errors). We also excluded all observations

	n	%
Sex		
Male	292928	44
Female	371557	56
School type		
High school	491802	74
Specialized high school	39852	6
Lycee/gymnasium	93831	14
Evening school	21311	3
Vocational school	17689	3
School location		
Countryside	196812	30
City	467673	70
	mean	**standard deviation**
n pupils in the school	38.7	27.8
n pupils in the classroom	19.3	7.5
proportion boys in the school	0.44	0.14
proportion boys in the classroom	0.44	0.17

Table 3: Descriptive statistics for the analytical sample

with missing values for any of the variables. The final estimation sample size was 664,485 pupils (out of original 738,459).[1] Descriptive statistics for the analytical sample is presented in Table 3.

The results of the OLS and quantile regressions (estimated at 0.1, 0.5, 0.9 and 0.95 quantiles) are presented in Table 4. The results of quantile regressions (estimated at every 0.05 quantile) for selected effects (sex, school type, school location and interaction effects between sex and school type and location) are also graphically presented in Figure 5. As we are interested in the gender differences in test scores rather than predictors of test scores, we mainly interpret the effects for sex and interaction effects between sex and other predictors.

To interpret the results, we look at the coefficients for regression equations presented in Table 4. We set the number of pupils and the proportions of boys in the school and the classroom at the mean level in our analytical sampe (i.e., at zero, according to our coding). Then, for example, the predicted expected male test score in ordinary high schools in the countryside is simply 10.08 (the intercept in the OLS model). For girls at the same values of other independent variables, the predicted test score in the OLS model is 10.31 (10.08+0.23). So according to our model, in an average ordinary high school in the countryside girls slightly outperform boys.

Note, however, that the effect varies throughout the distribution. It is easy to see this from the plot for the main effect of sex in Figure 5. Taken by itself, it shows how the effect of female gender changes throughout the distribution for ordinary high schools in the countryside. At 0.1 quantile, boys and girls perform about the same (the effect is zero). In other words, the thresholds that separate the bottom 10% of girls and the bottom 10% of boys from the rest of the male and female distributions are about the same. At larger quantiles girls do better than boys, with the maximum female advantage at 0.6 quantile. Then female advantage diminishes, and 0.95 quantiles for boys and girls are about the same. In high schools in the countryside at no point throughout the distribution boys perform at the test better than girls.

