Mind the Gap: Evidences from Gender Differences in Scores in Brazil

Juliana Guimarães*, Breno Sampaio†

July 21, 2008

Abstract

There is always been a claim that: "boys are good at counting, and girls are good at reading". Analysis of the data on the gender differences in mathematics achievement up to the present indicates that females are not consistently under-performing relative to males. Guiso, Monte, Sapienza and Singales (2008) analyzed data for 40 different counties and showed that culture explains most of the difference in maths, at least. Also, the gap in mathematics scores between boys and girls virtually disappears in countries with high levels of sexual equality. In this paper we want to investigate the gender score gap using data from Entrance Test Scores for Universidade Federal de Pernambuco (UFPE) in Brazil. Our results show that in fact, the gender gap in scores varies across subjects and also across quantiles of the conditional score distribution. Boys may show an advantage in Math, but the advantage is lower as we control for family background and academic variables. This shows that girls disadvantage in math may be in fact related to family background and cultural factors, as confirming the results by Guiso, Monte, Sapienza and Singales (2008). Girls have an advantage when it comes to language, and to some extent to Portuguese, confirming the results found on the international literature that girls out-perform boys in reading and writing.
Key words: Gender Score Gap, Quantile Regression, Brazil.

Resumo

Existe a percepção que "homens são bons em cálculo e mulheres são boas em leitura". Diversos estudos analisando desempenho de estudantes constatam que mulheres não apresentam desempenho inferior aos homens em matemática. Guiso, Monte, Sapienza e Singales (2008), analisando dados de 40 países, constatam que diferenças culturais podem explicar grande parte das diferenças de nota em matemática e que essa diferença quase desaparece em países menos discriminatórios em relação a gênero. Neste artigo, analisam-se diferenças de nota entre gêneros utilizando dados do vestibular da Universidade Federal de Pernambuco (UFPE). Os resultados mostram que a diferença de notas entre gênero varia de acordo com a disciplina, assim como esse efeito muda nos quantis da distribuição condicional da nota. Os homens têm melhor desempenho em matemática, porém esta vantagem diminui quando controla-se para background familiar e outros fatores. Isso mostra que o desempenho inferior das mulheres pode estar relacionado a fatores culturais, o que confirma em parte os resultados obtidos por Guiso et al (2008). As mulheres têm vantagem em línguas, o que confirma os resultados obtidos na literatura internacional onde as mulheres se sobressaem na leitura e na escrita.
Palavras-Chave: Diferencas de Desempenho entre Generos, Regressao Quantilica, Brasil.

JEL Cassification #: J01, J24, C13. *Área 11: Economia Social e Demografia Econômica*

*Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge,CB3 9DD, Tel: +44.01223.335272, jgc29@cam.ac.uk

†University of Illinois at Urbana-Champaign, Department of Economics, sampaio2@express.cites.uiuc.edu. All the remaining errors are our responsibility.

1 Introduction

There is always been a claim that girls perform significantly worse, on average, than boys do on mathematical tasks, activities and examinations. Basically, "boys are good at counting, and girls are good at reading". Some authors claim that in the research literature it is often taken for granted that there is a problem concerning gender and mathematics (Walkerdine, 1998). The explanations for the gender-gap on scores vary from biological to cultural differences explanations related to gender-biased environments (Baron-Cohen, 2003, Hausmann, Tyson and Zahidi, 2006).

In a recent paper at *Science,* Luigi Guiso from European University Insitute, Ferndinando Monte and Luigi Zingales, both from Chicago Univeristy and Paola Sapienza from Northwestern University, studied the gender differences in test performance across countries using data from the 2003 Program for International Student Assessment (PISA), that reports over 276 thousand 15-year-old students from 40 countries who took identical tests in mathematics and reading. They show that on average, girls scored 2% bellow boys in math. However, the gap was largest in countries with the least equality between the sexes (by any score), such as Turkey. Moreover, it vanished in countries such as Norway and Sweden, where gender inequality almost does not exist. The gender gap was also reversed in reading. On average girls out-performed boys by 6%. In Turkey girls reading scores were 25.1 higher and in Iceland 61.0 higher. An interesting fact was that where girls enjoyed the strongest advantage in reading, they exhibited the smallest disadvantage in math. Their results suggest that the gender gap in math, although historically in favor of boys, disappears in more gender-equal societies.

In this paper we want to investigate the gender score gap using data from Entrance Test Scores for Universidade Federal de Pernambuco (UFPE) in Brazil. Universidade Federal de Pernambuco (UFPE) is the major University in Northeast of Brazil[1]. The data set brings information on their standardized entrance test scores, personal characteristics, such as age, gender, race, religion, family background, and high school attended. The students are supposed to take identical tests on mathematics (algebra and geometry), portuguese (grammar, literature and writing), language (they can choose either Spanish, English or French), geography, history and political science. If they reach a certain score than they are eligible to perform the second part of the test that contains specific material depending on the college their are applying for (medical school, social science, engineering, etc.).

We took the scores on math, portuguese and language to assess the existence of a possible gender gap. Using a female indicator variable and variables on family background and related variables on academic achievement at school, we regressed them on log scores in order to check for the existence of gender gap. Moreover, we used both Ordinary Least Squares (OLS) and quantile regression in order to show differences in patterns on gender gap across the conditional score distribution.

