UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
DEPARTAMENTO DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

EDUARDO ANDRÉ TILLMANN

THE ROLE OF GENDER IN BRAZILIAN ACADEMIC ACHIEVEMENT: INEQUALITY AND PEER EFFECTS

Porto Alegre
2018

EDUARDO ANDRÉ TILLMANN

THE ROLE OF GENDER IN BRAZILIAN ACADEMIC ACHIEVEMENT: INEQUALITY AND PEER EFFECTS

Tese submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Doutor em Economia, com ênfase em Economia Aplicada.

Orientador: Prof. Dr. Flávio V. Comim

Porto Alegre

2018

Tillmann, Eduardo Andre
The Role of Gender in Brazilian Academic Achievement: inequality and peer effects / Eduardo Andre Tillmann. -- 2018.
79 f.
Orientador: Flavio Vasconcellos Comim.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, BR-RS, 2018.

1. Gender inequality. 2. Academic achievement in Brazil. 3. Peer effects. I. Comim, Flavio Vasconcellos, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

EDUARDO ANDRÉ TILLMANN

THE ROLE OF GENDER IN BRAZILIAN ACADEMIC ACHIEVEMENT: INEQUALITY AND PEER EFFECTS

Tese submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Doutor em Economia, com ênfase em Economia Aplicada.

Aprovada em: Porto Alegre, 23 de abril de 2018.

BANCA EXAMINADORA:

Prof. Dr. Flavio Vasconcellos Comim - Orientador
Universidade Federal do Rio Grande do Sul (UFRGS)

Prof. Dr. Hudson da Silva Torrent
Universidade Federal do Rio Grande do Sul (UFRGS)

Prof. Dra. Izete Pengo Bagolin
Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)

Prof. Dr. Sílvio Hong Tiing Tai
Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)

ABSTRACT

The aim of this research is to analyze the role of gender on scholastic achievement in Brazil, emphasizing inequalities and peer effects. These issues are analyzed in five chapters, including the introduction and the concluding remarks. We start by briefly describing gender inequality in terms of math and literacy achievements in Brazil, focusing, mainly, on 5th and 9th grade students. We observe that boys tend to outperform girls in math, a relation that reverses in literacy and, furthermore, that these inequalities increase in more advanced schooling years. The second chapter aims to investigate the factors associated with these differences, exploring students', teachers' and schools characteristics in two different types of decomposition methods, one that explores differences in mean achievement and another that assesses the entire test score distribution. The results indicate that despite boys and girls having similar family and socioeconomic characteristics, the main contributor towards the learning differences is the return of these characteristics in terms of test scores for each of the two genders, which, therefore, reduces the role of teachers' and schools in diminishing these inequalities. The third chapter deals with peer effects in 5th grade Brazilian public schools. It investigates the casual relationship between the proportion of girls at school and learning. We identify a positive relation between test scores and the proportion of girls in literacy and, mainly, in math, a subject that girls tend to be outscored by boys. This, in fact, draws attention to the fourth chapter, which seeks to elucidate the mechanisms behind this influence. We verify that the benefits of having a greater proportion of girls are mainly through improvements in student behavior, which reflects in less violence, greater teacher expectations over the student's academic future, and facilitates classroom progress. In terms of public policies, this research draws attention to gender as an important factor in the allocation of students and teachers within schools. Therefore, the consideration of our findings in the formulation and execution of policies can result in effective and low cost measures aimed at increasing scholastic achievement.

Keywords: Gender inequality. Academic achievement in Brazil. Peer effects.

RESUMO

O objetivo desta pesquisa é analisar o papel do gênero no aprendizado escolar no Brasil, enfatizando desigualdades e o efeito de pares. Essas questões foram analisadas em cinco capítulos, incluindo a introdução e as considerações finais. Inicia-se com uma breve descrição sobre a desigualdade de gênero em termos de aprendizado de matemática e português no Brasil considerando, principalmente, os alunos de 5º e 9º ano. Observa-se que os meninos tendem a se sair melhor em matemática que as meninas, uma relação que se inverte para português e, ainda, que estas desigualdades aumentam nos anos escolares mais avançados. O segundo capítulo visa investigar os fatores associados a estas diferenças, explorando características dos alunos, professores e escolas através de dois métodos diferentes de decomposição, uma que explora a diferença de média dos resultados e outra que analisa toda a distribuição de notas. Os resultados indicam que apesar de meninos e meninas possuírem características de contexto socioeconômico e familiares parecidas, o principal fator contribuinte para as diferenças de aprendizado está no retorno destas características em termos de nota para cada um dos dois gêneros, o que, portanto, reduz o papel do professor e da escola na diminuição destas desigualdades. O terceiro capítulo trata do efeito de pares no aprendizado do 5º ano das escolas públicas brasileiras. Ele investiga, de maneira causal, a relação entre a proporção de meninas na escola e o aprendizado. Identifica-se uma relação positiva entre notas e a proporção de meninas em português e, principalmente, em matemática, uma disciplina cujas meninas tendem a se sair piores do que os meninos. Isto, portanto, chama atenção para o tópico do quarto capítulo, que busca elucidar mecanismos por trás desta influência. Verifica-se, portanto, que o efeito positivo das meninas ocorre via comportamento, o que se reflete em menor violência, maior expectativa dos professores sobre o futuro escolar dos alunos e facilita o andamento da classe. Diante disso, em termos de políticas públicas, o trabalho chama atenção para o gênero como fator importante na alocação de alunos e professores dentro da escola. Assim, levar os resultados aqui apresentados em consideração na formulação e execução de políticas pode resultar em medidas efetivas e de baixo custo voltadas para o aumento do aprendizado escolar.

Palavras-chave: Desigualdade de gênero. Aprendizado no Brasil. Efeito de pares.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUCTION

Over the past decades, basic education in Brazil progressed by increasing the number of students in schools and on knowledge, as documented by national and international evaluation tests (SOARES; NASCIMENTO, 2011; SOARES; DELGADO, 2016). However, the country still lags considerably behind OECD average and performs worse than countries with similar spending per student (OECD, 216).

The search for improvements on school efficacy and efficiency in Brazil has led researchers to emphasize the impact of students' socioeconomic background (FRANCO; MENEZES-FILHO, 2008; BIONDI; FELÍCIO, 2007), school quality (SOARES; CANDIAN, 2007; SOARES et al.; 2016) and inequalities (SOARES; DELGADO, 2016; ALVES et al., 2016) on their education performance.

This thesis contributes to such literature by focusing on two determining aspects of the Brazilian education system, namely, gender and peer effects. In regard to gender, we first identify the factors associated with gender inequality on Math and Portuguese achievement. Then, with respect to peer effects, we examine the influence of the proportion of girls on test scores and, lastly, the possible channels through which the gender composition influences these test scores.

We focus on 5th and 9th grade students when assessing gender inequalities as these are the two population-based schooling grades surveyed by the Brazilian Ministry of Education and we seek to identify differences on these inequalities by schooling grades, as the related literature indicates that sex differences on academic achievement grows with school progress (BUCHMANN et al., 2008). As to peer effects, our emphasis is exclusively on 5th grade and at the school level, since we pursue a casual effect and, therefore, need to avoid biases arising from the fact that parents, teachers and school authorities may have some discretion in placing students in different classes within a grade. The two topics of this thesis, namely, gender inequality on achievement and peer effects are scarcely explored in the Brazilian literature. As to inequality, authors have discussed it in a range of aspects, such as wage differentials (KASSOUF, 1998; MADALOZZO 2009), occupational segregation (RASCHE JÚNIOR, 2015; MADALOZZO; ARTES, 2017), and on the decision between work and study (TILLMANN; COMIM, 2016). However, few studies have investigated gender differences on academic achievement. With regards to peer effects, apart from Pinto (2008), Firpo, Jales and

Pinto (2015), very few papers that discussed the issues of identification and estimation of peer effects.

Our results with regards to inequality indicate that girls and boys have very distinct educational production functions, that is, despite the fact that students of both genders have similar background characteristics, the way that they translate their background characteristics into achievement accounts for the majority of the test score differences in math and literacy. This effect, which is captured by the unexplained component of the Oaxaca-Blinder decomposition, favors girls in literacy and boys in math.

Moreover, our findings suggest that the proportion of girls positively affects Math and Portuguese test scores of boys and girls on 5th grade. Despite the fact that these effects are quite similar, they are slightly larger for Math, a subject that girls, on this level of education, tend to be outperformed by boys. This, in turn, suggests that the gender peer effects are not exclusively associated with spillovers that arise from having higher achieving peers. We further explore this topic on our third paper and identify that the benefits of having a greater proportion of girls are mainly through improvements in student behavior, which reflects less violence, greater teacher expectations over the student's academic future, and facilitates classroom progress.

Before we proceed to the three papers, the next subsection depicts the context of the gender inequality in academic achievement in Brazil.

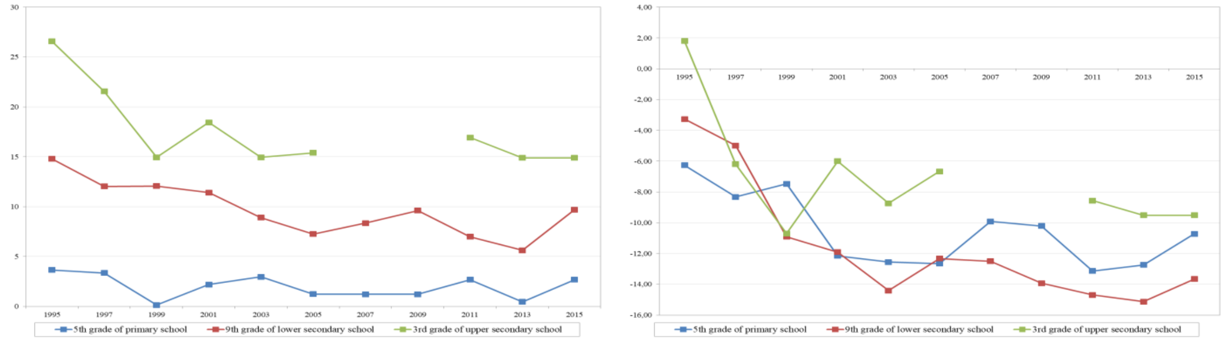
1.1 ACADEMIC ACHIEVEMENT AND GENDER IN BRAZIL

This subsection aims to briefly describe the gender differences in Brazilian elementary education. The idea is to motivate the reader and build the background on which the three papers that constitute this thesis are going to be based on. We use data from the biannual National Basic Education Assessment System (SAEB)[1], an assessment from the Brazilian Ministry of Education that gathers information on schools, students, teachers and principal, allowing policymakers and researchers to track and assist on the evolution of the educational system. We focus our overview mostly to the latest year of our sample, 2015, and to the 5th and 9th grades, which constitute the emphasis of this thesis.

[1] First created in the early 90's, it later changed its name in 2005, when it began to include two kinds of assessment: ANEB, which is a sample-based assessment keeping trends from the former SAEB, and Prova Brasil, a population-based assessment of public schools.

Figure 1.1, below, shows estimates of the gender gaps[2] in Math and Portuguese, that are the two subjects which the SAEB constantly evaluates in Brazilian students.

Figure 1.1 - Math and Portuguese gender gaps in Brazil (1995-2015).



a) Math gender gap. b) Portuguese gender gap.

Source: Elaborated by the authors using SAEB (1995-2015).

The gender differences on mean achievement are quite stark. In Math, as shown by Figure 1.1a, boys always outperform girls, and there is a clear distinction between the three different grades. That is, the more advanced the grade is, the greater is the gap. These findings are in accordance with many developed and developing countries, and with the literature that highlight the fact that the Math gap increases with schooling grades (BEDARD; CHO, 2010; BHARADWAJ et al., 2016; CONTINI et al., 2017). Also, Figure 1.1 shows that for Brazil, the differences are quite stable throughout the years, and show only a slight tendency of decreasing.

As for literacy, girls outperform boys on all years and grades, expect for the 3rd grade of upper secondary school on the first year of the assessment. This tendency of girls to outscore boys in language is also found in many other countries (HUSAIN; MILLIMET, 2009; BUCHMANN et al., 2008). The distinction between the grades, however, is less clear than in Math, and the tendency is towards an increase of the gender gap. Note that after 2005, the year the assessment became population-based, the grade differences became more distinct, being the widest gap, as in Math, observed for the last year of compulsory education and the narrowest, in contrast, is among 9th graders.

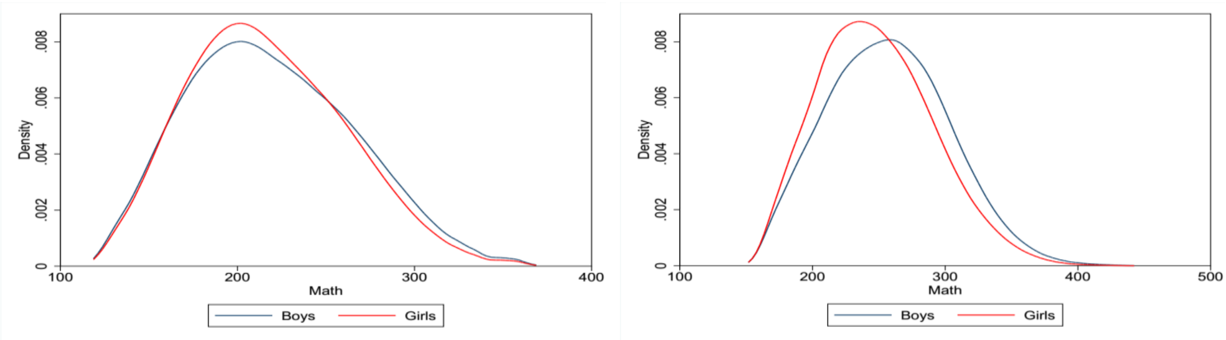
The literature that investigates gender inequalities in educational performance has put forward various explanations for these gender differences in math and literacy. School and family characteristics (BERTRAND; PAM, 2013; BAKER; MILLIGAN, 2016), as well as

[2] Calculated by the differences between boys and girls.

biological (CAHILL, 2006) and psychological factors (SPELKE, 2005; SAX; 2005; GOLSTEYN; SCHILS, 2014) have been associated with gender disparities in schooling outcomes. Moreover, Tinklin (2003), points to the role of peers in school, among the factors that also affect gender differences. A better assessment of this literature, however, extrapolates the goal of the present chapter, but will be the main focus of the remainder of this thesis.

Figure 1.2 and 1.3, looks beyond the mean and plots[3] the distribution of test scores for 5th and 9th grades in 2015[4].

Figure 1.2 - Math density distributions of 5th and 9th graders test scores in 2015.



a) 5th grade. b) 9th grade.

Source: Elaborated by the authors using SAEB (2015).

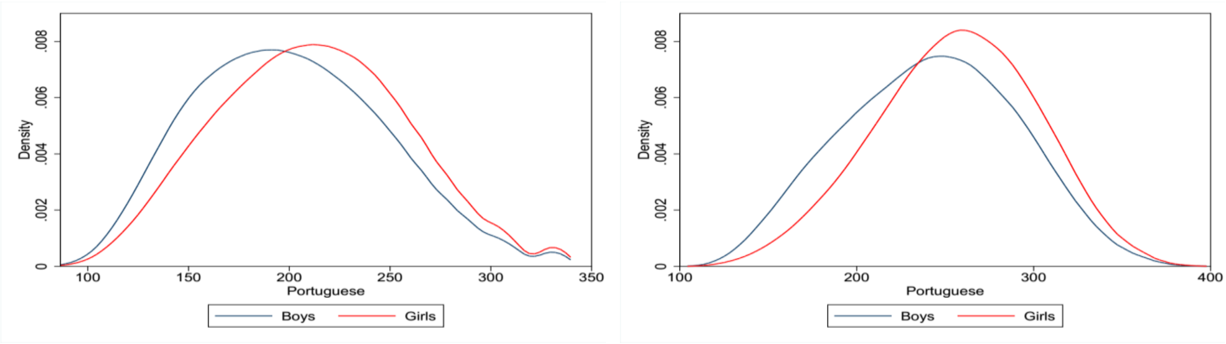
The shapes of the density distributions on Figure 1.2 indicate that boys tend to perform more unequally in Math than girls, on both grades. This is shown by the fact that the blue density curve, that represents boys test scores, is lower than the red, which represents girls. Nonetheless, as already described by Figure 1.1, the boys test scores density is slightly more to the right than the ones for girls, indicating a higher mean for male students.

The differences, however, are more accentuated when we look at the Portuguese test score distributions, as seen in Figure 1.3.

[3] We also perform t-tests, to assess whether the differences on the mean are statistically significant, and Kolgomorov-Smirnov equality of distribution tests on all shown distributions. All results indicate statistical significant differences at the 1% level.

[4] We focus only on 2015 because of its similarity with earlier years.

Figure 1.3 - Portuguese density distributions of test scores in 2015, 5th and 9th graders.



a) 5th grade b) 9th grade

Source: Elaborated by the authors using SAEB (2015).

Now, the density curves for the girls on both grades are the ones more to the right, in accordance with their relative superior performance. A comparison between the 9th grade densities for the two genders shows that, not only girls outperform boys, but since their density is taller and therefore more concentrated around the mean, their overall performance is better.

The setting depicted by the Math and Portuguese test scores density distributions is similar to the one described by Ellison and Swanson (2010) for the United States, where the gender gap in Math skills widens at the top of the distribution and, for languages, girls have advantage in reading and language across the full distribution.

Further, we further explore the gender gaps by calculating the boys-girls ratio in selected parts of the test score distribution, in such a way that we are able to better scrutinize gender differences for each subject and graded.

Table 1.1 - Ratios boys-girls by grade in Brazil (2015).

	Math			Portuguese		
	5th grade primary school	9th grade secondary school	3rd grade secondary school	5th grade primary school	9th grade secondary school	3rd grade secondary school
Enrollment	1.038	0.923	0.798	1.038	0.923	0.798
Bottom 1%	1.184	0.896	0.587	1.734	2.434	1.363
Bottom 5%	1.127	0.793	0.622	1.667	2.178	1.364
Below Median	0.999	0.766	0.627	1.281	1.168	0.886
Above Median	1.075	1.107	1.002	0.851	0.731	0.718
Top 5%	1.381	1.428	1.727	0.761	0.628	0.653
Top 1%	1.483	1.621	2.371	0.778	0.605	0.765

Source: Elaborated by the authors using SAEB (2015).

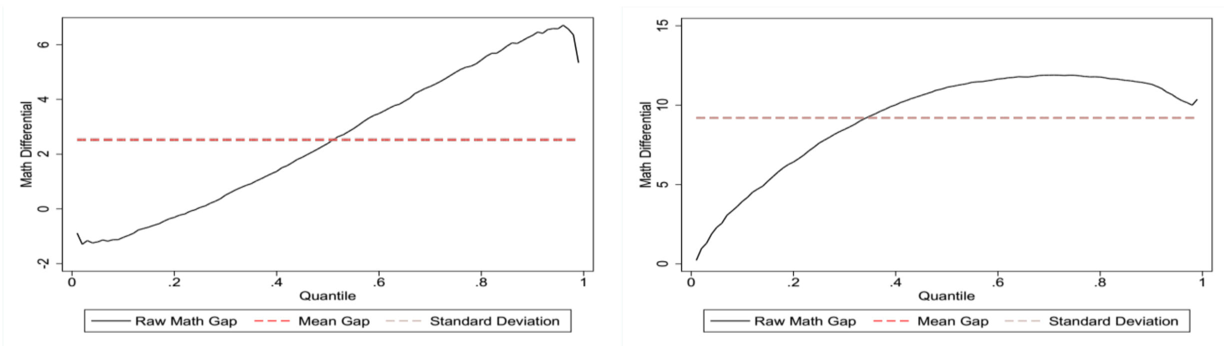
Table 1.1 shows, first, that the SAEB sample, at least in the number of girls and boys, becomes less balanced as the grades advance. This can be associated with the fact that in Brazil there is a high opportunity cost of schooling, especially for males, in such a way that they are more prone to leave school and start their work life earlier than girls (CORSEUIL et al., 2012).

With regards to achievement, Table 1.1 shows that the Math gap increases with the advancement of grades. More specifically, among the Top 1% performers in Math, there is almost 1½ boy for each girl on 5th grade. A number that increases to just below 2½ boys for each girl on the last grade of compulsory school. As to Portuguese, in contrast to Math, girls outperform boys and, therefore, are the majority on all upper tail of the test scores distributions. However, the difference is relatively more stable than the one for Math.

