UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Gender, Skills and Educational Outcomes

Gênero, Habilidades e Resultados Educacionais

Ana Paula Melo da Silva

Orientador: Prof. Dr. Marcos de Almeida Rangel

São Paulo - Brasil

2016

Prof. Dr. Marco Antonio Zago

Reitor da Universidade de São Paulo

Prof. Dr. Adalberto Américo Fischmann

Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Hélio Nogueira da Cruz

Chefe do Departamento de Economia

Prof. Dr. Márcio Issao Nakane

Coordenador do Programa de Pós-Graduação em Economia

ANA PAULA MELO DA SILVA

Gender, Skills and Educational Outcomes

Gênero, Habilidades e Resultados Educacionais

Dissertação apresentada ao Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito parcial para a obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Marcos de Almeida Rangel

Versão Corrigida

São Paulo - Brasil

2016

FICHA CATALOGRÁFICA
Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Silva, Ana Paula Melo da
Gender, skills and educational outcomes / Ana Paula Melo da Silva. – São Paulo, 2016.
61 p.

Dissertação (Mestrado) – Universidade de São Paulo, 2016.
Orientador: Marcos de Almeida Rangel.

1. Economia da educação 2. Economia da personalidade 3. Gênero 4. Habilidades não cognitivas 5. Desempenho educacional 6. Decomposição estatística I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 370.193

Agradecimentos

Agradeço ao meu pai e minha mãe, por terem me dado tanto suporte e incentivo pelas minhas escolhas acadêmicas. Sei que não foi fácil, mas obrigada por me incentivarem a nunca desistir de alcançar meus objetivos. À minha irmã, primas e primos. Agradeço às minhas tias Elisa, Elisângela e Eloisa por serem igualmente responsáveis pelo meu sucesso. Agradeço especialmente à minha tia Enilce, seu apoio abriu a principal porta para que eu esteja hoje encerrando meu mestrado na USP.

Ao meus orientadores, Marcos A. Rangel e Ricardo Madeira, por tudo que me ensinaram, pelas boas discussões e pela paciência em momentos mais críticos. Todo apoio foi fundamental para minha formação como economista e pesquisadora. Ao Professor Daniel Santos pela orientação voluntária e por todas as oportunidades que me possibilitou e que me abriram muitas portas na vida acadêmica. Agradeço pelos incentivos e pelo conhecimento compartilhado.

Ao apoio financeiro concendido pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por meio de bolsa de mestrado (proc. 2013/21099-5) e bolsa de estágio de pesquisa no exterior (BEPE, proc. 2014/08834-0).

Ao *Center for The Economics of Human Development (CEHD)* e ao Professor James J. Heckman pelo apoio e orientação durante o período de visita a Universidade de Chicago. Agradeço também aos meus colegas de trabalho nesse grupo, os quais se tornaram também bons amigos.

Agradeço aos amigos que fiz nessa cidade que hoje considero minha casa, que faziam os momentos mais difíceis parecem mais leves e agradáveis. Levarei essas amizades para a vida.

Resumo

Por que meninas e meninos têm desempenhos diferentes na escola? Nesta pesquisa, vamos examinar as disparidades de gênero no desempenho escolar em Linguagem e Matemática utilizando um novo conjunto de dados de 10.000 alunos do Sistema Público de Ensino de São Paulo, Brasil. O amplo conjunto de informações disponíveis nos permite testar a importância relativa das competências socioemocionais e incentivos à educação sobre os resultados educacionais de meninos e meninas. Nós traçamos um perfil socioemocional único de meninos e meninas no Brasil e encontramos diferenças de gênero significativas nas habilidades e incentivos à educação. Realizando um exercício de decomposição, descobrimos que a diferença de gênero nos resultados educacionais é significativamente explicada pelas diferenças nos níveis de insumos observados, especialmente incentivos à educação. Ainda, encontramos novas evidências sobre a importância dos incentivos à educação para explicar as diferenças de gênero em notas de sala de aula e resultados de testes padronizados.

Palavras-chaves: economia da educação, economia da personalidade, gênero, habilidades não-cognitivas, desempenho educacional, decomposição estatística.

Abstract

Why do girls and boys perform differently in school? In this paper, we examine the gender gap in educational achievement for Language and Mathematics using a novel dataset of 10,000 students in the Public School System of Sao Paulo, Brazil. The broad set of available information allows us to test the relative importance of skills and incentives to schooling on the educational outcomes of boys and girls. We summarize a unique socioemotional profile of boys and girls in Brazil and find significant gender differences in skills and incentives to schooling. Performing a decomposition exercise, we find that the gender gap in educational outcomes is significantly explained by differences in the levels of observed inputs, especially incentives to schooling. We bring new evidence on the importance of incentives to schooling in explaining the gender differences in both classroom grades an standardized test scores.

Key-words: economics of education, economics of personality, gender, non-cognitive skills, educational outcomes, statistical decomposition.

Sumário

1 Introduction

Within the last century, as women gained equal access to education, they have tended to attain more years of schooling and perform better than their male counterparts on other measures of school success. In 2010, women surpassed men in higher educational attainment in 67 of 120 countries (Becker, Hubbard & Murphy (2010)). Developing countries have recently become more similar to the developed world, with gender gaps in grade progression and in dropout rates evolving in favour of girls (Grant & Behrman (2010)).

Gender disparities appear early in school and girls in average earn higher classroom grades in every subject. However, in relation to test scores, girls systematically outscore boys in Language but, in Mathematics the difference favours Boys. Additionally, as shown in a recent report released by OECD (OECD (2015)), boys are expected to drop out of school earlier, a difference more striking for higher levels of educational attainment. Although boys and girls have similar levels of achievement in the PISA standardized tests, their level of expectations about their future educational attainment and occupational choices differs substantially.

This paper aims to understand what are the factors driving the divergence between girls' and boys' educational achievement in elementary school, both in tests scores and grades. We use unique data from a sample of 10,000 6th graders in poorly performing Brazilian public-schools. It combines data on standardized tests and teacher assessments with a detailed inventory of personality traits. We start from the evidence that, in compass with worldwide evidences, in our sample girls outperform boys in mathematics and language in classroom grades, but not necessarily in standardized exams: no gap is found in mathematics while a positive gap emerges in language. A unique feature of this research is that the rich set of available measures allows us to add evidence on their relative importance to explain the gender gap.

Our main estimates are derived from the decomposition of three measures of

achievement in which a significant gap is found: classroom language and mathematics grades and standardized test scores in language (Portuguese)[1]. Our chosen method is a pooled two-folded Oaxaca decomposition. We use alternative ways to proxy the relevant inputs: (i) measures of cognition, personality skills and incentives to schooling obtained from a survey; (ii) a factor on accumulated skills in a previous year estimated using a set of classroom grades and standardized scores in language, mathematics and physical education.

Overall, this research documents a novel educational profile of boys and girls in Brazil. Our results can be summarized as follows. First, we find evidence that teacher classroom grades are not aligned with test scores and that gender differences are significantly larger in teacher assessments. Second, in contrast to racial gaps for example, gender differences in achievement are not very sensitive to the inclusion of socioeconomic or parenting controls. Third, we find a significant difference in the distribution of skills and incentives for girls and boys[2]. Fourth, in our estimates, the measures related to incentives to schooling account for the largest portion of the explained component, followed by noncognitive skills. However, for all decomposed gaps, at least half of the difference remains unexplained by our observed measures, even when we control for a lagged measure of achievement.

We also test a few additional hypothesis. Our main estimates assume that boys and girls have the same technology of production, that is, they map inputs into outcomes in the same way[3]. We relax this assumption to test whether the coefficients differ between gender. A surprising finding is that we cannot reject the hypothesis of equality of the coefficients between the production functions of boys and girls - which corroborates the hypothesis in the theoretical framework. Also, another type of educational outcome we will analyse is a widely used measure in the literature since the work of Lavy (2008): the difference between

[1] No significant gap is found in mathematics test scores

[2] Differences in personality and general behavior between males and females are widely known in the literature. Although the gaps are modest, they are consistent and present across cultures (Jr., Terracciano & McCrae (2001), Hyde (2005)).

[3] In practice, we assume the coefficients of the regression equations for boys and girls for each educational outcome is the same.

the two measures of assessment, which is often referred to as Potential Bias. We test how sensitive the unexplained gender gap in these outcomes is to: 1) students' characteristics; 2) teacher characteristics. We find that the gap across exams is more sensitive to students' characteristics than to teachers's characteristics, weakening the possible interpretation of the residual as discrimination against boys, although we cannot rule out this possibility.

This research complements and relates to the literature on assessment and gender gap in school achievement by bringing them together with an emerging literature on personality economics.[4] The former shows the importance of noncognitive abilities in explaining life outcomes, including education. [5] Both cognitive and noncognitive skills, which are elements of an individual's personality[6], can change over the life cycle, although not in the same way. Since cognitive skills seem to stabilize earlier in life, it is more difficult to compensate low cognitive endowments at later ages (Cunha & Heckman (2008)). The idea that these soft skills are malleable makes this kind of research relevant to support future public policy.

The remainder of this research is organized as follows. Chapter 2 describes the database and the gender gap in achievement in our sample, addressing the main outcomes of interest. We review the literature on chapter 3. Chapter 4 develops a basic framework on task performance extending it to gender differences and in Chapter 5 we detail our empirical strategy. Descriptive statistics are shown in Chapter 6. Finally, we present the results in Chapter 7. We summarize our main findings in Chapter 8.

[4] View Borghans et al. (2008) and Almlund et al. (2011) for an extensive description of the Economics of Personality Literature.

[5] See Heckman & Rubinstein (2001); Heckman, Humphries & Kautz (2012); Heckman & Kautz (2014a); Heckman (2008), Cunha & Heckman (2009); Heckman, Stixrud & Urzua (2006); Carneiro, Heckman & Masterov (2005)

[6] See Roberts (2006) for a model of personality.

2 Boys and Girls in the Public School System of São Paulo: The Gender Gap in Achievement

2.1 The Data

Sao Paulo's Secretary of Education (SEE-SP) collects detailed information on the universe of students and teachers in the state's educational system[1]. We merged information at the individual level on matriculation, teachers' allocation to classrooms, transcript records, standardized tests and a survey on personality. Also, we were able to track students within the school system over the years. For this paper, we use information of students from 2009 (4th grade) and 2011 (6th grade).

The two measures of achievement we use are not graded in the same way, although both should measure the same set of competences described in the curriculum. Teachers, for example, are free to choose their methods to allocate students across the grading points, ranging from 1 to 10, but they are supposed to follow the pedagogical material provided by the state administration. On the other hand, our measure of standardized tests is the SARESP Evaluation System[2], administered annually statewide on all state-run public school students in certain grades, including 6th grade. The exam contains 30 multiple choice items, covering the basic curriculum and the scores are calculated using Item Response Theory methods. It is important to note that Standardized exams are taken in the same classroom the student takes classes regularly and a different teacher is allocated to supervise the students during the test. The grading is blind, done electronically. Another relevant difference between the two exams is that the score on the standardized test is never made available at the individual level to students, schools or parents. Besides the exam, students and parents also answer a survey on socioeconomic background, demographics, study habits, among other issues which is available to us to all 6th grade students that took the test in 2011.

[1] A detailed description of the database can be seen in Botelho, Madeira & Rangel (2015)
[2] Sistema de Avaliacao de Rendimento do Estado de Sao Paulo - SARESP)

Our survey on personality was administered to the students as part of a randomized controlled trial run in 2011 in poorly performing schools in the state's school system. Respondents of the survey are from targeted schools with the worst performance in SARESP exams. Since the program was based on tutorial sections, the schools should also have underutilized capacity that could be used to implement the tutorial sessions. Approximately 200 schools were offered the program but only 141 principals expressed interest in hosting it. Data collection took place in November 2011. We do not use the randomization information at any point. The personality survey has information at the individual level on the following topics: schooling and career expectations, impediments to schooling, personality measures, fluid intelligence and study habits.

Our final database is composed of all the above components. Our working sample consists of 10,082 students in 126 schools in 2011. From 2009, we can match information for 5,262 students in our sample.

2.2 The Gender Gap in Achievement in São Paulo

This research focuses mainly on the fact that girls and boys perform differently across types of assessment. For this analysis we normalize classroom grades and standardized exams to have mean 0 and standard deviation 1 and we do this transformation over the entire sample. Regarding achievement in classroom, girls outscore boys by 0.28 standard deviations in mathematics and 0.45 standard deviations in language. No gap is found on standardized test in mathematics while girls outscore boys in language tests by 0.34 standard deviations. The same results are found when the distributions of grades and standardized exams are compared (Figures 2 and 3).

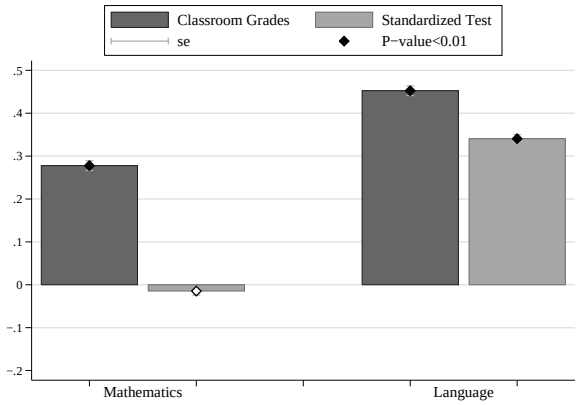


Figure 1 – Gender Gap in Math and Language Achievement Measured by Grades and Standardized Exams, in standard deviations

Note: This figure displays the gender gap in a sample of 6th graders students of São Paulo. All variables are normalized to have mean 1 and standard deviation 0. Errors are robust and clustered at the classroom level.