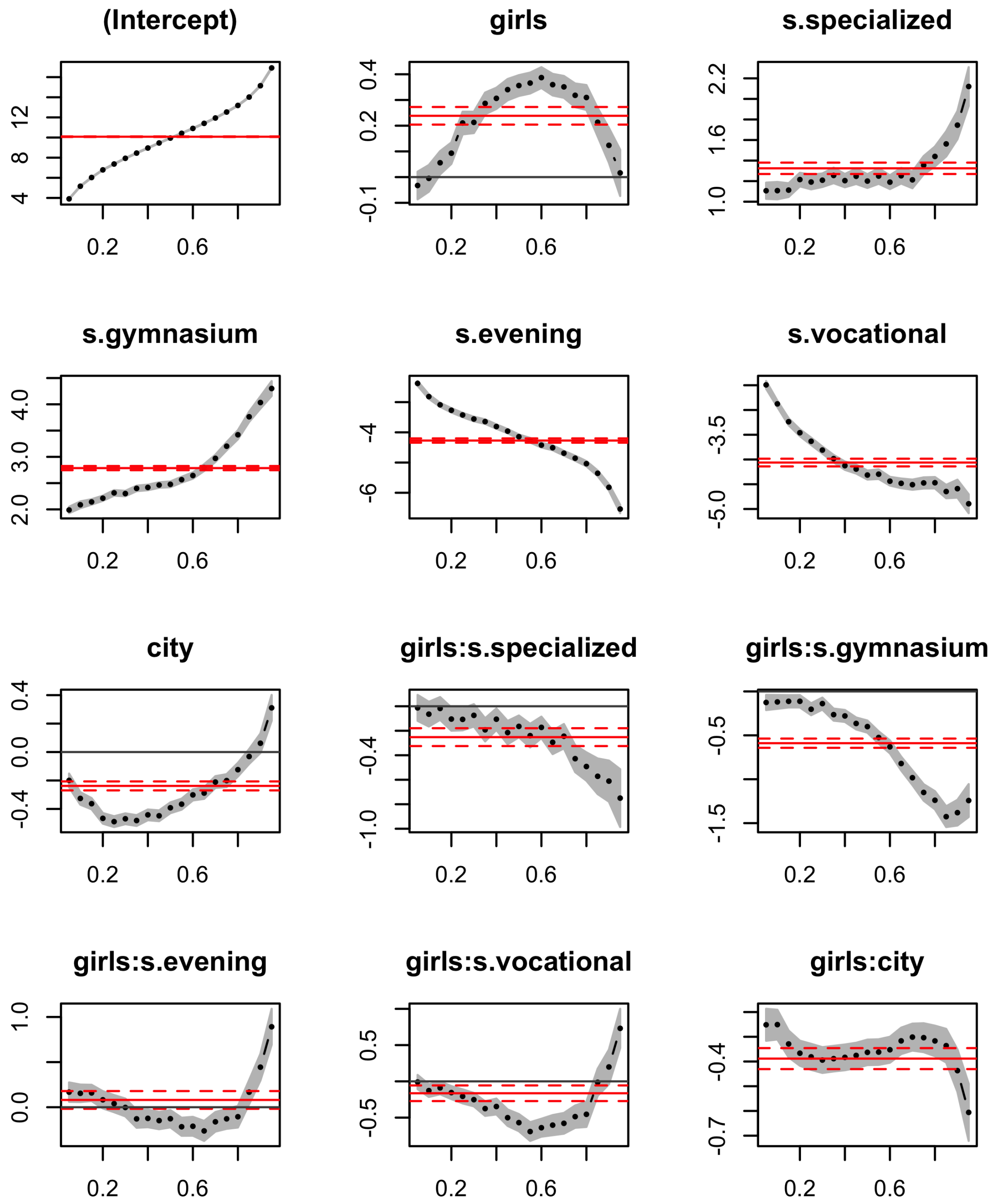
Once we move from the countryside to cities, the picture changes. In the OLS model the effect of female gender in high schools in the cities is -0.13 (0.23-0.39), holding the number of pupils and proportions of boys at their means. This is a weak effect, given the 30-point scale for test scores. However, the effect gets larger at the top of the distribution. See plots for the effects "girls" and "girls:city" in Figure 5. Starting from 0.8 quantile,

[1] We estimated the models without restricting the sample and the results were very similar. We also estimated the OLS model with fixed effects for Russia's 83 regions. Compared to the model that did not control for region, this influenced some of the main effects (in particular, for school location), but the interaction effects remained unchanged. As estimating quantile regression models with fixed effects for regions was computationally difficult, we presented the models without fixed effects.