The literature on gender differences in test scores is not conclusive as to when in the life course gender differences in performance emerge (LEAHEY; GUO, 2001; BAKER; MILLIGAN, 2016), and whether sex differences in test scores are changing over time (HYDE et al., 1990; HEDGES; NOWELL, 1995; FRYER; LEVITT, 2010). However, the literature has pointed out by that social and cultural gender-biased environments (GUISO et al., 2008; DICKERSON, MCINTOSH, VALENTE, 2015) are contributors to the gap. Moreover, Guiso et al. (2008) finds that the gender gap in math disappears in more gender-equal societies, because girls improve their test scores, while the gap on literature expands.

Lastly, we rank boys and girls by test score performance and compare the percentiles of each Math and Portuguese distribution in 2015.

Figure 1.4 - Quantile differences on Math achievement.



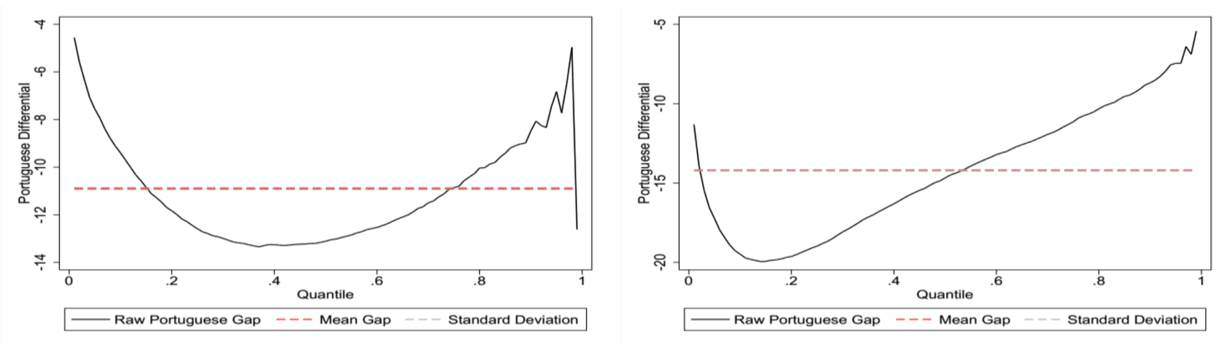
a) 5th grade.　　b) 9th grade.

Source: Elaborated by the authors using SAEB (2015).

Figure 1.4 indicates that among the low achievers in Math, boys and girls tend to perform very similarly. In 5th grade, among the bottom 20% of test scores, girls outperform boys, as shown by the values below zero on Figure 1.4a. This, therefore, reverses the usual gender gap, as depicted before on our analysis. However, as we further explore the differences among the high achieving students, the gap between boys and girls increases. When we look at Figure 1.4b, for 9th graders, even comparing lower performers, boys score higher than girls, and the gap, despite relatively stabilizing and even decreasing amongst the top 50% performers, is much greater than on 5th grade, as shown by the scale of both gap axis.

When we explore Portuguese test scores, as shown on Figure 1.5, the gap is more U-shaped, that is, lower and high achievers have more similar performance than the ones on the middle of the distribution. Note, however, that in comparison to the Math analysis, the gap is much wider in Portuguese, as shown by the scale on the graph.

Figure 1.5 - Quantile differences on Portuguese achievement.



a) 5th grade b) 9th grade.

Source: Elaborated by the authors using SAEB (2015).

This subsection gave an overview of the gender gap in Brazilian education. The differences described here strongly influenced the construction and assembly of this thesis. The following three papers build upon this description and further analyze the sources of inequities and the influence of gender composition on achievement. That is, in the first paper of this thesis we seek to identify the factors associated with gender inequality on the Brazilian education system. In the second, we explore if the assignment to a school with a higher proportion of girls influences test scores and, lastly, try to uncover possible channels through which the gender composition influences these test scores. The last chapter presents the concluding remarks.

2 ASSESSING THE GENDER GAP ON ACADEMIC ACHIEVEMENT IN BRAZIL

This first essay of the present thesis assesses the gender gap on academic achievement in Brazil.

2.1 INTRODUCTION

The issue of gender inequality is a concern among researchers and policymakers. Throughout much of the developing world, women tend to be disadvantaged in terms of job opportunities and wages (WORLD BANK, 2012). Nevertheless, researchers have identified the importance of women in improving economic efficiency in terms of growth, as well as human welfare (SCHULTZ, 2002).

In the scope of Brazil, authors have discussed the theme of gender inequality in a range of aspects, such as wage differentials (KASSOUF, 1998; MADALOZZO, 2009), occupational segregation (RASCHE JÚNIOR, 2015; MADALOZZO; ARTES, 2017), and on the decision between work and study (TILLMANN; COMIM, 2016). However, few studies have investigated gender differences on academic achievement.

Interestingly, recent studies have noted the inexistence of gender gaps in achievement at earlier ages, but that test scores differences arises and intensifies with age (FRYER JR; LEVITT, 2010). This, in turn, suggests that these gaps may not be related to intrinsic distinct characteristics between males and females, paving the way for researchers who seek to identify the main features behind the these differences. That is, the emergence of test score gaps between boys and girls might be related to distinctions in the learning process of both genders, and, therefore, associated with observable educational characteristics, such as those of family, teachers and schools.

In order to assess this possibility, we apply the Oaxaca-Blinder decomposition technique (OAXACA, 1973; BLINDER, 1973) that allow us to measure how much of the mean gap between girls and boys depends on differences in endowments, and how much of it is attributable to differences in the return of those endowments, namely in the unexplained component. An extension of this method is the Recentered Influence Function (RIF) proposed by Fortin et al. (2011) and Firpo et al. (2009) which allows us to study the contribution of specific covariates at different parts of the test scores distributions.

The core idea of these decomposition methods[1] is to explain the distribution of test scores by a set of factors that are correlated with educational achievement. The gap is, therefore, decomposed into a part that is due to gender differences in the magnitudes of the determinants of the test score in question, and gender differences on the effects of these determinants. This, in turn, help us to identify if the gap is related to differences between boys and girls in resources, or whether they are due to differences in the way boys and girls use these resources for school success. This can be an important guide for policies, since one that aims to diminish inequality between the two genders might have a limited effect in raising performance if it focuses only on the distribution of resources, and not on their use. This paper aims to investigate the factors associated with gender differentials in Math and Portuguese academic achievement in Brazil. We use the two already mentioned types of decomposition methods and apply them to a pool of four cohorts of 5th graders, and to a pool of 9th graders. The idea is to assess students, teachers and schools characteristics in order to identify the sources of these gaps, and to verify if there is any difference between the two grades.

We contribute to the existing literature, first, by exploring the factors associated with differences on gender achievement in Brazil. Also, we contribute by employing mean and quantile decomposition analysis to identify covariates contributing to the gender gap at various points in the test scores distribution.

Our results, overall, indicate that boys perform relatively better in math than girls, an advantage that reverses in literacy. Besides that, we highlight that boys and girls have very distinct educational production functions, as the unexplained or coefficient component of the Oaxaca-Blinder decomposition, which corresponds to the way they translate their background characteristics into achievement, is responsible for the greater bulk of both gaps. Also, from the aspects that we are able to measure, we identify that the students characteristics are the main constituents of this coefficient component. Moreover, by employing the extension of the Oaxaca-Blinder decomposition, we verify that in all percentiles boys and girls tend to perform relatively more equally in Math than in Portuguese, and on 5th grade than on 9th. In sum, our results show that regardless of the distributional statistic, the majority of the gender difference is related to how boys and girls translate their background characteristics into achievement, and not to their characteristics in itself.

[1] Two elucidative descriptions of decomposition methods can be found on Firpo (2017) and O'Donnel et al. (2008).

The rest of the paper is structured as follows. The next section presents our dataset and describes the two decomposition techniques we use. Section.3 presents the results, separated by each method. Section 4 contains the concluding remarks.

2.2 DATA DESCRIPTION AND METHODS

This subchapter describes the data and the methods we apply on this essay.

2.2.1 Data

This paper uses four cohorts of the Brazilian Ministry of Education assessment of Math and Portuguese learning for 5th graders and, also, for 9th graders in public schools (2009, 2011, 2013 and 2015). This assessment is biannual, and includes questionnaires for Principals, Teachers, Students and of School characteristics. The students' questionnaire includes detailed socioeconomic information and a Math and Portuguese test. The Teachers' questionnaire, besides having their background information, has also questions regarding their contract with the school, and other work related questions. We also match the dataset from these four assessments to the corresponding Brazilian Education Census, in order to obtain data on school infrastructure.

The variables we use on our estimates are based on other studies that try to identify student school success in Brazil, such as Franco and Menezes-Filho (2008), Biondi and Felício (2007), Soares and Candian (2007) and Soares et al. (2016). Therefore, the students characteristics[2] we are interested in are race, which is defined by white or non-white, if the student has the expected age for this grade, if lives with his mom and dad, if he or she works, if performs more than one hour of domestic tasks, if attended kindergarten, if he or she lives in an urban area, if he or she does the Math or Portuguese homework and if at least one of the parents have a graduate degree, we also include two indexes[3], a socioeconomic status index (SES), which takes into consideration the number of bedrooms, bathrooms, computers, cars and televisions in the student household, and an index for family involvement with the student, which takes into consideration if whether the parents attend school meetings, if they talk about what happened at school and, if they incentive the student to study, go to school

[2] Table 1A in the Appendix contains all variables we use in this study and their description.
[3] Both indexes are built from the first component of Principal Component Analysis, a procedure commonly used as a dimensionality reduction technique.

and read. The school characteristics are if at the beginning of the school year the students had their books and an infrastructure quality index which consists of if the school has a computer lab, a science lab, a sports gym, a library and a reading room. The teachers characteristics are sex, that is, if the teacher is female, if he or she has a graduate degree, if works in another activity besides teaching, if he or she has more than 10 years of experience, if works more than 40hrs a week and if whether he or she have a permanent teaching contract. We also include year dummies and for each of the five Brazilian regions in our estimations.

Table 2.1, below, presents descriptive statistics[4] of the number of students and test scores for 5th and 9th graders.

Table 2.1 - Descriptive Statistics of sample Math and Portuguese test scores.

		Students	Mean	Std Dev	10th	50th	90th
5th grade Math	Boys	1994489	221.42	48.50	159.34	219.42	286.62
	Girls	2075855	215.82	46.06	158.30	212.89	278.07
5th grade Portuguese	Boys	1999704	198.03	46.51	139.09	195.17	260.91
	Girls	2081876	206.37	46.63	146.09	204.85	268.60
9th grade Math	Boys	1640355	253.71	46.64	191.75	254.00	313.70
	Girls	1933266	244.70	44.40	187.63	243.04	303.09
9th grade Portuguese	Boys	1628734	238.68	47.59	174.94	238.83	301.38
	Girls	1920328	251.32	45.32	190.61	252.43	309.87

Source: Elaborated by the authors using SAEB (2009-2015).

As we can see from Table 2.1, the number of girls and boys are fairly well distributed in 5th grade, and there are a slight higher number of girls at 9th grade. Also, girls tend to perform worse than boys in Math, a difference that is greater at the highest grade. However, girls tend to outperform boys on language, a difference that is also increasing with the grade. All the percentiles tend to reproduce these patterns, the exception is the lowest percentile in Math for 5th graders, where the gender difference is almost negligible.

2.2.2 **Empirical Specifications**

Our primary objective is in how the characteristics of boys and girls influence their Math and Portuguese performance. In addition, it is also important to examine if there is any difference when we consider older students, by including a more advanced grade. In order to attend our goals, we first employ the traditional Oaxaca-Blinder decomposition (OAXACA,

[4] A more detailed version of this table is on Table 2A, on the Appendix. Descriptive statistics of the other variables used in this study are on Table 3A and 4A, also on the Appendix.

1973; BLINDER, 19743), a counterfactual decomposition technique, which estimates a linear education production function for each gender, such as:

$$Y_{ij} = \alpha_{j0} + \sum_{k=1}^{K} X_{ik}\beta_{jk} + e_{ji} \qquad 2.1$$

Where Y_i is the Math of Portuguese tests scores of each *i* individual and *j* gender, boys or girls, and, as discussed above, X_{ik} is the vector of the *K* covariates we use in this paper. Therefore, the overall mean difference in outcomes between boys and girls can be written as:

$$E[Y_{ib}/X_{ib}] - E[Y_{ig}/X_{ig}] = (\hat{\alpha}_b - \hat{\alpha}_g) + \left(\sum_{k=1}^{K} \bar{X}_{bk}\hat{\beta}_{bk} - \sum_{k=1}^{K} \bar{X}_{gk}\hat{\beta}_{gk}\right) = \Delta \qquad 2.2$$

If we add and subtract $\sum_{k=1}^{K} \bar{X}_{bk}\,\hat{\beta}_{gk}$, Δ can be rewritten as a sum of two different components:

$$\Delta = \sum_{k=1}^{K} (\bar{X}_{bk} - \bar{X}_{gk})\,\hat{\beta}_{gk} + \left[(\hat{\alpha}_b - \hat{\alpha}_g) + \sum_{k=1}^{K} \bar{X}_{bk}\,(\hat{\beta}_{bk} - \hat{\beta}_{gk})\right] \qquad 2.3$$

The first term on the right hand side of the equation is called the characteristics or explained effect, as it portrays the differences on the distributions of the covariates between the two genders. The second term is called the coefficient or unexplained effect, since it reflects differences on the returns of each characteristic, between boys and girls. Note, that both these elements rely on $\sum_{k=1}^{K} \bar{X}_{bk}\,\hat{\beta}_{gk}$, a counterfactual estimate or, in other words, how boys would score on their Math and Portuguese tests if they had the same estimated coefficients as girls.

One issue regarding this technique is due to the identification of the contribution of categorical variables. The problem is that the choice of the reference group affects the decomposition results[5]. We are able to contour this issue using normalized regressions, a method proposed by Yun (2005) that consists in expressing the effects as deviations from the mean.

---

[5] For more information see Jann (2008).

Another limitation of Oaxaca-Blinder decomposition, as discussed by Barsky et al. (2002), is that they may not provide consistent estimates of the coefficients and characteristics effect when the conditional mean is a non-linear function. This happens because the counterfactual mean will not be equal to $\sum_{k=1}^{K} \bar{X}_{bk} \hat{\beta}_{gk}$. One possible solution to this problem is to use a reweighting approach as in DiNardo, Fortin and Lemieux (1996) that reweight the sample of boys so that the distribution of their characteristics is similar to that of girls, using the following reweighting function:

$$\psi(X) = \frac{P(g = 1/X)/P(g = 1)}{P(g = 0/X)/P(g = 0)} \qquad 2.4$$

Where $P(g = 1/X)$ represents the probability of a student being a girl and, $P(g = 0)$ and $P(g = 1)$, are, respectively, the samples proportions of boys and girls. The reweighting factor is then applied to the boys sample to calculate the counterfactual test score distribution.

Firpo et al. (2009), Fortin et al. (2011) and Chernozhukov et al. (2013) expand the Oaxaca-Blinder decomposition of the means and provide a comprehensive approach to study the entire distribution function. Their method consists of two stages. First, by transforming the outcome variable, using influence functions, in order to obtain what they call Recentered Influence Function (RIF), which can be computed for most distributional statistics and, therefore, allows their decomposition to assess quantiles, variance, and even inequality measures. The second stage uses these estimates to generate Oaxaca-Blinder decomposition for the measures of interest.

The approach, in its simplest form, assumes that the conditional expectation of the $RIF(Y; Q_t)$ can be modelled as a linear function of the explanatory variables:

$$E[RIF(Y; Q_\tau)/X] = X\gamma + \varepsilon \qquad 2.5$$

Nonetheless, the RIF function can be defined as:

$$RIF(Y; Q_\tau) = Q_\tau + \frac{\tau - \boldsymbol{I}(Y \leq Q_\tau)}{f_Y(Q_\tau)} \qquad 2.6$$

Where $\boldsymbol{I}(Y \leq Q_\tau)$ is an indicator variable for whether the outcome variable is smaller or equal to the quantile Q_τ. Running a linear regression of $\boldsymbol{I}(Y \leq Q_\tau)$ on X is a distributional regression estimated at $y = Q_\tau$.

According to Fipo, et al. (2009) to run regressions of the RIF on the vector of covariates one should plug in the estimates of the sample quantile, $\hat{Q}_\tau$, and of the density at that point, $\hat{f}(\hat{Q}_\tau)$, into Equation 2.6. Letting the coefficients of the unconditional quantile regressions for each group be:

$$\hat{\gamma}_{g,\tau} = \left(\sum_{i \in J} X_i X_i^T\right)^{-1} \sum_{i \in J} \widehat{RIF}\left(N_{ji}; Q_{j,\tau}\right) X_i \qquad 2.7$$

We can write the equivalent of the Oaxaca-Blinder decomposition for any unconditional quantiles as:

$$\hat{\Delta} = \bar{X}_b\left(\hat{\gamma}_{b,\tau} - \hat{\gamma}_{g,\tau}\right) + \left(\bar{X}_b - \bar{X}_g\right)\hat{\gamma}_{g,\tau} \qquad 2.8$$

The right hand side of Equation 2.7 displays the coefficient and the characteristics effect. The identification of these effects by quantile, allow us to analyze, with greater detail the influence of the students, schools, teachers and families in Math and Portuguese achievement. However, the two already mentioned issues regarding the Oaxaca-Blinder decomposition, namely, the identification of categorical variables and the nonlinearity of the conditional expectation, may hold in RIF regressions, hence, we also apply normalized regressions and Dinardo, Fortin and Lemieux (1996) reweighting approach to perform this analysis.

2.3 RESULTS

This subchapter describes the results of the present essay.

2.3.1 Oaxaca-Blinder decomposition

In order to assess whether test score determination process differs according to gender, Table 2.2 shows how much of the difference in achievement between the two genders is explained by the changes in characteristics or due to the estimated coefficients. It is important to highlight that we calculate the difference with girls as the reference group, therefore the gap is positive if it favors boys and negative otherwise.

Table 2.2 - Oaxaca-Blinder decomposition 5th and 9th grade students in Brazil (2009-2015).

	5th grade		9th grade	
	Math	Portuguese	Math	Portuguese
Boys	221.419***	198.055***	253.714***	238.693***
	(0.131)	(0.111)	(0.124)	(0.117)
Girls	215.816***	206.411***	244.695***	251.339***
	(0.128)	(0.111)	(0.118)	(0.110)
Difference	5.603***	-8.356***	9.019***	-12.646***
	(0.068)	(0.064)	(0.082)	(0.077)
Explained	-0.032	-0.190***	0.232***	0.215***
	(0.028)	(0.028)	(0.025)	(0.026)
Students	-0.023	-0.165***	0.013	-0.029
	(0.020)	(0.020)	(0.019)	(0.020)
Teachers	0.002	0.002	-0.002	-0.001
	(0.003)	(0.003)	(0.002)	(0.002)
Schools	0.004	0.005*	0.015***	0.019***
	(0.002)	(0.003)	(0.004)	(0.003)
Regions	0.007	0.008	0.003	0.013
	(0.013)	(0.010)	(0.011)	(0.010)
Year	-0.022***	-0.041***	0.204***	0.212***
	(0.005)	(0.009)	(0.014)	(0.015)
Unexplained	5.635***	-8.166***	8.786***	-12.860***
	(0.063)	(0.057)	(0.078)	(0.073)
Students	1.663***	-0.575**	2.517***	3.604***
	(0.246)	(0.226)	(0.251)	(0.243)
Teachers	-0.014	-0.356***	-0.173	0.200
	(0.115)	(0.117)	(0.168)	(0.171)
Schools	0.071*	0.028	-0.003	-0.077*
	(0.039)	(0.038)	(0.037)	(0.039)
Regions	-0.370***	-0.286***	-0.961***	-0.709***
	(0.040)	(0.038)	(0.056)	(0.053)
Year	0.083***	0.001	-0.080***	-0.056***
	(0.009)	(0.009)	(0.017)	(0.018)
Intercept	4.202***	-6.978***	7.487***	-15.822***
	(0.266)	(0.248)	(0.319)	(0.311)
Observations	4070344	4056137	3573621	3538880

Robust standard errors clustered at the school level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

As we can see from Table 2.2, boys and girls have different educational production functions for Math and Portuguese. The gender gaps on achievement are mostly due to the unexplained, or coefficient component of the Oaxaca-Blinder decomposition, but have opposite signs when we compare the two subjects. Moreover, the mean differences on achievement are smaller for 5th graders. This conclusion is in accordance with the idea that the gap widens with age, as pointed out by Fryer and Levitt (2010) and Bedard and Cho (2010).