2 Mind the Gap

Gender score gap has always been a controversial issue. Analysis of the data on the gender differences in mathematics achievement up to the present indicates that females are not consistently underperforming relative to males. These results are different across age groups and across countries. Data for the UK show that girls did slightly better than boys in the national mathematics test

[1]Established in 1946, the Universidade Federal de Pernambuco is according to the Education Ministry the major University on the North/Northeast regions of Brazil. The courses are offered by 10 colleges of four different areas which consist of 67 Departments. The University offers 62 different undergraduate courses, 17 of them are lectured at night. The University also has 108 post graduate courses, among masters, PhD's and MBA's. In 2004 the University had 25,000 students registered (20,500 were undergrads and 4,500 were graduate students) and 1,647 professors.

for 14-years old trialed in 1991 (Hacket, 1993). From 1994 onwards, girls have equalled and then overtaken boys at age 16 in GCSE mathematics examination. It is worth noticing that in Britain from 1988 on, mathematics has been compulsory in maintained schools, up to the age of 16. After the age of 16 all education is voluntary, and hence differences in participation rates in mathematics emerge. So in overall terms, the British problem might be claimed by some authors to have shifted from one of female under-achievement to one of female under-participation (Wakerdine, 1998).

A study for over 5600 pupils in Hong-Kong in 1989 found that while boys attained better scores in geometry and the girls in mathematical manipulation problems, there were no significant gender differences in overall mathematical achievement (Cheung, 1989). Cumberbatch (1993) using the 1992 Barbados Secondary School Entrance Examination found that there was a significance gender difference in achievement in favor of girls, which was almost 14 per cent higher. At each point of comparison the girls scored higher, except for the top 10 per cent in achievement, where there was no difference. According to Walderdine (1998) there are five times as many female scientists in Latin American countries as there are in Anglo-Saxon, so the problem seems to be more severe in Anglo-Saxon countries than in some others.

Goldin, Katz and Kuziemko (2006) analyzing three different surveys for the U.S. , notice that girls achieved considerably higher grades in high school than did boys. Aptitude and achievement tests show a different pattern. Twelfth-grade math and reading achievement test scores for 1972 show that boys did far better in math, whereas girls did better in reading. By 1992 girls had widening their lead in reading and narrowed the gap with boys in math. According to their results, girls gained about 0.17 of a standard deviation in both math and reading from 1972 to 1992.

Guiso, Monte, Sapienza and Singales (2008) analyzed data for 40 different counties. Their results suggest that culture explains most of the difference in maths, at least. They show that the gap in mathematics scores between boys and girls virtually disappears in countries with high levels of sexual equality. Moreover, the gender gap is reversed in reading, and it got bigger as the sexes became more equal. On average, girls had reading scores that are over 32.0 higher than those of boys, in Turkey they scored 25.1 higher whereas in Iceland they out-performed boys by 61.0. Reinforcing the study by Cheung (1989) for Hong-Kong, they found that the one mathematical gap that did not disappear was the differences between girls and boys in geometry.

In a previous paper we were able to identify the gender gap in entrance test scores for a major University in Brazil (Guimarães and Sampaio, 2007). Our estimated coefficient for the gender variable shows that women tend to perform worse than men. Moreover this result is stronger for those on the upper quantiles. Therefore, not only women has lower entrance test scores (ETS) compared to their male counterparts but also, this is larger for students at upper quantiles of the conditional score distribution. These findings go in the opposite direction to those on the international empirical literature. We decided to look deeper and examine female performance among different University departments. The results showed that the gap tended to narrow on predominantly male departments such as the Engineering College. At Human Sciences and Social Sciences women at low quantiles have lower grades than men, this changes as we move along the conditional score distribution.

In this paper we decided to analyze the gender gap on scores from a different point of view. Based on the international literature on gender gap on scores we decided to separate the scores according to each different test the student take. Focusing on Math, Portuguese and Language we analyze female students performance on each individual test and investigate the possible existence of differences in scores across gender for our data.

3 Data

In this paper we use an unique data set on students entrance test scores at Universidade Federal de Pernambuco (UFPE) which is the major University in the Northeast of Brazil. The students were taking the entrance test (*vestibular*) in 2005. University student records data are very rich in characteristics of individuals, the colleges they are applying for and information about their previous school. Our dependent variable is their scores on the entrance test for Math, Portuguese and Language. The explanatory variables can be divided into, *Personal Information* such as and indicator for females, age, marital status, race, religion, number of children, parents schooling, parents employment status, family income, hours worked and *Academic History* such as school attended (so we can identify the type of school if private or public, catholic or other), if had lab classes, foreign language classes, preparation classes among others.

3.1 Summary Statistics

Table 1 brings information on summary statistics on the key variables of interest. Data on 56,723 students who took the Entrance Test to Universidade Federal de Pernambuco in 2005 was collected. Cases with missing values of variables included in the study are omitted. This leaves 54,877 students in the sample used in the statistical analysis covering more than 90% of all students with roughly equal numbers of males and females. The main explanatory variable is the students achievement on the entrance test for Math, Portuguese and Language that ranges from zero to 10.