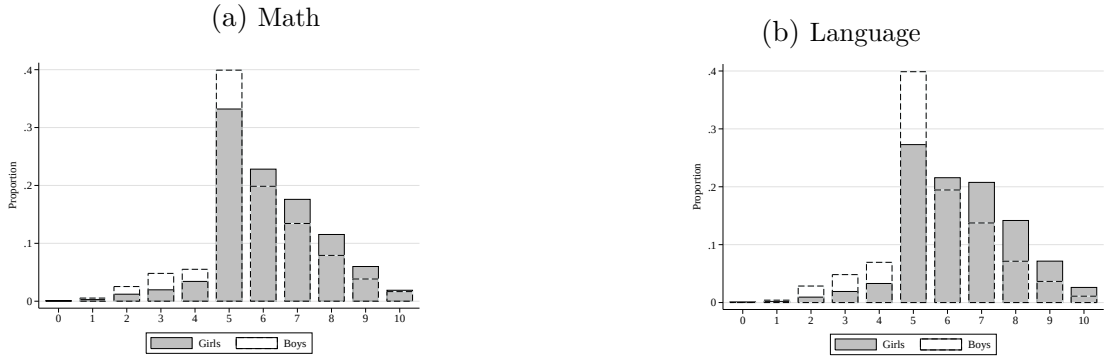


Figure 2 – Distributions of Teacher Assigned Grades in Language and Math, by gender

Note: Grades are assigned by a teacher from 0-10, with 5 as the passing grade.

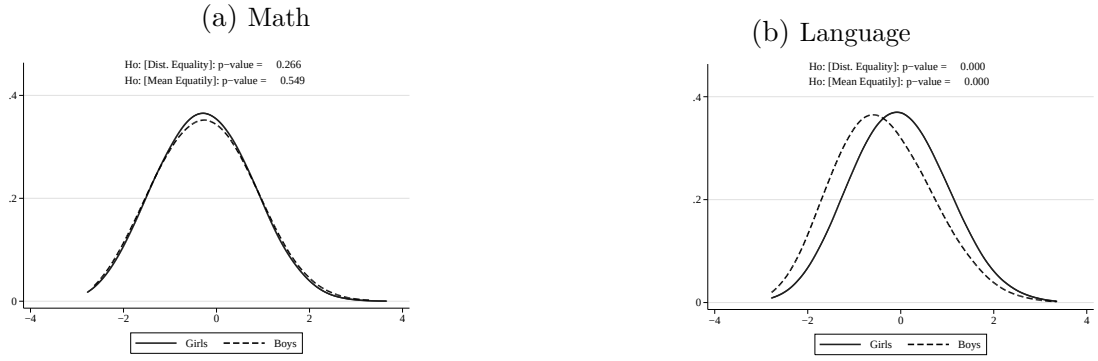


Figure 3 – Distribution of State Administered Standardized Exams in Language and Math, by gender

Note: Standardized Exams scores are reported by the state administration, responsible for the exam. Variables were normalized to have mean zero and standard deviation one.

3 The Gender Gap in Achievement: a review of the literature

Few other papers addressed the importance of skills and aspirations on the gender gap in educational outcomes, but our paper is the first one to test the relative importance of a large set of measurements on these variables to explain differences in both test scores and classroom grades in Language and Mathematics. Similar to our paper, Cornwell, Mustard & Parys (2013) also test the importance of noncognitive skills in explaining the gender differences in classroom grades relative to test scores for a sample of primary school students in the United States. They find that controlling for noncognitive skills eliminates the gender gap in most of their specifications. However, a limitation of their study is the use of a teacher report as a measure of the student's noncognitive skills [1]. So, the disappearance of the gap could be due either to students' abilities or to teacher' gender bias. This possible bias is likely to be included in the measure of personality used by the authors, shrinking the gap beyond to what is related to the students' abilities. As we will detail in Section 6, our measures on the students' abilities are self-reported.

The role of the aspirations and expectations in explaining the gender gap in educational outcomes was explored by Fortin, Oreopoulos & Phipps (2013) in a sample of high school students. Using longitudinal data, they decompose grades into sets of factors that changed differently between gender over time. Educational expectations account for the largest share compared to non-cognitive traits, family environment and labor market work during school. Our results corroborate this finding. We interpret aspirations as a set of characteristics that works as an incentive to performance. Kids that aspire towards a higher level of education are more motivated to perform well in school in order to effectively achieve the pursued goal.

Most of the studies closer in empirical strategy to ours focus on gender differences in test scores. For example, Fryer & Levitt (2010) find no evidence on gender differences in early stages of life, while in 5th graders girls underscore boys in more than 0.2 standard

[1] They use data from the 1998-99 ECLS-K cohort.

deviations in math (which is equivalent to 2.5 months of schooling). They find low support for theories related to low investment in math by girls or low parental expectations. Also, Golsteyn & Schils (2014) decompose the gender gap in standardized tests score in both math and language into cognitive and noncognitive skills for children in elementary school in the south of Netherlands. They find that both cognitive and non-cognitive skills play an important role in the gap, not only due to differences in the levels (endowment effect) but also in the returns to skills (parameter effect).

In general, classroom environments are considered more friendly and less competitive[2] than a standardized exam situation, even if no direct incentives are offered in the latter[3]. Relative to other situational inputs, Legewie & DiPrete (2012) use data on German schools to target the extent in which school context affects the gender gap in achievement. Schools create a pro-girls environment, shaping boys' and girls' learning orientations, with boys being more sensible to school-level inputs.

Furthermore, historical and persistent gender gaps give rise to discrimination hypothesis. A historical gender gap in favor of girls has been pointed by Voyer & Voyer (2014) as a reason for a teacher to give girls better grades, despite any difference in abilities. Likewise, Lavy (2008) evaluates a natural experiment in Israel to test for the presence of discrimination towards gender in public schools. He finds that the gender gap in all subjects is sensitive to teachers' characteristics, straightening the interpretation of the potential bias as discrimination. Lavy & Sand (2015) extend those conclusions by estimating the impact of potential teacher bias on future educational outcomes and occupational choices of boys and girls. These results are corroborated by Falch & Naper (2013) and Lindahl (2007) for a sample of students from Norway and Sweden, respectively, with results supporting Lavy's. Following a more accurate methodology, Hinnerich, Höglin & Johannesson (2011)

[2] Evidence on gender related reactions towards competition is mixed and not robust to cultural differences. According to Gneezy & Rustichini (2004), competition improves performance of boys but not of girls in a sample of Israeli children. This evidence could be related to Borghans, Meijers & Weel (2008b). The latter suggest individuals who score high in conscientiousness (or related facets) are already exerting a high level of effort. Yet, Dreber, Essen & Ranehill (2011) replicates the Israeli study and finds no evidence on gender differences in the response to competition.

[3] The standardized exam we use in this paper is administered by the state for school-level evaluation purposes and the students do not receive significant rewards for any level of achievement.

compare the same exam graded in a blind and a non-blind way. Their findings have no support for gender discrimination.

4 Theoretical Framework

Skills are an important input on the outcomes people produce in their life time. They are multiple and malleable and they relate not only to cognition but also to personality[1]. Cunha & Heckman (2007), Cunha & Heckman (2008), Heckman & Mosso (2014) describe how a person is characterized by skills (θ) and how these skills, along with effort (e), incentives (R) and situational aspects (A and Z) are related to outcomes. Consider a person (i) characterized by:

$$\theta_i = (\theta_i^C, \theta_i^N) \tag{4.1}$$

where θ_i^C is a vector of cognitive skills and θ_i^N is a vector of noncognitive skills. Let $j = \{1, ..., J\}$ index the measures on the set $\mathcal{J}$. Individuals are indexed by $i = \{1, ..., I\}$ in the set $\mathcal{I}$. We express in Equation 4.2 the outcome in a Task for individual $i(T_{i,j})$ as:

$$T_{i,j} = \phi_j(\theta_i^C, \theta_i^N, e_{i,j}, A_{i,j}) \tag{4.2}$$

$$e_{i,j} = \psi_j(\theta_i^C, \theta_i^N, R_{i,j}, Z_{i,j}) \tag{4.3}$$

where A and Z are sets of situational inputs that affect outcomes and effort (e), respectively. Any outcome can be seen as a task performance (e.g. labor market outcomes, achievement tests, answering a personality inventory). Effort is also characterized by the incentives ($R_{i,j}$) a person has to perform a task. In other words, an incentive can be interpreted as the expected reward given the information the individual has. Each individual allocates a finite amount of effort (E) across the performed tasks: $\sum_j^J e_j = E$.

Equation 4.2 shows how these skills, along with effort, incentives and constraints, generates a varied set of outcomes. Observe that some outcomes could require more

[1] See Almlund et al. (2011)

cognitive skills, like performing in a standardized test, while passing grades could require not only cognition but also noncognitive skills. Also, people can achieve the same level of output for different reasons. One can compensate for a lack of cognitive skills by applying more effort on a task or for being more goal oriented, for example. Borghans, Meijers & Weel (2008a) provides evidence on how the performance on an achievement test can vary according to the incentives the student faces and how the response to the incentive depends on skills.

Substituting (4.3) into (4.2), we have the reduced form:

$$T_{i,j} = \widetilde{\phi_j}(\theta_i^C, \theta_i^N, R_{i,j}, X_{i,j}) \tag{4.4}$$

where X is the combination of the two situational factors (A and Z).

As an extension of the model, we describe how boys and girls can perform differently in each kind of task as a function of skills, incentives and environmental inputs. Task performance across groups can emerge from differences in the inputs once we restrict the coefficients to be equal across gender.

Consider a boy and a girl performing the same task, as in Equation 4.5. We assume $\widetilde{\phi}_j^B(.) = \widetilde{\phi}_j^G(.) = \widetilde{\phi}_j(.)$ (*Assumption 1*), then differences in the outcome of the same task can arise only from differences in the levels of the inputs. We also allow for incentives to vary across tasks. Let the gender index $g \in \{G, B\}$ indicate if the equation is for Girls or Boys.

$$T_{i,j}^g = \widetilde{\phi}_j(\theta_i^{Cg}, \theta_i^{Ng}, R_{i,j}^g, X_{i,j}^g) \tag{4.5}$$

We represent T as a linear in the parameters equation(*Assumption 2*). We assume

the output in a task is linear in its arguments for every student in our sample:

$$T_{i,j} = \alpha_{C,j}\theta_i^C + \alpha_{N,j}\theta_i^N + \alpha_{R,j}R_{i,j} + X'_{i,j}\sigma_j + \epsilon_{i,j} \tag{4.6}$$

where ϵ is the component unobserved by the econometrician. It can be due to individual, family or classroom/school components. If we take the difference between two tasks (Equation 4.6), where indexes 1 and 2 indicate the type of test. We obtain:

$$T_{1i} - T_{2i} = \beta_C\theta_i^C + \beta_N\theta_i^N + \rho(R_i) + \vartheta(X'_i) + \xi_i \tag{4.7}$$

where $\rho(.)$ and $\vartheta(.)$ are unknown functions of incentives and situational inputs from both type of tasks, but under (*Assumption 1*) these functions are the same for boys and girls. Nonetheless, the more similar the two tasks are then closer to zero the coefficients are.

5 Empirical Strategy

To simplify the interpretation of the estimates, we apply a transformation to the variables. Consider M_i to be a vector of measures of interests for each individual i. T_i is a vector of the main outcomes: achievement test scores and classroom grades. We will estimate $\widetilde{M}$ and $\widetilde{T}$, which are M e T residualized with respect to a vector of controls, X, respectively, with exception of the gender indicator (G).

Then, consider $T_{i,j}$ to be a student's score in each type of assessment (j) indexed by $j = \{1...J\}$ on the set $\mathcal{J}$:

$$\widetilde{T}_{i,j} = \widetilde{M'}_i \gamma_j + \delta_j G_i + \nu_{i,j} \tag{5.1}$$

where $\widetilde{M} = (\widetilde{\theta}^C, \widetilde{\theta}^N, \widetilde{R}, \widetilde{S})$. S is a vector of classroom and teacher characteristics. This approach assumes the coefficients are the same between boys ($G = 0$) and girls ($G = 1$). That is, gender assignment does not determine how the outcomes are produced, leaving only differences in the levels of inputs to explain the gender gap in the outcomes. So, the first step of our empirical strategy is to estimate Equation 5.1 for each type of assessment.

Our second step is to decompose the gender gap estimated in the previous step to assess the relative importance of the inputs. In a decomposition, there are different ways to choose the counterfactual (see Fortin, Oreopoulos & Phipps (2015)) in order to split the gap into the explained and unexplained parts[1]. The method we chose is considered the most recommended to separate explained and unexplained gaps between males and females (Elder, Goddeeris & Haider (2010)). This method is also used by Fryer & Levitt (2010) for racial gaps and by Cornwell, Mustard & Parys (2013) to girls and boys in an exercise similar to ours.

From Equation 5.1, we take the expectations over $\widetilde{M}$ and G, an indicator of group

[1] This term is originally considered as structural differences in the production function. However, for the cases in which the group membership is determined by an unchangeable characteristic of the individuals, the literature call it as "unexplained component"or "discrimination".

membership. We assume that $\mathbb{E}[\nu|\widetilde{M}] = 0$, then for each outcome j, which is omitted for notation simplification, we have:

$$\mathbb{E}[\widetilde{T}_i|\widetilde{M}, G] = \widetilde{M}_i\gamma + \delta G_i \tag{5.2}$$

It follows that:

$$\begin{aligned} &\mathbb{E}[\widetilde{T}_i|\widetilde{M}, G = 1] - \mathbb{E}[\widetilde{T}_i|\widetilde{M}, G = 0] = \\ &= (\mathbb{E}[\widetilde{M}_i|G = 1]\gamma + \delta) - (\mathbb{E}[\widetilde{M}_i|G = 0]\gamma) \\ &= \underbrace{(\mathbb{E}[\widetilde{M}_i|G = 1] - \mathbb{E}[\widetilde{M}_i|G = 0])\gamma}_{\text{Explained Part}} + \underbrace{\delta}_{UnexplainedPart} \end{aligned} \tag{5.3}$$

This raw decomposition we perform allows us to indicate the relative role of each component in the gender gap in the outcomes of interest. Due to the use of non-experimental data, we cannot denote a casual interpretation to the estimated coefficients. Still we can give a relevant interpretation to the decomposed gender gap into the explained and unexplained parts. That is, we can interpret how changes in the levels of observed measures correlates to the differences in the outcomes.