In general, however, the results show that a greater part of the gender gap is not due to endowments and is more associated with the return of these characteristics, especially the ones related to student background and to the intercept. Interestingly, studies for developed countries have found a greater role of the endowments on achievement (LE; NGUYEN, 2018; GEVREK; SEIBERLICH, 2014; SOHN, 2012). International studies have also encountered a large part of the gender test score gap being explained by the intercept (LE; NGUYEN, 2018; GOLSTEYN; SCHILS, 2014). Our results show that the endowments part of the gender test score differences of 5^{th} grades is only statistically significant in Portuguese and, indicates that the individual characteristics of girls help them outperform boys at literacy. As to the unexplained component, the detailed decomposition, as shown in small letters of Table 2.2, highlights that the differences rely not only on the intercept, but also on the corresponding characteristics of students and schools. In literacy, all of the grouped variables show an improvement of girls relative to boys on achievement, therefore contributing to widen the gap. Nonetheless, the return to individual characteristics in Math contributes a better performance of boys relative to girls.

When we look at the unexplained component detailed by each variable, on Table 5A in the Appendix, both subjects have doing the subjects' homework and living in urban areas are the main contributors towards a relative better performance of boys, while work and attending kindergarten are the main contributors for girls. In this sense, Le and Nguyen (2018) also highlights the importance of pre-school on Australian gender test differences and it also contributes positively for girls' achievement. Moreover, investigating the differences on top performers Fortin et al. (2015) also finds that family environment and labor market work during school are important in accounting for gender achievement gap. Also, regardless of the subject, teacher sex, that is, if the teacher is female, also contributes towards girls' relative performance. This last result is consistent with the importance of gender on the dynamics between students and teachers, as highlighted by Dee (2006).

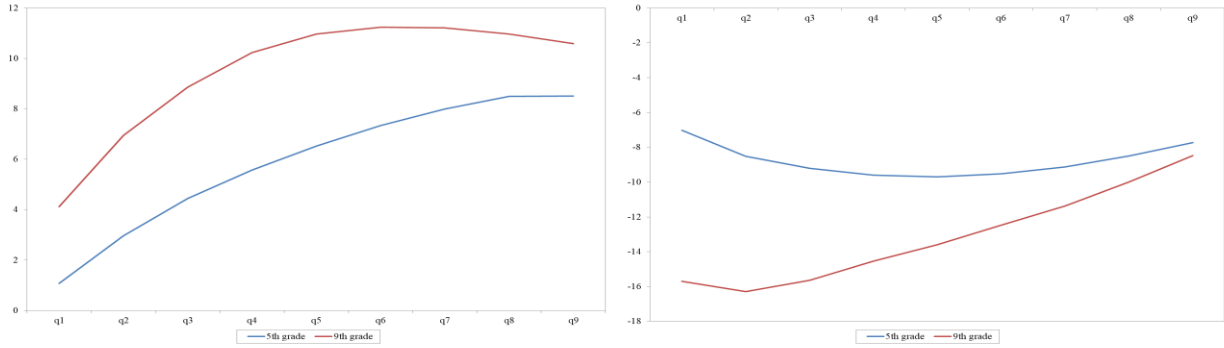
The results for the 9^{th} grade students keeps in line with the pattern described for 5^{th} graders, that is, again the unexplained component of the decomposition is responsible for the greater bulk of the gender gap. However, the explained part is statistically significant for both subjects, and both indicate an improvement towards boys relative to girls, especially due to the year effect and the schools characteristics. As to the unexplained component, the return of individual characteristics have a positive sign in literacy, meaning that, despite the fact that boys characteristics contribute to a relative better performance of them over girls in language, the effect of the intercept more than suppresses it. When we look at the contribution of each variable for both subjects, on Table 6A in the Appendix, the ones on the coefficients component that are most important for the relative mean performance of girls are having attended kindergarten and if she lives with her mom and dad, and, for boys, are if they are living in urban areas, are working, or if at least one of the parents graduated from university.

The results described here are in accordance with Biondi and Felício (2007), Franco and Menezes-Filho (2008) and Scorzafave and Ferreira (2011) who acknowledge that the characteristics of the students and of their families are the most relevant to explain academic performance in Brazil, while school, teachers and the principal characteristics display a minor role in student's achievement. Moreover, Menshawy, Oliver and Appleton (2004) who explore math gender inequalities in Middle East and North Africa countries, find results similar to ours for Syria and Tunisia. The next subsection further explores this same framework but using the RIF decomposition for percentiles of the test scores distributions.

2.3.2 RIF decomposition

After assessing the differences on mean academic achievement, we present the results capturing the gap along the distribution of the test scores at the two different grades. For that, we classify both girls and boys into percentiles of each test score distribution. Figure 2.1, shows the estimated differences, while Table 7A and 8A, on the Appendix, presents the detailed version of the estimated RIF decompositions.

Figure 2.1 - Quantile differences in Math and Portuguese achievement from RIF decomposition (2009-2015).



a) Math quantile difference   b) Portuguese quantile difference

Source: Elaborated by the authors using SAEB (2009-2015).

Figure 2.1 highlight that, regarding the difference in performance, boys and girls tend to perform more equally in Math, as shown by the scale of both graphs. Students in 5th grade have a less unequal performance on both subjects and on all percentiles than 9th grade students. This is shown by the fact that the blue line, which indicates 5th graders, is below the one for 9th grade in Math, where boys outperform girls. Also, since the gap is calculated by the difference between boys and girls, the line for 5th graders is above the red line on Portuguese in all percentiles.

As to the shapes of the lines, 9th grade gender differentials in Math, as described by the red line, is steeper, which also indicates a greater gap on the first percentiles, as compared to the blue line. Nonetheless, it also shows a slight decrease on the two last percentiles, indicating that the difference between the best girls and boys in Math tend to slightly decrease. In Portuguese, the line for 9th graders shows that the worst performing boys tend to perform much lower than the worst girls, a difference that tend to diminish as we explore further percentiles. With regard to 5th grade, the line for Portuguese is more U-shaped, indicating that the greatest difference in performance is more towards the middle, that is, the best and worst tend to perform relatively more equally.

Tables 7A, on the Appendix, detail the decomposition by each percentile. There, it is shown that for 5th grade Math, as the top panel of Table 7A shows, the unexplained component of Student characteristics is not statistically significant only in the first percentile. In Portuguese, at the lower panel of Table 7A, the students aggregated variable in the coefficients component changes its sign, showing that it contributes towards a relative better performance of girls at the lowest percentiles, and does the opposite on the highest quantiles. The main drivers of this change[6] are three variables, the intercept and having the ideal age for the grade, that in accordance with the aggregated variable also change sings, and if attended kindergarten, whose effect is higher and favors girls at lower percentiles.

Moreover, Table 8A describes the factors behind 9th grade gender differences by percentile. The results are very similar to the ones described by the Oaxaca-Blinder decomposition, and indicate a very uniform importance of students' characteristics at the unexplained component of the decomposition for both subjects. Also, and specifically for Math, there is an increasing importance of the intercept as main driver of the gender difference, indicating that our model is able to account for most of the gap at lower quantiles, but other factors explain a greater bulk of this gap at higher percentiles.

Similar to our results, Ellison and Swanson (2010) highlights that for the United States the gender gap in Math skills widens at the top of the distribution. Besides that, Robinson and Lubiensky (2010) also find that in reading, gaps favoring females generally narrow but widen

[6] The detailed decomposition by each variable is not shown due to space restrictions, but are available upon request from the authors.

among low-achieving students. This is an interesting contrast with the Brazilian case, as we have seen that differences in literacy narrow among bottom and top performers, the latter especially on 9th grade. Moreover, Le and Nguyen (2018) investigating Australian test scores from 3rd to 7th grade find, in contrast with our results, sparse significant gender differences on reading, but they also find a more distinguished difference favoring males in math over virtually the whole distribution and in all grades, especially at the upper end of the distribution. Nonetheless, in relation to these other authors, our findings show that the gender differences in the Brazilian case are much more pronounced. One possible explanation for that, according to Guiso et al. (2008), is the role that social and cultural gender-biased environments play in these sex differences, as he indicates, by investigating a range of countries, that the gender gap in math disappears in more gender-equal societies, because girls improve their test scores, while the gap on literature, which favors girls, expands.

Overall, regardless of the grade, our results indicate that the differences in achievement are not due to the fact that one of the genders comes from a better background, or have better teachers or study in better schools, but are associated with the return of these characteristics. Also, despite the fact that we are able to disentangle the gender gap and point out some important contributors of this difference with the covariates in our dataset, most of the gap is still due to the intercept, and therefore require additional investigations, a task we intend to continue performing on the next articles of this thesis.

2.4 CONCLUDING REMARKS

This paper aims at identifying the sources of gender differences in Math and Portuguese achievement in 5th and 9th grade. The inclusion of the more advanced grade is justified by the search for differences when considering older students. In order to attend our goals, we use a pool of four cohorts of 5th graders, and, separately, a pool of 9th graders, to employ two different types of decomposition methods on a dataset that includes students, teachers and schools characteristics.

Moreover, we apply the Oaxaca-Blinder decomposition technique that allow us to assess how much of the mean gap between girls and boys depends on differences in endowments, and how much of it is attributable to differences in the return of those endowments, namely in the unexplained component. The other decomposition method we use is an extension of the Oaxaca-Blinder that allows us to study the contribution of specific covariates at different parts of the test scores distributions.

The results we find indicate that the gap in Math favors boys, and the one in literacy favors girls. Besides that, we highlight that boys and girls have very distinct educational production functions, as the unexplained or coefficient component of the Oaxaca-Blinder decomposition, which corresponds to the greater bulk of both gaps, have a different sign for each subject. Moreover, from the aspects we are able to measure, we identify that the Students characteristics is the main constituent of this coefficient component. In sum, our results show that the majority of the gender difference is related to how boys and girls translate their background characteristics into achievement, and not to their characteristics in itself, and that this effect favors boys in Math and girls in Portuguese.

Also, despite the fact that the quantile decompositions results are very similar to the ones described by the Oaxaca-Blinder decomposition, however, the coefficient effect of the students' characteristics on 5^{th} grade Portuguese academic achievement, changes its sign after the 40^{th} percentile, showing that it contributes towards a relative better performance of girls at the lowest percentiles, and does the opposite on the highest quantiles. Therefore, indicating a change in the production function that is not captured by the mean decomposition. Nonetheless, with the percentile decomposition we are also able to identify that boys and girls tend to perform relatively more equally in Math than in Portuguese, and on 5^{th} grade than on 9^{th}.

Overall, regardless of the grade, our results indicate that the differences in achievement are not due to the fact that one of the genders comes from a better background, or have better teachers or study in better schools, but is associated with the return of these characteristics. Therefore, on the policy perspective, our study findings indicates that a more efficient policy aiming at the reduction of test score differences between boys and girls should look beyond resources. It also reduces the role played by teachers and schools in narrowing these test score differences. Moreover, despite the fact that we are able to disentangle the gender gap and point out some important contributors of this difference with the covariates in our dataset, most of the gap is still due to the intercept, and therefore require additional investigations, a task we intend to continue performing on the next articles of this thesis.

2.5 REFERENCES

BARSKY, R. BOUND, J.; CHARLES, K. K; LUPTON, J. P. Accounting for the black-white wealth gap: a nonparametric approach. **Journal of the American Statistical Association**, New York, v. 97, p. 663-673, 2002.

BEDARD, K.; CHO, I. Early gender test score gaps across OECD countries. **Economics of Education Review**, Cambridge, v. 29, n. 3, p. 348-363, 2010.

BIONDI, R. L.; FELÍCIO, F. **Atributos escolares e o desempenho dos estudantes**: uma análise em painel dos dados do SAEB. Brasília: INEP, 2007. (Texto para Discussão, 25).

BLINDER, A. S. Wage Discrimination: Reduced Form and Structural Estimates. **Journal of Human Resources**, Madison, v. 8, n, 4, p. 436-55, 1973.

DEE, T. S. The Why Chromosome: How a Teacher's Gender Affects Boys and Girls. **Education Next**, Cambridge, v. 6, n. 4, p. 68-75, 2006.

DINARDO, J.; FORTIN, N. M.; LEMIEUX, T. Labor Market Institutions and the Distribution of Wages 1973-1992: A Semiparametric Approach. **Econometrica**, Chicago, v. 64, p. 1001-1044, 1996.

ELLISON, G.; SWANSON, A. The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions. **Journal of Economic Perspectives**, Nashville, v. 24, n. 2, p. 109-28, 2010.

FIRPO, S.; FORTIN, N. M.; LEMIEUX, T. Unconditional Quantile Regressions. **Econometrica**, Chicago, v. 77, p. 953-973, 2009.

FIRPO, S. Identifying and measuring economic discrimination. **IZA World of Labor**, Bonn, v. 347, 2017.

FORTIN, N. M.; LEMIEUX, T; FIRPO, S. Decomposition Methods in Economics. In: ASHENFELTER, O.; CARD, D. (Eds.). **Handbook of Labor Economics**, Amsterdam, Elsevier, 2011.

FORTIN, N.; OREOPOULOS; P.; PHIPPS, S. Leaving Boys Behind: Gender Disparities in High Academic Achievement. **Journal of Human Resources**, Madison, v. 50, n. 3, 2015

FRANCO, A. M. P.; MENEZES-FILHO, N. A.; Os determinantes do aprendizado com dados de um painel de escolas do SAEB. In: Encontro Nacional de Economia, 2008. **Anais...** Salvador: Anpec, 2008.

FRYER JR, R.; LEVITT, S. D. An Empirical Analysis of the Gender Gap in Mathematics. **American Economic Journal: Applied Economics**, Washington, v. 2, n. 2, p. 210-40, 2010.

GEVREK, Z. E.; SEIBERLICH, R. R. Semiparametric Decomposition of the Gender Achievement Gap: An Application for Turkey. **Labour Economics**, Amsterdam, v. 31, p. 27-44, 2014.

GOLSTEYN, B, H.; SCHILS, T. Gender gaps in primary school achievement: a decomposition into endowments and returns to IQ and non-cognitive factors. **Economics of Education Review**, Cambridge, v. 41, p. 176-187, 2014.

GUISO, L.; MONTE, F.; SAPIENZA, P.; ZINGALES, L. Culture, gender and math. **Science**, Washington, v. 320, p. 1164-1165, 2008.

JANN, B. The Blinder-Oaxaca decomposition for linear regression models. **The Stata Journal**, College Station, v. 8, n. 4, p. 453-479, 2008.

KASSOUF, A. L. Wage gender discrimination and segmentation in the Brazilian labor market. **Economia Aplicada**. Ribeirão Preto, v. 2, n. 2, p. 243-269, 1998.

LE, H. T.; NGUYEN, H. T. The evolution of the gender test score gap through seventh grade: new insights from Australia using unconditional quantile regression and decomposition. **IZA Journal of Labor Economics**, Amsterdam, v. 7, n. 2, 2018.

MADALOZZO, R. **Market and Home Production**: Gender Differences in Brazil. São Paulo: Insper, 2009. (Insper Working Paper, 168**)**.

MADALOZZO, R.; ARTES, R. Escolhas profissionais e impactos no diferencial salarial entre homens e mulheres. **Cadernos de Pesquisa**, Brasília, v. 47, n. 163, 2017.

MENSHAWY, B.; OLIVER, M.; APPLETON, S**. Gender differentials in maths test scores in MENA countries**. Nottingham: CREDIT, 2004. (CREDIT Research Paper. 12/04).

OAXACA, R. Male-Female Wage Differentials in Urban Labor Markets. **International Economic Review**, Pennsylvania, v. 14, n. 3, p. 693-709, 1973.

O'DONNEL, O.; VAN DOORSLAER, E.; WAGSTAFF, A; LINDELOW, M. Explaining differences between groups: Oaxaca decomposition. **Analyzing Health Equity Using Household Survey Data**, Washington-DC: World Bank, 2008.

RASCHE JÚNIOR, F. **Privação de Liberdades Femininas:** uma análise da participação das mulheres nas áreas de ciência e tecnologia no Brasil entre 2004 e 2013. 2015. 73f. Undergraduate Thesis (Graduation in Economics) – Faculty of Economics, Rio Grande do Sul Federal University, Porto Alegre.

ROBINSON, J. P.; LUBIENSKI, S. T. The Development of Gender Achievement Gaps in Mathematics and Reading During Elementary and Middle School: Examining Direct Cognitive Assessments and Teacher Ratings. **American Educational Research Journal**, Washington, p. 1035, 2010.

SCHULTZ, T., P. Why Governments Should Invest more to Educate Girls. **World Development**, Oxford, v. 30, n. 2, p. 207-225, 2002.

SCORZAFAVE, L. G.; FERREIRA, R. A. Desigualdade de Proficiência no Ensino Fundamental Público Brasileiro: Uma Análise de Decomposição. **Revista EconomiA**, São Paulo, v. 12, n. 2, p. 337-359, 2011.

SOHN, K. A new insight into the gender gap in math. **Bulletin of Economic Research**, Hull, v. 64, n. 1, p. 135-155, 2012.

TILLMANN; E. T.; COMIM, F. Os determinantes da decisão entre estudo e trabalho dos jovens no Brasil e a geração Nem-Nem. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 46, n. 2, p. 47-78, 2016.

WORLD BANK. **World Development Report 2012**: gender equality and development. Washington: World Bank Publications, 2011.

YUN, M. A simple solution to the identification problem in detailed wage decompositions. **Economic Inquiry**, Long Beach, v. 43, n. 4, p. 766-772, 2005.

2.6 APPENDIX

Table 1A - Description of the variables included in the estimates.

Variable	Description
Qescz	index built from school has a computer lab, a science lab, a sports gym, a library and a reading room
Sesz	index built from number of bedrooms, bathrooms, computers, cars and televisions in the student household
Familiaz	index built from the parents attend school meetings, if they talk about what happened at school and, if they incentive the student to study, go to school and read
White	1 if student is white, 0 otherwise
Idealage	1 if has the ideal age for the grade, 0 otherwise
mom&dad	1 if student lives with his mom and dad, 0 otherwise
parentes_univdg	1 if at least one of the student parent has na university degree
domestic chores	1 if student reports doing household chores, 0 otherwise
Work	1 if student works, 0 otherwise
Urban	1 if student lives in an urban area, 0 otherwise
Begschool	1 if student attended kindergarten, 0 otherwise
Teachsex	1 if teacher is female, 0 otherwise
Teachhighdg	1 if teacher has university degree, 0 otherwise
Teachothact	1 if teacher works on another job besides teaching, 0 otherwise
Teachexper	1 if teacher has more than 10 years of experience, 0 otherwise
Teachhrs	1 if teacher works more than 40 hours, 0 otherwise
Hmwok	1 if student reports that he or she does his homework, 0 otherwise
Teachcontract	1 if teacher has a permanent job contrat with the school, 0 otherwise
Region	indicator variable for each of the five regions in Brazil (North, Northeast, Central-West, Southeast and South)
Bookatbeg	1 if teacher states that most of his/hers students had the classroom book at the beggining of the school year, 0 otherwise

Table 2A - Descriptive statistics of Math and Portuguese test scores over the years (2009-2015).