Covariates	0.1	0.5	0.9	0.95	OLS
Intercept	5.16 (0.02)	9.94 (0.02)	15.15 (0.03)	16.92 (0.04)	10.08 (0.02)
Girls	0.00 (0.03)	0.36 (0.02)	0.12 (0.04)	0.02 (0.05)	0.23 (0.02)
School type (ref. high school)					
Specialized	1.11 (0.05)	1.20 (0.04)	1.74 (0.08)	2.12 (0.11)	1.32 (0.03)
Lycee/gymnasium	2.09 (0.04)	2.48 (0.03)	4.03 (0.07)	4.30 (0.09)	2.79 (0.102)
Evening	−2.82 (0.04)	−4.14 (0.04)	−5.82 (0.06)	−6.54 (0.08)	−4.27 (0.04)
Vocational	−2.88 (0.04)	−4.31 (0.05)	−4.59 (0.11)	−4.89 (0.12)	−4.06 (0.05)
City	−0.33 (0.03)	−0.39 (0.02)	0.06 (0.04)	0.31 (0.06)	−0.23 (0.02)
n pupils in school (st.)	0.14 (0.01)	0.29 (0.01)	0.67 (0.03)	0.80 (0.03)	0.36 (0.01)
Proportion of boys in school (st.)	−0.07 (0.02)	−0.18 (0.02)	−0.25 (0.03)	−0.24 (0.05)	−0.18 (0.01)
n pupils in classroom (st.)	0.47 (0.01)	0.47 (0.01)	0.46 (0.02)	0.46 (0.03)	0.48 (0.01)
Proportion of boys in classroom (st.)	0.05 (0.02)	0.30 (0.02)	0.58 (0.04)	0.64 (0.05)	0.35 (0.01)
Interaction effects:					
Girls*specialized	−0.06 (0.06)	−0.16 (0.05)	−0.61 (0.10)	−0.75 (0.14)	−0.25 (0.04)
Girls**Lycee/gymnasium*	−0.12 (0.05)	−0.40 (0.04)	−1.38 (0.09)	−1.24 (0.11)	−0.59 (0.03)
Girls*evening	0.15 (0.06)	−0.13 (0.06)	0.44 (0.09)	0.89 (0.12)	0.08 (0.06)
Girls*vocational	−0.13 (0.05)	−0.57 (0.07)	0.20 (0.14)	0.73 (0.16)	−0.16 (0.07)
Girls*city	−0.25 (0.04)	−0.36 (0.03)	−0.44 (0.05)	−0.61 (0.07)	−0.39 (0.03)
Girls * n school	−0.01 (0.02)	−0.03 (0.02)	−0.04 (0.03)	−0.01 (0.04)	−0.02 (0.01)
Girls * pr. boys in school	−0.02 (0.02)	−0.05 (0.02)	−0.27 (0.04)	−0.29 (0.05)	−0.12 (0.02)
Girls * n classroom	−0.03 (0.01)	−0.09 (0.02)	−0.19 (0.03)	−0.20 (0.04)	−0.12 (0.01)
Girls * pr. boys in classroom	0.10 (0.02)	0.13 (0.02)	0.27 (0.04)	0.32 (0.06)	0.16 (0.02)
n	664,485	664,485	664,485	664,485	664,485

Table 4: OLS and quantile regressions of USE test scores on school- and classroom-level variables. Standard errors (estimated with the kernel estimate of the sandwich method as provided in the package `quantreg` in R) are in parentheses. Quantitative variables were standardized with mean of zero and standard deviation of one

Figure 5: Quantile regression of USE math test score on sex and school characteristics



The dots represent the estimates from quantile regression estimated at every 0.05 percentile; the grey areas represent 95% confidence intervals. The red lines show the OLS estimate with 95% confidence intervals. The model also includes all other predictors from Table 4.

both curves on the plots go down, indicating larger female disadvantage. For example, 0.95 quantile for girls in urban high schools is predicted to be 16.64 (conditional on the number of pupils and proportions of boys being set at their mean level), while for boys it is 17.23, so that the difference is 0.59.

The main effects for school type establish a clear academic hierarchy. *Lycees* and *gymnasiums* have the best results in the USE math test, followed by specialized schools and then ordinary high schools. Pupils from evening and vocational schools perform much worse.

The interaction effects between sex and school type show that the size of the gender gap in math depends on the type of school. For example, while in cities the difference between male and female mean test scores in ordinary high schools is 0.13, in specialized schools it is 0.41 and in *lycees* and *gymnasiums* 0.75. In evening schools the mean test scores are about equal (the effect of female sex is -0.08) and in vocational schools boys perform somewhat better (0.32). All these gender differences were calculated with the OLS coefficients presented in Table 4.

A richer pattern can be noticed, though, once we look at the top of the distribution. Visual analysis of plots "girls", "girls:city", "girls:s.specialized" and "girls:s.gymnasium" shows that at the top of the distribution all curves go down, especially steeply for the interaction effect "girls:s.gymnasium".