Another outcome we are interested is the difference between the two types of assessments (Equation 7.2). As in Lavy (2008), we will test the sensitivity of this gap across exams to the inclusion of the inputs as controls for the students' abilities and incentives. We will also allow this difference to vary across teacher characteristics. Both approaches will shed light on whether this difference is an indicative of differences in behaviour between boys and girls or an indicative of teacher's discrimination against boys.

$$\widetilde{T}_{j,i} - \widetilde{T}_{l,i} = \widetilde{M'}_i\beta + \delta G_i + \nu_i \tag{5.4}$$

6 Measures on Inputs

In this section, we show descriptive statistics of the students in our sample and how we built measures on cognitive and personality skills, incentives to schooling and classroom characteristics (peers' and teachers') as well as socioeconomic status, background and parental information.

6.1 Estimating a Linear Factor Model

When dealing with personality assessments, Factor Models are useful for estimating latent traits based on observed measures. It seeks the smallest number of constructs able to reproduce the original data (Gorsuch (2003)). Let $n = \{1, ..., N\}$ index the measures on the set A and $l = \{1, ..., L\}$ the factors on the set B. Individuals are indexed by $i = \{1, ..., I\}$ in the set C. We can express each measurement available for individual $i(M_{i,n})$ as:

$$M_{i,n} = \beta' F_i + \epsilon_{i,n} \quad , i = 1, ..., I \quad n = 1, ..., N \tag{6.1}$$

where $\beta = (\beta_1, ..., \beta_L)$ is a vector of real numbers, the factor loadings. $F_i = (F_{i1}, ...F_{iL})$ is the set of latent factors. The aim is to recover the latent variables that explain the maximum variance on the original data at the same time that it reduces the number of variables available ($L < N$). $\epsilon_{i,n}$ is the measurement error (or uniqueness) in each measure.

Our chosen method to extract the factors is Maximum Likelihood with oblique quartimin rotation. It extracts factors with maximum correlations of measures within a factor, at the same time it allows the factors to be minimally correlated. We perform Factor Analysis on two sets of measures in order to estimate latent abilities ($M_{i,n}$): (i) Self-reported measures of behaviour n of a person i obtained from a Standard Individual Assessment - Personality Survey; (ii) the set of subject specific grades on Mathematics, Language and Physical Education as well as standardized test scores on mathematics and language to account for additional characteristics not included in the measures of step (i).

We describe the results for (i) and (ii) in the next subsections.

6.2 Measures from the Standard Individual Assessment - Personality Survey

Our measures on personality are obtained either using factor analysis (Self-Efficacy, Goal Orientation, Self-Esteem and Self-Control) or building a linear scale of the items (Cognition and Locus of Control). We use Exploratory Factor Analysis (EFA) to determine the number of latent factors. We assume that the measures present both convergent and discriminant validity, that is, there are subsets of variables highly correlated within a cluster and poorly correlated across clusters. From EFA results, we decided to retain 4 factors, which together explain 72% of the variance in the data. We apply an oblique quartimin rotation to the resulting loading matrix to have a clear interpretation of each factor. In Table 1, we present the items with the highest loadings in each factor retained and the respective Cronbach's Alpha [1] along with the interpretation we gave to each of them.

[1] The Cronbach's Alpha is a widely used measure of the internal consistency reliability of a set of measures.

Table 1 – Personality Factors: Description and Internal Consistency

Goal Orientation	*Self-Efficacy*
New ideas and plans distract me from previous ones.	If I know some problem can happen, I can think of something to avoid it.
I set life goals but later I choose a different one.	I can solve most of my problems if I put effort in it.
I have difficulties to stay focused on projects that take more than a few months to finish.	I can remain calm under difficult situations because I believe in my abilities.
Cronbach's Alpha (All): 0.86	Cronbach's Alpha (All): 0.86
Self-Worth	*Self-Control*
Sometimes I feel useless.	Sometimes I can't stop doing something even knowing it's wrong.
Sometimes I feel worthless.	Sometimes I act without thinking of the consequences or alternatives.
I think I am a loser.	I do things that are fun even if they are not good for me.
Cronbach's Alpha (All): 0.84	Cronbach's Alpha (All): 0.74

Notes: The sample of questions in this table intended to capture a student's level of grit, self-esteem, self-efficacy, peer relations and impulsiveness. Given the high correlation within these facets, we applied Factor Analysis to recover the latent measures explaining the variance across the observed measures. We choose the Maximum Likelihood method combined with an oblique Quartimin rotation. The scree test indicated 4 factors to be retained which we interpreted as Goal Orientation, Self-Efficacy, Self-Control and Self-Worth. The table display the items with the highest loadings in each factor and the corresponding Cronbach's Alpha for the complete set of items.

(i) *Self-Control:* the ability to control impulsive feelings - factor score;

(ii) *Self-Worth:* degree in which the individual considers himself a person of value - factor score;

(iii) *Self-Efficacy:* degree in which the individual believes in her capacity to face and resolve problems - factor score;

(iv) *Goal Orientation:* the ability of the individual to target and pursue long term goals - factor score;

(v) *Locus of Control:* the degree in which an individual believes he can control what happens in his life, the higher the score the more intern is his locus of control - linear scale.

Items representing each construct are described in Table 1. Locus of Control is calculated using a linear scale of the items. We detail information on each measure in the following subsections. Figures 4-5 show the density estimates for each of the variables on cognitive and noncognitive skills.

Cognitive and Personality Skills

We use a measure of fluid intelligence as a proxy for Cognitive Skills. This measure is calculated using a linear scale of 6 cognitive related items[2] that were administered to the students in 2011 as part of the survey on personality. Figure 4 shows the density estimate of cognitive skills for boys and girls. The scale was normalized to have mean zero and standard deviation one. Distribution comparisons indicate that Boys and Girls differ in the distribution of cognitive abilities - boys score slightly better on the tails while girls score better in the middle of the distribution - although we cannot reject the hypothesis of mean equality. The evidence of no gender differences in the mean non-verbal reasoning ability up to the age of 14, , measured by standard progressive matrices, is supported by a metanalysis on 57 studies (Lynn & Irwing (2004)).

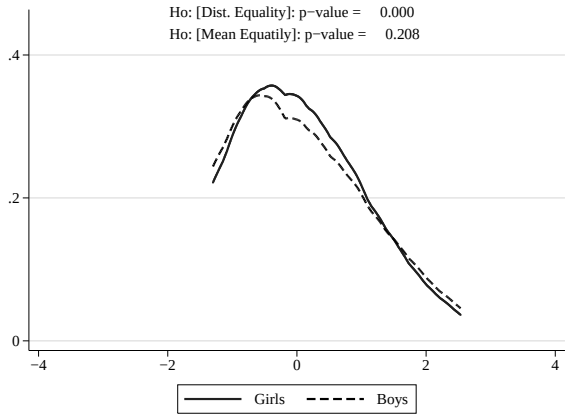


Figure 4 – Density Estimate of Cognitive Ability, by gender

Note: This figure display the Cognitive Distribution between boys and girls. We use a t-test for the mean difference and a kolmogorov-smirnov test for the difference in the distribution.

Regarding the measures on socioemotional skills, the densities plots in Figure 5 show the gender differences in the distribution of the variables. Girls scored significantly higher in all measures, except in self-efficacy which no significant difference was found at a 5% significant level[3].

[2] The selection of the items was inspired by the Raven's Matrices, which is considered a measure of non-verbal reasoning ability. The same test was also used in the survey instruments of the *Mexican Family Life Survey*

[3] The evidence in the literature is diverse. See Huang (2013), Duckworth & Seligman (2006) and Cooper, Burger & Good (1981) for evidences on gender differences in socioemotional skills.

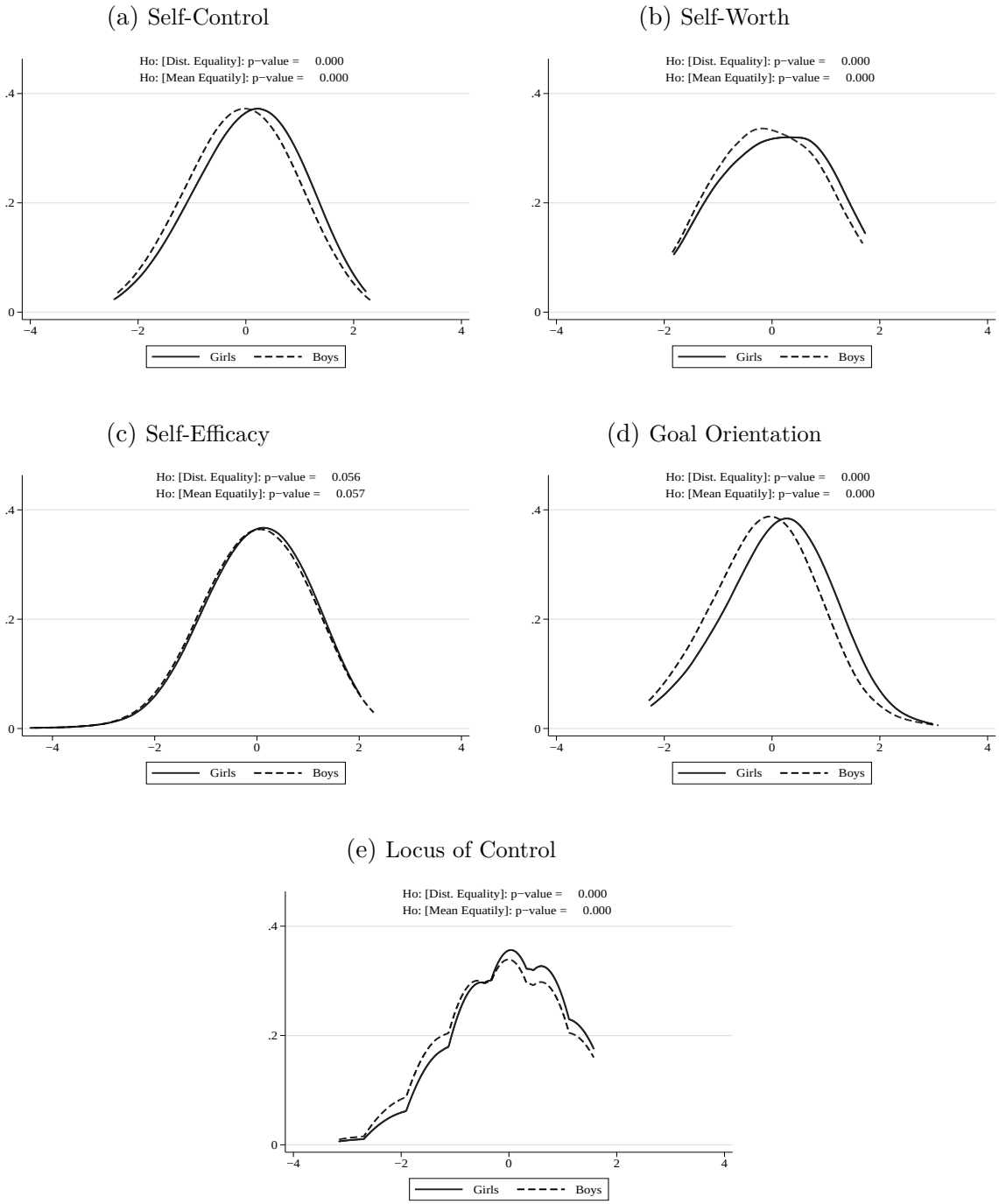


Figure 5 – Density Estimates of Personality Measures, by gender

Note: These graphs display Kernel Density Estimates. P-Values are reported for the Kolmogorov-Smirnov Test for Equality of Distributions and the t-test for Equality of Means. All variables are normalized to have mean 0 and standard deviation 1.

Incentives: Aspirations for Schooling and Ease to Accumulate Education

"Ease to Accumulate Education" is a factor score derived from a set of items included in the Personality Survey that aimed to capture perceived incentives to accumulate education from the student's perspective. Using factor analysis, we decided to retain two of the factors underlying the set of measures which we interpret as Ease to Accumulate Education. Each of these measures are related to: (i) Resources: the student does not see costs of materials or low returns to schooling as an impediment to keep studying, (ii) Peers: the student does not see peer influence as an impediment to keep studying. Figure 6 displays the density estimate of the two components of the Ease to Accumulate Education. Girls also report greater ease in accumulating education in both components. The Cronbach's Alpha is 0.79 for the Resources component and 0.74 for the peers component.

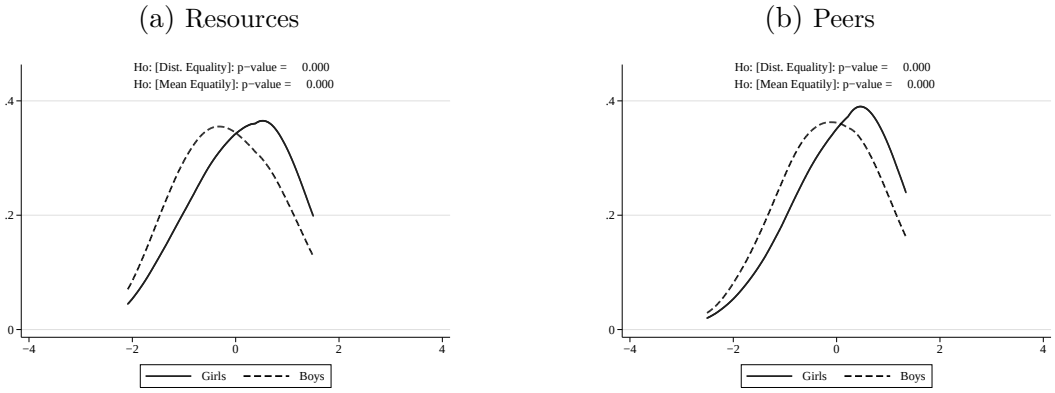


Figure 6 – Density Estimate of Ease to Accumulate Education, by gender

Note: These graphs display a Kernel density estimates. P-Values are reported for the Kolmogorov-Smirnov Test for Equality of Distributions and the t-test for Equality of Means. Variables are normalized to have mean 0 and standard deviation 1.

