MOX-Report No. 23/2023

The Determinants of Mathematics Achievement: A Gender Perspective Using Multilevel Random Forest

Bertoletti, A.; Cannistrà, M.; Diaz Lema, M.; Masci, C.; Mergoni, A.; Rossi, L.; Soncin, M.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it https://mox.polimi.it

The Determinants of Mathematics Achievement: A Gender Perspective Using Multilevel Random Forest

Alice Bertoletti [1], Marta Cannistrà [2], Melisa Diaz Lema [2], Chiara Masci [3], Anna Mergoni [4], Lidia Rossi [2] and Mara Soncin [2]

February 24, 2023

[1] Joint Research Center, European Commission, 41092 Seville, Spain; alice.bertoletti@ec.europa.eu
[2] Department of Management, Economics and Industrial Engineering, Politecnico di Milano, 20133 Milan, Italy; marta.cannistra@polimi.it (M.C.); melisalucia.diaz@polimi.it (M.D.L.); lidia.rossi@polimi.it (L.R.)
[3] MOX, Department of Mathematics, Politecnico di Milano, 20133 Milan, Italy; chiara.masci@polimi.it
[4] Faculty of Economics and Business, KU Leuven, 3000 Leuven, Belgium; anna.mergoni@kuleuven.be

Abstract

This paper investigates the determinants of mathematics performance by gender, exploiting a multilevel random forest approach. OECD PISA 2018 data from 28 European countries are employed to explore the performance of male and female students as a function of students' family characteristics, their attitudes towards education, and class and school environment. Results show that the gender gap in favour of boys persists in most European countries. However, teacher and school practices like fostering student reading and creating a cooperative environment allow mitigating the influence of family background in countries without gender gap. Policy implications to foster performance equality are provided.

Keywords mathematics achievement; comparative analysis; gender gap; random forest.

1 Introduction and Motivation

Equality of opportunity across individuals is a matter of primary importance in the political agenda of worldwide economies (Dunnzlaff et al., 2011). To reach this objective, a fair educational system is a necessary step, given that higher educational levels are associated with higher wages, better health, and higher well-being level. In particular, gender inequality represents an unresolved question that ranges from reduced women's participation in the labour market to salary gaps and gender stereotyping in career choice (Education et al., 2012). It is particularly evident how women are structurally under-represented in science, technology, engineering, and math (STEM) careers, making this an untapped opportunity to expand employability and innovation capacity (Beede et al., 2011).

Gender studies have traditionally traced back these gender differences to disparities in educational outcomes (Evans et al., 2020). While girls tend to outperform boys in reading (van Hek et al., 2019), the gap in mathematics is structurally in favour of boys in most European countries (Education et al., 2012; Contini et al., 2017). Despite the net difference in mathematics being usually smaller than the gap in reading, the amplification effects in terms of different career choices and salary gap in favour of men are relevant enough to make this a question of primary relevance (UNICEF et al., 2020). Lower math achievement leads, in many cases, to lower participation of females in STEM majors at university (Card and Payne, 2021). In turn, this easily translates into gender gaps in the labour market and occupational choices in disfavour of women (Machin and Puhani, 2003; Piazzalunga, 2018; Bertocchi and Bozzano, 2020). The paper addresses this issue by focusing on student achievement in mathematics and investigating the causes leading to disparities across gender.

Extant studies tend to explain the gender gap in education by looking at differences in the level of the main determinants of school achievement between boys and girls (Buchmann et al., 2008; Figlio et al., 2019). However, the decomposition analyses available in the literature generally reveal that differences in the key factors predicting learning achievement (such as household resources, parents' and teachers' support, family expectations, and career motivation) can only partially explain the gender gap in education (Gevrek et al., 2020; Munir and Winter-Ebmer, 2018). While cultural and societal dimensions can play a relevant role (Else-Quest et al., 2010; Giuliano, 2020), unexplained educational differences between males and females may also be associated with structural differences in the way in which some key factors influence student performance across gender. Thus, the present paper relaxes the baseline assumption that the determinants of educational achievement have the same impact across gender by modelling the determinants of boys' and girls' performance separately. Moreover, the study adopts an international approach, by exploring the European countries as empirical context. In particular, the research addresses the following question: *How do the determinants of mathematics achievement differ between male and female students and among European countries?*

The paper explores the OECD Programme for International Student Assessment (PISA) 2018 dataset for 28 European countries (i.e., the EU countries with available data, plus the UK, Iceland, and Switzerland). The data provided by PISA refers to 15-year-old students and, therefore, can be employed to examine the gender gap in a crucial moment of education, corresponding to the last year of compulsory school in most countries.