Our sample consists of students, which are 20 years old on average, most of them single and still living with their parents. The majority of our students classifies themselves as white or *pardos*[2]. More than 50% of the students are catholics, 21% are protestants, 11% declared themselves as atheists and less than 1% are jewish.

The average monthly family income is R$1,620.00 (around 4.5 minimum wages). Notice however that the income distribution across families is very unequal as shown by the standard deviation. Parents have around 11 years of education. Almost 60% of students' fathers are working while 50% of their mothers have paid jobs. On average 27% of the students are working around 1.6 hours per day.

When it comes to the education system, 52% of the sample comes from private schools while 48% of them studied on the public education system. Public School students have lower scores compared to private ones, 3.9 against 4.7 (Table 2). Students were queried about the access of educational resources. In our sample 34% of the students have access to internet, 36% have additional lab classes and only 4% of the students have extra foreign languages classes.

In Brazil, the Education Ministry offers an alternative education method for those individuals who have either drop off or did not have the chance to go to school when they should have. Those are individuals that have usually a large distortion age/grade, sometimes even illiterate adults. This alternative method is called *Supletivo* (Supplementary) and offers short-term courses with a condensed material for different grades. The students can have for instance, middle school diploma in a one-year course. In our data set we have the information if the student graduated from Supletivo or not, 3% of our sample got a high-school Supletivo degree. Next section analyzes some basic statistics for boys and girls separately.

[2]Due to interbreeding of races (blacks and whites, natives and whites and blacks and natives)which happens to be stronger in the Northeast, the individual classifies himself (herself) as brown or *pardos*.

Table 1: Summary Statistics

Variables	**Mean**	**Stand. Dev.**
Observations	54,877	
Female	0.558	0.497
Age	20.538	5.402
Married	0.068	0.252
Number of Children	0.118	0.460
Minority	0.147	0.354
Catholics	0.573	0.495
Monthly Family Income (R$)[1]	1,620.454	2,072.072
Living with Family	0.803	0.398
Father Schooling	11.607	4.403
Mother Schooling	11.780	4.457
Years in Public School	3.426	4.345
Reading	0.287	0.452
Private Classes	0.400	0.490
Internet User	0.347	0.476
Working Father	0.586	0.492
Foreign Language	0.042	0.201
Lab. Classes	0.360	0.480
Hours Worked	1.690	2.921
Human Sciences	0.196	0.397
Engineering	0.099	0.299
Health	0.231	0.421
Supletivo	0.034	0.182
Tests Taken	1.864	0.996
Experiencia	0.039	0.194
Scores		
· Math	4.088	2.045
· Language	6.104	2.747
· Portuguese	5.093	1.731

Note: [1] 1 Brazilian Real = 0,3501 US Dollar.

3.2 Gilrs going for and at the University

When we analyze the data for boys and girls separately we notice that females out numbered boys on the Entrance Tests. However first semester GPA shows that a lower number of girls actually made their way through. This is probably behind the major reason why girls do worse than boys on average in all math, language and Portuguese at the entrance test, but actually out-performed them at first year GPA in all five colleges. Those girls that made their way through the University are the best ones and as such will do better than boys once their are there. The raw data either for entrance tests or for GPAs are actually showing a selectivity process at the University.

Looking separately to the numbers for boys and girls we notice some differences. Girls families' have lower income compared to boys', as well as less educated mothers' that have on average on year less of education. Girls tend to work less than boys, and are less likely to be married and have children. When queried about reading, girls on average read more than boys. Boys on our sample however, tend to have more foreign language classes as well as lab classes.

When it comes to scores, the raw data show that boys perform better in all three subjects, the gap being lower in Portuguese however. As shown by the standard deviations boys scores show a higher dispersion, compared to girls .

The gender gap vanishes when girls enter the University as shown by first year GPA. But as pointed before one has to recongnize a on going selectivity process where only the best girls make their way through the University. Here we are focusing only on the scores at the entrance test.

Table 2: Summary Statistics

Variables	**Male**	**Female**
Age	20.979	20.190
	(5.777)	(5.060)
Married	0,082	0,058
	(0,274)	(0,233)
Number of Children	0,136	0,103
	(0,498)	(0,427)
Family Income (R$)[1]	1,783.843	1,493.191
	(2,175.611)	(1,987.403)
Father Education	11.908	11.372
	(4.432)	(4.365)
Mother Education	12.014	11.597
	(4.474)	(4.436)
Years in Public School	3,286	3,538
	(4,305)	(4,374)
Reading	0,230	0,331
	(0,421)	(0,471)
Internet User	0,394	0,310
	(0,489)	(0,463)
Foreign Language	0,050	0,036
	(0,218)	(0,186)
Lab. Classes	0,379	0,346
	(0,485)	(0,476)
Hours Worked	2.030	1.421
	(3.113)	(2.729)
Supletivo	0,047	0,024
	(0,211)	(0,154)
Scores		
· Math	4.580	3.703
	(2.163)	(1.858)
· Language	4.580	3.703
	(2.163)	(1.858)
· Portuguese	5.113	5.076
	(1.734)	(1.729)
N	25,097	31,625

Note: Standard Deviation presented in parentheses. [1] 1 Brazilian Real = 0,3501 US Dollar.