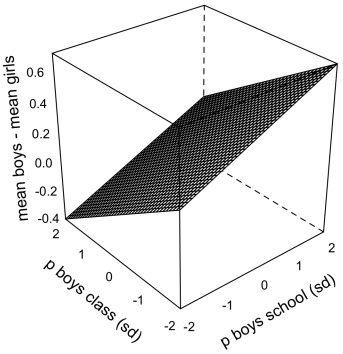
As mentioned before, the difference between boys and girls at 0.95 quantile in ordinary urban high schools is 0.59. However, in urban specialized schools it increases to 1.34 and in urban *lycees* and *gymnasiums* to 1.83. These are substantial differences. This shows that in schools with the highest level of performance and higher socio-economic status of pupils the top 10% of boys strongly outperform the top 10% of girls, while the gender difference in mean scores remains smaller.

Evening and vocational schools show a different pattern. There is not much gender difference in test performance in any part of the distribution. At 0.95 quantile in urban evening schools girls outperform boys by 0.3, and in urban vocational schools by 0.14.

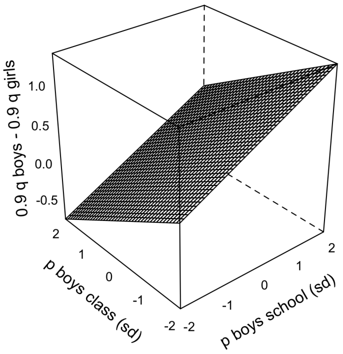
In the countryside male advantage at the top of the distribution in specialized schools and *lycees* and *gymnasiums* becomes somewhat smaller (compared to cities).

So far all these estimates were given for schools and classrooms of average size with the average proportion of boys. We also estimate the interaction effects between sex and school size, classroom size and the proportion of boys in school and in classroom (coefficients shown in Table 4).

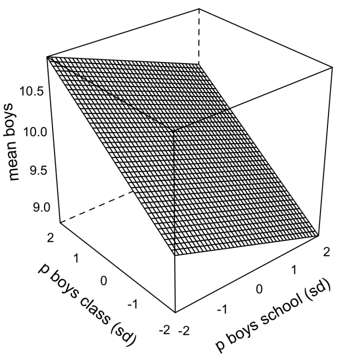
There is virtually no effect of the school size on the gender differences in test scores (after controlling for the classroom size). In bigger classes, girls perform slightly worse than boys, although the effect is small. The OLS estimate for the interaction effect between female gender and the number of pupils in the classroom is -0.12 for one standard deviation change in the



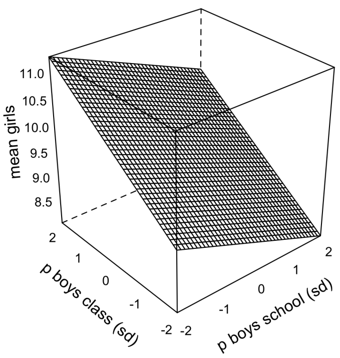
(a) gender differences in means



(b) gender differences at 0.9 quantile



(c) effect on mean male score



(d) effect on mean female score

Figure 6: Conditional effects of the proportion of boys in school and in classroom. Estimated from the model presented in table 4. Other variables set at the following levels: high school, city, school and classroom of mean size. Proportions of boys were standardized with mean of zero and standard deviation of one

number of pupils in the class (controlling for school size). In other words, girls tend to perform slightly worse (and boys tend to perform slightly better) in the schools where the same number of pupils are divided into a smaller number of classes and classes are larger.

The coefficients for the proportions of boys in school and in classroom jointly indicate that in classes with a higher proportion of boys girls perform relatively better, while in more female classes they perform relatively worse. To illustrate this point, we produced a figure that shows the joint effects of the proportions of boys in school and in classroom on the difference in means between boys and girls, on the gender difference at 0.9 quantiles and also on male and female mean test scores (see Figure 6). Other variables in the model were set at the following values: city, ordinary high school of average size (number of pupils in the school about 39), average classroom size (number of pupils in the class about 19).