"Aspirations to Schooling" is a self-reported measure that indicates, if left to the will of the student, whether the student would drop out school, complete primary school, complete high school or complete college. As shown in Figure 7, girls are willing to accumulate more education than boys are, have higher schooling expectations and are more likely to aspire towards a college education.

The interpretation of educational aspirations and expectations as input in the production function of achievement is also applied by Fortin, Oreopoulos & Phipps (2015).

We extend this interpretation to the ease to accumulate education also as an input. We interpret these two variables together as incentives to schooling (or to education), which operates as an intrinsic motivation to allocate effort to excel in education. For example, the more education a person wants to achieve, the higher the motivation (and incentives) she has to perform well in school. The same follows for ease to accumulate education.

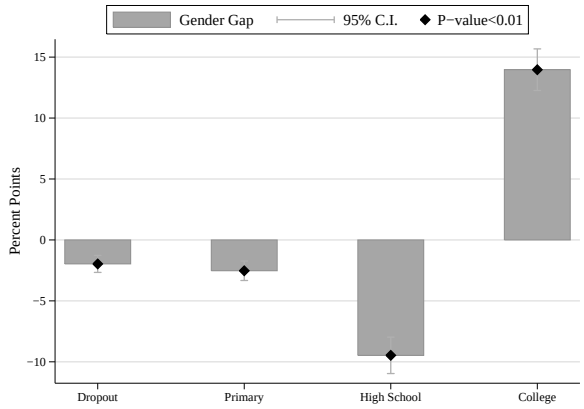


Figure 7 – Gender Gap in the Proportion of Students by Schooling Expectation

Note: This graph display the gender gap in the proportion of students choosing an expected level of education. The bars indicate the percentage points difference between the proportion of girls and boys choosing each level.

6.3 Using Administrative Records to Estimate a Latent Factor: Accumulated Skills in 2009

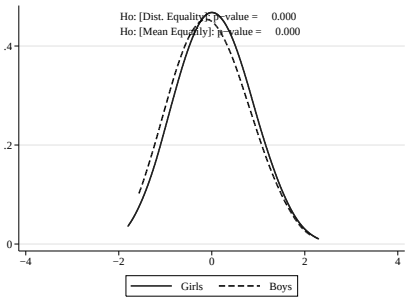
We build on the idea developed in Heckman & Kautz (2014b) to justify the use of lagged grades and tests scores as a measure of skills. This approach is based on the assumption that if a behaviour depends on skills then it is itself a valid measure of those skills (Kautz & Zanoni (2014)). The task performance framework sets the basis for this interpretation. Since grades can be interpreted as the outcome of a task, they are together explained by a common factor: the student's ability. We estimate a factor model using observed grades and tests scores in Language and Mathematics from 2009. Using EFA, we conclude that two factors are sufficient to be retained.

To guarantee the interpretation of these two factors, we use Confirmatory Factor

Analysis to estimate their distributions. First, we assume that scores on achievement tests depend only on one factor and that grades depend only in the second one. Second, we set the scale of the factor by assuming that the loading of the first factor is one for Mathematics Test Scores and that the loading for the second factor is 1 for Language grades. Third, we also assume that the factors are independent. Under these assumptions, the first ("Standardized Test Skills") measures the students' characteristics that are captured by tests scores and the second ("Classroom Skills") captures what is measured by grades that is not already measured by test scores. For example, performance in classroom might depend on "showing up on time to classes" or the "collaborative profile" the student reveals to the teacher.

It is important to recognize that the latent factor on "Classroom Skills" can include not only a student's ability but also a component capturing a common perception about the student of all three teachers of each subject we use to compute this measure: Mathematics, Language and Physical Education. That is, it comprehends a large set of abilities perceived by the teacher as necessary to perform well in the classroom as well as any *common* stereotyping pattern they might include in their gradings.

(a) Standardized Test Skills in 2009



(b) Classroom Skills in 2009

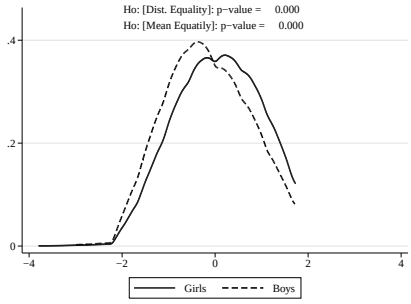


Figure 8 – Density Estimates of Accumulated Skills in 2009, by gender

Note: This figure displays the Distributions of Accumulated Skills in 2009. We use a t-test for the mean difference and a Kolmogorov-smirnov test for the difference in the distributions.

6.4 Additional Measures

Additionally to the variables described above, we control for a set of other observable variables capturing the parenting characteristics and socioeconomic background. "Parenting" is measured by a set of variables capturing the parental involvement in a child's education (see Table 2). Measures are both from administrative records and from a survey voluntarily answered by parents of the students who took the SARESP exam in 2011. It includes: years spent in private schools, kindergarten attendance and daily activities regarding school (if parents help kids with homework, if they talk to kids about school, if they talk to teacher, go to parent-teacher conferences, other activities or none). Girls were more likely to have attended preschool than boys. Parents of girls attend more school meetings, talk more to their kids and help them more with homework.

Table 2 – Summary of Variables, by gender

	Mean Girls	Mean Boys	Diff.	P-Value
Years in Priv. School	0.039	0.045	-0.006	0.344
Preschool	0.681	0.591	0.090	0.000
Parent Involvement in Kid's Education:				
School Meeting	0.632	0.544	0.088	0.000
Talk to Kid	0.402	0.357	0.045	0.000
Talk to Teacher	0.289	0.283	0.005	0.548
Help with Homework	0.268	0.222	0.046	0.000
Other	0.063	0.044	0.019	0.000
Do not participate	0.016	0.015	0.001	0.644

Note: Years in Private School is a discrete variable indicating the number of years the student spent in a private institution from 2006 to 2011. This information is obtained from administrative records. All other variables are available for the kids who took the Saresp Exam whose parents answered the survey. P-Values are from a t-test for mean equality.

Our database also contains additional measures on Socioeconomic Background (Mother's Education, Income, Family Structure and occupation status of the male and female primary caretaker), Classroom Characteristics (mean ability level and ease to accumulate education of the peers[4], share of girls, class size) and Teachers Characteristics (teacher's education, race and whether the student was assigned to the same teacher in

[4] This was calculated after the factor scores estimated by individuals. We then estimate variables for each classroom using a simple average.

the previous year). Information on family background is obtained from the SARESP Survey. We also add as controls whether the student is on track with grade level and the student's race. This information is obtained from administrative records. See Table 3 for a summary of the variables. The differences we notice in the proportion of boys and girls by socioeconomic background, specially in lower levels of socioeconomic status, could be explained by a survival bias against boys. It means boys are more likely to dropout or to have to repeat grades, and the ones we observed are more likely to be those who neither drop out nor repeat a grade.

Table 3 – Proportion of Boys and Girls by Socioeconomic Status

	Mean				Mean		
	Boys	Girls	P-Value		Boys	Girls	P-Value
Student:				*Mother's Education*			
White	0.357	0.335	0.015	Primary	0.394	0.454	0.000
Black	0.095	0.122	0.000	HS	0.214	0.246	0.000
Brown	0.538	0.532	0.511	College	0.037	0.041	0.390
Other Race	0.010	0.011	0.411				
Family Structure				*Income Levels*			
Intact Family	0.647	0.670	0.012	1	0.187	0.216	0.000
Single Mom	0.127	0.126	0.823	2	0.174	0.202	0.000
				3	0.121	0.126	0.439
				4	0.039	0.039	0.855
				5	0.007	0.010	0.103

Note: This table shows the proportion of Boys and Girls by Student's characteristics, Mother's Education, Family Structure and Income Level. Primary includes all levels below primary school graduation. High School includes incomplete High School and College includes incomplete college. Income Levels are as follows: 1 = Less than R$850.00, 2 = From R$850.00 to R$ 1,275.00, 3 = from R$ 1,275.00 to R$2,125.00, 4 = from R$2,125.00 to R$4,250.00, 5 = More than R$4,250.00. Information was collected in 2011. P-Values are calculated from a t-test for mean equality.

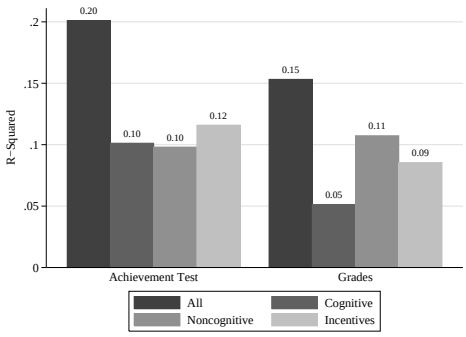
Overall, the description of the gender gap developed in this section is unique in Brazil. With this, we established a potential reason why boys and girls perform differently in classroom and standardized exams.

7 Results

7.1 Baseline Results

As pointed out in Section 4, standardized exams and grades are both outcomes from a combination of skills, incentives and other inputs. Figure 9 shows evidence on how both measures are explained by cognitive and personality traits, but in different extensions[1]. The variance of the two exams is decomposed into Cognitive, Non-Cognitive Skills and Incentives to Schooling, showing how incentives are as relevant as the other two set of skills to explain the variance.

(a) Language



(b) Mathematics

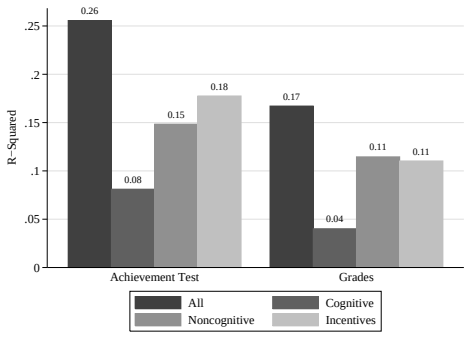


Figure 9 – Decomposing the Explained Variance in a Student Assessment into Skills and Incentives to Schooling

Note: The figure shows the r-squared from the regression of each type of test on Noncognitive Skills, Fluid Intelligence, Incentives and all together. We use the Mathematics and Language tests of the Standardized Test administered in 2011 and Final Mathematics and Language grades assigned by teachers in 2011.

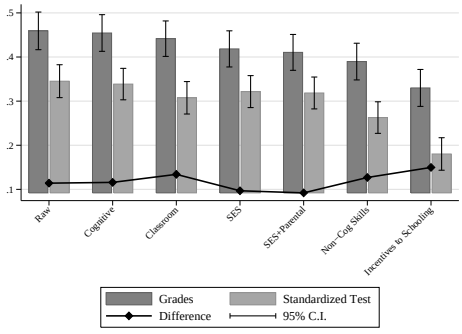
Before displaying the main results for the oaxaca decomposition, we explore the explanatory power of each set of variables used to explain the gender gap in educational achievement. In Figure 10, we estimate variations of the Equation 5.1 by adding each set of covariates at a time. We estimate 7 different equations for each type of assessment (grades and test scores), both for Language and Math. In both Figures, we also report the difference between the two types of assessments. All variables are standardized to have

[1] This exercise is the same as in Borghans et al. (2011). Without adding measures on incentives to schooling, the results are very aligned with the ones found in that study.

mean 0 and standard deviation 1 and they were residualized in relation to socioeconomic background and parenting information.

This exercise shows that the gender gap is more sensitive to noncognitive skills and incentives to schooling than to cognitive skills, classroom variables and socioeconomic and parenting background. Different from what is usually found in the literature for racial gaps, for example, girls and boys differ relatively less in these socioeconomic and school characteristics than blacks and whites do. On the other hand, as shows in Section 6, boys and girls present significantly differences in socioemotional skills and mainly in the incentives to schooling (educational aspirations and ease to accumulate education). These results are well aligned with the ones found by Cornwell, Mustard & Parys (2013) and Fortin, Oreopoulos & Phipps (2015).

(a) Language



(b) Mathematics

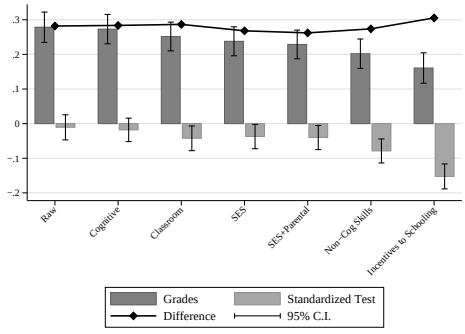


Figure 10 – Gap in Achievement Measured by Grades and Standardized Exams, by set of controls

Note: This figure displays the gender gap in Classroom Grades and Standardized Exams as well as in the difference between the two types of assessment. See Table 6 for a detailed description of the set of variables used in this exercise.

At this point, we can summarize two relevant informations inferred: [1] Although they are supposed to cover the same set of abilities and educational curriculum, teacher grades are not aligned with test scores for both subject; [2] The gender gap in both types of assessments is more sensitive to noncognitive skills and incentives to schooling than it is to socioeconomic status (SES) or classroom characteristics. Both findings are aligned with the literature (See Cornwell, Mustard & Parys (2013) and Duckworth & Seligman (2006)).

We now follow to estimate the full specification and to point the relative importance of skills and incentives to explain the gender gap in grades and standardized exams. In Column 2 and 5 of Table 4, we display estimates of Equation 5.1 for Mathematics grades and Standardized Test Scores, respectively. Column 1 displays the gender gap[2] after we account for socioeconomic background and parenting information.