To ensure homogeneity in structural characteristics, the analyses are carried out by classifying countries into three categories, i.e., the ones with a gender gap in favour of boys, the ones with a gender gap in favour of girls, and the ones with no gap. A multilevel random forest (Pellagatti et al., 2021), where student and country levels are considered, is implemented separately in the three groups of countries. We follow a random forest approach because its flexibility adapts well to the educational context, in which several input variables co-exist in the same environment (Masci et al., 2018). More specifically, while more classic linear multilevel models are able to estimate only linear associations between covariates and the response, this technique, by relaxing any a priori parametric assumption, performs well in presence of several interactions among predictors and allows to discover the most likely relationship between the variables. At the same time, the multilevel approach allows modelling the heterogeneity between countries and to disentangle the variability given at student and country levels. The empirical results indicate the existence of structural differences in some relevant determinants of math achievement between boys and girls, such as perception of cooperation and reading attitudes. The way through which factors influence math performances of males and females is also strongly related to the geographic area in which pupils are studying. To this extent, the paper presents useful evidence to design specific policy actions for enhancing gender equality in education and the labour market.

The remainder of the paper is organised as follows. Section 2 revises the literature on the determinants of the gender gap in mathematics and presents the conceptual framework. Section 3 describes the data and the methodology used for the empirical analyses. Then, the results and their discussion are presented in Sections 4 and 5, while final implications and conclusions are reported in Section 6.

2 Literature Review

2.1 Evidence of Gender Gap in Student Performance

The existence and persistence of a gender gap in mathematics in favour of boys has been demonstrated by multiple studies over time (Frye and Levitt, 2010; González de San Román and De La Rica, 2012; Contini et al., 2017; Borgonovi et al., 2018). What is particularly striking from the current literature is the absence of a gap between boys and girls when they enter school, while it becomes larger with the years of schooling (Borgonovi et al., 2018; Mejias et al., 2021). Frye and Levitt (2010) show that the gap in mathematics increases from 0 to 0.2 standard deviations after 6 years of education. This gap becomes particularly pronounced after males and females leave compulsory schooling and enter in post-compulsory education and the labour market, with important effects on students' educational trajectories and opportunities. Multiple possible explanations have been attempted, ranging from less involvement in maths for girls to low parental expectations, but the determinants of such a phenomenon are still highly debated (Bouffard and Hill, 2005; Levine et al., 2005; Frye and Levitt, 2010).

The Programme for International Student Assessment (PISA), an international survey of 15-year-old students among OECD countries, has often been employed to study the extent of the gender gap internationally. Gender gap in mathematics performance remained broadly stable between PISA 2012 and PISA 2015 (OECD, 2016), showing, if anything, a small reduction of boys' advantage in mathematics. In 2012, boys outperformed girls in mathematics in 38 of the 65 participating countries by an average of 11 score points (across OECD countries) (OECD, 2019a), while in 2018 boys significantly outperformed girls in 32 of the 79 participating countries by an average of 6 points. Interestingly, in 2018, 14 countries showed an opposite gender gap in mathematics (OECD, 2019a). Among these economies, Finland represents the European country where girls obtained the highest scores with respect to boys in mathematics, on average. On the opposite, in 2018, the largest difference in favour of boys has been observed in Colombia, where boys scored around 20 points higher than girls. Among the countries with a high gap, between 15 and 18 points, Italy is the only European country (OECD, 2019a; Contini et al., 2017). In 43 out of 64 countries and economies, the gender gap in mathematics performance in favour of boys did not change significantly between 2009 and 2018. Notably, in Finland, Greece, Iceland, Luxembourg, the Netherlands, and Switzerland, the narrowing of the gender gap in mathematics performance observed in 2018 assessment is due to a significant decline in boys' performance in mathematics (OECD, 2019a).