The figures show that the higher is the proportion of boys in a school, the lower are the mean test scores and the higher is the gender difference in scores in boys' favour. The effect of the proportion of boys in a classroom is exactly the opposite. Note, however, that when added in the model together these two variables show the effects of gender stratification between classes within schools. In schools with the same proportion of boys some classes have a higher proportion of boys compared to the average school level, and other classes have a higher proportion of girls. Classes with a higher proportion of boys have higher mean test scores and at the same time lower male advantage in scores (actually, according to the model, in mainly male classes in mainly female schools girls would perform better than boys). On the other hand, in classes with a higher proportion of girls (compared to school average) mean test scores are lower and at the same time male advantage is larger. Note that the effects at 0.9 quantile are somewhat more pronounced than the effect in the OLS model. However, in general the effects are not very large, especially given that the proportions of boys in school and in class covary ($r = 0.41$).

Our analysis shows that there are substantial differences in the size of the gender gap in mathematical performance depending on a number of school- and classroom contextual characteristics. Boys are more likely to outperform girls in math in cities, in schools of better quality, in larger classes and in classrooms with a higher proportion of girls compared to the school average. These effects are more pronounced at the top of the test score distribution than at the mean. For example, the predicted boys' advantage in the mean test score in urban *lycees* and *gymnasiums* (schools and classrooms of mean size, with the proportion of boys in school 0.44 and the proportion of boys in classroom 0.27) is 0.91 (at 0.95 quantile it is 2.15). At the same time, the expected gender difference in means in small high schools in the countryside (one classroom in the final year that consists of four boys and six girls) is about 0.37 in girls' favour (0.23 at 0.95 quantile).

4 Discussion

Our analysis yields several conclusions. First, we find with the Russian data that the gender difference in mean mathematical test scores is negligible ($d = 0.05$). However, boys show a greater variance of test scores than girls (VR=1.12). Both findings are consistent with the results previously reported in the studies based on US and international data. The size of d that we find in Russia is exactly the same as in a recently published cross-national meta-analysis (Lindberg et al., 2010).

Gender equality in mean mathematical performance in the USA, Russia and a number of other countries disproves a popular myth that boys are on average more talented in math than girls. While some studies show that boys did better than girls in math in the 1990s (Penner, 2008; Marks, 2008), more recent data do not allow to make this conclusion (Else-Quest et al., 2010; Kane and Mertz, 2012; Lindberg et al., 2010).

However, a greater male variability in mathematical test scores that we find with the Russian data is a more robust phenomenon reported in many studies (Hedges and Nowell, 1995; Machin and Pekkarinen, 2008; Hyde et al., 2008). It is mainly the result of a higher proportion of boys at the top of the test score distribution (see also Ellison and Swanson, 2010). Another study reported that a greater male variance in intelligence test scores was found as early as age 3 to 10 (although among two-year old children girls showed a high mean and a higher variance than boys) (Arden and Plomin, 2006). It is unlikely that gender stereotyping had an effect on the variance of children's abilities at that early age. Arden and Plomin (2006) suggest genetic, environmental and evolutionary explanations for greater male variability, an idea that goes back to Darwin.

Note that we only have data on about 60% of children in the age cohort, mainly those who completed the 11th grade in high school. Children with lower academic performance often choose the vocational track after the 9th grade. Most of them do not take USE. The vocational track is more common for boys than for girls, and the ratio of boys to girls in our data is about 8 to 10. Hence the results are probably biased against girls, and the size of this bias is unclear. However, the bias is likely to affect means to a high extent than variances, as pupils from the top of the test score distribution rarely choose the vocational track.

The effect of gender differences in transition rates to the 10th grade is confirmed with the analysis at the regional level. We find correlation between regional differences in the gender ratio in our data and d so that the more boys stay in high school compared to other regions, the smaller is d (in some regions it is negative). On the other hand, there is no effect of the gender ratio on VR.

Besides, we find that in more urbanized Russian regions male advantage in mean test scores is higher. We further explore the effects of contextual characteristics on the gender gap in math with the analysis at the individual level, applying quantile regression to model the effects throughout the distribution.