Table 4 – Regression results for Mathematics grades and Test Scores

	Mathematics Grades in 2011			**Mathematics Test Scores in 2011**		
Dependent	(1)	(2)	(3)	(4)	(5)	(6)
Girl	0.218***	0.113***	0.0906***	-0.0504***	-0.176***	-0.161***
	(0.0218)	(0.0205)	(0.0254)	(0.0177)	(0.0171)	(0.0200)
Fluid Intelligence		0.133***	0.0726***		0.210***	0.104***
		(0.0103)	(0.0136)		(0.00914)	(0.0104)
Goal Orientation		0.0663***	0.0445***		0.0524***	0.0249
		(0.0136)	(0.0170)		(0.0138)	(0.0158)
Self-Efficacy		0.0932***	0.0435***		0.0528***	-0.0148
		(0.0121)	(0.0162)		(0.0120)	(0.0151)
Self-Worth		0.0360***	0.0215		0.0695***	0.0105
		(0.0108)	(0.0137)		(0.0107)	(0.0123)
Self-Control		0.117***	0.0789***		0.0187	0.0156
		(0.0131)	(0.0171)		(0.0127)	(0.0140)
Locus of Control		0.0693***	0.0472***		0.0911***	0.0416***
		(0.00967)	(0.0127)		(0.00921)	(0.0106)
Educational Expectations		0.124***	0.106***		0.0999***	0.0649***
		(0.0123)	(0.0173)		(0.0126)	(0.0152)
Ease to Acc Educ (Peers)		-0.00821	-0.0413*		0.0111	-0.0154
		(0.0186)	(0.0234)		(0.0180)	(0.0213)
Ease to Acc Educ (Goods)		0.102***	0.0454**		0.177***	0.0738***
		(0.0173)	(0.0222)		(0.0179)	(0.0222)
Standardized Test Skills 2009			0.338***			0.733***
			(0.0355)			(0.0302)
Classroom Skills 2009			0.232***			0.0655***
			(0.0247)			(0.0215)
Teacher/Classroom Controls	No	Yes	Yes	No	Yes	Yes
Effort	No	Yes	Yes	No	Yes	Yes
Constant	-0.110***	-0.0749	0.155	0.0226	-0.123	0.0383
	(0.0228)	(0.155)	(0.196)	(0.0156)	(0.0998)	(0.114)
Observations	10,379	10,082	5,262	10,379	10,082	5,262
R-squared	0.013	0.126	0.265	0.001	0.181	0.428

Note: This table shows results for the regressions for the measures of achievement. All variables are residualized in relation to socioeconomic background and parenting. Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

All measures of skills and incentives are statistically significant but they correlates with the outcomes differently, as we assumed in the theoretical framework. Fluid Intelligence has a higher correlation with test scores (0.210 sd) than with grades (0.133 sd). On the

[2] As described before, this is not the raw gap since the variables are residualized in relation to socioeconomic variables.

other hand, grades are more correlated with socioemotional measures and incentives than tests scores are. This finding is consistently supported by the literature on personality and education (Borghans et al. (2011)). More thoroughly, Self-control predicts mathematics grades in magnitude (0.117 sd) compared with fluid intelligence. Interestingly, we find no significant correlation between self-control and test scores. Educational Expectations outdoes every socioemotional variable in predicting mathematics grades and test scores, being more important to the first. In addition, the measure of "Ease to Accumulate Education - Goods"[3] is as important to explain grades as self-efficacy and self-control and it is the second most important variable explaining test scores, after cognition.

The gender coefficient is our measure of unexplained gap, which remains significant for both subjects although in opposite directions. Girls consistently outperform boys in classroom grades but not in standardized exams. Compared to Column 1 in Table4, controlling for skills and incentives reduces the gap for grades but it increases the gap in standardized exams. In order to control for potential unobserved variables biasing the coefficient, we control for accumulated skills in 2009 and results are reported in Columns 3 and 6 of Table 4[4]. The two variables captures additional abilities that are important to perform in both types of assessment. As it can be noted when comparing to Columns 2 and 3, all coefficients loose a fraction of their predictive power when we include these variables but they remain significant in most cases.

In Table 5, we estimate Equation 5.1 for Language Grades and Test scores. For this subject, "Ease to Accumulate Education (Goods)" outdoes all other measures in predicting the outcomes, for both grades and test scores. It is followed by Educational Expectations in predictive power for grades but not for the standardized test. IQ continues to be strongly correlated with test scores. As in Mathematics, Self-control outdoes IQ in predicting Language grades and it is not significantly correlated with test scores. In

[3] which indicates whether resources are not an obstacle to pursue the student's Expected Level of Schooling.

[4] Note that the sample size is reduced in half after the inclusion of the lagged variable. We tested whether the sample differ in observables and we conclude that the samples are comparable. See Appendix A for the comparison results.

relation to the gender coefficient, the gaps are in favour of girls for both grades and tests scores and they are larger than the ones found in Mathematics. To control for further unobservable variables, we also add the two measures of accumulated skills in 2009 which reduces both gaps, but also do not eliminate them.

Table 5 – Regression results for Language grades and Test Scores

	Language Grades in 2011			**Language Test Scores in 2011**		
Dependent	(1)	(2)	(3)	(4)	(5)	(6)
Girl	0.391***	0.288***	0.255***	0.301***	0.140***	0.122***
	(0.0214)	(0.0201)	(0.0242)	(0.0186)	(0.0176)	(0.0209)
Fluid Intelligence		0.107***	0.0436***		0.168***	0.0683***
		(0.00974)	(0.0119)		(0.00941)	(0.0118)
Goal Orientation		0.0458***	0.0453**		0.0929***	0.0607***
		(0.0157)	(0.0186)		(0.0133)	(0.0158)
Self-Efficacy		0.107***	0.0655***		0.116***	0.0551***
		(0.0118)	(0.0160)		(0.0115)	(0.0135)
Self-Worth		0.0506***	0.0309**		0.0928***	0.0561***
		(0.0103)	(0.0133)		(0.0101)	(0.0119)
Self-Control		0.116***	0.0888***		0.0193	-0.0144
		(0.0126)	(0.0162)		(0.0122)	(0.0132)
Locus of Control		0.0750***	0.0556***		0.103***	0.0640***
		(0.00912)	(0.0113)		(0.00868)	(0.0101)
Educational Expectations		0.117***	0.0718***		0.0969***	0.0698***
		(0.0127)	(0.0155)		(0.0123)	(0.0165)
Ease to Acc Educ (Peers)		0.00453	-0.0214		0.0733***	0.0386*
		(0.0173)	(0.0220)		(0.0178)	(0.0205)
Ease do Acc Educ (Goods)		0.122***	0.0709***		0.173***	0.0774***
		(0.0158)	(0.0213)		(0.0164)	(0.0199)
Stardardized Test Skills 2009			0.272***			0.737***
			(0.0340)			(0.0316)
Classroom Skills 2009			0.212***			0.0435*
			(0.0262)			(0.0222)
Teacher/Classroom Controls	No	Yes	Yes	No	Yes	Yes
Effort	No	Yes	Yes	No	Yes	Yes
Constant	-0.196***	0.165	0.124	-0.154***	-0.314***	-0.148
	(0.0237)	(0.157)	(0.206)	(0.0162)	(0.0999)	(0.109)
Observations	10,379	10,082	5,262	10,379	10,082	5,262
R-squared	0.042	0.161	0.261	0.027	0.238	0.463

Note: This table shows results for the regressions for the measures of achievement. All variables are residualized in relation to socioeconomic background and parenting. Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

From this second block of estimates, we can summarize our additional findings: [3] Grades and standardized exams both depend on cognition, socioemotional skills and incentives but in different extensions. Cognitive abilities are more important to explain tests scores than grades while socioemotional skills predict grades in a proportion larger then test scores, with exception of Self-Worth and Locus of Control; [4] Educational

Expectations and Ease to Accumulate Education predict test scores in Mathematics and Language in a higher magnitude than every socioemotional skills; [5] Controlling for cognition, socioemotional skills and incentives reduce in more than half the unexplained gap in Mathematics and Language grades but does not eliminate it. Concerning Mathematics tests score, controlling for these variables actually increases the gender gap in three times in favour of boys. For language Grades, the inclusion of controls has a more modest impact in reducing the unexplained gap than for Mathematics.

7.2 Decomposition Results

In the results showed in Tables 4 and 5, the inclusion of socioemotional and incentives variables reduces the unexplained gap for all outcomes but Mathematics Tests Scores. To understand the relative importance of these variables in explaining the gender gap in both types of exam, we decompose the gender gap in both subjects. We will only decompose Mathematics Grades, Language Grades and Test Scores because they present a significant raw gap (Figure 10). Thus, we decompose the results estimated in Columns 2 and 4 from Tables 4 and 5 by estimating Equation 5.3. We also decompose the estimates in Columns 3 and 6 which include the accumulated skills to the decomposition. We test the aggregated effect of the variables in the following sets characteristics: (1) Skills; (2) Incentives; (3) Classroom and Teacher characteristics; (4) Accumulated Skills in 2009. See Table 6 for a description of each variable included in our aggregated estimates. All variables are standardized to have mean 0 and standard deviation 1 and they are residualized in relation to socioeconomic background and parenting information.

Table 6 – Description of Variables: Decomposition Estimates

	Included Variables	*Data Source*
Cognitive Skills	Score on a 6-items test on fluid intelligence	Personality Survey
Noncognitive Skills	Goal Orientation Self-Efficacy Self-Control Self-Worth Study Time (effort)[5]	Personality Survey
Accumulated Skills in 2009	Factor scores for Standardized Tests Skills and Classroom Skills	Administrative Records and Saresp
Incentives to Schooling	Educational Expectations Ease to Accum. Education (Peers) Ease to Accu. Education (Resources)	Personality Survey
Classroom Effects	Teacher Characteristics Mean Level of Skills in the classroom Class size Share of Girls	Administrative Records Personality Survey

In the first estimate, Figure 11, we decompose the outcomes into Skills, Incentives and Classroom Characteristics. Results show the relative importance of each observed input in explaining the gender gap in classroom grades, for both Math (46%) and Language Grades (24%) and standardized test scores (50%). Incentives to schooling explains the largest proportion of the gaps. For Language Test, this component accounts for 29%.

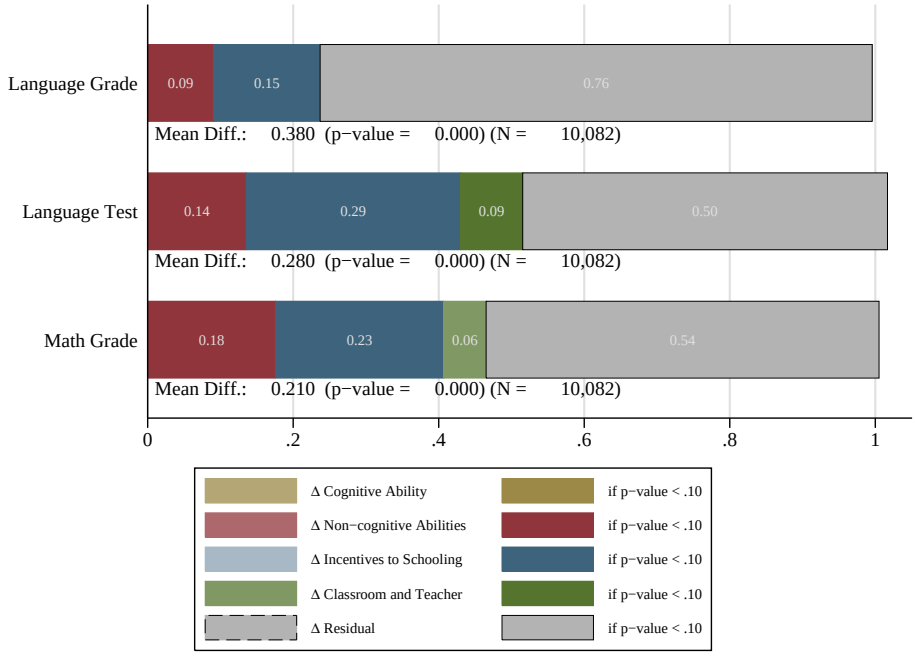


Figure 11 – Cognitive and Personality Skills, Educational Expectations, Ease to Accumulate Education and Parental Investments as Mediators of the Gender Gap in Educational Achievement

Note: We display results for Teacher-Assigned Grades and the Difference between Grades and Standardized Test in Math. Factors, Grades and Scores are normalized to have mean zero and standard deviation one.

Table 7 displays detailed results for these decompositions. Notice that, although Cognitive Skills have one of the highest predictive power on all outcomes, since girls and boys do not have a significant mean difference in this variable, it does not represent a significant share in the decomposition in any of the specifications. Within the set of Non-Cognitive Abilities, Goal Orientation accounts for most of the effect in the three outcomes (2.45%, 7.06% and 6.72% for Language Grades, Language Scores and Math Grades respectively). Self-control has a significant effect for Language (3.21%) and Math Grades(5.83%), but not for Language Scores. The importance of Self-Control to explain gender differences in classroom assessments relates to Duckworth & Seligman (2006)[6]. As in Fortin, Oreopoulos & Phipps (2015), Educational Expectations explain the gender gap in all outcomes in a magnitude comparable or higher than the measures on socioemotional skills.

[6] Our measures are directly correlated to theirs (grit and self-control)

The new evidence we find relates to the importance of Ease to Accumulate Education which outdoes every other characteristics for all grades. This is a variable that captures, for example, the necessity of the individual to work to complement the family's income or the perception of a low return to schooling. Boys perceive a higher value in activities outside the school, either for their benefits or for family need. For Language Test and Mathematics Grades, it explains 16.86% and 13.23%, more than twice the share explained by Goal Orientation in both outcomes. For Language Grades, it explains 8.73%, which is four times bigger than that explained by Goal Orientation. The sub-score in Goods corresponds to roughly all the effects of this component, indicating that girls perceive less costs to pursue further education than boys do.