Table 3: Summary Statistics: First Semester GPA

Variables	**Male**	**Female**
GPA x Social Sciences	7.714	8.088
	(0.960)	(0.857)
GPA x Human Sciences	7.770	7.984
	(1.069)	(0.799)
GPA x Engineering	6.116	6.572
	(2.051)	(1.859)
GPA x Health	7.704	7.950
	(0.904)	(0.782)
N	2,960	2,504

Note: Standard Deviation presented in parentheses.

4 Methodology

The estimations are performed using both OLS techniques and quantile regression[3] . Notice that Y is the dependent variable, entrance test scores in Math, Portuguese and Language (estimated separately), and X a matrix of explanatory variables including and indicator for female, personal characteristics, family background, school attended among others. In the same way that $\hat{\beta}_{\mathrm{OLS}}$ minimizes the sum of the loss function, $(log(Y_i) - X_i'\beta)^2$, $\hat{\beta}(\tau)$ minimizes the sum of the following linear loss function $\rho_\tau(log(Y_i) - X_i'\beta)^4$. Thus the τ-th conditional quantile function is given by

$$Q_{\mathrm{log(Y_i)}}(\tau|X) = X'\beta(\tau) \tag{1}$$

The conditional quantile function will give a family of functions, one for each τ, which provides a more complete characterization of the relationship between $\log(Y_i)$ and X compared to the one given by OLS regression, which concentrates on the first conditional moments (Arias, Hallock and Sosa (2001)). In addition, Koenker and Portnoy (1996, pp. 36-42) show that the quantile functions have, in general, the same robustness properties to outlying observations as the ordinary τ-th sample quantiles. These robustness properties are very important when the distribution of the disturbance term deviates from the Gaussian distribution. With the quantile model the entrance test scores can be influenced by personal characteristics in different ways at different parts of the distribution.

Finally, the slope parameters of the family of estimated quantile functions provides a way to test for the presence of heteroskedasticity in the model (Koenker and Bassett, 1982). For instance, if some slope coefficients are changing with τ then this is indicative of some form of heteroskedasticity. Therefore, the use of quantile regression technique allow us to address the heterogeneity of the unobservable effects in an informative and constructive way. Formally, Koenker and Bassett (1982) propose a Wald-type statistic to test if the slope parameters are equivalents for different quantiles[5].

Quantile regression has been used extensively in the economics literature to analyze gender wage differentials, returns to education and income inequality, and recently to examine students achievements[6].

5 Results

Tables 4 to 6 present the results for the specifications using the *OLS* estimation procedure. The sign of the female indicator variable varies across subject but are mostly statistically significant apart for the Portuguese score estimations for the last specification only.

Quantitatively, we have from column (1) that math test scores for girls are on average about 21 percent lower than boys. However, once we control for individual and family characteristics, such as mother's years of schooling, family income, and religion[7], this effect decreases from 21 to 19 percent (see column (3)). It decreases further to 17 percent once we control for another set of variables, such as academic variables relating type of school (if public or private), the college the student is

[3]We do not discuss quantile regression in detail. Instead we will just comment on some important properties of this approach, which are useful for this study. We suggest the works of Koenker and Basset (1978), Koenker and Portnoy (1997) and Hallock and Koenker (2001) as comprehensive sources of how to understand quantile regression.

[4]$\rho_\tau(u) = u\tau - uI(u \leq 0)$ where $I(u \leq 0)$ is an indicator function.

[5]See Koenker and Bassett (1978) and Koenker and Portnoy (1996) for more details.

[6]See for instance Eide & Showalter (1998), Ng & Pinto (2003), Bassett *et al* (2002), Kremer & Levy (2003), Smith and Naylor (2005)and Birch and Miller (2006), Guimarães (2007), Guimarães and Sampaio (2007).

[7]We include five indicator variables for religion and the Catholic religion is the reference group. They are: Protestant, Jewish, Atheist, Afro-Religions, and others.

applying for and working conditions (students' and family) (see column (4)). Observe also that the R squared indicates that the most complete specification (column (4)) captures about 18 percent of the total variability in the math entrance test scores. In our most parsimonious specification, in column (1), the female dummy explains about three percent of total variability in math score performances.

By analyzing Table 5 and 6 we notice a very different scenario from the one shown for math scores. At Table 5 the female indicator variable shows that girls out-performed boys in language tests by 3% on average. When we introduce family background variables in our specification the gender gap goes up to 5% in favor of girls. In the more comprehensive specification, including family background and academic variables, girls scores are on average 3.4% higher than boys.

Table 6 describe the results using the Portuguese scores as dependent variable. The results here are not that clear. In the more parsimonious specification, the female indicator variable shows that girls are performing worse than boys in the Portuguese test, which includes grammar and literature, but the gap is relatively low. When we introduced family background variables the girls disadvantage vanishes, and they out-perform boys with scores 1% higher. Again, in the more comprehensive specification, including family background and academic performance variables, the female indicator variable is negative but shows a very small coefficient. The R square indicates that the most complete specification captures about 19 percent of the variability in Portuguese scores. In the smallest specification the female dummy explanation power is very limited, however.