Looking at the different performance levels, boys are generally over-represented at both the bottom and the top of the performance distributions in mathematics (OECD, 2019a). In many countries, girls' scores in the first decile of the distribution of mathematics performance are higher than boys' scores, meaning that the lowest-performing girls score above the lowest-performing boys in their countries. However, the largest differences are observed at the top of the distribution of mathematics performance, where an important male-oriented gender distributional imbalance among high achievers emerges (Zhou et al., 2017; Breda et al., 2018).

2.2 Conceptual Framework

The factors affecting the achievement of students have been widely studied for a long time (De Witte and Kortelainen, 2013). In particular, Chaman et al. (2014) presents a review of the factors affecting math performances on secondary education students. He considers in particular mathematic anxiety, attitude towards math, parental involvement, gender, and cultural differences. The present research focuses on three categories of determinants impacting students' performance, which might have heterogeneous impacts on male and female students. The categories relate to (i) student's family characteristics, (ii) student's perceptions and attitudes, and (iii) class and school environment.

Among student's family characteristics, the home environment, ranging from parents' attitude toward education to socioeconomic status, plays an important role in shaping students' achievement of girls and boys (Bertocchi and Bozzano, 2020). Steinthorsdottir and Sriraman (2008) have found that the involvement and support of families have a different effect on boys and girls: while boys benefit from a family context with high parent pressures, female students benefit more from parents showing interest in their school activities. Future plans and ambitions expressed by students are also important determinants explaining gender differences in student performance (Steinthorsdottir and Sriraman, 2008). Parents' preference for boys may also explain a gender disparity in the school support provided to their children (Dossi et al., 2021). The socioeconomic status has a greater impact on the PISA results in mathematics for female pupils, determining a higher gender gap for disadvantaged students (Schleicher, 2019). Moreover, girls' performance is usually better in families with working mothers, suggesting that gender identities are transmitted from mothers to daughters (González de San Román and De La Rica, 2012). Confirming this result, Brenøe and Lundberg (2018) have found that girls benefit more from maternal education and employment than boys.

Second, the students' perceptions and attitudes, such as well-being and personal interests may explain a substantial part of the students' performance in mathematics (Marsh and Martin, 2011). In relation to COVID-19, studies on psychological aspects are gaining attention (Wang et al., 2022). Evidence shows that boys usually report a greater self-efficacy compared to girls (Close and Shiel, 2009). Performances being equal, female students tend to underestimate their mathematical abilities than their male fellows (Sikora and Pitt, 2019), and this affects their cognitive performance, motivation and attitudes, as well as future career perspectives (Aiello et al., 2021). Similarly, girls seem to be more anxious about mathematical problems and in implementing mathematical thinking (Close and Shiel, 2009). Halpern and Ikier (2002) argue that this could be linked to the fact that boys have a greater experience of using math in their everyday life, compared to girls. However, females' anxiety about math may be related to additional factors (Caviola et al., 2022) such as low levels of confidence and self-perception (Cvencek et al., 2014; Pajares, 2005) or gender stereotypes regarding STEM and math achievement (Flore and Wicherts, 2015; Starr and Simpkins, 2021; Tomassini, 2021). Finally, the positive attitude towards reading is not only an important predictor of reading performance, but it is also related to mathematics achievement, as a measure of the positive attitude of students towards learning. In this respect, (Ajello et al., 2018) demonstrate how girls are advantaged in mathematics items with a high reading demand, independent of their level of reading literacy.