		Students	Avg Score	Std dev	Min	Max
2009						
5th grade Math	Boys	468324	212.27	47.68	93.34	358.31
	Girls	470985	208.19	44.95	93.34	358.31
5th grade Portuguese	Boys	466341	185.77	43.14	87.13	331.29
	Girls	469176	193.72	44.12	87.13	331.29
9th grade Math	Boys	365504	248.84	45.96	124.74	413.39
	Girls	446152	238.15	43.53	124.75	413.39
9th grade Portuguese	Boys	352521	234.04	46.52	100.56	380.71
	Girls	430762	246.02	44.38	100.56	380.71
2011		Students	Avg Score	Std dev	Min	Max
5th grade Math	Boys	491972	219.16	48.03	90.13	338.18
	Girls	521440	213.02	45.65	90.13	338.18
5th grade Portuguese	Boys	484191	192.08	44.65	77.20	339.46
	Girls	513096	201.97	45.03	77.20	339.46
9th grade Math	Boys	308888	252.14	47.57	107.23	398.27
	Girls	399159	244.90	45.01	108.18	398.27
9th grade Portuguese	Boys	312510	234.70	46.74	103.46	380.83
	Girls	403840	249.02	44.26	105.23	380.83
2013		Students	Avg Score	Std dev	Min	Max
5th grade Math	Boys	441025	225.14	50.55	78.92	341.25
	Girls	475526	219.20	48.38	78.92	341.25
5th grade Portuguese	Boys	434354	201.45	47.77	84.89	330.69
	Girls	468369	209.63	48.15	84.89	330.69
9th grade Math	Boys	447323	251.94	47.45	129.25	414.71
	Girls	507211	244.87	45.64	129.25	414.71
9th grade Portuguese	Boys	447461	236.99	47.53	124.00	379.16
	Girls	508727	250.23	46.37	124.00	379.16
2015		Students	Avg Score	Std dev	Min	Max
5th grade Math	Boys	593168	226.98	46.73	120.99	366.45
	Girls	607904	221.00	44.39	120.99	366.45
5th grade Portuguese	Boys	585396	209.25	46.47	88.73	337.29
	Girls	599997	216.49	45.89	88.73	337.29
9th grade Math	Boys	518640	259.21	45.11	154.03	440.14
	Girls	580744	248.56	42.88	154.03	440.14
9th grade Portuguese	Boys	516242	245.58	48.05	106.74	395.28
	Girls	576999	257.28	45.07	106.74	395.28

Source: Elaborated by the authors using SAEB (2009-2015).

Table 3A - Descriptive statistics of 5th grade covariates.

Variable	Boys	Girls	Variable	Boys	Girls
qschool	0,028	0,026	teach_grad	0,904	0,904
	(0,994)	(0,996)		(0,295)	(0,295)
sesz	0,013	0,009	teach_postg	0,591	0,592
	(0,974)	(0,983)		(0,492)	(0,492)
familyz	0,083	0,092	teach_othactivity	0,414	0,414
	(0,822)	(0,871)		(0,493)	(0,493)
White	0,372	0,370	teach_+10yrsexper	0,701	0,701
	(0,483)	(0,483)		(0,458)	(0,458)
ideal_age	0,843	0,844	teach_+40hrs	0,553	0,553
	(0,363)	(0,363)		(0,497)	(0,497)
mom&dad	0,697	0,697	port_hmwork	0,956	0,966
	(0,460)	(0,460)		(0,205)	(0,182)
parents_univdg	0,150	0,148	math_hmwork	0,978	0,979
	(0,357)	(0,355)		(0,145)	(0,144)
domestic chores	0,413	0,413	teach_contract	0,743	0,743
	(0,492)	(0,492)		(0,437)	(0,437)
work	0,072	0,071	North	0,089	0,088
	(0,258)	(0,257)		(0,285)	(0,283)
urban	0,916	0,915	Northeast	0,226	0,228
	(0,278)	(0,279)		(0,418)	(0,420)
kindergarten	0,790	0,791	Southeast	0,452	0,452
	(0,408)	(0,407)		(0,498)	(0,498)
book at begin	0,731	0,731	South	0,150	0,150
	(0,443)	(0,444)		(0,357)	(0,357)
teach_sex	0,088	0,088	Central-West	0,083	0,082
	(0,284)	(0,284)		(0,275)	(0,275)

Standard deviation in parentheses.

Source: Elaborated by the authors using SAEB (2009-2015).

Table 4A - Descriptive statistics of 9th grade covariates.

Variable	Boys	Girls	Variable	Boys	Girls
Qschool	0,010	0,007	teach_grad	0,962	0,962
	(1,003)	(0,999)		(0,192)	(0,192)
Sesz	-0,071	-0,079	teach_postg	0,649	0,648
	(0,959)	(0,967)		(0,477)	(0,478)
Familyz	0,046	0,039	teach_othactivity	0,461	0,461
	(0,905)	(0,945)		(0,498)	(0,499)
White	0,381	0,379	teach_+10yrsexper	0,641	0,641
	(0,486)	(0,485)		(0,480)	(0,480)
ideal_age	0,552	0,558	teach_+40hrs	0,560	0,557
	(0,497)	(0,497)		(0,496)	(0,497)
mom&dad	0,640	0,641	port_hmwork	0,949	0,957
	(0,480)	(0,480)		(0,219)	(0,203)
parents_univdg	0,110	0,105	math_hmwork	0,940	0,940
	(0,312)	(0,307)		(0,237)	(0,238)
domestic chores	0,631	0,630	teach_contract	0,716	0,716
	(0,482)	(0,483)		(0,451)	(0,451)
Work	0,201	0,205	North	0,083	0,080
	(0,400)	(0,403)		(0,276)	(0,271)
Urban	0,923	0,920	Northeast	0,259	0,263
	(0,267)	(0,271)		(0,438)	(0,440)
kindergarten	0,859	0,859	Southeast	0,444	0,444
	(0,348)	(0,348)		(0,497)	(0,497)
book at begin	0,736	0,731	South	0,138	0,138
	(0,441)	(0,444)		(0,345)	(0,345)
teach_sex	0,325	0,322	Central-West	0,076	0,075
	(0,468)	(0,467)		(0,264)	(0,263)

Standard deviation in parentheses.

Source: Elaborated by the authors using SAEB (2009-2015).

Table 5A - Oaxaca-Blinder 5th grade detailed decomposition.

	Math	Portuguese		Math	Portuguese
	Explained			Unexplained	
qescz	0.002	0.003*	qescz	0.005***	0.001
	(0.001)	(0.002)		(0.002)	(0.001)
sesz	0.022***	0.019***	sesz	0.011***	0.002***
	(0.008)	(0.007)		(0.003)	(0.001)
familiaz	-0.029***	-0.038***	familiaz	-0.045***	-0.071***
	(0.004)	(0.005)		(0.006)	(0.005)
white	0.004**	0.004*	white	0.019	-0.048***
	(0.002)	(0.002)		(0.018)	(0.018)
idealage	-0.003	-0.004	idealage	0.016	0.111**
	(0.009)	(0.009)		(0.056)	(0.053)
livesmomdad	0.000	-0.000	livesmomdad	-0.097***	-0.075***
	(0.001)	(0.001)		(0.025)	(0.023)
escresp	0.002***	0.003***	escresp	0.230***	0.226***
	(0.001)	(0.001)		(0.054)	(0.058)
domwork	0.002	0.002	domwork	0.243***	0.144***
	(0.006)	(0.006)		(0.011)	(0.011)
work	-0.010*	-0.012*	work	-0.806***	-1.481***
	(0.005)	(0.007)		(0.092)	(0.080)
urban	0.004	0.006	urban	0.969***	0.397**
	(0.003)	(0.004)		(0.181)	(0.158)
begschool	-0.011*	-0.008	begschool	-0.473***	-0.564***
	(0.006)	(0.005)		(0.041)	(0.041)
teachsex	-0.000	0.000	teachsex	-0.241***	-0.306***
	(0.001)	(0.001)		(0.080)	(0.079)
teachhighdg	0.001	0.001	teachhighdg	0.076	-0.113
	(0.002)	(0.001)		(0.078)	(0.082)
teachothact	0.000	0.000	teachothact	0.019*	0.016
	(0.001)	(0.001)		(0.010)	(0.010)
teachexper	0.000	0.000	teachexper	-0.006	-0.013
	(0.001)	(0.001)		(0.031)	(0.028)
teachhrs	-0.000	-0.000	teachhrs	0.032***	0.021***
	(0.000)	(0.001)		(0.007)	(0.006)
mathhmwok	-0.004*	-0.135***	mathhmwok	1.597***	0.784***
	(0.002)	(0.004)		(0.145)	(0.124)
teachcontract	0.001	0.001	teachcontract	0.107***	0.039
	(0.002)	(0.001)		(0.040)	(0.035)
region	0.007	0.008	region	-0.370***	-0.286***
	(0.013)	(0.010)		(0.040)	(0.038)
year	-0.022***	-0.041***	year	0.083***	0.001
	(0.005)	(0.009)		(0.009)	(0.009)
bookatbeg	0.002	0.002	bookatbeg	0.066*	0.028
	(0.002)	(0.002)		(0.039)	(0.038)
			Intercept	4.202***	-6.978***
				(0.266)	(0.248)

Robust standard errors clustered at the school level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Region is a grouped variable for the five regions of Brazil, and its base category is the Southeast region. Year is also a grouped variable for the four years of analysis and its benchmark is 2009.

Source: Elaborated by the authors using SAEB (2009-2015).

Table 6A - Oaxaca-Blinder 9th grade detailed decomposition.

	Math	Portuguese		Math	Portuguese
	explained			Unexplained	
qescz	-0.005*	-0.002	Qescz	0.000	0.001
	(0.003)	(0.002)		(0.001)	(0.001)
sesz	0.033***	0.030***	Sesz	0.061***	0.082***
	(0.007)	(0.007)		(0.006)	(0.007)
familiaz	-0.001***	0.003***	Familiaz	-0.045***	-0.063***
	(0.000)	(0.001)		(0.003)	(0.004)
white	0.014***	0.014***	White	-0.028	-0.042**
	(0.005)	(0.005)		(0.017)	(0.017)
idealage	-0.101***	-0.103***	Idealage	-0.072***	-0.098***
	(0.013)	(0.013)		(0.013)	(0.014)
livesmomdad	-0.001*	0.000	livesmomdad	-0.202***	-0.237***
	(0.001)	(0.000)		(0.022)	(0.022)
escresp	0.036***	0.035***	Escresp	1.102***	1.216***
	(0.004)	(0.003)		(0.079)	(0.077)
domwork	0.005*	0.004	Domwork	-0.110***	0.025
	(0.003)	(0.003)		(0.017)	(0.017)
work	0.009***	0.019***	Work	0.452***	0.856***
	(0.002)	(0.003)		(0.062)	(0.058)
urban	0.011***	0.016***	Urban	1.270***	1.208***
	(0.003)	(0.005)		(0.210)	(0.192)
begschool	0.003	0.005	Begschool	-0.401***	-0.331***
	(0.003)	(0.003)		(0.058)	(0.061)
teachsex	-0.005***	-0.003***	Teachsex	-0.020	-0.035
	(0.001)	(0.001)		(0.022)	(0.029)
teachhighdg	0.000	0.000	teachhighdg	-0.349**	0.109
	(0.001)	(0.000)		(0.165)	(0.169)
teachothact	-0.000	-0.000	teachothact	-0.007	-0.008
	(0.000)	(0.000)		(0.005)	(0.006)
teachexper	0.000	-0.000	teachexper	0.006	0.025
	(0.001)	(0.001)		(0.025)	(0.023)
teachhrs	0.002***	0.003***	Teachhrs	0.024***	0.020**
	(0.001)	(0.001)		(0.009)	(0.008)
mathhmwok	0.004	-0.052***	mathhmwok	0.489***	0.987***
	(0.004)	(0.002)		(0.101)	(0.127)
teachcontract	0.001	-0.000	teachcontract	0.173***	0.089**
	(0.001)	(0.001)		(0.045)	(0.036)
regiao	0.003	0.013	Regiao	-0.961***	-0.709***
	(0.011)	(0.010)		(0.056)	(0.053)
year	0.204***	0.212***	Year	-0.080***	-0.056***
	(0.014)	(0.015)		(0.017)	(0.018)
bookatbeg	0.020***	0.021***	bookatbeg	-0.003	-0.078**
	(0.003)	(0.003)		(0.037)	(0.039)
			Intercept	7.487***	-15.822***
				(0.319)	(0.311)

Robust standard errors clustered at the school level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Region is a grouped variable for the five regions of Brazil, and its base category is the Southeast region. Year is also a grouped variable for the four years of analysis and its benchmark is 2009.
Source: Elaborated by the authors using SAEB (2009-2015).

Table 7A - RIF detailed decomposition for 5th graders.

	5th grade Math								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Boys	159.340***	178.152***	192.928***	206.307***	219.422***	232.963***	247.492***	264.187***	286.617***
	(0.130)	(0.133)	(0.143)	(0.147)	(0.153)	(0.155)	(0.157)	(0.161)	(0.168)
Girls	158.268***	175.187***	188.479***	200.736***	212.893***	225.615***	239.502***	255.691***	278.108***
	(0.112)	(0.117)	(0.128)	(0.137)	(0.147)	(0.154)	(0.156)	(0.163)	(0.179)
Difference	1.072***	2.966***	4.448***	5.571***	6.529***	7.348***	7.990***	8.495***	8.510***
	(0.112)	(0.094)	(0.094)	(0.088)	(0.093)	(0.091)	(0.095)	(0.099)	(0.119)
Explained	-0.035	-0.042	-0.044	-0.043	-0.040	-0.035	-0.030	-0.023	-0.015
	(0.025)	(0.028)	(0.030)	(0.032)	(0.033)	(0.032)	(0.031)	(0.029)	(0.026)
Student	-0.038**	-0.038*	-0.037*	-0.033	-0.028	-0.022	-0.015	-0.007	0.002
	(0.018)	(0.020)	(0.022)	(0.022)	(0.023)	(0.022)	(0.022)	(0.020)	(0.018)
Teachers	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Schools	0.003*	0.004*	0.004	0.004	0.004	0.005	0.004	0.004	0.003
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Region	0.013	0.011	0.009	0.008	0.007	0.006	0.004	0.003	0.001
	(0.011)	(0.013)	(0.015)	(0.015)	(0.016)	(0.016)	(0.015)	(0.013)	(0.011)
Year	-0.015***	-0.020***	-0.023***	-0.024***	-0.025***	-0.025***	-0.026***	-0.025***	-0.023***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Unexplained	1.107***	3.007***	4.493***	5.615***	6.569***	7.383***	8.020***	8.519***	8.524***
	(0.109)	(0.090)	(0.089)	(0.083)	(0.087)	(0.085)	(0.090)	(0.095)	(0.117)
Student	0.033	1.551***	2.147***	1.987***	2.571***	2.530***	2.643***	2.488***	2.073***
	(0.499)	(0.375)	(0.344)	(0.328)	(0.323)	(0.323)	(0.336)	(0.347)	(0.431)
Teachers	0.115	0.059	0.262	-0.048	0.015	0.184	-0.124	-0.323**	-0.467**
	(0.234)	(0.204)	(0.182)	(0.161)	(0.179)	(0.172)	(0.157)	(0.160)	(0.185)
Schools	0.123	0.210***	0.133**	0.083	0.099**	-0.023	0.030	0.056	-0.031
	(0.111)	(0.068)	(0.054)	(0.052)	(0.050)	(0.057)	(0.060)	(0.070)	(0.058)
Region	-0.464***	-0.462***	-0.372***	-0.327***	-0.281***	-0.342***	-0.391***	-0.359***	-0.394***
	(0.065)	(0.055)	(0.057)	(0.054)	(0.055)	(0.055)	(0.058)	(0.061)	(0.072)
Year	0.102***	0.097***	0.086***	0.076***	0.079***	0.100***	0.090***	0.069***	0.060***
	(0.020)	(0.014)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.015)	(0.014)
Intercept	1.198**	1.554***	2.238***	3.843***	4.086***	4.934***	5.772***	6.589***	7.283***
	(0.572)	(0.426)	(0.384)	(0.363)	(0.350)	(0.348)	(0.348)	(0.352)	(0.425)

Robust standard errors clustered at the school level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

Table 7A. (continuation)

	5th grade Portuguese								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Boys	139.097***	156.024***	169.963***	182.653***	195.188***	208.141***	222.226***	238.647***	260.940***
	(0.101)	(0.117)	(0.122)	(0.131)	(0.134)	(0.136)	(0.137)	(0.139)	(0.144)
Girls	146.112***	164.538***	179.162***	192.244***	204.889***	217.651***	231.352***	247.150***	268.663***
	(0.118)	(0.123)	(0.127)	(0.129)	(0.131)	(0.133)	(0.130)	(0.130)	(0.128)
Difference	-7.015***	-8.514***	-9.198***	-9.592***	-9.700***	-9.510***	-9.126***	-8.503***	-7.724***
	(0.100)	(0.096)	(0.089)	(0.087)	(0.088)	(0.088)	(0.088)	(0.106)	(0.116)
Explained	-0.218***	-0.233***	-0.232***	-0.224***	-0.212***	-0.199***	-0.179***	-0.156***	-0.127***
	(0.024)	(0.029)	(0.031)	(0.032)	(0.032)	(0.032)	(0.031)	(0.029)	(0.025)
Student	-0.217***	-0.226***	-0.218***	-0.202***	-0.183***	-0.164***	-0.139***	-0.114***	-0.085***
	(0.019)	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)	(0.022)	(0.020)	(0.018)
Teachers	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Schools	0.003*	0.004**	0.005*	0.005*	0.006*	0.006*	0.006*	0.006*	0.005*
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Region	0.014	0.014	0.013	0.011	0.009	0.007	0.005	0.003	0.001
	(0.009)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)	(0.010)	(0.008)
Year	-0.021***	-0.028***	-0.034***	-0.041***	-0.046***	-0.051***	-0.053***	-0.054***	-0.050***
	(0.006)	(0.008)	(0.009)	(0.010)	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)
Unexplained	-6.797***	-8.281***	-8.966***	-9.367***	-9.488***	-9.310***	-8.947***	-8.347***	-7.597***
	(0.097)	(0.092)	(0.083)	(0.081)	(0.082)	(0.082)	(0.083)	(0.102)	(0.113)
Student	-9.359***	-5.823***	-2.356***	-0.360	1.158***	2.426***	3.295***	4.247***	3.848***
	(0.483)	(0.414)	(0.378)	(0.337)	(0.341)	(0.310)	(0.285)	(0.332)	(0.330)
Teachers	-0.353	-0.337*	-0.484***	-0.455***	-0.327*	-0.224	-0.147	-0.252	-0.538
	(0.226)	(0.195)	(0.170)	(0.162)	(0.167)	(0.155)	(0.152)	(0.226)	(0.336)
Schools	-0.096	-0.092	-0.102*	0.011	0.068	0.059	0.063	0.119*	0.206***
	(0.066)	(0.061)	(0.060)	(0.054)	(0.050)	(0.050)	(0.061)	(0.067)	(0.058)
Region	-0.334***	-0.430***	-0.463***	-0.440***	-0.326***	-0.312***	-0.240***	-0.112*	-0.076
	(0.061)	(0.057)	(0.054)	(0.053)	(0.053)	(0.055)	(0.055)	(0.063)	(0.070)
Year	-0.097***	-0.116***	-0.084***	-0.058***	-0.013	0.033***	0.068***	0.105***	0.114***
	(0.014)	(0.013)	(0.013)	(0.012)	(0.011)	(0.012)	(0.013)	(0.015)	(0.015)
Intercept	3.442***	-1.483***	-5.478***	-8.065***	-10.048***	-11.292***	-11.987***	-12.453***	-11.150***
	(0.561)	(0.464)	(0.410)	(0.384)	(0.374)	(0.331)	(0.302)	(0.336)	(0.389)

Robust standard errors clustered at the school level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

Table 8A - RIF detailed decomposition for 9th graders.