Our results for the OLS estimations show that on average girls perform better in language, and this result is robust across different specifications. Boys perform better in math, but the gap lowers as we introduce a more comprehensive set of controls that capture differences in family background and academic performance. For the Portuguese scores that includes tests for grammar and literature, the result is not clear, but mostly, the female indicator variable is either positive or very small and negative, when statistically significant. Therefore, there is no statistical evidences of an average gender gap in Portuguese tests. The gap is more clear for math and language, being in favor of boys for math, and in favor of girls for language.

Table 4: Determinants of Math Score - OLS Estimation

Variables	**(1)**	**(2)**	**(3)**	**(4)**
Gender (female=1)	-0.216* (-47.09)	-0.192* (-41.16)	-0.171* (-34.52)	-0.169* (-34.32)
Age		-0.005* (-7.95)	-0.011* (-13.87)	-0.011* (-13.87)
Married		0.056* (4.22)	0.065* (4.89)	0.062* (4.71)
Number of Children		0.012 (1.41)	0.047* (6.04)	0.047* (6.06)
Minority		-0.043* (-6.75)	-0.039* (-6.10)	-0.039* (-6.12)
Catholics		-0.012* (-2.54)	-0.021* (-4.37)	-0.018* (-3.90)
Monthly Family Income		4.28E-5* (35.05)	2.88E-5* (21.66)	2.97E-5* (22.39)
Living with Family		-0.026* (-3.87)	-0.031* (-4.59)	-0.033* (-4.91)
Father Schooling		9.45E-3* (13.70)	3.69E-3* (5.20)	3.29E-3* (4.65)
Mother Schooling		13.10E-3* (19.20)	6.98E-3* (9.88)	6.81E-3* (9.68)
Years in Public School			-6.31E-3* (-9.50)	-7.52E-3* (-11.30)
Reading Habit (yes=1)			0.014* (2.61)	0.015* (2.72)
Private Classes			0.014* (2.78)	0.013* (2.60)
Internet User			0.087* (15.02)	0.084* (14.48)
Working Father			0.013* (2.60)	0.013* (2.54)
Foreign Language			0.133* (11.95)	0.131* (11.81)
Lab. Classes			0.065* (12.57)	0.054* (10.50)
Hours Worked			-0.001 (-1.26)	-0.001 (-1.47)
Supletivo			-0.104* (-7.20)	-0.097* (-6.75)
Tests Taken			0.079* (28.83)	0.077* (28.36)
Experiencia			0.010 (0.76)	0.011 (0.86)
Human Sciences			-0.029* (-4.61)	-0.028* (-4.43)
Engineering			0.208* (26.89)	0.200* (25.92)
Health			0.046* (7.51)	0.045* (7.50)
Federal and Military Schools	NO	NO	NO	YES
Intercept	1.4053* (407.64)	1.184* (64.99)	1.234* (61.69)	1.245* (62.39)
N. of Observ.	54,433	48,765	46,168	46,168
R^2	0.039	0.124	0.177	0.184

T-Statistics are presented in parentheses, using heteroskedasticity consistent standard errors. * indicates significant at the 99 percent confidence level, ** at the 95 percent, and *** at the 90 percent.

Table 5: Determinants of Language Score - OLS Estimation

Variables	**(1)**	**(2)**	**(3)**	**(4)**
Gender (female=1)	0.033* (6.63)	0.049* (9.36)	0.033* (5.97)	0.034* (6.10)
Age		-4.83E-3* (-6.50)	-0.018* (-19.39)	-0.018* (-19.38)
Married		5.31E-3 (0.35)	0.030** (1.98)	0.029** (1.90)
Number of Children		-0.056* (-6.05)	0.015 (1.55)	0.015 (1.55)
Minority		-0.020* (-2.66)	-0.019* (-2.60)	-0.019* (-2.60)
Catholics		-0.050* (-9.56)	-0.052* (-9.99)	-0.051* (-9.79)
Monthly Family Income		1.74E-5* (13.62)	8.09E-6* (5.88)	8.49E-6* (6.16)
Living with Family		-0.027* (-3.58)	-0.026* (-3.42)	-0.027* (-3.53)
Father Schooling		9.57E-3* (12.44)	4.31E-3* (5.47)	4.17E-3* (5.28)
Mother Schooling		9.23E-5* (12.02)	2.75E-3* (3.50)	2.67E-3* (3.40)
Years in Public School			-8.63E-3* (-11.29)	-9.11E-3 (-11.80)
Reading Habit (yes=1)			0.030* (5.14)	0.030* (5.16)
Private Classes			0.062* (11.51)	0.062* (11.46)
Internet User			0.047* (7.54)	0.045* (7.33)
Working Father			9.73E-3*** (1.72)	9.57E-3*** (1.69)
Foreign Language			0.210* (22.77)	0.210* (22.71)
Lab. Classes			0.024* (4.29)	0.020* (3.51)
Hours Worked			-6.11E-3* (-5.89)	-6.22E-3* (-6.00)
Supletivo			-0.116* (-6.65)	-0.114* (-6.48)
Tests Taken			0.128* (43.28)	0.128* (43.04)
Experiencia			-0.112* (-7.57)	-0.111* (-7.52)
Human Sciences			0.022* (3.18)	0.023* (3.27)
Engineering			8.47E-4 (0.09)	-2.86E-3 (-0.31)
Health			0.012*** (1.77)	0.012*** (1.76)
Federal and Military Schools	NO	NO	NO	YES
Intercept	1.677* (438.65)	1.576* (73.91)	1.723* (74.43)	1.728* (74.61)
N. of Observ.	53,869	48,292	45,739	45,739
R^2	8.00E-4	0.039	0.105	0.106

T-Statistics are presented in parentheses, using heteroskedasticity consistent standard errors. * indicates significant at the 99 percent confidence level, ** at the 95 percent, and *** at the 90 percent.