Third, to better understand the role of schools and teachers, it is relevant to consider variables on teacher behaviour and school characteristics (i.e., class and school environment). Previous research has shown how teachers' beliefs and expectations about student performance differ depending on students' gender-leading to learning gaps, usually in favour of male students and especially regarding STEM subjects (Jaremus et al., 2020; Mizala et al., 2015; Rainey et al., 2019). Rainey et al. (2019) find that active teaching environments may positively impact students' sense of belonging and desire to continue in STEM. Bertocchi and Bozzano (2020) also point out that female students can be encouraged and engaged in studying STEM subjects by the presence of a female teacher, who may be seen as a role model and could set up curricula that are more attractive to girls. Finally, teh school environment represents a key factor in affecting gender differences in student performance, and previous studies have shown that the school peer pressure and expectations not only are very different between boys and girls, but also influence differently student behaviours and performance (Steinthorsdottir and Sriraman, 2008). Moreover, Gibbs (2010) stresses the role of school curricula in enlarging the gender gap in disfavour of girls, particularly because of a content change in mathematics topics over years, which increasingly focus on topics that tend to favour boys (like spatial and logical items).

3 Data and Methods

3.1 Data and Variables' Selection

The empirical analyses are based on the PISA 2018 dataset, which provides internationally comparable data on the educational achievement of 15-year-old students, together with several background information on students, schools, and families. PISA 2018 is the last wave available, allowing exploring the most recent information on students' achievement. As mentioned in Section 1, by analysing the math achievement of 15-year-old students, we can provide evidence of the gender gap in the last years of compulsory school. The results are, thus, particularly significant since the gender gap found at this educational stage is more likely to affect the future job career of secondary-school participants. For the same reason, we focus exclusively on students enrolled in general track schools, without considering the ones attending a vocational track. In this way, we can provide detailed evidence on the students who are more likely to attend universities and, therefore, who would be potentially more affected by a gap in math achievement during their educational path.

Based on the conceptual framework presented in Section 2.2, we study the influence of three categories of variables on the mathematics achievement of male and female students. More specifically, the three groups of variables concern the student's family characteristics, student's perceptions and attitudes, and class and school environment. All the indicators are based on PISA 2018 questionnaire and are described in Table 1.

Table 1: Description of the student-level variables.

Variable	PISA Code	Type	Description
Student's family characteristics			
Mother edu	ST005 and ST006	cat	Indicate the highest level of education achieved by the mother and it is based on the questions ST005 and ST006. 0 = primary education not completed, 1 = complete primary education, 2 = complete lower secondary education, 3 = complete upper secondary education, 4 = complete post-secondary non tertiary education, 5 = complete tertiary education, 6 = complete postgraduate education
Parent support	EMOSUPS	num	Standardized indicator of parents' emotional support. It was constructed by PISA on the base of question ST123 and it ranges between −2.447 and 1.035.
ESCS	ESCS	num	Standardized index of economic, social and cultural status, derived by PISA, based on the parents' highest level of education (PARED), parents' highest occupational status (HISEI), and home possessions (HOMEPOS), including Books in the home
Foreign language	ST022Q01TA	0/1	Language that the students speak at home. 0 = same language as at school, 1 = different language
ICT resources	ICTRES	num	Standardized indicator of ICT home possessions. It ranges between −3.968 and 3.612
Student's perceptions and attitudes			
Fear failure	GFOFAIL	num	Standardized indicator of the fear of failure of the student. It is based on question ST183 and it ranges between −1.894 and 1.891
Feel awkward *	ST034Q04TA	0/1	Indicator based on the sentence 'I feel awkward and out of place in my school'. 0 = (strongly) disagree with the sentence, 1 = (strongly) agree with the sentence
Feel outsider *	ST034Q01TA	0/1	Indicator based on the sentence 'I feel like an outsider (or left out of things) at school'. 0 = (strongly) disagree with the sentence, 1 = (strongly) agree with the sentence
Self confidence *	ST034Q05TA	0/1	Indicator based on the sentence 'Other students seem to like me.' 0 = (strongly) disagree with the sentence, 1 = (strongly) agree with the sentence
Sociable *	ST034Q02TA	0/1	Indicator based on the sentence 'I make friends easily at school.' 0 = (strongly) disagree with the sentence,1 = (strongly) agree with the sentence
Enjoyment reading	ST175Q01IA	cat	Time spent by the students reading for enjoyment. 0 = no time, 1 = less than 30 min per day, 2 = between 30 and 60 min per day, 3 = between 1 and 2 h, 4 = more than 2 h
Class and school environment			
Teach support (global)	TEACHSUP	num	Standardized indicator of teacher support, constructed by PISA on the base of question ST100. It ranges between −2.743 and 1.341
Discipline language class	DISCLIMA	num	Standardized indicator of disciplinary climate in the language-of-instruction lessons, provided by PISA. It is based on ST097 and it ranges between −2.712 and 2.034.
Longest book	ST154Q01HA	cat	Number of pages of the longest text the student had to read for school. 1 = one page or less, 2 = between 2 and 10 pages, 3 = between 11 and 50 pages, 4 = between 51 and 100 pages, 5 = between 101 and 500 pages, 6 = more than 500 pages
Class size	CLSIZE	num	Number of students in the class, it ranges between 13 and 53.
Perception cooperation	PERCOOP	num	Cooperation climate perceived by students, it is a standardized indicator computed by PISA based on question ST206. and it ranges between −2.143 and 1.676