We then add the two measures of Accumulated Level of Skills to the decomposition performed in Figure 11 to account for unobservable characteristics that are important to performance in both kind of task. As it can be seen in Figure 12, the gender gap in abilities in 2009 explains 32% of the difference in Math in 2011, but the previous gap is less important for Language. Accounting for these measures increases our explained part of the gap to 33%, 55% and 59% for Language grades, Language test scores and Mathematics grades, respectively. In Table 7, we also displays detailed results of the decomposition after we include these two variables. Notice that some of the effects in other variables are lower, indicating that the inclusion of the accumulated skills netted out some of their effects, although none of them looses significance.

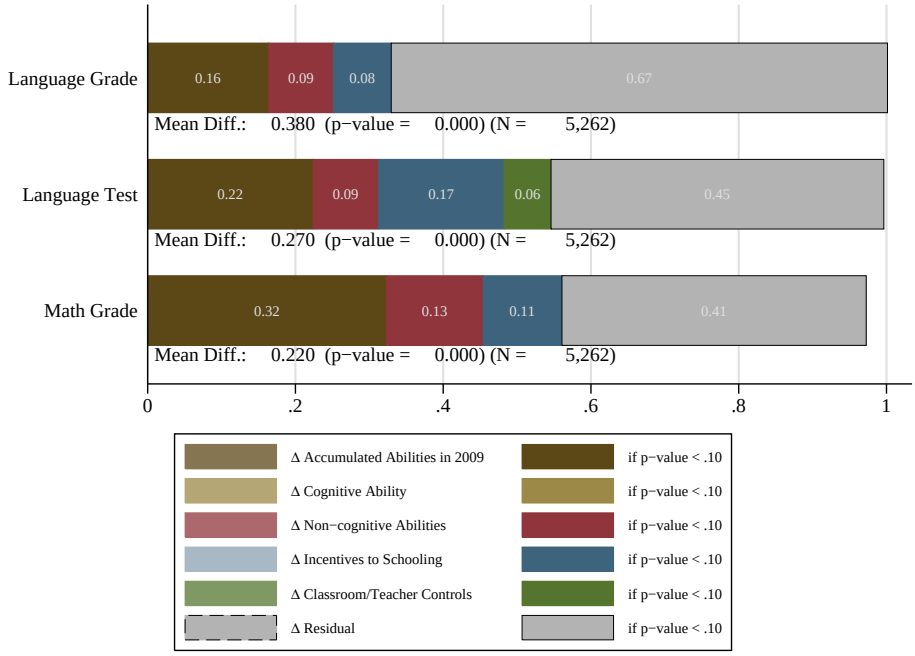


Figure 12 – Accumulated Skills in $(t-1)$, Cognitive and Personality Skills, Educational Expectations, Ease to Accumulate Education and Parental Investments as Mediators of the Gender Gap in Educational Achievement

Note: We display results for Teacher-Assigned Grades and the Difference between Grades and Standardized Test in Math. Factors, Grades and Scores are normalized to have mean zero and standard deviation one.

Table 7 – Detailed Decomposition of the Gender Gap in Math and Language, for grades and standardized test scores

	Lang.: Grades				**Lang.: Tests**				**Math: Grades**			
	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
Gender Gap (s.d.)	0.38	0.00	0.38	0.00	0.28	0.00	0.27	0.00	0.21	0.00	0.22	0.00
Standardized Test Skills	.	.	5.02	0.00	.	.	19.12	0.00	.	.	10.78	0.00
Classroom Skills	.	.	11.35	0.00	.	.	3.28	0.03	.	.	21.53	0.00
Fluid Intelligence	-0.09	0.87	0.03	0.93	-0.20	0.87	0.06	0.93	-0.21	0.87	0.08	0.93
Goal Orientation	2.56	0.00	2.32	0.02	7.06	0.00	4.38	0.00	6.72	0.00	3.94	0.02
Self-Efficacy	0.53	0.31	0.67	0.15	0.78	0.31	0.80	0.15	0.83	0.31	0.77	0.19
Self-Worth	0.84	0.01	0.59	0.10	2.10	0.00	1.51	0.02	1.09	0.02	0.71	0.21
Self-Control	3.21	0.00	2.85	0.00	0.73	0.13	-0.65	0.32	5.84	0.00	4.37	0.00
Locus of Control	2.03	0.00	2.06	0.00	3.78	0.00	3.34	0.00	3.39	0.00	3.02	0.00
Educational Expectations	5.62	0.00	3.58	0.00	6.33	0.00	4.90	0.00	10.81	0.00	9.09	0.00
Ease to Acc. Edu (Peers)	0.29	0.82	-1.38	0.39	6.28	0.00	3.49	0.06	-0.94	0.68	-4.59	0.10
Ease to Acc Educ (Goods)	8.73	0.00	5.60	0.00	16.86	0.00	8.61	0.00	13.23	0.00	6.19	0.07
Teacher Characteristics	-0.10	0.53	-0.05	0.82	0.19	0.28	0.13	0.54	-0.46	0.12	-1.46	0.09
Classroom Characteristics	0.72	0.58	-0.02	0.99	8.34	0.00	6.23	0.00	6.31	0.01	2.82	0.37
Unexplained Gap	75.86	0.00	67.17	0.00	50.13	0.00	45.05	0.00	54.00	0.00	41.18	0.00
Observations	10,082		5,262		10,082		5,262		10,082		5,262	

Note: This table shows results for the decomposition exercises displayed in the Figures 11 and 12. All variables are residualized in relation to socioeconomic background and parenting. P-values are for the test for significance of the coefficients.

We can now add additional findings to our summary: [6] Cognitive Skills have an insignificant effect on all outcomes, since girls and boys do not have a significant mean difference in cognition. [7] Goal Orientation is the most important sociomotional skills to explain the gender differences in all outcomes. [8] The new evidence is found in relation to the importance of Ease to Accumulate Education which outdoes the effect of every other variable for all grades. This means that Girls have a lower level of perceived costs to pursue further education than boys do.

7.3 The gender gap among low and top achievers

In the exercise described above, we focused on the mean results. What about the students on the tails of the distributions? Does the gender gap follow the same pattern? To address this question, in Figure 13 we estimate the gender gap in the raw measures of achievement for the 25% highest and 25% lowest performers. Comparing to the gaps in the whole sample (Figure 1), the gaps differs in a great amount for classroom grades. Among top achievers, boys and girls are not graded differently by the teacher in Mathematics while the difference increases for Language. For low achievers, however, teachers grade girls better than boys but the gap is significantly smaller than in the whole sample. For test scores, the results are not strongly different from the whole sample, although the gender gap in mathematics test score for top achievers turns significantly in favour of boys.

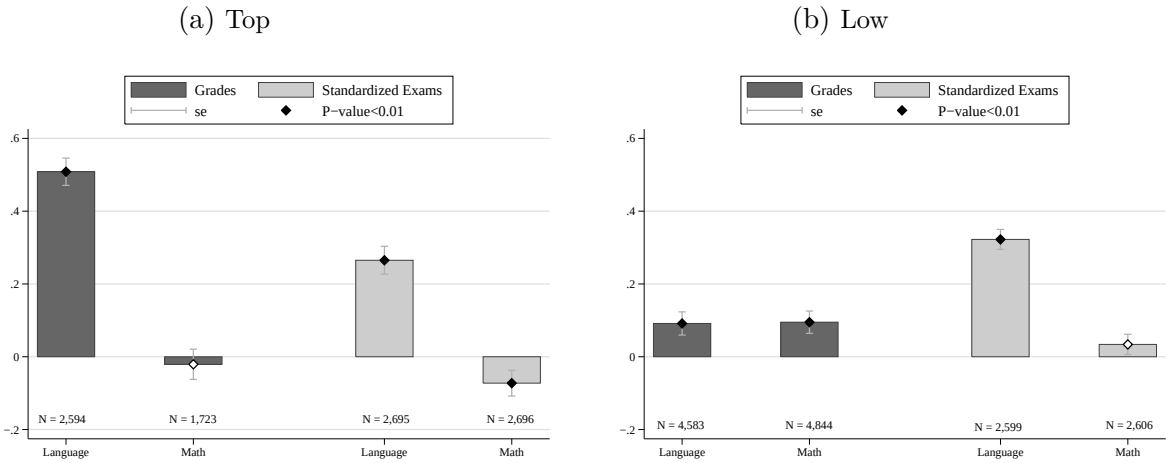


Figure 13 – Achievement Gender Gap in Language and Math: Grades and Standardized Exams, for top and low performers

Note: This figure displays the gender gap in both Classroom Grades and Standardized Test Scores for Language and Mathematics. In Figure (a) We compare the top 25% of boys in each type of achievement male distribution with the top 25% of girls in the female distributions. Sample (a): [Math Grades: Girls = 1,026, Boys = 697; Math Test Scores: Girls = 1,362, Boys = 1,334; Language Grades: Girls = 1,261, Boys = 1,333; Language Test Scores: Girls = 1,362, Boys = 1,333]. In Figure (b) We compare the bottom 25% of boys in each type of achievement male distribution with the bottom 25% of girls in the female distribution. Sample (b): [Math Grades: Girls = 2,106, Boys = 2,738; Math Test Scores: Girls = 1,320, Boys = 1,286; Language Grades: Girls = 1,772, Boys = 2,811; Language Test Scores: Girls = 1,312, Boys = 1,287].

How sensitive are those gaps in relation to the addition of individual controls? In Tables 8 and 9, we estimate Equation 5.1 restricting the sample to top and low achievers in each subject. For both top and low achievers, the gender gap in mathematics is fully explained by observed measures while for test scores the difference rises as significant and in favour of boys for both quartiles. The same does not hold for Language assessments. Among low achievers, the difference in Language Grades is close to zero but still significant, while the gap among top achievers remains closely insensitive to the inclusion of the students characteristics. For Test Scores, the gender gap in favour of girls for both quartile of achievement remains mostly unexplained.

Table 8 – Regression Results for Mathematics Grades and Test Scores: Top and Low performers

	Mathematics Grades			**Mathematics Test Scores**		
Dependent	(1)	(2)	(3)	(4)	(5)	(6)
Top 25% Performers						
Girl	-0.0510**	-0.0638**	-0.0417	-0.0935***	-0.102***	-0.111***
	(0.0252)	(0.0252)	(0.0333)	(0.0199)	(0.0207)	(0.0261)
Low 25% Performers						
Girl	0.0416**	0.0218	0.0161	-0.0143	-0.0529***	-0.0697***
	(0.0183)	(0.0183)	(0.0227)	(0.0147)	(0.0165)	(0.0237)
Accumulated Abilities in 2009	No	No	Yes	No	No	Yes
Cognitive Skills	No	Yes	Yes	No	Yes	Yes
Non-cognitive Skills	No	Yes	Yes	No	Yes	Yes
Incentives to Schooling	No	Yes	Yes	No	Yes	Yes
Classroom/Teacher Controls	No	Yes	Yes	No	Yes	Yes
Constant	1.514***	1.560***	1.439***	1.147***	1.169***	1.009***
	(0.0223)	(0.130)	(0.165)	(0.0159)	(0.0853)	(0.108)
Observations	1,723	1,704	929	2,695	2,656	1,391
R-squared	0.003	0.075	0.125	0.009	0.084	0.218

Note: This table shows results for the regressions for achievement measures by top and bottom quartiles of the distributions. All variables are residualized in relation to socioeconomic background and parenting. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9 – Regression Results for Language Grades and Test Scores: Top and Low performers

	Language Grades			**Language Test Scores**		
Dependent	(1)	(2)	(3)	(4)	(5)	(6)
Top 25% Performers						
Girl	0.464***	0.429***	0.433***	0.241***	0.205***	0.228***
	(0.0223)	(0.0229)	(0.0329)	(0.0199)	(0.0206)	(0.0287)
Low 25% Performers						
Girl	0.0512**	0.0304	0.0468**	0.288***	0.260***	0.250***
	(0.0202)	(0.0196)	(0.0236)	(0.0171)	(0.0179)	(0.0238)
Accumulated Abilities in 2009	No	No	Yes	No	No	Yes
Cognitive Skills	No	Yes	Yes	No	Yes	Yes
Non-cognitive Skills	No	Yes	Yes	No	Yes	Yes
Incentives to Schooling	No	Yes	Yes	No	Yes	Yes
Classroom/Teacher Controls	No	Yes	Yes	No	Yes	Yes
Constant	0.961***	1.039***	0.736***	0.991***	0.838***	0.747***
	(0.0215)	(0.133)	(0.162)	(0.0156)	(0.100)	(0.134)
Observations	2,594	2,554	1,248	2,690	2,654	1,387
R-squared	0.157	0.214	0.289	0.048	0.139	0.269

Note: This table shows results for the regressions for achievement measures by top and bottom quartiles of the distributions. All variables are residualized in relation to socioeconomic background and parenting. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

In summary, [9] Classroom gaps change substantially along the distribution. Among top achievers, the raw gender gap in mathematics is not significant. After controlling for observed characteristics, we fully explain the gender gap in Math grades, which does not hold for the whole sample. On the other hand, girls in the top quartile still benefit largely from the classroom environment, a gap that remains largely insensitive to the inclusion of controls on observables. [10] On the other hand, the gap in mathematics test scores remains significant and in favour of boys for math. Yet, the gender gap in language test scores remains closely unexplained by the inclusion of controls.

7.4 Do boys and girls have different technology of production?

Up to this point, in our main specification, we assumed that girls and boys map inputs into outcomes in the same way, that is, that they have the same technology of production for each type of assessment. Now, we relax this hypothesis to test whether girls and boys differ in their productivities. In Table 10 and 11, we present the results for Equation 7.1 and the p-value for the test for the equality of the coefficients,

$$\widetilde{T}_{mig} = \widetilde{M'}_{mig}\gamma_{mg} + \nu_{mig} \tag{7.1}$$

where m is the type of assessment and $g \in \{G, B\}$ indicates whether the equation is restricted to the sample of girls or boys.