Table 6: Determinants of Portuguese Score - OLS Estimation

Variables	**(1)**	**(2)**	**(3)**	**(4)**
Gender (female=1)	-6.97E-3** (-2.06)	0.015* (4.33)	-6.61E-3*** (-1.87)	-5.56E-3 (-1.57)
Age		-9.60E-4** (-2.03)	-9.62E-3* (-16.61)	-9.60E-3* (-16.55)
Married		0.025* (2.61)	0.044* (4.64)	0.043* (4.49)
Number of Children		-0.025* (-4.58)	0.023* (4.15)	0.023* (4.13)
Minority		-0.030* (-6.19)	-0.026* (-5.46)	-0.026* (-5.47)
Catholics		-0.063* (-18.34)	-0.067* (-20.03)	-0.066* (-19.69)
Monthly Family Income		2.64E-5* (30.22)	1.56E-5* (17.10)	1.61E-5 (17.75)
Living with Family		-0.032* (-6.40)	-0.026* (-5.40)	-0.028* (-5.69)
Father Schooling		9.18E-3* (18.05)	4.59E-3* (8.99)	4.34E-3* (8.52)
Mother Schooling		10.99E-3* (21.67)	5.27E-3* (10.26)	5.18E-3* (10.10)
Years in Public School			-5.11E-3* (-10.50)	-5.86E-3* (-11.93)
Reading Habit (yes=1)			0.067* (18.12)	0.068* (18.28)
Private Classes			0.043* (12.29)	0.043* (12.17)
Internet User			0.071* (17.97)	0.069* (17.50)
Working Father			8.49E-3** (2.32)	8.36E-3* (2.29)
Foreign Language			0.114* (15.43)	0.113* (15.34)
Lab. Classes			0.041* (11.37)	0.035* (9.54)
Hours Worked			-7.91E-3* (-11.85)	-8.02E-3* (-12.02)
Supletivo			-0.092* (-7.96)	-0.088* (-7.62)
Tests Taken			0.102* (52.96)	0.101* (52.58)
Experiencia			0.011 (1.26)	0.012 (1.37)
Human Sciences			0.045* (10.05)	0.046* (10.23)
Engineering			0.041* (7.12)	0.036* (6.28)
Health			0.055* (12.84)	0.054* (12.82)
Federal and Military Schools	NO	NO	NO	YES
Intercept	1.564* (614.52)	1.356* (98.16)	1.412* (94.87)	1.418* (95.21)
N. of Observ.	54,859	49,152	46,542	46,542
R^2	1.00E-4	0.092	0.188	0.193

T-Statistics are presented in parentheses, using heteroskedasticity consistent standard errors. * indicates significant at the 99 percent confidence level, ** at the 95 percent, and *** at the 90 percent.

After analyzing girls performance on average on three different tests we now look at the results for different quantiles of the conditional scores distributions.

5.1 Quantile Results

In this subsection we analyze the quantile regression estimates. Since quantile regression procedures produce one point estimation for each quantile, for the sake of space, we focus only on the coefficient of the indicator variable *gender*[8].Quantile regression provides the appropriate tool to determine whether there are any difference in marginal responses of the Entrance Test Score to boys and girls.

These plots show the quantile regression (QR) estimates, for the gender variable for the Math, Language and Portuguese Tests, as well as the 95% confidence intervals. For comparison reasons, the least square estimate is presented by the dotted horizontal line.

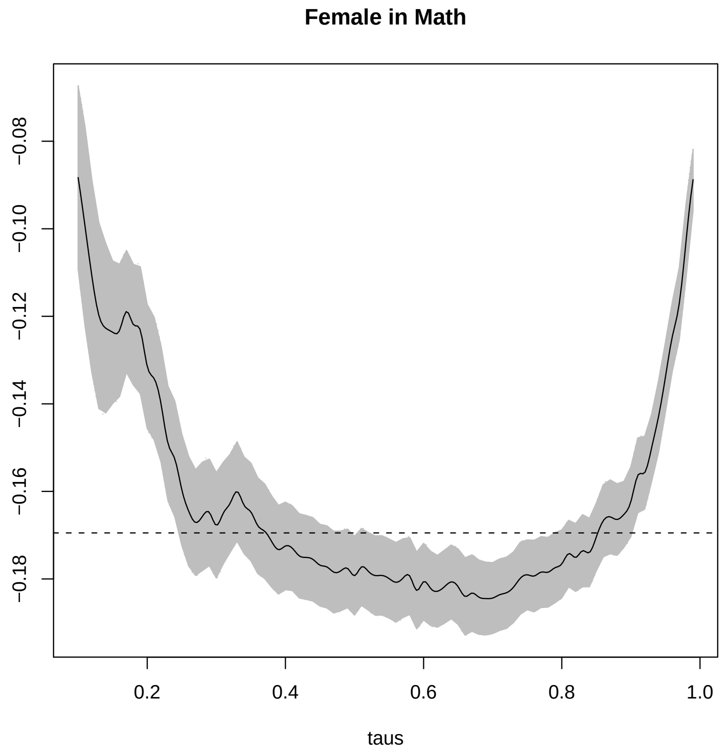

Figure 3: The above plots present the quantile regression estimates for the female variable having log of math scores as the dependent variable . The solid lines are the quantile estimates and the shade the 95% confidence intervals. The dotted line presents the Ordinary Least Square estimates.