Note: Variables marked with * were originally Likert scale questions from 1 to 5, here dichotomized by the authors. Values from 1 to 3 were assigned 0, otherwise 1.

Student's family characteristics include: (i) the level of education of the mother (ST005 and ST006), (ii) the perceived support of the parents (EMOSUPS), (iii) an index of the socio-economic background of the student (ESCS), (iv) a binary variable indicating whether the pupil speaks a foreign language at home (ST022Q01TA), and (v) a binary variable indicating whether there are ICT resources at home (ICTRES).

Regarding student's perceptions and attitudes, we consider variables that describe students' fear of failure (GFOFAIL), student's feeling awkward (ST034Q05T4) or outsider (ST034Q01TA), student's perception of being liked (or not) by other students (ST034Q05TA), and student's ability to make friends easily (ST034Q02TA). Moreover, as described in Section 2.2, attitude towards reading is included here as it potentially explains math scores' differences between genders (ST175Q01IA).

Lastly, on the class and school environment, we consider variables concerning how much teachers are supportive towards students (TEACHSUP), how students perceive the teacher to be able to maintain discipline in the class (DISCLIMA), the number of pages of the longest book students had to read for school purposes (ST154Q01HA), the class size (CLSIZE), and the perceived climate of cooperation in the school (PERCOOP).

The indicators described in Table 1 are available for 28 European countries. Despite some relevant countries (such as Sweden and Norway) having been excluded from the study for problems with data availability, the analyses can provide a comprehensive overview of the European area. On the other hand, additional indicators that could potentially explain math achievement have not been considered because they report missing values for several European countries.

3.2 Methodology

The aim of our analysis is to investigate the mechanisms that determine the heterogeneity in students' performance across gender. We are interested in exploring the gender educational gap within countries and in identifying which variables are associated with females' and males' performance, within different contexts. To this end, our methodological approach consists of two steps. In the first step, we identify three categories of European countries: countries where males perform on average better than females (*Group 1*), countries where there is no evidence of a gender gap (*Group 2*), and countries where females perform on average better than males (*Group 3*). For each country, we perform a parametric two-sample t-test for comparing the means of males and females performances and, standing on the p-value, we assign the country to one of the three categories (see Table 4 in Appendix A for details).

The three categories represent three different social contexts and, in the second step, our aim is to investigate, separately for each of them, which are the most important determinants of students' scores for boys and girls, respectively. To this end, for each category of countries $c =$ {Group 1, Group 2, Group 3} and for each gender $g =$ {Female, Male}, we perform a multilevel random forest (Pellagatti et al., 2021) in which we consider students (level 1), nested within countries (level 2). For each student i of gender g, attending a school in country j within category c, the model takes the following form:

$$y_{ij,gc} = f_{gc}(\mathbf{x}_{ij,gc}) + b_{j,gc} + \epsilon_{ij,gc}$$

where $y_{ij,gc}$ is the math PISA test score of student i; $\mathbf{x}_{ij,gc}$ is the set of student level covariates relative to student i; $f_{gc}(\cdot)$ identifies the random forest term; $b_{j,gc} \sim \mathcal{N}(0, \sigma^2_{gc})$ is the random intercept relative to country j; and $\epsilon_{ij,gc} \sim \mathcal{N}(0, \omega^2_{gc})$ is the error term.