The analysis shows that the size and direction of the gender gap in math vary depending on school type. In schools of better quality with more advanced curricula (specialized schools, *lycees* and *gymnasiums*) boys perform at the USE math test better than girls, with the gender gap being larger at the top of the distribution. On the other hand, in ordinary high schools, evening and vocational schools gender differences either in means or at the top of the distribution are much smaller or even reversed.

Why is it so? One possible answer is simply selection. Children are not distributed randomly across different types of schools. Specialized schools, *lycees* and *gymnasiums* select more talented pupils who also tend to come from families of higher socio-economic status. Some schools in this group only take pupils after primary school or even at a later stage when their abilities can be tested. If we assume that boys have a more dispersed distribution of ability before coming to school (either naturally or as a result of early-life environmental effects) and hence are more represented at the top of the distribution and better schools attract the best students, then it is not surprising that we only find substantial gender differences in certain types of schools.

However, it is doubtful that selection works perfectly and most children with better mathematical abilities get into schools with more advanced training. Cleary, the USE test score is not a good measure of pre-school mathematical abilities. Still, while the majority of pupils scoring at the USE test 25 and higher came from specialized schools and in particular *lycees* and *gymnasiums*, 33% of these pupils were from ordinary high schools.

Another explanation is school effect. Legewie and DiPrete (2012) demonstrated a causal effect of classroom's average socio-economic status on the gender gap in reading performance. In classes with a higher average socio-economic status of students the advantage of girls in reading test scores was smaller. The proposed mechanism for this effect was male peer culture that in classes with lower average SES discourages the investment of efforts into the learning process among boys.

Another possible mechanism is gender stereotyping, either internalized or enforced by teachers and peers. In schools with advanced curricula in math the requirements for girls may be lower if they are considered to be naturally less able. As a result of gender stereotyping, girls themselves may feel less able for advanced mathematical training.

The gender differences in USE math test scores also depend on school location. Controlling for the type of school, in urban schools male advantage is larger, with the effect being stronger at the top of the distribution. As with school type, selection can be a possible explanation. If children who are naturally more talented in math are more likely to live in cities and boys have a higher variance of abilities than girls before school, then we would expect a greater gender gap favouring boys in urban schools even without any additional effect of location. Note that pupils in urban schools do perform at the test better than pupils in the countryside, although the difference in means is small (0.44).

Another explanation is an additional effect of school quality. School type is not the only indicator of school quality, and we would expect *lycees* and *gymnasiums* in cities to be better than in the countryside. Then the peer effects mechanism proposed by Legewie and DiPrete (2012) can contribute to the observed effect. There may also be a separate effect of location. Entwisle et al. (1994) suggest that boys' school performance can be affected by resources available in the neighbourhood. For example, in the countryside boys can do more housework than in cities and this can affect their performance in school.

Given the cross-sectional design of our study, we are unable to separate convincingly the effects of selection and school context. Thus, our findings remain mainly descriptive and the mechanisms we suggest cannot be tested. Further studies are required to demonstrate if and to what extent there is a causal effect of school context on the gender gap in test performance in Russia and to clarify the mechanisms for it. Note that the effects of selection and school context may be at work simultaneously and interact. If male variance of pre-school math abilities is greater than female and there are more boys at the top of the distribution, these differences may be further amplified in school as a result of differential treatment of girls and boys by teachers and gender differences in mathematics self-concept.

We also establish the association between the gender gap in math, the size of the classroom and the gender ratios in the school and the classroom. The analysis shows the when the same number of pupils in the school is divided into a smaller number of classrooms (so that the classrooms are larger), mean test scores are higher and the male advantage in test scores is larger. When within the same school there exists gender stratification across classrooms (so that some classrooms have a higher proportion of boys than others), classrooms with a higher proportion of boys demonstrate higher mean test scores and at the same time lower male advantage in performance. In contrast, classes with a higher proportion of girls have lower mean scores and a greater gender gap in scores favouring boys. With our research design, we are unable to establish causality for these effects and to find a credible interpretation for them. Further research is required to provide more evidence (perhaps of qualitative nature) on the association between gender ratios in the classroom, number of pupils and the gender differences in mathematics performance.