[8]Upon request we provide the quantile estimation for the other parameters

Figure 3 shows the quantile plot for the female indicator variable having the Math scores as dependent variable. Notice that despite the fact that girls perform worse than boys in Math the gap varies across the conditional score distribution. Girls at the middle of the conditional score distribution suffer from a higher gap compared to girls at the upper and lower tail. That is, girls in the middle of the conditional distribution scored about 18% lower than boys in math. These gap is lower, around 10%, for girls with lower and higher conditional scores.

In Figure 4 we notice that the higher performance of girls at language tests diminishes and vanishes as we move up on the conditional score distribution. That is, among lower scores students, girls perform better on language tests, conditional to family background and academic variables. Girls enjoy higher advantage up to the 6^{th} quantile, when the conditional gap in Language scores vanishes.

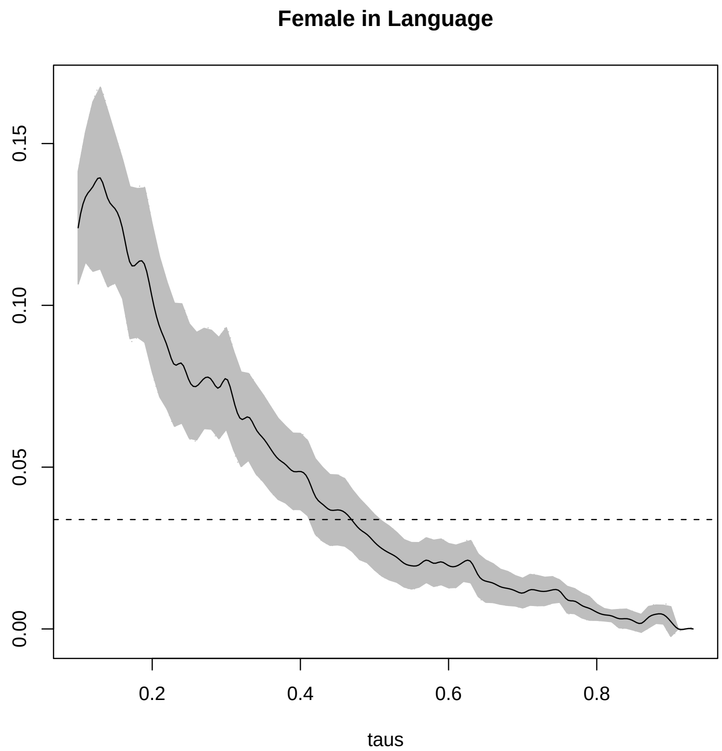

Figure 4: The above plots present the quantile regression estimates for the female variable having log of portuguese scores as the dependent variable . The solid lines are the quantile estimates and the shade the 95% confidence intervals. The dotted line presents the Ordinary Least Square estimates.

The quantile plots for the female indicator for Portuguese tests show a different picture from the OLS estimations. The quantile estimates show that girls in fact enjoy a stronger advantage in grammar, literature and writing, compared to boys, up to the 4^{th} quantile of the conditional score distribution. After that, there is no statistical evidence of a female advantage on Portuguese scores, conditional to other characteristics.

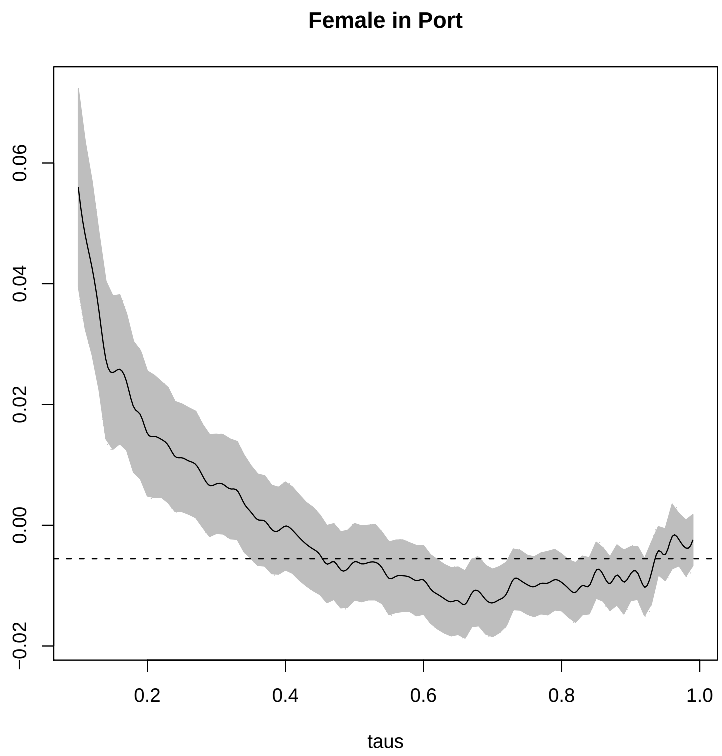

Figure 5: The above plots present the quantile regression estimates for the female variable having log of language scores as the dependent variable . The solid lines are the quantile estimates and the shade the 95% confidence intervals. The dotted line presents the Ordinary Least Square estimates.