We adopt this modelling for two main reasons. First, the multilevel approach allows us to take into account the countries as grouping factor and to estimate the heterogeneity in student performances net of any structural differences across countries. Educational systems could significantly influence students' differences in performance by gender, and thus it is relevant to estimate determinants within countries. For instance, by performing an empirical analysis on 32 countries, Ayalon and Livneh (2013) show that the between-countries variation in the gender gap in mathematics can be explained by the different levels of standardisation of the national educational systems. In addition, Cascella et al. (2021) show that gender differences in mathematics can be attributed to different socio-cultural and economic factors that can vary among countries and regions. Similarly, González de San Román and De La Rica

(2012) and Cuevas-Ruiz et al. (2020) state that girls' performance is better in societies where gender equality is valued. For these reasons, after the partition of the European countries within the three categories, there is still a component that varies across countries of the same category that we can quantify. Therefore, given the estimates of the variance of the random effects $\hat{\sigma}^2_{gc}$ and of the error term $\hat{\omega}^2_{gc}$, we compute the Percentage of Variability explained by the Random effects (PVRE) as $\frac{\hat{\sigma}^2_{gc}}{\hat{\sigma}^2_{gc}+\hat{\omega}^2_{gc}}$, that represents the percentage of the unexplained variability in student performance explained at country level. By comparing this quantity across categories of countries and across genders, we can explore the relevance of the country component by gender. In particular, for both genders, the estimate of the coefficient $b_{j,gc}$ quantifies the effect of country j on its female and male student performances, respectively.

Second, the random forest approach allows us to estimate the effect of the covariates in a flexible and interpretable way (Schiltz et al., 2017; Masci et al., 2018). This is fundamental given the numerous predictors that we would like to consider and their potential non-linear association with the response. Parametric multilevel models require a priori knowledge to choose their parametric form and often result to be too restrictive when covariates have different types of relationships and interactions with the response. Indeed, they basically capture only relationships that have the pre-specified functional form. With respect to them, random forest allows handling a higher number of-potentially correlated-covariates and easily modelling their interactions and their different associations with the response. Random forest is an ensemble of regression trees (Breiman, 2001; Lewis, 2000; Friedman et al., 2001; James et al., 2013) and, given a response variable and a set of covariates, it computes an importance ranking of the covariates by measuring the ability of each covariate to improve the estimation. For each covariate, this measure, labelled as $\%IncMSE$, is computed from permuting Out-Of-Bag (OOB) data in the following way: for each tree of the random forest, the Mean Square Error (MSE) on the OOB portion of the data is recorder; the same is then done after permuting the covariate; the difference between the two are then averaged over all trees, and normalised by the standard deviation of the differences. Besides the importance ranking of the covariates, we can further investigate the effect of each covariate on the response variable by means of Partial Dependence Plots (PDPs). For each covariate, the PDP represents the net effect of the covariate on the response, after averaging out the effect of all other covariates.

4 Results

4.1 Preliminary Results: Country Groups Based on Gender Gap'S Direction

As underlined in Sections 1 and 3.2, the gender gap in mathematics can differ importantly among countries. These disparities are linked to substantial heterogeneity in socioeconomic and cultural characteristics across European regions, as well as differences in educational systems. While cross-country disparities can be taken into account by the random intercept in the multilevel random forest, the work aims at exploring how the results differ across the three groups of countries that we have identified (i.e., Groups 1, 2, and 3). The analyses are, thus, performed separately for the three different groups. Table 4 in Appendix A provides an overview of the PISA math scores by gender for all the countries considered in the analysis and the p-values resulting from the two-sample t-test, indicating if the distributions of scores are statistically different across countries. The map in Figure 1 displays the selected countries divided into the three groups. Group 1 is the most numerous group, with 20 countries. The large number of countries in this group stresses the urgency of addressing the gender gap in disfavour of girls in most European countries. Moreover, preliminary results indicate that scores' distributions are not statistically different across gender (Group 2) in five European countries: the Czech Republic, Switzerland, Slovakia, Poland, and Lithuania. Finally, Group 3 gathers the countries where females perform significantly better than males in mathematics, which are only three: Finland, Iceland, and Malta.