	9th grade Math								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Boys	191.754***	212.176***	227.777***	241.407***	254.003***	266.266***	279.111***	293.644***	313.698***
	(0.145)	(0.144)	(0.143)	(0.141)	(0.138)	(0.136)	(0.134)	(0.137)	(0.159)
Girls	187.640***	205.233***	218.908***	231.166***	243.029***	255.018***	267.890***	282.673***	303.102***
	(0.101)	(0.113)	(0.119)	(0.125)	(0.131)	(0.136)	(0.139)	(0.144)	(0.159)
Difference	4.114***	6.943***	8.869***	10.240***	10.974***	11.247***	11.221***	10.971***	10.596***
	(0.138)	(0.121)	(0.112)	(0.102)	(0.098)	(0.099)	(0.098)	(0.100)	(0.114)
Explained	0.224***	0.247***	0.246***	0.247***	0.242***	0.234***	0.227***	0.219***	0.232***
	(0.022)	(0.025)	(0.027)	(0.028)	(0.028)	(0.028)	(0.027)	(0.026)	(0.026)
Student	-0.040**	-0.029	-0.016	-0.003	0.011	0.023	0.035*	0.049**	0.072***
	(0.018)	(0.020)	(0.021)	(0.022)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)
Teachers	-0.001	-0.001	-0.001	-0.002	-0.002	-0.003	-0.003	-0.004	-0.004
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Schools	0.011***	0.014***	0.015***	0.016***	0.017***	0.017***	0.017***	0.016***	0.018***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Region	0.017*	0.013	0.009	0.006	0.003	-0.001	-0.003	-0.006	-0.010
	(0.009)	(0.011)	(0.012)	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)	(0.011)
Year	0.237***	0.250***	0.239***	0.229***	0.215***	0.198***	0.181***	0.164***	0.156***
	(0.018)	(0.017)	(0.016)	(0.015)	(0.014)	(0.013)	(0.012)	(0.011)	(0.011)
Unexplained	3.890***	6.696***	8.623***	9.994***	10.732***	11.013***	10.995***	10.752***	10.364***
	(0.137)	(0.119)	(0.109)	(0.098)	(0.093)	(0.094)	(0.094)	(0.096)	(0.111)
Student	2.662***	2.987***	3.113***	2.954***	2.622***	2.133***	2.174***	2.108***	2.473***
	(0.461)	(0.403)	(0.381)	(0.338)	(0.322)	(0.310)	(0.301)	(0.307)	(0.328)
Teachers	0.271	-0.060	-0.178	-0.144	-0.315*	-0.252	-0.053	-0.441**	-0.251
	(0.598)	(0.272)	(0.231)	(0.215)	(0.189)	(0.182)	(0.193)	(0.190)	(0.222)
Schools	0.047	-0.002	0.027	0.015	-0.017	-0.048	-0.067	-0.039	0.032
	(0.070)	(0.060)	(0.055)	(0.053)	(0.051)	(0.047)	(0.047)	(0.048)	(0.056)
Region	-0.964***	-1.098***	-1.058***	-1.004***	-1.014***	-1.000***	-1.029***	-0.968***	-0.901***
	(0.085)	(0.080)	(0.080)	(0.077)	(0.073)	(0.071)	(0.072)	(0.074)	(0.083)
Year	0.024	0.027	-0.061**	-0.121***	-0.207***	-0.259***	-0.204***	-0.086***	0.072***
	(0.037)	(0.030)	(0.026)	(0.023)	(0.022)	(0.022)	(0.021)	(0.021)	(0.025)
Intercept	1.850**	4.842***	6.779***	8.293***	9.663***	10.439***	10.174***	10.177***	8.939***
	(0.923)	(0.518)	(0.467)	(0.423)	(0.393)	(0.369)	(0.364)	(0.357)	(0.400)

Robust standard errors clustered at the school level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

Table 8A. (continuation)

	9th grade Portuguese								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Boys	174.949***	195.522***	211.605***	225.768***	238.854***	251.875***	265.523***	281.096***	301.387***
	(0.129)	(0.138)	(0.146)	(0.142)	(0.139)	(0.138)	(0.137)	(0.136)	(0.134)
Girls	190.644***	211.806***	227.245***	240.304***	252.444***	264.350***	276.888***	291.091***	309.867***
	(0.136)	(0.131)	(0.130)	(0.127)	(0.125)	(0.123)	(0.120)	(0.116)	(0.113)
Difference	-15.695***	-16.284***	-15.640***	-14.537***	-13.590***	-12.475***	-11.366***	-9.995***	-8.480***
	(0.132)	(0.121)	(0.120)	(0.108)	(0.101)	(0.099)	(0.094)	(0.098)	(0.104)
Explained	0.151***	0.183***	0.211***	0.228***	0.242***	0.249***	0.252***	0.248***	0.230***
	(0.023)	(0.027)	(0.029)	(0.030)	(0.030)	(0.030)	(0.029)	(0.027)	(0.024)
Student	-0.110***	-0.093***	-0.071***	-0.049**	-0.027	-0.007	0.012	0.030	0.045***
	(0.020)	(0.022)	(0.023)	(0.023)	(0.022)	(0.022)	(0.021)	(0.019)	(0.017)
Teachers	0.000	-0.000	-0.001	-0.001	-0.001	-0.001	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Schools	0.018***	0.021***	0.022***	0.022***	0.021***	0.021***	0.020***	0.019***	0.017***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Region	0.031***	0.029***	0.024**	0.018	0.013	0.009	0.003	-0.000	-0.003
	(0.009)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)	(0.009)
Year	0.212***	0.226***	0.237***	0.237***	0.235***	0.229***	0.218***	0.200***	0.173***
	(0.016)	(0.018)	(0.018)	(0.018)	(0.017)	(0.016)	(0.015)	(0.014)	(0.012)
Unexplained	-15.846***	-16.467***	-15.851***	-14.764***	-13.831***	-12.725***	-11.617***	-10.242***	-8.710***
	(0.130)	(0.118)	(0.116)	(0.104)	(0.097)	(0.095)	(0.089)	(0.095)	(0.101)
Student	0.669	3.304***	4.872***	4.979***	5.213***	5.176***	4.580***	4.044***	3.697***
	(0.496)	(0.427)	(0.407)	(0.346)	(0.312)	(0.300)	(0.279)	(0.302)	(0.308)
Teachers	0.174	0.674**	0.272	0.164	0.286	0.216	0.386*	0.197	0.136
	(0.352)	(0.330)	(0.322)	(0.277)	(0.210)	(0.194)	(0.197)	(0.197)	(0.212)
Schools	-0.236***	-0.163**	-0.055	-0.073	-0.040	-0.009	0.016	0.066	-0.006
	(0.069)	(0.065)	(0.062)	(0.057)	(0.053)	(0.051)	(0.051)	(0.052)	(0.056)
Region	-0.557***	-0.924***	-1.109***	-1.056***	-0.946***	-0.815***	-0.714***	-0.472***	-0.347***
	(0.090)	(0.082)	(0.081)	(0.076)	(0.072)	(0.074)	(0.068)	(0.070)	(0.078)
Year	-0.373***	-0.287***	-0.177***	-0.085***	0.010	0.080***	0.113***	0.144***	0.120***
	(0.039)	(0.034)	(0.031)	(0.026)	(0.023)	(0.022)	(0.022)	(0.022)	(0.024)
Intercept	-15.522***	-19.070***	-19.654***	-18.694***	-18.354***	-17.373***	-15.998***	-14.221***	-12.311***
	(0.645)	(0.598)	(0.576)	(0.480)	(0.387)	(0.367)	(0.350)	(0.366)	(0.378)

Robust standard errors clustered at the school level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

3 GENDER PEER EFFECTS IN BRAZILIAN ELEMENTARY SCHOOLS

The second essay of the present thesis assesses the gender peer effects in Brazilian elementary schools.

3.1 INTRODUCTION

The idea that the interaction among students influences behavior and the learning environment, thus, affecting school productivity is quite intuitive. The magnitude of this phenomenon is of great interest among educators, researchers and policymakers who are seeking to enhance the efficiency of schools by designing classrooms or schools in such a way that the grouping of peers takes into account ability levels and the background of the students.

There is a growing body of literature in Economics that contributes to this debate[1]. In the applied field, probably the greatest concern is with providing credible estimates of these influences, given the difficulty to disentangle the effects of belonging to a group from the reasons for belonging to it. In other words, the assignment of students to classes and schools are hardly random, tainting the formation of peers with sorting and self-selection, which turns the pursuit for reliable estimations into a steep challenge.

Aside from this difficulty, several authors have been able to make important contributions to this literature[2], first by assessing if the mean achievement of schoolmates influences individual performance and, more recently, to the role of differences in peer characteristics such as race, ethnicity and gender. In Brazil, however, this literature is still scarce, being Pinto (2008), Firpo, Jales and Pinto (2015), Raposo (2015) and Vianna (2017) a few of the exceptions.

This paper aims to contribute to this existing literature by exploring if the assignment to a school with a higher proportion of girls influences 5th grade Math and Portuguese public school student achievement in Brazil. We also investigate if whether this effect is influenced by school size, students' socioeconomic background and by nonlinearities in the proportion of girls.

[1] See Lazear (2001), Hanushek et al. (2003) and Angrist (2014).

[2] See Sacerdote (2011), Epple and Romano (2011) and Sacerdote (2014) for a review of the theoretical and applied literatures.

Our database also allows us to identify those schools that tend to overly allocate boys and girls to different classrooms, enabling us to assess whether this aspect of the gender composition within schools also further influences the achievement of boys and girls. This analysis, which also concerns the disputed literature of single-sex and coeducational schools (STRAIN, 2013; LEE et al., 2014) allow us to further contribute in the policy perspective, since the gender composition of classrooms is more easily managed by the public school authorities than the gender composition at the time of the enrollment.

Our findings suggest that the proportion of girls positively affects Math and Portuguese test scores of boys and girls. Despite the fact that these effects are quite similar, they are slightly larger for Math, a subject that girls tend to be outperformed by boys. This, in turn, suggests that the gender peer effects are not exclusively associated with spillovers that arise from having higher achieving peers. Furthermore, we found evidence that these estimates do not rely solely on the size of the school, or on the socioeconomic background of the school's students. Lastly, we find that girls' achievements are higher when they attend schools with classes that tend to segregate the two genders. This effect, however, does not hold for boys, implying that despite both genders profit from coexisting, girls benefit even more if clustered together.

This paper is structured as follows. The next section describes our empirical strategy, discusses the data and the construction of the samples used throughout the analysis. Section 3 starts exhibiting evidences of the validity of our identification strategy, than presents our main estimates of the gender peer effects and the possibility of heterogeneous and non-linear effects. Section 4 presents the concluding remarks.

3.2 EMPIRICAL STRATEGY

One of the main challenges in estimating peer effects is that schools and classrooms are not formed randomly, in such a way that the composition of schools and classes may reflect, among other factors, the family background of students. This may occur, for example, when affluent parents are more active in the search of the best school for their children, or when principals organize the composition of classes so that better performers or students with the same socio-economic background can study together. Many other classroom and school characteristics can influence their composition, and some of them may not be observed by the researcher. These factors, therefore, act as confounders in the estimation of peer effects, since they may influence peer composition and the outcomes of the students in a given school.

The identification strategy used in this paper builds on the works of Hoxby (2000) and Lavy and Schlosser (2011), relying on school level gender peer effects and using school fixed effects to control for sorting and self-selection of students across schools. The idea behind this strategy, according to Hoxby (2000), is that the variation in adjacent cohorts' peer composition within a grade within a school is idiosyncratic and beyond the easy management of parents and schools. This happens because of the difficulty associated with predicting the gender composition of a specific cohort, and to the fact that it is expensive for the parents to react to this variation by changing schools, as opposed to the classroom composition.

Therefore, by using repeated observations on schools in a school fixed-effects framework we are able to control for unobserved and unchanging characteristics that are related to both achievement and peer gender composition. That is, we account for sorting and selection, paving the way for a causal estimate of the peer composition on the student test score.

Our dataset comprises of four years of the national representative Brazilian assessment of Math and Portuguese learning for 5^{th} grade public school students merged with the Brazilian Education Census, where we obtain the proportion of girls at each school, the main variable of interest, which measures gender peer effects, and exclude the few schools that use an exam as base to their admission criteria, in order to avoid potential bias.

According to the description above, we estimate the following reduced-form equation separately for boys and girls:

$$y_{ist} = \alpha_s + \tau_t + \lambda P_{st} + \beta x_{ist} + \theta \bar{X}_{(-i)st} + \varepsilon_{ist} \qquad 3.1$$

Where *i* denotes the individuals, *s* the schools, and *t* time. y_{ist} is the normalized Math or Portuguese score of each student; α_s and τ_t are the school and time effects, respectively; P_{st} is the gender peer effect variable, which corresponds to the proportion of girls in 5^{th} grade for each school; x_{ist} are the individual characteristics, which comprises of race, which is defined by white or non-white, a socioeconomic status index (SES), the number of people living in student household, and dummies for the cases where the father lives with the student, and if at least one of the parents graduated from university, as well as the student's school enrollment and enrollment squared; $\bar{X}_{(-i)st}$ are the school averages, excluding the own student, of the individual characteristics; ε_{ist} is the error term, which is composed of a school-specific random element that allows for any type of correlation within observations of the same school across time and an individual random element.

It is worth noting that the SES index is built using the first component of Principal Component Analysis, a procedure commonly used as a dimensionality reduction technique (JOLLIFFE, 2002). It takes into consideration the number of bedrooms, bathrooms, computers, cars and televisions in the student household. With regard to missing[3] observations in any of these variables, we imputed the correspondent classroom mean of the characteristic, and added indicator dummies for these missing values in the model estimation. This procedure accounts for the possibility of non-random missing values and, also, avoids drastic reductions in the sample size (ALLISON, 2002).

The main coefficient of interest is λ, which can be understood as the average treatment effect of having more female peers in a given school. In order for this interpretation to hold, changes in the unobserved factors that could affect the student's achievement must be uncorrelated with changes in the proportion of females within a school. Also, there must be enough variation in the percentage of females at the different cohorts to enable a precise estimate of the gender peer effects on the achievement. Tests regarding these two concerns are shown on the first part of Section 3.3.

3.2.1 Data

This paper uses four cohorts of the Brazilian Ministry of Education assessment of Math and Portuguese learning for 5th grade public school students (2009, 2011, 2013 and 2015). This assessment is biannual, and also includes questionnaires for Principals, Teachers, Schools and Students. The latter, besides comprising of a test for each subject, collects information on the background of the students, and it is where we gather most of the variables used as controls in our estimates. We match the dataset from these four assessments to the corresponding Brazilian Education Census, in order to obtain data on enrollment and on the percentage of girls at the 5th grade school level, which is the variable of interest in the study.

We focus on 5th grade as we are pursuing a casual effect and need to reduce the influence of dropouts and school retention, two factors associated with gender that play an important role in the Brazilian school system, especially at more advanced school grades (OECD, 2016). The final sample only takes into account students in mixed gender regular

[3] A total of 11% of the individuals in our sample had missing observations that needed to be imputed in order to build the SES index.

education schools that appear in all the years of the assessment. We also exclude schools that base their admission criteria on an exam, as this is a source of selection, which could bias our estimates. Also, in order to avoid changes in the gender composition that might be a result of structural changes in the school, we drop schools that have an annual enrollment lower than 10 students, and those that experienced a change in enrollment of 80% or more between two consecutive years.

Table 3.1 - Descriptive statistics by cohort.

Cohort	Students	Schools	% of girls	Average Math Score		Average Portuguese Score	
				Females	Males	Females	Males
2009	1,605,822	23,689	0.482	205.87	207.58	191.11	181.03
2011	1,491,958	23,689	0.480	209.55	212.81	198.37	185.44
2013	1,338,303	23,689	0.479	212.66	213.82	203.32	190.82
2015	1,369,161	23,689	0.481	218.53	221.36	213.31	202.68
All	5,805,244	23,689	0.481	211.38	213.61	201.03	189.52

Source: Elaborated by the authors using SAEB (2009-2015).

Table 3.1[4] shows the sample size and descriptive statistics by cohort for the main variables of interest. It can be seen that the sample has almost 6 million students, fairly distributed across the years, and in 23,689 schools. The mean proportions of girls in all of the cohorts are slightly below 50%, without any time trend. It can also be seen that on average, boys tend to outperform girls in Math, while the gap in score favors women in literacy.

3.3 RESULTS

This subsection displays the results we found for the present thesis.

3.3.1 Evidence on the validity of the identification strategy

The identification strategy used in this paper relies on cohort-to-cohort gender composition variability within schools to obtain precise estimates of the gender peer effects on achievement. Therefore, in order to examine if there is sufficient variation within schools, Table 3.2 report the variance decomposition of the proportion of females for both Math and

[4] Descriptive statistics for the rest of the variables are in Table 1A in the Appendix.

Portuguese, highlighting that 78% of the variance in the Math sample and 82% in Portuguese is within schools.

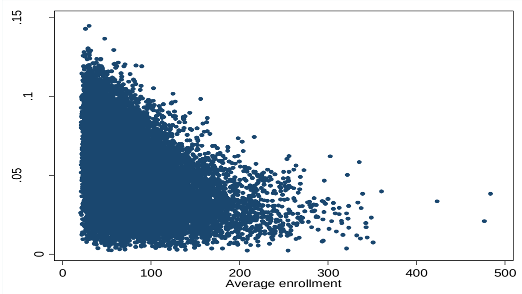
Table 3.2 - Decomposition of variance in the proportion of female students.

	Math			Portuguese		
	Sum of squares	Share of total	Df	Sum of squares	Share of total	df
Between	1288761.7	22.11%	23686	1032691.5	17.74%	23688
Within	4539005.4	77.89%	5779269	4789048.1	82.26%	5780156
Total	5827767.1		5802955	5821739.6		5803844

Source: Elaborated by the authors using SAEB (2009-2015).

In addition to this substantial within school variance, Figure 3.1 depicts the correlation between the standard deviation of the proportion of girls in each school and its average enrollment. This enables a better evaluation if there is a specific size of school that is responsible for this variability. The graph, in turn, shows that despite the fact that smaller schools account for the majority of the variance, there is still significant variability in schools whose average 5th grade enrollment is up to 300 students.

Figure 3.1 - Standard deviation of the proportion of females and school average enrollment.



Source: Elaborated by the authors using SAEB (2009-2015).

Another concern addressed by the identification strategy is to whether this within school variation in the proportion of female students is indeed random. If somehow unobserved characteristics, as well as, characteristics of the students, parents and schools influence the gender composition of a cohort within a school, the estimated peer effects will be biased. In this regard, we assume that parents are not able to predict the gender composition of their child's cohort and hence, are not able to respond to it. This is corroborated by the fact that we use a national representative sample of public schools, which corresponds to 80% of the enrollment in elementary schools in Brazil, and that the majority of these schools have the place of residence as main admission criteria. Both these features diminish the possibility that any unobserved characteristic will influence the gender composition of a cohort within a given school.

Aside from that, in Table 3.3, we test whether the proportion of female students within schools is correlated with any of the control variables in our final model. We perform this balancing test by performing separate regressions of the treatment variable on each of the controls using School Fixed Effects (SFE) and Ordinary Linear Squared (OLS), as a benchmark for comparison.

Table 3.3 - Balancing tests for the schools' proportion of female students and student characteristics.

	OLS	SFE		OLS	SFE
Race	0.001102^{***}	-0.000012	parents univdg	0.002938^{***}	-0.000062
	(0.000157)	(0.000067)		(0.000157)	(0.000071)
num people	-0.000663^{***}	0.000038	father	-0.000092	-0.000077
	(0.000057)	(0.000030)		(0.000097)	(0.000057)
Sesz	0.002049^{***}	-0.000038	enrollment	0.000027^{***}	0.000005
	(0.000104)	(0.000030)		(0.000004)	(0.000009)

Robust standard errors clustered at the school level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
All regressions include year dummies.
Source: Elaborated by the authors using SAEB (2009-2015).

Observing the coefficients from each estimate, it becomes clear that under Ordinary Least Squares almost all observable characteristics are correlated with the proportion of females, the exception being the presence of the student's father at home. Nonetheless, these correlations became statistically insignificant when we look within schools, with the addition of School Fixed Effects. Thus, once we include school dummies in the model, most of the sorting and selection are accounted for, thus, enabling a proper identification of the effect of the proportion of girls on achievement.

3.3.2 Estimates of gender peer effects in Brazilian elementary schools.

After analyzing the feasibility of the identification strategy in the last subsection, Table 3.4 reports the effects of the proportion of female peers on 5th grade achievement in Math and Portuguese. Each cell in the table shows the estimated coefficient, and its corresponding standard deviation, from separate regressions for boys and girls.

Columns 1-3 report the effects of the proportion of females in Math and Portuguese on the achievement of girls, while Columns 4-6 report these effects on boys. Three different specifications are considered, in order to assess how sensitive these estimates are to the control of school, individual and cohort characteristics. Columns 1 and 4 shows OLS estimates with only year dummies included as controls. Columns 2 and 5 include school fixed effects as controls, reducing drastically the size of the coefficients and their standard deviations. This decline, as on Table 3.3, suggests that selection and sorting across schools play a significant role in the estimation of peer effects and, therefore, controlling for these characteristics avoid estimation bias. The most complete specification, with the inclusion of individual and mean cohort characteristics as controls, are seen in Columns 3 and 6, and further reduces the size of the estimated coefficients and, to a less substantial degree, their standard deviation.

Table 3.4 - Estimation of the effect of proportion of female students on achievement.