References

Arden, R. and R. Plomin. 2006. "Sex differences in variance of intelligence across childhood." *Personality and Individual Differences* 41:39–48.

Baker, D.P. and D.P. Jones. 1993. "Creating gender equality: Cross-national gender stratification and mathematical performance." *Sociology of Education* 66:91–103.

Buchmann, C., T.A. DiPrete, and A. McDaniel. 2008. "Gender inequalities in education." *Annu. Rev. Sociol.* 34:319–337.

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* 24:109–128.

Else-Quest, N.M., J.S. Hyde, and M.C. Linn. 2010. "Cross-national patterns of gender differences in mathematics: A meta-analysis." *Psychological Bulletin* 136:103–127.

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

Fryer, R. and S. Levitt. 2010. "An empirical analysis of the gender gap in mathematics." *American Economic Journal: Applied Economics* 2:210–40.

Geary, D.C. 1996. "Sexual selection and sex differences in mathematical abilities." *Behavioral and Brain Sciences* 19:229–246.

Guiso, L., F. Monte, P. Sapienza, and L. Zingales. 2008. "Culture, gender, and math." *Science* 320:1164–1165.

Handcock, M.S. 2013. *Relative Distribution Methods*. Los Angeles, CA. Version 1.6-2. Project home page at urlhttp://www.stat.ucla.edu/ handcock/RelDist.

Handcock, M.S and M. Morris. 1999. *Relative Distribution Methods in the Social Sciences*. Springer.

Hao, L. and D.Q. Naiman. 2007. *Quantile Regression*. Thousand Oaks: Sage.

Hedges, L.V. and A. Nowell. 1995. "Sex differences in mental test scores, variability, and numbers of high-scoring individuals." *Science* 269:41–45.

Hyde, J.S., E. Fennema, and S.J. Lamon. 1990. "Gender differences in mathematics performance: a meta-analysis." *Psychological Bulletin* 107:139–155.

Hyde, J.S., S.M. Lindberg, M.C. Linn, A.B. Ellis, and C.C. Williams. 2008. "Gender similarities characterize math performance." *Science* 321:494–495.

Hyde, J.S. and J.E. Mertz. 2009. "Gender, culture, and mathematics performance." *PNAS* 106:8801–8807.

Kane, J.M. and J.E. Mertz. 2012. "Debunking myths about gender and mathematics performance." *Notices of the AMS* 59:10–21.

Kimura, D. 2000. *Sex and Cognition*. Cambridge: MIT press.

Koenker, R. 2013. *quantreg: Quantile Regression*. R package version 4.96.

Leahey, E. and G. Guo. 2001. "Gender differences in mathematical trajectories." *Social Forces* 80:713–732.

Legewie, J. and T.A. DiPrete. 2012. "School context and the gender gap in educational achievement." *American Sociological Review* 77:463–485.

Lindberg, S.M., J.S. Hyde, J.L. Petersen, and M.C. Linn. 2010. "New trends in gender and mathematics performance: A meta-analysis." *Psychological Bulletin* 136:1123.

Machin, S. and T. Pekkarinen. 2008. "Global sex differences in test score variability." *Science* 322:1331–1332.

Marks, G.N. 2008. "Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries." *Oxford Review of Education* 34:89–109.

Penner, A.M. 2008. "Gender differences in extreme mathematical achievement: An international perspective on biological and social factors." *American Journal of Sociology* 114:S138–S170.

Pope, D.G. and J.R. Sydnor. 2010. "Geographic variation in the gender differences in test scores." *Journal of Economic Perspectives* 24:95–108.

Rosstat. 2012. *Regions of Russia. Social and economic indicators 2012 [Regiony Rossii: Sotsialno-ekonomicheskie pokazateli 2012]*. Russian Statistical Office.

Spelke, E.S. 2005. "Sex differences in intrinsic aptitude for mathematics and science? A critical review." *American Psychologist* 60:950–958.

Alexey Bessudnov (corresponding author)
Department of Sociology, Centre for Advanced Studies,
National Research University – Higher School of Economics
3 Kochnovsky proezd Moscow 125319 Russia
abessudnov@hse.ru