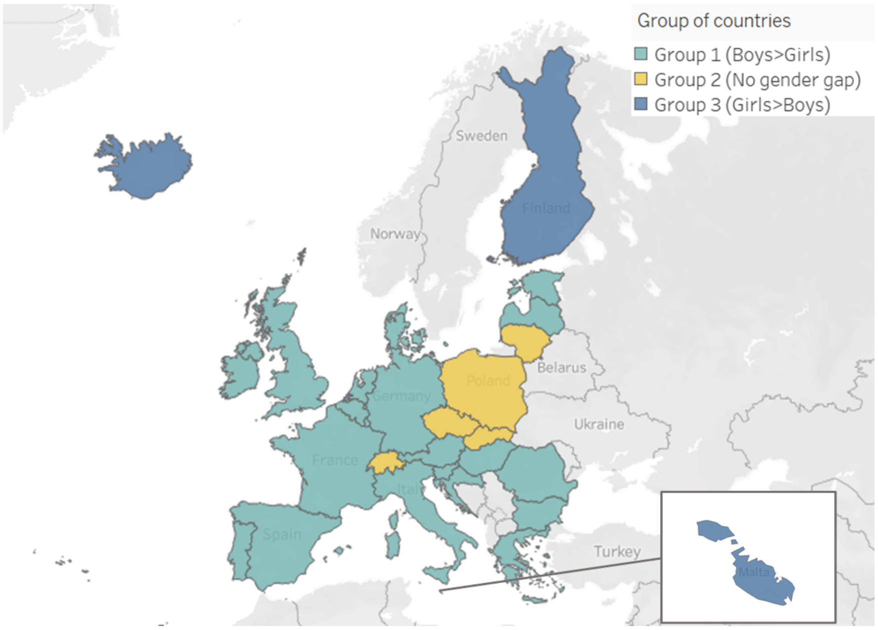


Figure 1: Selected countries coloured standing on their assignment to the three groups based on gender gap in PISA math scores.

Table 2 displays the descriptive statistics of selected variables divided by gender and group and stresses different patterns. Parental support is perceived to a larger extent by countries belonging to Group 1, where females perform worse than males. In this group, girls have a higher perceived cooperation climate than males, a higher perceived discipline in the class, and a more pronounced perceived parental support. This indicates that girls tend, in general, to be more positive about the school and home environment, and this attitude possibly leads to positive spill-overs. However, on average, girls also have a higher fear of failure-revealing that positive attitudes regarding the environment are not translated into better self-esteem. Other characteristics for which male and female populations differ are the ones related to ICT and reading. In particular, boys report having greater access to ICT at home, while girls read more and, mostly, enjoy more reading.

4.2 Main Results: Multilevel Random Forest

In this section, we describe the main findings emerging from the multilevel random forest models. In particular, six models, obtained by grouping countries based on their gender gap and by gender, are computed. For each model, the country effect and student-level variables' importance are shown in Table 3 and reported in detail in the Appendix A (see Figures 5 and 4). Table 3 shows the ranking's position of the covariates and the related Inc%MSE for each model (by gender and by context's country group). To facilitate the reading, results about student and country levels are presented separately in the following sections.

Table 2: Descriptive statistics of the student-level variables, stratified by gender and country groups.