	Females Proportion of Females in Cohort			Males Proportion of Females in Cohort		
	(1)	(2)	(3)	(4)	(5)	(6)
Math	0.890***	0.293***	0.272***	1.031***	0.266***	0.254***
	(0.031)	(0.016)	(0.016)	(0.033)	(0.018)	(0.017)
Portuguese	0.887***	0.307***	0.283***	0.967***	0.253***	0.241***
	(0.029)	(0.015)	(0.015)	(0.029)	(0.016)	(0.015)
Year effects	✓	✓	✓	✓	✓	✓
SFE		✓	✓		✓	✓
Individ. controls			✓			✓
Cohort controls			✓			✓

Robust standard errors clustered at the school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

Focusing on the more complete specification, the estimated coefficients show that both females and males tend to perform better in each of the subjects when they are in classes with a higher proportion of females. All estimates are statistically significant and indicate that an additional 10% in the proportion of girls would lead to increases of 2.72 and 2.83 percent of a standard deviation on the average of the Math and Portuguese tests for women, and, for boys, the rise would be of 2.54 of a standard deviation for Math and 2.41 for Portuguese. Therefore, overall, the effect of having more female peers is slightly larger among girls than boys.

It is important to highlight that, despite being quite similar, a comparison between the subjects for boys, shows that the peer effects are slightly larger for Math, a subject that boys tend to outperform girls, as we have seen on Table 3.1. This suggests that the gender peer effects are not solely associated with spillovers that arise from having higher achieving peers.

In terms of comparison, our results are moderate when compared to other estimates of gender peer effects. For example, Hoxby (2000) found that a 10% increase in the proportion of females on Texas elementary schools leads to scores 1 to 2% of a standard deviation higher in Math and English. Lavy and Schlosser (2011) despite not finding any significant impact of the proportion of female students on 5th grade Israeli students in language tests, encountered that a 10 percentage point increase of female students raises average Math score by 3.7 percentage points of a standard deviation for girls and 2.18 for boys. While Cabezas (2010) found that the same increase in the proportion of girls leads to a 0.53 percent of a standard deviation higher Math scores for Chilean students.

The next chapter of this thesis tries to explore the reasons behind this increase on students' grades. Mainly, the results we find, that are also partially shared by Lavy and Schlosser (2011), is that a greater proportion of girls is associated with improvements in student behavior, which reflects less violence, greater teacher expectations over the student's academic future, and facilitates classroom progress.

On a broader view, our results are similar in magnitude to the ones found in assessing the impact of school violence, as identified by Dalcin (2016), who estimated that school violence can reduce achievement by up to 2.6% of a standard deviation in Math and 2.2% in Portuguese. Nonetheless, other educational policies have a much higher impact, such as an additional year of compulsory education that increases 13% and 8% of a standard deviation in the scores of Math and Portuguese, respectively (ZANON, 2017), and attending daycare that raise Math scores from 28% to 42% of a standard deviation, depending on the mother's education (PINTO; SANTOS; GUIMARÃES, 2016).

3.3.3 Heterogeneous effects.

In order to gain further insights on the extent of the gender peer effects, in Table 3.5 we explore heterogeneous effects of the proportion of girls across different dimensions. First, we investigate differences by school size, by stratifying our sample into quartiles of the enrollment variable, from lowest (q1) to highest (q4), and running the baseline model for each subsample. This can be viewed as providing additional evidence that the mean effect we are capturing does not merely apply to small schools.

Second, we calculate the mean of the Socioeconomic Status index for each school and stratify the sample into quartiles of this variable's distribution[5], in order to investigate whether the background of the students within each school interfere with the effect of the proportion of female students.

Table 3.5 - Heterogeneous effects of the proportion of female students by school size and school student socioeconomic status.

	School Size							
	q1		q2		q3		q4	
	Female	Male	Female	Male	Female	Male	Female	Male
Math	0.170***	0.139***	0.248***	0.195***	0.295***	0.319***	0.549***	0.570***
	(0.023)	(0.025)	(0.029)	(0.031)	(0.036)	(0.039)	(0.051)	(0.056)
Schools	10234	10234	6213	6213	4425	4425	2815	2815
Students	709556	741489	712078	738637	715243	735802	715882	734269
Portuguese	0.182***	0.123***	0.258***	0.218***	0.304***	0.304***	0.562***	0.506***
	(0.022)	(0.022)	(0.028)	(0.028)	(0.034)	(0.036)	(0.047)	(0.048)
Schools	10235	10235	6213	6213	4425	4425	2816	2816
Students	709647	741565	712116	738673	715283	735843	716166	734552
	School Students Socioeconomic Status							
	q1		q2		q3		q4	
	Female	Male	Female	Male	Female	Male	Female	Male
Math	0.291***	0.236***	0.267***	0.248***	0.267***	0.232***	0.223***	0.237***
	(0.031)	(0.034)	(0.031)	(0.034)	(0.034)	(0.037)	(0.034)	(0.036)
Schools	7240	7240	6058	6058	5308	5308	5081	5081
Students	606713	618249	641300	644589	653577	660683	658660	669845
Portuguese	0.301***	0.222***	0.259***	0.276***	0.265***	0.217***	0.228***	0.202***
	(0.029)	(0.030)	(0.030)	(0.031)	(0.032)	(0.033)	(0.032)	(0.033)
Schools	7240	7240	6059	6059	5308	5308	5082	5082
Students	606756	618263	641534	644834	653612	660725	658736	669895

Robust standard errors clustered at the school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Regressions also includes year effects, school fixed effects, individual and cohort means as controls.
Source: Elaborated by the authors using SAEB (2009-2015).

The estimations by school size, on the upper part of Table 3.5, indicate that higher proportions of female students lead to increases in achievement in all school sizes, for both Math and Portuguese. Also, by comparing the columns for each higher quartile of the

[5] We exclude from these estimates all students with missing values for the SES index.

enrollment distribution, the effect is increasing for both genders. The fact that the highest quartiles have the greatest impact is especially interesting, since there isn't much of a difference in the proportion of female students for each quartile[6]. This, therefore, provides evidence that larger schools, despite the higher standard deviations, are also important to the estimation of the mean effect.

The bottom columns on Table 3.5, indicate that the effect is similar and positive on all quartiles of the school mean of the students' socioeconomic status variable, for both Math and Portuguese. Hence, suggesting that the effect of the gender composition does not strongly rely on the socioeconomic background of the school's students. The estimates for women, however, are lower when the school has the highest level of student background, for both Math and Portuguese, while the results for men do not seem to have a clear pattern as we move from the lowest to higher quartiles.

We also investigate nonlinearities in the proportion of females on Table 3.6, following Lavy and Schlosser (2011), by replacing the treatment variable by quartile identifier dummies of the proportion of girls. Therefore, the source of variation necessary for the identification of these non-linear effects consists in the dynamics of the schools in switching from each quartile in the different years of our sample. Table 2A on the Appendix report summary statistics on each quartile and a matrix with information on the extent to which schools switch from quantile to quantile. It shows that there is substantial variance in the quantiles, for example, only 342 schools are in the first quantile on all four years of our sample, while 9,852 appears on the first and second quantiles during the four year period. Overall, of the total[7] 71,067 possible changes, 57,349 (80.7%) occurred.

[6] The first quartile has, on average, 47.82% of females, while the last has 48.30%.

[7] Total possible changes are calculated multiplying the total number of schools (23,689) with the three remaining years it could switch quantiles.

Table 3.6 - Nonlinear estimates of the effect of the proportion of female students on achievement.

	q2		q3		q4	
	Female	Male	Female	Male	Female	Male
Math	0.012***	0.017***	0.033***	0.029***	0.043***	0.040***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Portuguese	0.016***	0.015***	0.030***	0.024***	0.046***	0.037***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Range	0.442 - 0.481		0.481 - 0.519		0.519 – 1	
Mean	0.463		0.499		0.557	

Robust standard errors clustered at the school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regressions also includes year effects, school fixed effects, individual and cohort means as controls.
Source: Elaborated by the authors using SAEB (2009-2015).

The results on Table 3.6 indicate that the effect of the proportion of females increases as schools switch to higher quartiles, as compared to the first. The effects do not seem to differ by subject, as the estimates for each quartile and gender are very similar. A move from the first to the second quartiles increases Math achievement by 0.012 and 0.017 percentage points of a standard deviation, respectively, for girls and boys, while in Portuguese, this increase would be of 0.016 for girls and 0.015 for boys.

Nonetheless, as on Hoxby (2000), for Texas elementary students, and Lavy and Schlosser (2011), for high schools in Israel, the greatest estimated effect on achievement is from moving to the highest quartile, where girls are the strict majority. In our case, this raises Math scores by 0.043 percentage points of a standard deviation for girls and 0.04 for boys, while the Portuguese scores are raised by 0.046 and 0.037 percentage points of a standard deviation.

3.3.4 Gender clustering schools

Lastly, we ask whether the school’s classroom formation interfere with the main effect of the proportion of girls. In order to do that, we drop from our sample all the schools with only one classroom, and build an indicator variable for those schools that have a greater than the mean standard deviation of its gender composition across classrooms[8]. That is, we identify those schools that tend to cluster girls in different classrooms from boys. Therefore, by interacting this indicator variable with the proportion of girls we are able to go beyond the

[8] Descriptive statistics of this variable is found on Table 3A, on the Appendix.

mean treatment effect and assess if the gender composition of the school's classrooms interferes with it.

Table 3.7 - Impact of school classroom gender composition on achievement.

	Math		Portuguese	
	Female	Male	Female	Male
proportion of girls	0.270***	0.270***	0.284***	0.253***
	(0.018)	(0.019)	(0.017)	(0.017)
clustering school * proportion of girls	0.018***	-0.008	0.015***	-0.004
	(0.004)	(0.005)	(0.004)	(0.004)
Schools	22822	22822	22824	22824
Students	2692765	2785016	2693210	2785436

Robust standard errors clustered at the school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regressions also includes year effects, school fixed effects, individual and cohort means as controls.
Source: Elaborated by the authors using SAEB (2009-2015).

As we can see on Table 3.7, increasing the proportion of girls in schools that tend cluster girls and boys in different classrooms boosts achievement only for females. That is, increasing the number of girls where they already tend to be together further enhances the impact on achievement by 0.018 of a standard deviation in Math and 0.015 in Portuguese. Therefore, for girls, the total effect of having more females in schools that tend to cluster them together is around 0.3 standard deviations for both, Math and Portuguese. As for boys, the coefficient for the interaction with clustering schools is negative, but nonsignificant, indicating that a larger number of girls in such schools do not have an extra effect on achievement.

Nonetheless, our findings imply that despite the fact that boys profit from coexistence, indicating that the gender composition also influences achievement, girls benefit even more if clustered together, thereby increasing school efficiency. Other authors in the peer effects literature find positive effects of grouping girls together. For example, Lu and Anderson (2015), who explore micro-environments in Chinese middle schools, find that being surrounded by classroom desks occupied by female peers increases female's test scores, but it can have a potentially negative effect on males.

This is an interesting result regarding public policies, since Brazilian public schools arrange students within classrooms mostly by taking into consideration only the age and the achievement of the students, and not their gender. It also adds on the previous results of this paper, since the gender composition of classrooms is more easily managed by the public school authorities than the enrollment gender composition. Furthermore, one can point out that public policies may not be only interested in school efficiency, but also in equity. In this case, because clustering girls further benefit them, it could help narrow the gap in Math scores, as shown by the previous chapter of this thesis.

Among the reasons why girls accrue extra benefits from being clustered together, as we already mentioned, is that a greater proportion of girls is associated with improvements in student behavior. Moreover, the literature also indicates that boys and girls have distinct learning processes, which affect teacher-pupil relationships, curricular content and assessment methods (TINKLIN et al., 2001) and these might further influence test scores.

Moreover, we can confront our results to the literature regarding single-sex and coeducational schools. Jackson (2017) finds an increase in achievement, among other positive effects, for boys and girls from moving from a coeducational school to a single-sex school in Trinidad and Tobago. Yet, this is not a consensual result, as Strain (2013) finds evidence that offering single-sex classrooms to North Carolina students reduces the performance in mathematics, without any effect on reading scores. However, and closer to our case, Lee et. al (2014) find that Korean male students attending single-sex classes within coed schools score 0.10 of a standard deviation below male students in mixed-gender classes, without any significant effect for females. Interestingly, they also find that male students attending single-sex schools outperform their counterparts in mixed-gender classes.

As it can be noticed, this is a hotly disputed literature. Unfortunately, our dataset do not allow us to move further and try to compare outcomes between single-sex and coeducational schools in Brazil, which therefore, remains a challenge for future works. However, our findings highlight that, at least once we consider coeducational schools, the formation of the gender mix of classes are important to increase school efficiency and equity, as it benefits especially women.

3.4 CONCLUDING REMARKS

In this paper we measured empirically if the assignment to a class with a higher proportion of girls influences 5th grade Math and Portuguese public school student achievement in Brazil. We investigated if whether this effect is influenced by school size, students' socioeconomic background and by nonlinearities in the proportion of girls. We have also identified the schools that tend to cluster females and males in different classrooms, which allowed us to assess whether this aspect of the classroom formation interfere with the main effect of the proportion of girls.

Our estimates suggest that the proportion of girls positively affects Math and Portuguese test scores of boys and girls. In such a way that an additional 10% in the proportion of girls would lead to increases of 2.72 and 2.83 percent of a standard deviation on the average of the Math and Portuguese tests for women, and, for boys, the increase would be of 2.54 of a standard deviation for Math and 2.41 for Portuguese.

These results for the effect of the proportion of girls, overall, are quite similar but, interestingly, for boys the larger coefficient is in Math, a subject that they tend to outperform girls. This, in turn, suggests that the gender peer effects are not solely associated with spillovers that arise from having higher achieving peers.

Investigating heterogeneous effects, we stratified our sample into quartiles of the enrollment variable, in order to inquire if school size would influence the main estimates. The results indicate that the effect does not rely solely on the greater variation of the proportion of girls existent on small schools, as all of the quartiles displayed a positive and statistically significant influence of the proportion of girls on achievement. We have also stratified our sample into quartiles of the student's socioeconomic index, identifying that the effect of the proportion of girls does not rely strongly with the background of the students. As to the nonlinearities, we found that the effect of the proportion of girls on achievement is increasing with higher proportions of females, and is especially strong when females are the majority within the school.

Lastly, our results from the interaction of the mean treatment effect and those schools that tend to group girls and boys in different classrooms, indicate that it is not only having more girls in a school that is beneficial to achievement, but the actual coexistence between males and females also plays an important role in the improvement, especially for women. This, therefore, is not entirely shared by the literature on peer effects, and a more definitive evaluation of whether coeducational or single-sex schools are better for student achievement remains a challenge for future works.

Yet, our findings highlight that once we consider coeducational schools, the formation of the gender mix of classes is important to increase school efficiency, as it benefits especially girls. This is an interesting result, especially in a country like Brazil, where the schooling system is composed of mostly coed schools that allocate students within classrooms predominantly by age and achievement, and not their gender. Moreover, since our findings regards the reallocation of students, taking these findings into consideration can result in an effective and low cost measure aimed at increasing achievement.

3.5 REFERENCES

ALLISON, P. D. **Missing Data**. Pennsylvania: University of Pennsylvania, 2002.

ANGRIST, J. The perils of peer effects. **Labour Economics**, Amsterdam, v. 30, p. 98-108, 2014.

CABEZAS, V. **Gender peer effects in school: Does the gender of school peers affect student achievement?** 2010. 156f. Thesis (PhD in Economics of Education) – Faculty of Economics, Columbia University, New York.

DALCIN, A. K. **Uma análise da relação entre violência na escola e proficiência dos alunos**. 2016. 101f. Dissertation (MS in Applied Economics) – Faculty of Economics, Rio Grande do Sul Federal University, Porto Alegre.

EPPLE, D.; ROMANO, R. E. Peer effects in education: A survey of the theory and evidence. **Handbook of social economics**, Amsterdam, v. 1, p. 1053-1163, 2011.

FIRPO, S.; JALES, H.; PINTO, C. Measuring peer effects in the Brazilian school system. **Applied Economics**. Washington, v. 47, n. 32, 2015.

HANUSHEK, E. A.; KALIN, J. F.; MARKMAN, J. M.; RIVKIN, S. G. Does peer ability affect student achievement? **Journal of Applied Econometrics**. v. 18, n. 5, p. 527-544, 2003.

JOLLIFFE, I. T. **Principal Component Analysis** 2nd Edition. New York: Springer Series in Statistics, 2002.

HOXBY, C. **Peer effects in the classroom**: Learning from gender and race variation. Cambridge: NBER, 2000. (NBER Working Papers**,** 7867).

JACKSON, K. **The Effect of Single-Sex Education on Test Scores, School Completion, Arrests, and Teen Motherhood**: Evidence from School Transitions. 2017. Cambridge: NBER. (NBER Working Paper**,** 22222).

LAVY, V.; SCHLOSSER, A. Mechanisms and impacts of gender peer effects at school. American Economic Journal: **Applied Economics**, Washington, v. 3, n. 2, p. 1-33, 2011.

LAZEAR, E. P. Teacher incentives. **Swedish Economic Policy Review**. Stockholm, v. 10, n. 2., p. 179-214, 2001.

LEE, S.; LESLEY, J. T.; WOO, S.; LIM, K. **All or Nothing? The Impact of School and Classroom Gender Composition on Effort and Academic Achievement**. Cambridge: NBER, 2014. (NBER Working Paper, 20722).

LU, F.; ANDERSON, M. L. Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups. **Journal of Labor Economics**, Chicago, v. 33, n. 1, p. 91-122, 2015.

OECD. **Brazil Country Note** – Results from PISA 2015. 2016. Available at: http://www.oecd.org/brazil/PISA-2015-Brazil.pdf. Acess: abril 2018.

PINTO, C. Semiparametric estimation of peer effect in classrooms: Evidence for Brazilian schools in 2003. 2008. Available at: https://www.economics.uci.edu/files/docs/econoseminar/w10/pinto.pdf . Acess: march 2017.

PINTO, C. C. X.; SANTOS, D.; GUIMARÃES, C. The impact of daycare attendance on Math Test Scores for a Cohort of Fourth Graders in Brazil. **The Journal of Development Studies**, London, v. 53, n. 9, 2017.

RAPOSO, I. P. A. **O papel da rede de amizades e da formação aleatória de turmas por faixa etária sobre o desempenho escolar**. 2015. 102f. Thesis (PhD in Regional and Urban Economics) – Faculty of Regional and Urban Economics, Pernambuco Federal University, Recife.

SACERDOTE, B. Peer effects in education: How might they work, how big are they and how much do we know thus far? **Handbook of the Economics of Education**. Amsterdam, v. 3, p. 249-277, 2011.

SACERDOTE, B. Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward? **Annual Review of Economics**. Danvers, v. 6, p. 253-272, 2014.

STRAIN, M. R. Single-sex classes & student outcomes: Evidence from North Caroline. **Economics of Education Review**. Cambridge, v. 36, p. 73-87, 2013.

TINKLIN, T.; CROXFORD, L.; DUCKLIN, A.; FRAME, B. **Gender and Pupil Performance in Scotland's Schools**. Edinburgh: University of Edinburgh. 2001.

VIANNA, F. R. **O efeito dos pares sobre o desempenho escolar de alunos do ensino fundamental**. 2017. 81f. Dissertation (MS in Applied Economics) – Faculty of Economics, Juiz de For a Federal University, Juiz de Fora.

ZANON, D. **Aumento do tempo na educação formal e performance dos estudantes: evidências de curto e médio prazo**. 2017. 70f. Dissertation (MS in Applied Economics) – Faculty of Economics, Rio Grande do Sul Federal University, Porto Alegre.

3.6 APPENDIX

Table 1A - Descriptive statistics.

Cohort	Avg SES		Avg father		Avg parents univ dg		Avg race		Avg num. of people	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
2009	0,020	0,109	0,704	0,721	0,192	0,209	0,382	0,381	3,691	3,699
2011	0,025	0,108	0,805	0,826	0,212	0,235	0,369	0,367	3,686	3,698
2013	0,026	0,098	0,688	0,705	0,220	0,241	0,373	0,371	3,923	3,892
2015	0,023	0,075	0,704	0,721	0,255	0,273	0,347	0,348	3,900	3,875
All	0,023	0,098	0,726	0,744	0,219	0,238	0,368	0,367	3,793	3,785

Source: Elaborated by the authors using SAEB (2009-2015).

Table 2A - Descriptive statistics of the quartiles of the proportion of females.

	q1	q2	q3	q4
Range	0 - 0.442	0.442 - 0.481	0.481 - 0.519	0.519 - 1
Mean	0.405	0.463	0.499	0.557
Median	0.413	0.463	0.500	0.548
Students	1451387	1454291	1448863	1450703
School Transition Across Quartiles				
	q1	q2	q3	q4
q1	342	9852	9473	10687
q2		83	8475	9439
q3			56	9423
q4				205

Source: Elaborated by the authors using SAEB (2009-2015).