Tables 10 and 11, for most of the inputs and for both type of assessments, we cannot reject the hypothesis of equality of the coefficients between the equations for the restricted sample of boys and girls. A few exceptions can be noted, though. Boys benefit more from self-efficacy in Mathematics grades (Block 1) - which is not significant after we control for accumulated skills in 2009. In block (2), girls benefit more from the share self-control that is not already accounted by the lagged measures. In relation to tests scores, in the model represented in the block (4), boys are more productive in the use of their fluid intelligence and locus of control - netted out the effect of accumulated skills. For Language, the results are aligned with Math, with a few differences. Girls benefit more from IQ than boys in Language grades and test scores. Also, boys have advantage on Test Scores for Self-Control and Self-efficacy. However, in general, [11] with a few exceptions, boys and girls do not significantly differ in their coefficients for both subjects and types of assessment, which corroborates our initial hypothesis of equality in the production functions.

Table 10 – Regression Results by Gender and Test of Equality of the Coefficients

	(1)			(2)			(3)			(4)		
Dependent Variables	**Math Grades**			**Math Grades**			**Math Test Scores**			**Math Test Scores**		
	Boys	Girls	*P-Value*	Boys	Girls	*P-Value*	Boys	Girls	*P-Value*	Boys	Girls	*P-Value*
Fluid Intelligence	0.125***	0.138***	.560	0.0631***	0.0861***	.412	0.225***	0.191***	.077	0.134***	0.0735***	**.009**
	(0.0134)	(0.0137)		(0.0181)	(0.0177)		(0.0123)	(0.0126)		(0.0152)	(0.0145)	
Goal Orientation	0.0909***	0.0484**	.166	0.0663**	0.0215	.251	0.0763***	0.0396**	.173	0.0371	0.0154	.490
	(0.0215)	(0.0200)		(0.0283)	(0.0255)		(0.0197)	(0.0184)		(0.0238)	(0.0210)	
Self-Efficacy	0.138***	0.0498***	**.001**	0.0795***	0.00771	.053	0.0770***	0.0360**	.094	-0.00308	-0.0241	.492
	(0.0193)	(0.0178)		(0.0264)	(0.0228)		(0.0176)	(0.0164)		(0.0222)	(0.0188)	
Self-Worth	0.0397**	0.0284*	.629	0.0302	0.0161	.662	0.0612***	0.0720***	.613	-0.00158	0.0221	.364
	(0.0169)	(0.0158)		(0.0227)	(0.0203)		(0.0154)	(0.0146)		(0.0191)	(0.0167)	
Self-Control	0.0998***	0.136***	.159	0.0323	0.122***	**.010**	0.0369**	0.0143	.332	0.0135	0.0225	.754
	(0.0194)	(0.0185)		(0.0258)	(0.0237)		(0.0178)	(0.0171)		(0.0217)	(0.0195)	
Locus of Control	0.0709***	0.0718***	.962	0.0376**	0.0589***	.413	0.104***	0.0722***	.070	0.0631***	0.0161	**.033**
	(0.0138)	(0.0138)		(0.0185)	(0.0182)		(0.0126)	(0.0127)		(0.0156)	(0.0150)	
Study Time (Effort)	-0.00600	-0.0304*	.263	-0.0148	0.0353*	.074	-0.0286**	-0.0673***	.074	-0.000394	0.00849	.727
	(0.0153)	(0.0164)		(0.0202)	(0.0210)		(0.0140)	(0.0151)		(0.0170)	(0.0173)	
Educational Expectations	0.128***	0.117***	.703	0.112***	0.0940***	.629	0.0885***	0.119***	.237	0.0513**	0.0808***	.353
	(0.0179)	(0.0218)		(0.0237)	(0.0285)		(0.0164)	(0.0201)		(0.0200)	(0.0234)	
Ease to Acc Educ (Peers)	0.00872	-0.0176	.501	-0.0217	-0.0638*	.414	-0.0225	0.0340	.118	-0.0344	-0.00340	.442
	(0.0279)	(0.0272)		(0.0361)	(0.0346)		(0.0254)	(0.0251)		(0.0304)	(0.0285)	
Ease to Acc Educ (Goods)	0.0908***	0.103***	.728	0.0438	0.0576*	.786	0.186***	0.173***	.701	0.0791***	0.0788***	.993
	(0.0270)	(0.0264)		(0.0356)	(0.0343)		(0.0247)	(0.0244)		(0.0299)	(0.0283)	
Standardized Test Skills				0.373***	0.300***	.401				0.664***	0.797***	.208
				(0.0400)	(0.0378)					(0.0337)	(0.0311)	
Classroom Skills				0.229***	0.239***	.575				0.0871***	0.0390*	.660
				(0.0307)	(0.0282)					(0.0258)	(0.0232)	
Teacher/Classroom Controls	Yes	Yes		Yes	Yes		Yes	Yes		Yes	Yes	
Constant	-0.0979	0.137		-0.106	0.528***		-0.222**	-0.165		-0.163	0.0868	
	(0.104)	(0.110)		(0.151)	(0.150)		(0.0951)	(0.102)		(0.127)	(0.124)	
Observations	5,166	5,337		2,525	2,775		5,166	5,337		2,525	2,775	
R-squared	0.126	0.110		0.270	0.256		0.201	0.165		0.421	0.445	

Note: This table shows results for the regressions for achievement measures by gender. P-Values are for the coefficient difference test between the equations for boys and girls. All variables are residualized in relation to socioeconomic background and parenting. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 11 – Regression Results by Gender and Test of Equality of the Coefficients

	(1)			(2)			(3)			(4)		
Dependent Variables	**Language Grades**			**Language Grades**			**Language Test Scores**			**Language Test Scores**		
	Boys	Girls	*P-Value*	Boys	Girls	*P-Value*	Boys	Girls	*P-Value*	Boys	Girls	*P-Value*
Fluid Intelligence	0.0818***	0.122***	**.030**	0.0387**	0.0513***	.595	0.146***	0.184***	**.025**	0.0422***	0.0942***	**.008**
	(0.0129)	(0.0137)		(0.0175)	(0.0175)		(0.0120)	(0.0122)		(0.0150)	(0.0142)	
Goal Orientation	0.0701***	0.0269	.133	0.0485*	0.0357	.720	0.0984***	0.0987***	.989	0.0331	0.0840***	.093
	(0.0206)	(0.0200)		(0.0273)	(0.0253)		(0.0193)	(0.0179)		(0.0234)	(0.0205)	
Self-Efficacy	0.146***	0.0794***	**.011**	0.0868***	0.0431*	.210	0.151***	0.0905***	**.011**	0.0589***	0.0583***	.984
	(0.0185)	(0.0178)		(0.0255)	(0.0227)		(0.0173)	(0.0159)		(0.0218)	(0.0184)	
Self-Worth	0.0435***	0.0567***	.535	0.0217	0.0400**	.527	0.0877***	0.0909***	.869	0.0684***	0.0418**	.289
	(0.0162)	(0.0158)		(0.0219)	(0.0202)		(0.0151)	(0.0141)		(0.0188)	(0.0164)	
Self-Control	0.119***	0.117***	.942	0.0914***	0.0941***	.936	0.0583***	-0.00866	**.004**	0.0109	-0.0326*	.102
	(0.0186)	(0.0185)		(0.0249)	(0.0236)		(0.0174)	(0.0165)		(0.0214)	(0.0191)	
Locus of Control	0.0593***	0.0883***	.109	0.0379**	0.0739***	.140	0.0977***	0.110***	.476	0.0515***	0.0780***	.210
	(0.0133)	(0.0138)		(0.0179)	(0.0181)		(0.0124)	(0.0123)		(0.0154)	(0.0147)	
Study Time (Effort)	-0.00220	-0.0288*	.245	0.0174	0.0267	.754	-0.0695***	-0.0796***	.607	-0.0501***	-0.0164	.153
	(0.0146)	(0.0164)		(0.0196)	(0.0209)		(0.0137)	(0.0146)		(0.0168)	(0.0169)	
Educational Expectations	0.108***	0.132***	.381	0.0551**	0.0956***	.281	0.0849***	0.109***	.352	0.0603***	0.0828***	.476
	(0.0172)	(0.0218)		(0.0229)	(0.0283)		(0.0161)	(0.0194)		(0.0197)	(0.0229)	
Ease to Acc Educ (Peers)	-0.0325	0.0408	.066	-0.0569	0.00356	.237	0.0508**	0.0834***	.336	0.00418	0.0664**	.118
	(0.0267)	(0.0272)		(0.0349)	(0.0343)		(0.0249)	(0.0243)		(0.0299)	(0.0278)	
Ease to Acc Educ (Goods)	0.145***	0.100***	.252	0.103***	0.0517	.294	0.181***	0.170***	.746	0.114***	0.0444	.065
	(0.0259)	(0.0264)		(0.0344)	(0.0341)		(0.0242)	(0.0236)		(0.0295)	(0.0276)	
Standardized Test Skills				0.254***	0.279***	.110				0.702***	0.764***	.946
				(0.0386)	(0.0375)					(0.0331)	(0.0305)	
Classroom Skills				0.197***	0.225***	.459				0.0367	0.0523**	.951
				(0.0296)	(0.0280)					(0.0254)	(0.0227)	
Constant	0.293***	0.414***		0.185	0.353**		-0.380***	-0.126		-0.229*	0.0436	
	(0.0997)	(0.110)		(0.146)	(0.149)		(0.0932)	(0.0984)		(0.125)	(0.121)	
Observations	5,166	5,337		2,525	2,775		5,166	5,337		2,525	2,775	
R-squared	0.136	0.124		0.225	0.239		0.222	0.221		0.429	0.481	

Note: This table shows results for the regressions for achievement measures by gender. P-Values are for the coefficient difference test between the equations for boys and girls. All variables are residualized in relation to socioeconomic background and parenting. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

7.5 The Gender Gap across Exams

Another outcome of our interest is the puzzling significant and constant difference between boys and girls across types of assessments and subjects. A similar exercise was implemented by Lavy (2008) in which they interpret the difference across exams as a potential bias in favour of girls. In their paper, they test several hypothesis on why this differences emerges in every subject, that could arise for reasons ranging from differences in the students' behaviour to teacher discrimination against boys.

In Figure 10, the black line represents a variable that is the difference between the scores in both types of assessment. The positive and significant raw gap is measured by the following equation:

$$\widetilde{T}_{1,i} - \widetilde{T}_{0,i} = \alpha G_i + \eta_i \tag{7.2}$$

Note that this Equation is equivalent to a differences-in-difference setting in that our covariates have the same effect over the two types of assessments - effects that cancel each other when we take the difference between the two types of assessment. That is, it accounts for the fixed effects at the student level. We will relax this assumptions in two ways: first, we will allow the outcome to depend on our covariates and second, we will allow α to depend also on teachers' characteristics. With this two different specifications, we can test for two hypothesis: 1) Are these gaps sensitive to the inclusion of the students' characteristics? 2) Are these gaps sensitive to the teachers' characteristics?

In Figure 10, the gender gap in the difference between the two exams is positive and significant in all estimates displayed from Equation 7.3, where we allow the covariates to affect the outcomes differently. In Figure 10, we show the variation in the estimates by each set of controls added separately and this does not change the gap substantially in

any of the cases.

$$\widetilde{T}_{1,i} - \widetilde{T}_{0,i} = \widetilde{M'}_i\theta + \alpha G_i + \eta_i \tag{7.3}$$

To test the overall effect of the inclusion of the covariates in the unexplained gap, Table 12 and 13 show the results of the estimation of the Equation 7.3 for Mathematics and Language, respectively. The results point that the gap in Math is not sensitive to the inclusion of the controls in Column 2, indicating that at least in relation to our observable variables, there is no evidence of abilities and incentives influencing the gap in a different way. Even the inclusion of accumulated abilities marginally affects the coefficient for Math[7]. On the other hand, the inclusion of the set of controls on students' abilities and incentives alters significantly the coefficient on the gender variable for Language, increasing the gap (Columns 2 and 3) in Table 13 indicating that accumulated skills correlate with language assessments differently, directly affecting the differences between the two scores.

[7] A t-test on the coefficients between the equations could not reject the null hypothesis that the coefficients are equal.

Table 12 – Regression results for Differences across Exams, for Mathematics

Independent Variables	(1)	(2)	(3)
Girl	0.268***	0.289***	0.252***
	(0.0223)	(0.0214)	(0.0289)
Fluid Intelligence		-0.0769***	-0.0313**
		(0.0105)	(0.0157)
Goal Orientation		0.0139	0.0196
		(0.0157)	(0.0206)
Self-Efficacy		0.0404***	0.0583***
		(0.0138)	(0.0196)
Self-worth		-0.0335***	0.0111
		(0.0120)	(0.0165)
Self-control		0.0979***	0.0633***
		(0.0150)	(0.0206)
Locus of Control		-0.0218**	0.00551
		(0.0109)	(0.0146)
Study Time		0.0237*	0.00788
		(0.0134)	(0.0183)
Educational Expectations		0.0242	0.0406*
		(0.0147)	(0.0210)
Ease to Acc. Edu (Peers)		-0.0193	-0.0260
		(0.0208)	(0.0262)
Ease to Acc. Edu (Goods)		-0.0749***	-0.0284
		(0.0199)	(0.0259)
Standardized Test Skills			-0.395***
			(0.0404)
Classroom Skills			0.167***
			(0.0289)
Teacher/Classroom Controls	No	Yes	Yes
Constant	-0.133***	0.0483	0.116
	(0.0239)	(0.170)	(0.224)
Observations	10,379	10,082	5,262
R-squared	0.018	0.044	0.075

Note: This table shows results for the regressions for the difference across achievement measures. All variables are residualized in relation to socioeconomic background and parenting. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 13 – Regression results for Differences across Exams, for Language

Independent Variables	(1)	(2)	(3)
Girl	0.0904***	0.148***	0.134***
	(0.0221)	(0.0209)	(0.0266)
Fluid Intelligence		-0.0610***	-0.0247*
		(0.0105)	(0.0145)
Goal Orientation		-0.0472***	-0.0154
		(0.0158)	(0.0209)
Self-efficacy		-0.00896	0.0104
		(0.0141)	(0.0186)
Self-worth		-0.0422***	-0.0252
		(0.0122)	(0.0167)
Self-control		0.0965***	0.103***
		(0.0137)	(0.0192)
Locus of Control		-0.0278***	-0.00839
		(0.0106)	(0.0143)
Study Time (Effort)		0.0603***	0.0540***
		(0.0113)	(0.0158)
Educational Expectations		0.0198	0.00204
		(0.0145)	(0.0203)
Ease to Acc. Edu (Peers)		-0.0688***	-0.0600**
		(0.0211)	(0.0264)
Ease to Acc. Edu (Goods)		-0.0514***	-0.00650
		(0.0196)	(0.0250)
Standardized Test Skills			-0.464***
			(0.0405)
Classroom Skills			0.168***
			(0.0312)
Teacher/Classroom Controls	No	Yes	Yes
Constant	-0.0420*	0.480***	0.272
	(0.0242)	(0.166)	(0.207)
Observations	10,379	10,082	5,262
R-squared	0.002	0.035	0.075

Note: This table shows results for the regressions for the difference across achievement measures. All variables are residualized in relation to socioeconomic background and parenting. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Following Lavy (2008), the variation of the gap in relation to teacher characteristics could corroborate the hypothesis of gender discrimination towards boys. In Table 14 and 15, we show the estimates for this exercise. We split the sample for teacher's gender, race and whether the teacher already taught the student in a previous year. We do not find the gap across exams in mathematics to be sensitive to teacher characteristics, after

controlling for observable abilities, incentives and classroom characteristics. The results do not corroborate Lavy's. Then, in summary, [12] the gap across exams remains largely unexplained. [13] Language grades are more sensitive to the students' characteristics than mathematics. [14] We do not find evidence towards a hypothesis of gender discrimination, once the gap is not sensitive to the teachers' characteristics.