6 Conclusion

In this paper we study whether the gender score gap in math, language and reading also holds for Brazil. We use an unique and rich data set on students Entrance Test Scores at Universidade Federal de Pernambuco (UFPE) to evaluate quantitative differences in scores between boys and girls. The dataset has information on individual and school characteristics, and family background. In a recent paper Guiso, Monte, Sapienza and Singales (2008) analyzed data for 40 different countries. They show that the gap in mathematics scores between boys and girls virtually disappears in countries with high levels of sexual equality. Moreover, the gender gap is reversed in reading, and it got bigger as the sexes became more equal. Basically, their results show that the cultural factors are key on girls performance.

Here we used both OLS and quantile regression to study the existence of score gaps in math, language and portuguese. The OLS results show that boys on average have a advantage in math. Girls perform better at language tests on average. The results for the portuguese scores are not that clear.

The quantile results show a richer picture compared to OLS. The advantage in math for boys is indeed robust but is strongest for the middle of the distribution. That is, girls with both lower and higher conditional math scores, suffer from lower gap, compared to those in the middle of the conditional score distribution.

The advantage in language scores for girls is perceived up to the middle of the distribution. From the middle up to the end of the conditional score distribution, the gap in language vanishes. Despite the fact that OLS results could not show significant differences in Portuguese, the quantile results show that in fact, girls out-perform boys up to the 4^{th} quantile of the conditional score distribution.

Therefore, our results show that in fact, the gender gap in scores varies across subjects and also across quantiles of the conditional score distribution. Boys may show an advantage in Math, but the advantage is lower as we control for family background and academic variables, showing that girls disadvantage in math may be in fact related to family background and cultural factors, as showed by Guiso, Monte, Sapienza and Singales (2008). The R squared indicates that the most complete specification (Table 4, column (4)) captures about 18 percent of the total variability in the math entrance test scores. Boys advantage vanishes when it comes to language, and to some extent to portuguese, confirming the results found on the international literature that girls out-perform boys in reading and writing.

7 References

Arias, O., Kevin Hallock and Walter Sosa-Escudero (2001), "Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression using twins data". Empirical Economics.20, 7-40.

Baron-Cohen, S. (2003) The Essential Difference: Men and women, and the Extreme Male Brain, (Allen Lane, London).

Bassett, G. W., Tam, Mo-Yin, Knight, K. (2002). Quantile Models and estimators for data analysis. Metrika, 55, 17-26.

Birch, E. R., Miller, P. W. (2006). Student Outcomes at University in Australia: A Quantile Regression Approach. Australian Economic Papers, 45, n. 1, 1-17.

Cheung, K. (1989) Gender differences in the junior secondaty (grade 7) mathematics curriculum in Hong Kong, *Educational Studies in Mathematics*, 20 (1), p. 97-103.

Cumberbatch, G. (1993) An investigation into the effects of gender and teacher perception on pupils' performance in mathematics in the secondary schools entrance examination in Barbados, MA Thesis, University of the West Indies, Cave Hill Barbados.

Edie, E., Showalter, M. H. (1998). The effect of School quality on student performance: A quantile regression approach. Economics Letters, 58, 345-350.

Guimarães, J. (2007)."Changes in the Earnings Distribution in Brazil, 1986-1995: Human Capital and Institutional Factors", *mimeo.*

Guimares, J. F., Sampaio, B. (2007). The Influence of Family Background and Individual Characteristics on Entrance Tests Scores of Brazilian University Students. Anais do XXXV Encontro Nacional de Economia - ANPEC - Associao Nacional dos Centros de Psgraduao em Economia, Recife.

Guiso, L. F. Monte, P. Sapienza and L. Zingales. (2008) Culture, Gender and Math, *Science* , 320 (30), p. 1164-1665.

Hallock, Kevin and Roger Koenker (2001) "Quantile Regression", The Journal of Economics Perspectives,15(4), 143-156.

Hausmann, R , L.D. Tyson, S. Zahidi. (2006) The Global Gender Gap Report, *World Economic Forum, Geneva, Switzerland,.*

Koenker, R., and Bassett, G. (1978). "Regression Quantiles". Econometrica, 46 (1), 33- 50.

Koenker, Roger and Stephen Portnoy, (1997). "Quantile Regression", Office of Research Working Paper n.97-0100 College of Commerce and Business Administration, University of Illinois at Urbana-Champaign.

Kremer, M. and Levy, D. (2003), "Peer Effects and Alcohol Use among College Graduates", National Bureau of Economic Research Working Paperno.9873, National Bureau of Economic Research, Cambridge, United States.

Smith, J., Naylor, R. (2005). Schooling Effects on subsequent University Performance: evidence from the UK university population. Economics of Education Review, 24, 549-562.

Walkerdine, V. (1998) Counting girls Out. Girls and Mathematics. Studies in Mathematics Education Series: 8. University of Exeter, UK.