Table 3A - Proportion of girls for clustering and nonclustering schools.

Cohort	Proportion of girls clustering schools		Proportion of girls noncluster. schools	
	Female	Male	Female	Male
2009	0.488	0.476	0.488	0.477
2011	0.487	0.473	0.487	0.475
2013	0.485	0.471	0.487	0.473
2015	0.488	0.474	0.489	0.476
All	0.487	0.474	0.488	0.475
Schools	21,606		20,602	
Students	2,845,888		2,959,356	

Source: Elaborated by the authors using SAEB (2009-2015).

4 CLASSROOM CLIMATE THE INFLUENCES OF GIRLS

In the previous chapter of this thesis, we found evidence that schools with a higher proportion of girls show higher scores of Math and Portuguese in Brazil. Moreover, we also identified that these effects are larger for Math, a subject that girls tend to be outperformed by boys, thereby, suggesting that the gender peer effects are not exclusively associated with spillovers that arise from having higher achieving peers. Hence, now, we investigate the influences of the gender composition on student behavior and on the relationship between the students and their teachers.

4.1 INTRODUCTION

There are two main ways that the composition of a classroom might influence learning. First, by the congestion effects, which are negative externalities created when one student impedes the learning of all other classmates (LAZEAR, 2001). This effect, in the gender context, can take place when a more disruptive boy is replaced by a girl, or when girls exert such an influence that reduces the chances of boys to misbehave. Second, teachers may treat boys and girls differently, in such a way that this influences grading, the content and the organization of what is taught (DEE, 2006; LAVY, 2008; CORNWELL et al., 2011).

We, therefore, explore a national representative sample of Brazilian 5th grade schools in order to assess if the gender composition, measured by the proportion of girls at school, influences teachers' expectations over the students' academic progress, and their perception of the classroom climate, violence, as well as their pedagogical methods. Despite not being able to distinguish among the two mentioned different paths through which the peer effects can take place, our analysis helps to elucidate the drivers of the positive influences that girls have on other student's achievements.

The literature on gender differences in school outcomes emphasizes that girls begin school with more advanced social and behavioral skills, and that this advantage grows over time, reducing the costs of schooling, explaining, in part, why girls tend to outnumber boys at higher levels of education (DIPRETE, JENNINGS, 2009; BECKER et al., 2010). Moreover, having a poor family background tends to be especially harmful for boys, as even when compared to their siblings, they perform worst on tests and have higher rates of absence and behavioral problems (AUTOR et al., 2016; CHETTY et al., 2016). In a closely related paper, Lavy and Schlosser (2011) find evidence of non-cognitive benefits of higher proportions of girls on Israeli students' outcomes, such as violence and disruptive behavior. However, besides focusing on a different set of questions, their estimates are based on students' views about the classroom climate, which may or may not be aligned with the teachers' perception.

Our results indicate that the benefits of having a greater proportion of girls are mainly through improvements in student behavior, which reflects less violence, greater teacher expectations over the student's academic future, and facilitates classroom progress. These results suggest new ways of increasing school efficiency, in a country that already performs below OECD average, and worse than countries with similar expenditure per student (OECD, 2016).

This paper is structured as follows. The next section describes our empirical strategy, discusses the data and the construction of the sample used throughout the analysis. Section 3 starts exhibiting evidences of the validity of our identification strategy, than presents our main estimates of gender composition on classroom climate, teacher's expectations and student behavior. Section 4 presents the concluding remarks.

4.2 IDENTIFICATION STRATEGY

As in the previous chapter of this thesis, we use school fixed-effects to assess the influence of the proportion of girls at the school level, but this time with the different measures of classroom climate and student behavior taken from the teachers' questionnaire. The basic idea behind the school level fixed-effects identification strategy is to explore the fact that there is some idiosyncratic and exogenous variation in the gender peer composition of the adjacent 5th grade cohorts within a school. That is, it uses repeated observations on schools to control for unobserved and unchanging characteristics that are related to both measures of classroom climate and gender composition. Additionally, we also include average teacher and cohort characteristics to the control variables in order to provide a proper identification of the relationship between proportion of girls and classroom climate, as they might act as confounders of the casual effect we are trying to estimate.

In accordance with the above, we estimate the following reduced-form equation:

$$\bar{y}_{st} = \alpha_s + \tau_t + \lambda P_{st} + \theta \bar{X}_{st} + \varepsilon_{st} \qquad 4.1$$

Where s denotes the schools, and t time. y_{st} is the school average of the indicator variable built from the teachers' answers to relevant questions in the inquiry; α_s and τ_t are the school and time effects, respectively; P_{st} is the gender peer effect variable, which corresponds to the proportion of girls in 5th grade for each school; $\bar{X}_{st}$ consists of the average school cohort controls, also used in the previous chapter, which are race (white or non-white), a socioeconomic status index (SES), the number of people living in the student household, and dummies for the cases where the father lives with the student, and if at least one of the parents graduated from university, as well as the school enrollment and enrollment squared; it also contains the school averages of the teacher characteristics, that is, if whether the teacher has a graduate degree, a postgraduate degree, if he works in any other activity besides teaching, if he has more than 10 years of experience being a teacher, works more than 40 hours per week and a dummy indicating if he has a permanent teaching contract with the school; ε_{st} is the error term.

4.2.1 Data

This paper uses four cohorts of the Brazilian Ministry of Education assessment of Math and Portuguese learning for 5th grade public school students (2009, 2011, 2013 and 2015). This assessment is biannual, and includes questionnaires for principals, teachers, schools and students. In order to gather the variables used as control in our estimates, we use the Students questionnaire to build the school cohort means, as described above, and the Teachers questionnaire to obtain their expectations and beliefs about students' progress in school, as well as their perception of student behavior, violence, work satisfaction and on the pedagogical methods. We also match the dataset from these four assessments to the corresponding Brazilian Education Census, in order to obtain data on enrollment and on the percentage of girls at the 5th grade school level, which is the variable of interest in the study.

The final sample only takes into account Math and Portuguese teachers in mixed gender regular education schools that appear during all the years of the assessment. We also exclude schools that base their admission criteria on an exam, as this is a source of selection, which could bias our estimates. Also, in order to avoid changes in the gender composition that might be a result of structural changes in the school, we drop schools that have an annual enrollment lower than 10 students, and those that experienced a change in enrollment of 80% or more between two consecutive years.

Table 4.1 - Descriptive statistics by cohort.

Cohort	Schools	Teachers	% of girls
2009	20,824	55,686	0.481
2011	20,824	89,020	0.479
2013	20,824	65,390	0.477
2015	20,824	88,992	0.481
All	20,824	299,088	0.480

Source: Elaborated by the authors using SAEB (2009-2015).

Table 4.1 shows sample size, the number of schools, teachers and the average proportion of girls by cohort. It can be seen that the sample has almost 300 thousand teachers, in 20,824 schools. The variation in the number of teachers on each cohort is explained by missing valid questionnaires, as the number of students is fairly stable across the years[19]. The mean proportions of girls in all of the cohorts are slightly below 50%, without any noticeable time trend.

As to the independent variables used in this study, we gathered over thirty questions in the teachers' questionnaire that relate to classroom climate, separating them by topics, which are: teacher expectations; teacher performance and job satisfaction; teacher perceptions about student behavior; teacher violence exposure; and, teacher pedagogical methods. Since they are mostly yes or no questions, we transform them into dummy variables and calculate the yearly school average of each, creating an indicator that stands between 0 and 1[20].

4.3 RESULTS

This subchapter displays the results of the present essay.

4.3.1 Evidences on the validity of the identification strategy.

We use the same identification strategy as in the third chapter of the present thesis. As mentioned and described before, this strategy relies on cohort-to-cohort gender composition variability within schools to obtain precise estimates of the gender peer effects. Since now we are interested on the distinct measures of classroom climate, this subsection presents complementary tests needed to further validate this identification.

[19] See Table 3.1 from the previous chapter of this thesis.
[20] Descriptive statistics for these variables are in Table 1A in the Appendix.

First, in order to examine if there is sufficient variation within schools, Table 4.2 report the variance decomposition of the proportion of females on our final sample, highlighting that 67% of the variance is within schools.

Table 4.2 - Decomposition of variance in the proportion of female students.

	Sum of squares	Share of total	Df
Between	373.013	32.90%	20823
Within	760.831	67.10%	278264
Total	1133.844		

Source: Elaborated by the authors using SAEB (2009-2015).

In addition to this substantial within school variance, the correlation between the standard deviation of the proportion of girls[21] in each school and its average enrollment, shows that smaller schools account for the majority of the variance, but there is still significant variability in schools whose average 5th grade enrollment is up to 300 students.

Another concern with the identification strategy is to whether this within school variation in the proportion of female students is indeed random. If somehow unobserved characteristics, as well as, characteristics of the students, teachers, parents and schools influence the gender composition of a cohort within a school, the estimated peer effects will be biased. In this regard, we assume that parents are not able to predict the gender composition of their child's cohort and hence, are not able to respond to it. This is corroborated by the fact that we use a national representative sample of public schools, which corresponds to 80% of the enrollment in elementary schools in Brazil, and that the majority of these schools have the place of residence as main admission criteria. Both these features diminish the possibility that any unobserved characteristic will influence the gender composition of a cohort within a given school.

In addition, on Table 4.3 we test whether the proportion of female students within schools is correlated with any of the teachers' characteristics we use in our final model. We perform this balancing test by estimating separate regressions of the treatment variable on each of the controls using School Fixed Effects (SFE) and Ordinary Linear Squared (OLS), as a benchmark for comparison.

[21] Not shown due to its similarity to Figure 3.1 on the previous chapter.

Table 4.3 - Balancing tests for the proportion of female students and teacher characteristics.

	OLS	SFE		OLS	SFE
graduate dgr	0.002544***	-0.000191	+10yrs experience	0.002023***	0.000456
	(0.000407)	(0.000315)		(0.000383)	(0.000320)
Postgraduate	0.000982***	0.000135	type of contract	0.001767***	0.000368
	(0.000375)	(0.000316)		(0.000360)	(0.000298)
other activity	0.000280	-0.000180	+ 40hr/week	-0.000104	-0.000236
	(0.000358)	(0.000300)		(0.000352)	(0.000284)

Robust standard errors clustered at the school level in parentheses * p < 0.10, ** p < 0.05, *** p < 0.01. All regressions include year dummies.
Source: Elaborated by the authors using SAEB (2009-2015).

Observing the coefficients from each estimate, it becomes clear that under Ordinary Least Squares most of the observable teacher characteristics are correlated with the proportion of females, the exception being having another activity besides teaching and working more than 40 hours per week. Nonetheless, these correlations became statistically insignificant when we look within schools, with the addition of School Fixed Effects. Thus, showing that once we include school dummies in the model, most of the sorting and selection are accounted for, thus, enabling a proper identification of the effect of the proportion of girls.

4.3.2 Teacher expectations

After analyzing the feasibility of the identification strategy in the last subsection, we present the results of our estimates. As mentioned, we focus on over thirty questions in the teachers' questionnaire that relate to classroom climate, separating them by topics, which are: teacher expectations; teacher performance and job satisfaction; teacher perceptions about student behavior; teacher violence exposure; and, teacher pedagogical methods.

Table 4.4 reports the effects of the proportion of female students on 5th grade teacher expectations about his classroom students, namely, if whether he or she feels that more than half of his 5th grade students are going to graduate from elementary school, high school and will attend a college degree. Each cell in the table shows the estimated coefficient, and its corresponding standard deviations, from separate regressions. We also present OLS estimates on Column 1, and a version of our SFE model on Column 2, without controlling for teachers and cohort characteristics.

Table 4.4 - Estimation of the effect of proportion of female students on teachers expectations.

	(1)	(2)	(3)
Finish elementary school	0.064***	0.047***	0.041**
	(0.017)	(0.016)	(0.016)
Finish high school	0.145***	0.089***	0.086***
	(0.020)	(0.021)	(0.022)
Attend college	0.256***	0.143***	0.147***
	(0.028)	(0.028)	(0.028)
Year effects	✓	✓	✓
SFE		✓	✓
Teacher controls			✓
Cohort controls			✓

Robust standard errors clustered at the school level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

Table 4.4, indicates that regardless of the complexity of the model we estimate, the proportion of girls in school tend to influence positively all measures of teacher expectations regarding his students' progress. The effects seem stronger for finishing high school and attending college, leading, respectively, to 0.041 and 0.147 increases on average teacher's expectance. In order to assess if these expectations are affected by the gender of the teacher, as evidenced by Dee (2007), we separate our sample and estimate two distinct models[22] for each gender, but no statistical significant difference was found.

These results, in our understanding, provide the first set of evidence that teaching environment improves with the proportion of girls. A conclusion shared with Lavy and Schlosser (2011), who found positive spillovers from female peers in the quality of the teacher-student relationship. However, since our database refers to teachers' opinions, the results should not be taken as definite, as they may be reflecting teachers' stereotypes, which may constitute an important factor on student academic lives and even have long lasting consequences (FENNEMA et al., 1990; FIGLIO, 2005; LAVY, 2008; LAVY, SAND, 2015). Despite the fact that we cannot rule this possibility out entirely, we have reasons to believe that at least part of this effect we are capturing is due to a better class climate, as it will become clear below, when we analyze other measures of this same issue.

[22] We estimated conditional logit models, grouped by school and controlling for teachers and cohort characteristics.

4.3.3 Teachers' perception about student behavior

In order to measure the relationship between gender and student behavior, we use four questions about the teachers' perception that learning problems in his classroom occur due to student low self-esteem, disinterest, indiscipline or absenteeism. The results are shown on Table 4.5.

Table 4.5 - Estimation of the effect of the proportion of female students on teacher perceptions about student behavior.

	(1)	(2)	(3)
student low self-esteem	-0.143***	-0.061**	-0.060**
	(0.027)	(0.026)	(0.026)
student disinterest	-0.076***	-0.041**	-0.044**
	(0.016)	(0.018)	(0.019)
student indiscipline	-0.245***	-0.195***	-0.194***
	(0.027)	(0.026)	(0.026)
student absenteeism	-0.218***	-0.088***	-0.096***
	(0.027)	(0.028)	(0.029)
Year effects	✓	✓	✓
SFE		✓	✓
Teacher controls			✓
Cohort controls			✓

Robust standard errors clustered at the school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

The estimates indicate that the proportion of girls is associated with an improvement in all measures of student behavior, especially indiscipline and absenteeism. This evidence not only reinforces the ones from Table 5.4, above, but it is also consistent with Figlio (2007) and Blank e Shavit (2016), who emphasize that a disruptive classroom climate have negative effects on the learning process and lower achievement. In Brazil, instructional time is an important issue. Bruns, Evans and Luque (2011) point out that in Brazilian schools time spent on instruction is below 66%, while time spent on classroom management is much higher than in OECD countries. This, therefore, consists of one of the biggest challenges the country needs to overcome in the pursuit of a better education.

In accordance with our results are the literature on the gender differences in behavioral and non-cognitive skills, such as Bertrand and Pan (2013), Becker et al. (2010) and DiPrete and Jennings (2009), who point out that girls begin school with more advanced social and behavioral skills, an advantage that grows over time, having long term consequences on education.

4.3.4 Teacher performance and job satisfaction

To further investigate the effects of the gender composition, we consider three questions about the teachers' performance and job satisfaction. That is, if whether he or she was able to cover more than 80% of the syllabus during this school's calendar year, and if she feels that the learning problems among her students are due to her job stress, which makes difficult to plan and prepare her class, or if whether she feels dissatisfied and discouraged by her profession. The relationship between these answers and the proportion of girls are shown below, on Table 4.6.

Table 4.6 - Estimation of the effect of proportion of female students on teachers' performance and job satisfaction.

	(1)	(2)	(3)
+80% syllabus	0.407***	0.155***	0.150***
	(0.030)	(0.026)	(0.027)
teacher work burnout	0.008	0.043*	0.038
	(0.025)	(0.025)	(0.025)
teacher dissatisfaction	-0.043*	0.008	0.008
	(0.024)	(0.025)	(0.026)
Year effects	✓	✓	✓
SFE		✓	✓
Teacher controls			✓
Cohort controls			✓

Robust standard errors clustered at the school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

The results from our estimates indicate that a greater percentage of girls at school positively affect teachers capacity to cover the syllabus, as indicated by the teachers' school average response to the inquiry of whether they managed to teach more than 80% of the syllabus. This result builds on the evidence that girls lead to a better classroom climate and, therefore, combined with Table 4.4 results, provides a clearer evidence of the positive classroom climate spillovers of having more girls. This leads us to further investigate the reasons behind this positive relationship between classroom climate and gender composition, as we look at student behavior and violence indicators in the next two subsections.

Nonetheless, we do not find evidence in our model that relates the proportion of girls to teacher work satisfaction or job stress and burnout. This is in contrast with Lavy and Schlosser (2011), who focus on Israel educational data, and find a negative relationship between them.

4.3.5 Teacher violence exposure

We also explore teachers' exposure to violence at school. First, by compiling the answers over a more direct connection, that is, if the teacher was a victim of theft, armed robbery, or if he or she had his life threatened, or was threatened by a student inside the school. In another part of the inquiry, we are able to gather information about the exposure to violence through the students' conduct, exploring answers to if the students attended class under the effect of alcohol, drugs or either carrying melee weapons or fire arms.

Table 4.7 - Estimation of the effect of proportion of female students on teachers' violence exposure.

	(1)	(2)	(3)		(1)	(2)	(3)
life-threat.	-0.011	-0.008	-0.011	stud. alcohol	-0.035***	-0.016	-0.021
	(0.007)	(0.012)	(0.012)		(0.011)	(0.015)	(0.015)
Theft	-0.009	-0.008	-0.005	stud. drugs	-0.025	-0.025*	-0.027**
	(0.010)	(0.015)	(0.016)		(0.017)	(0.013)	(0.014)
armd robbery	0.012*	0.013	0.012	stud. weapon	-0.036***	-0.019	-0.026**
	(0.007)	(0.012)	(0.013)		(0.009)	(0.013)	(0.013)
threatened	-0.106***	-0.090***	-0.092***	std fire arms	-0.009	-0.006	-0.008
	(0.013)	(0.017)	(0.017)		(0.008)	(0.011)	(0.012)
Year effects	✓	✓	✓	Year effects	✓	✓	✓
SFE		✓	✓	SFE		✓	✓
Teach. cont.			✓	Teach. cont.			✓
Cohort cont.			✓	Cohort cont.			✓

Robust standard errors clustered at the school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

Our results indicate that a higher proportion of girls is associated with lower exposure to violence, as measured by the negative relationship with teachers' average response of being a victim of threats by their students and by having less students attending class carrying a melee weapon or under the effect of drugs. These results are closely related to the literature on the gender differences in behavioral and non-cognitive skills, where boys tend to be more disruptive (BERTRAND, PAN, 2013), in such a way that, as highlighted in the previous subsections, girls tend to increase classroom climate.

Nonetheless, violence is a major problem in Brazil, studies that relate it to school outcomes have found that it directly reduces achievement, and is also associated with higher teacher and principal absenteeism, turnover, and to temporary school closings, which exacerbate its negative direct impact (DALCIN, 2016; SEVERNINI; FIRPO, 2010; MONTEIRO; ROCHA, 2013). Other factors related to violence in schools are its surroundings, students' background and teacher quality (TAVARES; PIETROBOM, 2016; BECKER; KASSOUF, 2012).

Overall, the results regarding our measures of student behavior indicate that a higher share of females in school generate improvements in student behavior, which reflects in less violence, greater teacher expectations over the student's academic future, and improves the teaching environment. Aligning this result with the previous chapter of this thesis, we generate evidences that the gender mix should be directly taken into consideration when school authorities decide classroom composition. The same can be said about the teachers allocation, as those who are better prepared to deal with behavioral problems could be assigned to classes that have a higher proportion of boys. This is an interesting result, especially in a country like Brazil, where most of the schools allocate students within classrooms by taking into consideration only the students' age and achievement, and not their gender. Moreover, on a wider perspective, the country should welcome policies aimed towards boys, as they are associated with a range of issues, regarding behavior, violence and school progress.

4.3.6 Teacher pedagogical methods.

Lastly, we investigate if there is any relationship between the proportion of girls and some key pedagogical methods applied by the teachers. For that, we use conditional logit models, grouped by the schools in order to assess separately if Math and Portuguese teachers implement any one of the distinct pedagogical methods evaluated in the inquiry.