Table 14 – Regression results for Differences across Exams, for Mathematics and by teacher characteristics

		Teacher Characteristics								
	ALL	Female	Male	P-value	White	Black	P-value	Same Teacher	Diff. Teacher	P-Value
Girl	0.289***	0.269***	0.324***	0.215	0.279***	0.378***	0.115	0.295***	0.285***	0.841
	(0.021)	(0.026)	(0.034)		(0.022)	(0.058)		(0.046)	(0.0235)	
Student Abilites	Yes	Yes	Yes		Yes	Yes		Yes	Yes	
Incentives	Yes	Yes	Yes		Yes	Yes		Yes	Yes	
Classroom/Teacher	Yes	Yes	Yes		Yes	Yes		Yes	Yes	
Observations	10,082	6,185	3,866		8,406	1,542		1,801	8,281	

Note: This table shows results for the regressions for achievement measures by teacher's characteristics. All variables are residualized in relation to socioeconomic background and parenting. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 15 – Regression results for Differences across Exams, for Language and by teacher characteristics

		Teacher Characteristics								
	ALL	Female	Male	P-value	White	Black	P-value	Same Teacher	Diff. Teacher	P-Value
Girl	0.148***	0.127***	0.183***	0.178	0.150***	0.146***	0.947	0.153***	0.148***	0.935
	(0.021)	(0.026)	(0.034)		(0.023)	(0.053)		(0.049)	(0.023)	
Student Abilites	Yes	Yes	Yes		Yes	Yes		Yes	Yes	
Incentives	Yes	Yes	Yes		Yes	Yes		Yes	Yes	
Classroom/Teacher	Yes	Yes	Yes		Yes	Yes		Yes	Yes	
Observations	10,082	6,185	3,866		8,406	1,542		1,801	8,281	

Note: This table shows results for the regressions for achievement measures by teacher's characteristics. All variables are residualized in relation to socioeconomic background and parenting. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

8 Final Comments

This research brought together the literature on the economics of personality and economics of education by bringing evidence on the relative importance of Skills, both cognitive and noncognitive, and Schooling Incentives to explain the gender gaps in educational achievement. We use a unique database that allows us to describe a detailed profile of boys and girls in the state of Sao Paulo, in Brazil. Since we do not address a causal relation between skills and outcomes, this study opens up further discussion regarding the gender gap in educational outcomes. Still we can summarize informative findings.

We find that boys and girls differ significantly in relevant inputs to educational achievement. This difference translates into achievement gaps, both in classroom and standardized exams. That indicates teacher grades are not aligned with state test scores for both subjects as it was supposed to be. Regarding these differences, the gap in both types of assessments are more sensitive to socioemotional skills and incentives to schooling than it is to socioeconomic background or classroom characteristics. A new evidence this research shed light on was that educational expectations and ease to accumulate education predict test scores in Mathematics and Language in a higher magnitude than every socioemotional skills. Overall, although we were able to explain a large share of the explained gap, the difference between exams remains largely unexplained and insensitive to teachers' characteristics, weakening discrimination hypothesis. Further investigation is necessary in order to address what are the elements driving the gender gap in educational achievement.

References

ALMLUND, M. et al. Personality psychology and economics. In: HANUSHEK, E. A.; MACHIN, S.; WöSSMANN, L. (Ed.). *Handbook of the Economics of Education*. Amsterdam: Elsevier, 2011. v. 4, p. 1–181. 13, 20

BECKER, G. S.; HUBBARD, W. H. J.; MURPHY, K. M. Explaining the worldwide boom in higher education of women. *Journal of Human Capital*, v. 4, n. 3, p. pp. 203–241, 2010. 11

BORGHANS, L. et al. The economics and psychology of personality traits. *Journal of Human Resources*, v. 43, n. 4, p. 972–1059, Fall 2008. 13

BORGHANS, L. et al. Identification problems in personality psychology. *Personality and Individual Differences*, v. 51, n. 3: Special Issue on Personality and Economics, p. 315–320, 2011. 35, 38

BORGHANS, L.; MEIJERS, H.; WEEL, B. T. The Role Of Noncognitive Skills In Explaining Cognitive Test Scores. *Economic Inquiry*, v. 46, n. 1, p. 2–12, 01 2008. 21

BORGHANS, L.; MEIJERS, H.; WEEL, B. ter. The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry*, v. 46, n. 1, p. 2–12, January 2008. 18

BOTELHO, F.; MADEIRA, R.; RANGEL, M. Racial Discrimination in Grading: Evidence from Brazil. 2015. 14

CARNEIRO, P.; HECKMAN, J. J.; MASTEROV, D. V. Labor market discrimination and racial differences in pre-market factors. *Journal of Law and Economics*, v. 48, n. 1, p. 1–39, April 2005. 13

COOPER, H. M.; BURGER, J. M.; GOOD, T. L. Gender differences in the academic locus of control beliefs of young children. *Journal of Personality and Social Psychology*, v. 40, n. 3, 1981. 28

CORNWELL, C.; MUSTARD, D. B.; PARYS, J. V. Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School. *Journal of Human Resources*, v. 48, n. 1, p. 236–264, 2013. 17, 23, 36

CUNHA, F.; HECKMAN, J. Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, v. 43, n. 4, 2008. 20

CUNHA, F.; HECKMAN, J. J. The technology of skill formation. *American Economic Review*, v. 97, n. 2, p. 31–47, May 2007. 20

CUNHA, F.; HECKMAN, J. J. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, v. 43, n. 4, p. 738–782, Fall 2008. 13

CUNHA, F.; HECKMAN, J. J. The economics and psychology of inequality and human development. *Journal of the European Economic Association*, v. 7, n. 2–3, p. 320–364, April 2009. Presented as the Marshall Lecture, European Economics Association, Milan, Italy, August 29, 2008. 13

DREBER, A.; ESSEN, E.; RANEHILL, E. Outrunning the gender gap—boys and girls compete equally. *Experimental Economics*, v. 14, n. 4, p. 567–582, November 2011. 18

DUCKWORTH, A. L.; SELIGMAN, M. E. P. Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement scores. *Journal of Educational Psychology*, v. 98, n. 1, p. 198–208, 2006. 28, 36, 42

ELDER, T. E.; GODDEERIS, J. H.; HAIDER, S. J. Unexplained gaps and Oaxaca-Blinder decompositions. *Labour Economics*, v. 17, n. 1, p. 284–290, January 2010. 23

FALCH, T.; NAPER, L. R. Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review*, v. 36, n. C, p. 12–25, 2013. 18

FORTIN, N. M.; OREOPOULOS, P.; PHIPPS, S. *Leaving Boys Behind: Gender Disparities in High Academic Achievement*. [S.l.], 2013. 17

FORTIN, N. M.; OREOPOULOS, P.; PHIPPS, S. Leaving Boys Behind: Gender Disparities in High Academic Achievement. *Journal of Human Resources*, v. 50, n. 3, p. 549–579, 2015. 23, 30, 36, 42

FRYER, R. G.; LEVITT, S. D. An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, v. 2, n. 2, p. 210–40, April 2010. 17, 23

GNEEZY, U.; RUSTICHINI, A. Gender and competition at a young age. *American Economic Review*, v. 94, n. 2, p. 377–381, 2004. 18

GOLSTEYN, B.; SCHILS, T. *Gender gaps in primary school achievement. A decomposition into endowments and returns to IQ and non-cognitive factors*. [S.l.], 2014. 18

GORSUCH, R. L. Factor analysis. In: _____. [S.l.: s.n.], 2003. (Handbook of Psychology, v. 2), p. 143–64. 25

GRANT, M. J.; BEHRMAN, J. R. Gender Gaps in Educational Attainment in Less Development Countries. *Population and Development Review*, v. 36, n. 1, March 2010. 11

HECKMAN, J. J. Schools, skills and synapses. *Economic Inquiry*, v. 46, n. 3, p. 289–324, July 2008. 13

HECKMAN, J. J.; HUMPHRIES, J. E.; KAUTZ, T. *The GED and the Role of Character in American Life*. Unpublished book manuscript, University of Chicago, Department of Economics. 2012. 13

HECKMAN, J. J.; KAUTZ, T. Achievement tests and the role of character in American life. In: HECKMAN, J. J.; HUMPHRIES, J. E.; KAUTZ, T. (Ed.). *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. Chicago: University of Chicago Press, 2014. 13

HECKMAN, J. J.; KAUTZ, T. Fostering and measuring skills: Interventions that improve character and cognition. In: HECKMAN, J. J.; HUMPHRIES, J. E.; KAUTZ, T. (Ed.). *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. Chicago, IL: University of Chicago Press, 2014. p. 341–430. 31

HECKMAN, J. J.; MOSSO, S. The economics of human development and social mobility. *Annual Review of Economics*, v. 6, n. 1, p. 689–733, 2014. 20

HECKMAN, J. J.; RUBINSTEIN, Y. The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, v. 91, n. 2, p. 145–149, May 2001. 13

HECKMAN, J. J.; STIXRUD, J.; URZUA, S. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, v. 24, n. 3, p. 411–482, July 2006. 13

HINNERICH, B. T.; HöGLIN, E.; JOHANNESSON, M. Are boys discriminated in Swedish high schools? *Economics of Education Review*, v. 30, n. 4, p. 682–690, August 2011. 18

HUANG, C. Gender differences in academic self-efficacy: a meta-analysis. *European Journal of Psychology of Education*, v. 28, p. 1–35, 2013. 28

HYDE, J. S. The Gender Similarities Hypothesis. *American Psychologist*, September 2005. 12

JR., P. C.; TERRACCIANO, A.; MCCRAE, R. Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, v. 81, n. 2, p. 322–331, 2001. 12

KAUTZ, T.; ZANONI, W. Measuring and fostering non-cognitive skills in adolescents: Evidence from chicago public schools and the onegoal program. Unpublished manuscript, University of Chicago, Department of Economics. 2014. 31

LAVY, V. Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, v. 92, n. 10-11, p. 2083–2105, October 2008. 12, 18, 24, 53, 56

LAVY, V.; SAND, E. *On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases*. [S.l.], 2015. 18

LEGEWIE, J.; DIPRETE, T. School Context and the Gender Gap in Educational Achievement. *American Sociological Review*, v. 20, n. 10, p. 1–23, January 2012. 18

LINDAHL, E. *Comparing teachers' assessments and national test results – evidence from Sweden*. [S.l.], 2007. 18

LYNN, R.; IRWING, P. Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, v. 32, p. 481–498, 2004. 28

OECD. *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. PISA OECD Publishing, 2015. 11

ROBERTS, B. W. Personality development and organizational behavior. In: STAW, B. (Ed.). *Research on Organizational Behavior*. Oxford, UK: Elsevier Science/JAI Press, 2006. v. 27, cap. 1, p. 1–41. 13

VOYER, D.; VOYER, S. D. Gender Differences in Scholastic Achievement: A Meta-Analysis. *Psychol. Bulletin*, v. 140, p. 1174–1204, 2014. 18

Appendix A

Table A.1 – Comparing the 2011 Sample with the restricted 2009 Sample

	Mean			
	Sample 2011	Sample 2009	Difference	*P-Value*
Language Grades 2011	0.020	-0.020	0.039	0.020
Math Grades 2011	0.015	0.036	-0.021	0.224
Language Test Scores 2011	-0.189	-0.192	0.002	0.895
Math Test Scores 2011	-0.242	-0.244	0.002	0.912
Fluid Intelligence	0.005	-0.003	0.008	0.645
Goal Orientation	0.004	-0.022	0.026	0.102
Self-Efficacy	0.002	0.019	-0.017	0.293
Self-Worth	0.005	-0.007	0.012	0.449
Self-Control	0.007	-0.007	0.014	0.363
Locus of Control	0.007	-0.007	0.013	0.424
Study Time	-0.003	0.012	-0.015	0.288
Educational Expectations	0.000	0.008	-0.008	0.507
Ease to Acc. Educ (Peers)	0.011	0.000	0.010	0.483
Ease to Acc. Educ (Goods)	0.012	-0.002	0.014	0.366

Note: This table compares the mean in observables variables, residualized in relation to socioeconomic background and parenting, between the full sample from 2011 and the restricted sample in 2009. P-values are for the test for the mean difference between the two samples