Table 4.8 - Relationship between the proportion of girls and pedagogical methods utilized by teachers, separated by subject.

Math Teachers		Portuguese Teachers	
familiar situations for students	0.135	copy texts from books or blackboard	-0.114
	(0.157)		(0.139)
reinforce procedures and rules	0.243	discuss texts from papers & magaz.	-0.036
	(0.163)		(0.123)
discuss solutions	0.209	use papers & magazines for grammar	0.033
	(0.206)		(0.122)
memorize rules to solve exercises	0.186	read chronicles, poetry & novels	-0.114
	(0.151)		(0.141)
content from papers & magazines	-0.058	use poetry and novels for grammar	0.060
	(0.125)		(0.131)
discuss methods	-0.056	reinforce grammar concepts	0.171
	(0.207)		(0.128)
try new actions to solve exercises	0.184		
	(0.135)		
Year effects	✓	Year effects	✓
SFE	✓	SFE	✓
Teacher controls	✓	Teacher controls	✓
Cohort controls	✓	Cohort controls	✓

Robust standard errors clustered at the school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Elaborated by the authors using SAEB (2009-2015).

The estimated results indicate that there is no influence of the proportion of girls in any of the teaching methods inquired, regardless of the subject. Therefore, we provide evidence that teaching methods are not necessarily related to gender class or school composition. However, the fact that teachers do not change the way they conduct a class based on gender, does not eliminate the possibility that teachers can play an important role in shaping gender interactions or even display stereotype bias, which can contribute to the gender gap on achievement (DEE, 2007; FENNEMA et al.,1990). These latter effects, despite consisting in interesting research questions, remain a challenge for future work, as it extrapolates the objective of this paper.

4.4 CONCLUDING REMARKS

This paper aimed at identifying the influences of girls on classroom climate and student behavior. It builds upon the evidence that the proportion of girls at school contributes mostly towards other students' Math achievement, a subject that girls tend to be outperformed by boys, which, in turn, suggests that the gender peer effects goes beyond the traditional concept of spillovers that arise from having higher achieving peers.

Exploring a national representative sample of Brazilian 5th grade schools, we find that the benefits of having a greater proportion of girls are mainly through improvements in student behavior, which reflects less violence, greater teacher expectations over the student's academic future, and facilitates classroom progress. The association of boys with these issues also highlights the need for policies aimed towards the improvement of their condition. Moreover, we did not find any evidence of the relationship between different teaching methods and the proportion of girls. This, however, does not eliminate the possibility that teachers might influence gender interactions and produce stereotype bias.

In terms of educational public policies, our findings indicate that, since the proportion of girls positively influences behavior, and therefore achievement, the gender composition of classrooms and schools should be taken into consideration in the decision regarding the placement of low achievers and student with behavioral problems. Moreover, teachers who are better prepared to deal with behavioral problems could be assigned to classes that have a higher proportion of boys.

In sum, this paper brings forth evidences that the gender mix should be directly taken into consideration when school authorities decide classroom composition and the allocation of teachers to classrooms. This could result in a low cost measure aimed at increasing school efficiency, in a country that already performs poorly on achievement tests, even when compared to similar countries.

4.5 REFERENCES

AUTOR, D. H.; FIGLIO, D. N.; KARBOWNIK, K.; ROTH, J.; WASSERMAN, M. **Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes**.Munich: CESifo, 2016. (CESifo Working Paper**,** 5925).

BECKER, G. S.; HUBBARD, W. H. J.; MURPHY, K. M. **Explaining the Worldwide Boom in Higher Education of Women**. Washington: IMF, 2010. (IMF Working Paper Series, 2010-09).

BECKER, K. L.; KASSOUF, A. L. Violência nas escolas: uma análise da relação entre o comportamento agressivo dos alunos e o ambiente escolar. In: ENCONTRO NACIONAL DE ECONOMIA, 40., Porto de Galinhas, 2012. **Anais...** Porto de Galinhas: ANPEC, 2012.

BERTRAND, M.; PAN, J. The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. **American Economic Journal: Applied Economics**. Washington, v. 5, n. 1, p. 32-64, 2013.

BLANK, C.; SHAVIT, Y. The Association Between Student Reports of Classmates' Disruptive Behavior and Student Achievement. **American Educational Research Association**. Washington, v. 2, n. 3, 2016.

BRUNS, B.; EVANS, D.; LUQUE, J. **Achieving World Class Education in Brazil**: The Next Agenda. Washington: World Bank Publications, 2011.

CHETTY, R.; HENDREN, N.; LIN, F.; MAJEROVITZ, J.; SCUDERI, B. **Childhood Environment and Gender Gaps in Adulthood.** Cambridge: NBER, 2016. (NBER Working Paper, 21936).

CORNWELL, C. M.; MUSTARD, D. B.; VAN PARYS, J. **Non-cognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments**: Evidence from Primary School. Munich: IZA, 2011. (IZA Discussion Paper, 5973).

DALCIN, A. K. **Uma análise da relação entre violência na escola e proficiência dos alunos**. 2016. 101f. Dissertation (MS in Applied Economics) – Faculty of Economics, Rio Grande do Sul Federal University, Porto Alegre.

DEE, T. S. The Why Chromosome: How a Teacher's Gender Affects Boys and Girls. **Education Next**. Cambridge, v. 6, n. 4, p. 68-75, 2006.

DEE, T. S. Teachers and the Gender Gaps in Student Achievement. **The Journal of Human Resources**. Madison, v. 42, n. 3, p. 528-554, 2007.

DIPRETE, T. A.; JENNINGS, J. L. **Social/Behavioral Skills and the Gender Gap in Early Educational Achievement**. New York: CPRC, 2009. (CPRC Working Paper, 09-08).

FENNEMA, E.; PETERSON, P. L.; CARPENTER, T. P.; LUBINSKI, C. A. Teachers' attributions and beliefs about girls, boys, and mathematics. **Educational Studies in Mathematics**.Dordrecht, v. 21, n. 1, p. 55-69, 1990.

FIGLIO, D. Names Expectations and the Black-White Test Score Gap. Cambridge: NBER, 2005. (NBER Working Paper, 11195).

FIGLIO, D. Boys Named Sue: Disruptive Children and their Peers. **Education Finance and Policy**. Cambridge, v. 2, n. 4, 2007.

FIRPO, S.; JALES, H.; PINTO, C. Measuring peer effects in the Brazilian school system. **Applied Economics**. Washington, v. 47, n. 32, 2015.

LAVY, V. Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. **Journal of Public Economics**. Amsterdam, v. 92, p. 2083-2105, 2008.

LAVY, V.; SAND, E. **On The Origins of Gender Human Capital Gaps**: Short and Long Term Consequences of Teachers' Stereotypical Biases. Cambridge: NBER, 2015. (NBER Working Paper, 20909).

LAVY, V.; SCHLOSSER, A. Mechanisms and impacts of gender peer effects at school. American Economic Journal: **Applied Economics**, Washington, v. 3, n. 2, p. 1-33, 2011.

LAZEAR, E. P. Teacher incentives. **Swedish Economic Policy Review**. Stockholm, v. 10, n. 2., p. 179-214, 2001.

MONTEIRO, J. ROCHA, R. **Drug Battles and School Achievement**: Evidence from Rio de Janeiro's Favelas.2013.Caracas: CAF. (CAF Working Papers**,** 2013/05).

OCDE. **Brazil Country Note** – Results from PISA 2015. 2016. Available at: http://www.oecd.org/brazil/PISA-2015-Brazil.pdf. Access: march 2018.

SEVERNINI, E.; FIRPO, S. P. **The relationship between school violence and student proficiency**. São Paulo: EESP-FGV, 2009. (EESP-FGV Texto para discussão**,** 236).

TAVARES, P. A.; PIETROBOM, F. C. Fatores associados à violência escolar: evidências para o Estado de São Paulo. **Estudos Econômicos**.São Paulo, v. 46, n. 2, 2016.

4.6 APPENDIX

Table 1A - Descriptive statistics of teachers' perception of classroom climate.

Variable/Topic	Mean	Std. Dev.	Min	Max
Teacher Expectations				
Attend college	0.420	0.402	0	1
Finish elementary school	0.769	0.388	0	1
Finish high school	0.703	0.402	0	1
Teacher Performance and Job Satisfaction				
80% Syllabus	0.473	0.403	0	1
Teacher work burnout	0.299	0.358	0	1
Teacher dissatisfaction	0.273	0.343	0	1
Teachers' Perception of Student Behavior				
Student low self-esteem	0.692	0.356	0	1
Student indifference	0.870	0.255	0	1
Student indiscipline	0.644	0.374	0	1
Student absenteeism	0.413	0.388	0	1
Teacher Violence Exposure				
Teacher victim of life-threatening situation	0.026	0.136	0	1
Teacher victim of theft	0.048	0.168	0	1
Teacher victim of armed robbery	0.030	0.153	0	1
Teacher threatened by student	0.089	0.230	0	1
Student attended class under alcoholic influence	0.093	0.252	0	1
Student attended class under drugs influence	0.048	0.180	0	1
Student carrying melee weapon in class	0.049	0.174	0	1
Student carrying fire arms in class	0.041	0.173	0	1
Teacher Pedagogical Methods				
Copy texts from books or blackboard	0.680	0.383	0	1

Variable/Topic	Mean	Std. Dev.	Min	Max
Use familiar situations for students	0.823	0.298	0	1
Discuss texts from papers & magazines	0.682	0.366	0	1
Use papers & magazines for grammar	0.653	0.375	0	1
Read chronicles, poetry & novels	0.759	0.347	0	1
Use poetry and novels for grammar	0.678	0.377	0	1
Reinforce grammar concepts	0.658	0.382	0	1
Reinforce procedures and rules	0.843	0.283	0	1
Discuss solutions	0.893	0.243	0	1
Memorize rules to solve exercises	0.744	0.356	0	1
Content from papers & magazines	0.594	0.397	0	1
Discuss methods	0.896	0.241	0	1
Try new actions to solve exercises	0.728	0.354	0	1

Source: Elaborated by the authors using SAEB (2009-2015).

5 CONCLUDING REMARKS AND POLICY IMPLICATIONS

This thesis aimed at analyzing the role of gender on scholastic achievement in Brazil, emphasizing inequalities and peer effects. It starts describing the issue of gender inequality, especially among 5^{th} and 9^{th} graders, illustrating that sex differences in test scores in Brazil usually benefit boys relative to girls in math, the opposite of what happens in literacy. Also, this brief description highlights that these gender differences is wider in 9^{th} grade than in 5^{th} grade.

This, therefore, is the main topic of our first paper, in the second chapter of the thesis, which aims at identifying the factors associated with gender test scores differences in Brazil. It finds that most of these differences are related to the background characteristics of the students, but it is not due to differences in the characteristics themselves, as it is more related to the way boys and girls translate their background characteristics into achievement, as identified by the unexplained component of the Oaxaca-Blinder decomposition. We are also able to verify that in all percentiles boys and girls tend to perform relatively more equally in Math than in Portuguese, and on 5^{th} grade than on 9^{th}, even after controlling for students, teachers and schools characteristics. On the policy perspective, this first paper indicates that a more efficient policy aiming at the reduction of gender test score differences should look beyond resources. It also reduces the role played by teachers and schools in narrowing these test score differences. However, despite the fact that we are able to disentangle the gender gap and point out some important contributors of this difference with the covariates in our dataset, most of the gap is still due to the intercept, and therefore require additional investigations.

The second paper of this thesis, in chapter three, investigates gender peer effects among 5^{th} grade public school students in Brazil. It analyzes if the assignment of a student to a school with a higher proportion of girls influences Math and Portuguese achievement. It also investigates if this effect is influenced by school size, students' socioeconomic background, nonlinearities in the proportion of girls and it also looks at classroom gender composition, identifying those schools that tend to overly allocate boys and girls to different classes in order to verify if this also influences achievement. Our findings indicate a positive causal relationship between the proportion of girls and achievement for both genders and on both subjects, especially for Math. Moreover, we also find that girls' achievements are higher when they attend schools with classes that tend to segregate the two genders, implying that despite both genders profit from coexisting, girls benefit even more if clustered together. Therefore, it is not only having more girls in a school that is beneficial to achievement, but the actual coexistence between males and females also plays an important role in test score improvement, especially for women.

On the third paper of this thesis, in chapter four, we investigate the mechanisms through which girls positively affect test scores. For that, we relate school average teacher expectations over the students' academic progress, and their perception of the classroom climate, violence, as well as their pedagogical methods with the school gender composition. Our findings highlight that the benefits of having a greater proportion of girls at school are mainly through improvements in student behavior, which reflects in less violence, greater teacher expectations over the student's academic future, and facilitates classroom progress. In terms of public policies, the last two chapter of this thesis brings forth evidences that are capable of increasing school efficiency and equity in test scores, as it benefits especially girls. As the proportion of girls positively influences behavior, and therefore achievement, the gender composition of classrooms and schools should be taken into consideration in the decision regarding the placement of low achievers and student with behavioral problems. Moreover, teachers who are better prepared to deal with behavioral problems could be assigned to classes that have a higher proportion of boys. Nonetheless, this association of boys with violence and other behavioral issues also highlights the need for policies aimed towards the improvement of their conditions.

These are interesting results, especially in a country like Brazil, where the schooling system mostly allocate students within classrooms by age and achievement, and not their gender. Therefore, taking these findings into consideration can result in an effective and low cost measure aimed at increasing achievement.

REFERENCES

ALVES, M. T. G.; SOARES, J. F.; PEREIRA, F. Desigualdades educacionais no ensino fundamental de 2005 a 2013: hiato entre grupos sociais. **Revista Brasileira de Sociologia**, Sergipe, v. 4, n. 7, p. 49-82, 2016.

BAKER, M.; MILLIGAN, K. Boy-girl differences in parental time investments: Evidence from three countries. **Journal of Human Capital**, Chicago, v. 10, n. 4, p. 399-441, 2016.

BEDARD, K.; CHO, I. Early gender test score gaps across OECD countries. **Economics of Education Review**, Cambridge, v. 29, n. 3, p. 348-363, 2010.

BERTRAND, M.; PAN, J. The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. **American Economic Journal: Applied Economics**, Washington, v. 5, n. 1, p. 32-64, 2013.

BHARADWAJ, P.; DE GIORGI, G.; HANSEN, D.; NEILSON, C. The gender gap in mathematics: Evidence from Chile. **Economic Development and Cultural Change**, Chicago, v. 65, n. 1, 2016.

BIONDI, R. L.; FELÍCIO, F. Atributos escolares e o desempenho dos estudantes: uma análise em painel dos dados do SAEB. São Paulo: INEP, 2007. (Texto para Discussão, 25).

BUCHMANN, C.; DIPRETE, T. A.; MCDANIEL, A. Gender inequalities in education. **Annual Review Sociology**. v. 34, p. 319-337, 2008.

CAHILL, L. Why sex matters for neuroscience. **Nature Reviews Neuroscience**. v. 7, p. 477-484, 2006.

CONTINI, D.; TOMMASO, M. L. D.; MENDOLIA, S. The Gender Gap in Mathematics Achievement: Evidence from Italian Data. **Economics of Education Review**, Cambridge v.58, p. 32-42, 2017.

CORSEUIL, C.; FOGUEL, M. N.; GONZAGA, G.; RIBEIRO, E. P. The effects of Youth Training Program on Youth Turnover in Brazil. São Paulo: REAP – Rede de Economia Aplicada, 2012. (REAP Working Paper, 042).

DICKERSON, A.; MCINTOSH, S.; VALENTE, C. Do the maths: An analysis of the gender gap in mathematics in Africa. **Economics of Education Review**, Cambridge, v. 46, p. 1-22, 2015.

ELLISON, G.; SWANSON, A. The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions. **Journal of Economic Perspectives**. Nashville, v. 24, n. 2, p. 109-28, 2010.

FIRPO, S.; JALES, H.; PINTO, C. Measuring peer effects in the Brazilian school system. **Applied Economics**, Washington, v. 47, n. 32, 2015.

FRANCO, A. M. P.; MENEZES-FILHO, N. A.; Os determinantes do aprendizado com dados de um painel de escolas do SAEB. In: ENCONTRO NACIONAL DE ECONOMIA, 36., Salvador, 2008. **Anais...** Salvador: ANPEC, 2008.

FRYER JR, R.; LEVITT, S. D. An Empirical Analysis of the Gender Gap in Mathematics. **American Economic Journal: Applied Economics**. v. 2, n. 2, p. 210-40, 2010.

GOLSTEYN, B, H.; SCHILS, T. Gender gaps in primary school achievement: a decomposition into endowments and returns to IQ and non-cognitive factors. **Economics of Education Review**, Cambridge, v. 41, p. 176-187, 2014.

GUISO, L.; MONTE, F.; SAPIENZA, P.; ZINGALES, L. Culture, gender and math. **Science** v. 320, p. 1164-1165, 2008.

HEDGES, L. V.; NOWELL, A. Sex differences in mental test scores, variability, and numbers of high-scoring individuals. **Science**. v. 269, n. 5220, p. 41-45, 1995.

HUSAIN, M.; MILLIMET, D. L. The mythical 'boy crisis'? **Economics of Education Review**. v. 28, n. 1, p. 38-48, 2009.

HYDE, J. S.; FENNEMA, E.; RYAN, M.; FROST, L. A.; HOPP, C. Gender comparisons of mathematics attitudes and affect: a meta-analysis. **Psychology of women quarterly**, New York, v. 14, n. 3, p. 299-342, 1990.

KASSOUF, A. L. Wage gender discrimination and segmentation in the Brazilian labor market. **Economia Aplicada**, São Paulo, v. 2, n. 2, p. 243-269, 1998.

LEAHEY, E.; GUO, G. Gender differences in mathematical trajectories. **Social forces**. Chapel Hill, v. 80, n. 2, p. 713-732, 2001.

MADALOZZO, R. Market and Home Production: Gender Differences in Brazil. São Paulo: Insper, 2009. (Insper Working Paper, 168).

MADALOZZO, R.; ARTES, R. Escolhas profissionais e impactos no diferencial salarial entre homens e mulheres. **Cadernos de Pesquisa**, Brasília, v. 47, n. 163, 2017.

OECD. **Brazil Country Note** – Results from PISA 2015. Available at: http://www.oecd.org/brazil/PISA-2015-Brazil.pdf. 2016. Acess: march 2017.

PINTO, C. Semiparametric estimation of peer effect in classrooms: Evidence for Brazilian schools in 2003. 2008. Available at: https://www.economics.uci.edu/files/docs/econoseminar/w10/pinto.pdf . Acess: march 2017.

RASCHE JÚNIOR, F. **Privação de Liberdades Femininas: uma análise da participação das mulheres nas áreas de ciência e tecnologia no Brasil entre 2004 e 2013**. 2015. 73f. Undergraduate Thesis (Graduation in Economics) – Faculty of Economics, Rio Grande do Sul Federal University, Porto Alegre.

SAX, L. **Why gender matters**: What parents and teachers need to know about the emerging science of sex differences. New York: Doubleday. 2005.

SOARES, J. F.; CANDIAN, J. O efeito da escola básica brasileira: As evidências do PISA e do SAEB. **Revista Contemporânea de Educação**, Rio de Janeiro, v. 2, n. 4, 2007.

SOARES, J. F.; DELGADO, V. M. S. Medida das desigualdades de aprendizado entre estudantes de ensino fundamental. **Estudos em Avaliação Educacional**, Brasília, v. 27, n. 66, 2016.

SOARES, S. S. D.; NASCIMENTO, P. A. M. M. Evolução do desempenho cognitivo dos jovens brasileiros no Pisa. **Cadernos de Pesquisa**, Brasília, v. 42, n. 145, p. 68-87, 2012.

SOARES, J. F.; ALVES, M. T. G.; XAVIER, F. P. Effects of Brazilian schools on student learning. **Assessment in Education: Principles, Policy & Practice**, Oxford, v. 23, n. 1, p. 75-97, 2016.

SPELKE, E. S. Sex differences in intrinsic aptitude for mathematics and science? A critical review. **American Psychologist**. Washington, v. 60, n. 9, p. 950-958, 2005.

TILLMANN; E. T.; COMIM, F. Os determinantes da decisão entre estudo e trabalho dos jovens no Brasil e a geração Nem-Nem. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 46, n. 2, p. 47-78, 2016.

TINKLIN, T. Gender differences and high attainment. **British Educational Research Journal**, London, v. 29, n. 3, p. 307-325, 2003.