	Group 1		Group 2		Group 3	
	Female	**Male**	**Female**	**Male**	**Female**	**Male**
PISA score	495.133	504.811	505.922	506.394	500.243	491.831
	(86.050)	(92.307)	(92.648)	(97.538)	(85.268)	(94.960)
Student's family characteristics						
Mother edu	3.918	3.880	3.951	3.892	4.234	4.168
	(1.558)	(1.651)	(1.352)	(1.481)	(1.387)	(1.531)
Parent support	0.128	−0.025	−0.057	−0.204	0.183	0.047
	(0.976)	(0.990)	(1.006)	(1.008)	(0.983	(0.998)
ESCS	0.061	0.073	0.058	0.027	0.374	0.377
	(0.945)	(0.964)	(0.904)	(0.923)	(0.832)	(0.879)
Foreign language	1.135	1.141	1.107	1.117	1.216	1.229
	(0.341)	(0.348)	(0.309)	(0.321)	(0.412)	(0.420)
ICT resources	0.048	0.065	−0.006	0.033	0.373	0.432
	(0.918)	(0.953)	(0.832)	(0.912)	(0.830)	(0.911)
Student's perceptions and attitudes						
Fear failure	0.136	−0.263	0.153	−0.268	0.234	−0.284
	(0.979)	(0.953)	(0.958)	(0.935)	(0.993)	(0.972)
Feel awkward	1.170	1.168	1.210	1.219	1.220	1.206
	(0.375)	(0.374)	(0.407)	(0.414)	(0.414)	(0.404)
Feel outsider	1.155	1.159	1.215	1.236	1.186	1.176
	(0.361)	(0.365)	(0.411)	(0.425)	(0.389)	(0.381)
Self confidence	1.842	1.847	1.765	1.765	1.800	1.817
	(0.365)	(0.360)	(0.424)	(0.424)	(0.400)	(0.387)
Sociable	1.758	1.803	1.709	1.753	1.722	1.793
	(0.429)	(0.398)	(0.454)	(0.431)	(0.448)	(0.405)
Enjoyment reading	2.476	1.855	2.515	1.843	2.210	1.751
	(1.297)	(1.102)	(1.311)	(1.131)	(1.185)	(1.036)
Class climate and features						
Teach support	−0.028	−0.014	−0.107	−0.078	0.144	0.144
	(1.000)	(1.015)	(0.959)	(1.015)	(0.932)	(1.010)
Discipline language class	0.053	−0.029	0.157	0.078	−0.024	−0.090
	(1.034)	(1.084)	(1.037)	(1.116)	(0.966)	(1.010)
Longest book	3.820	3.831	3.816	3.681	3.668	3.645
	(1.459)	(1.469)	(1.445)	(1.504)	(1.454)	(1.488)
Class size	25.606	25.487	23.354	22.791	20.869	20.760)
	(7.144)	(7.243)	(5.790)	(5.829)	(4.241)	(4.244)
Perception cooperation	−0.036	−0.056	0.021	−0.016	0.263	0.251
	(0.357)	(0.349)	(0.399)	(0.403)	(0.377)	(0.378)
Sample Size	69,303	65,406	15,738	15,645	8919	9201

Group 1 = countries where males perform better than females; Group 2 = countries with no gap; Group 3 = countries where females perform better than males. Summary statistics are reported in terms of mean and standard deviation (in brackets).

4.2.1 Student Level

Figure 2 provides a visual overview of the results of the multilevel Random Forest for all the six models, displaying the student-level variables in order of importance and the respective value of the percentage increment of MSE ($Inc\%MSE$). Results in Table 3, together with the plots reported in Figure 4 in Appendix A, show that, in terms of importance, the first five variables are able to explain a major part of variability in the response in all the models, whereas the other covariates have a limited and similar value of Inc%MSE. Therefore, to improve the visibility and support a smooth interpretation of the results, only the five most important variables are displayed in the ribbon chart (but the complete list can be found in Table 3).

Table 3: Ranking of the student level variables according to the Random Forest variable importance, for the six model specifications.

	Group 1		**Group 2**		**Group 3**	
	Female	**Male**	**Female**	**Male**	**Female**	**Male**
Student's family charactersitics						
Mother edu	5	4	6	6	5	5
	(22.773)	(21.640)	(15.511)	(15.713)	(10.499)	(11.609)
Parent support	14	16	12	12	8	8
	(4.813)	(3.509)	(6.810)	(7.941)	(5.725)	(9.433)
ESCS	1	1	1	1	1	1
	(53.111)	(50.948)	(58.689)	(59.206)	(33.781)	(30.771)
Foreign language	15	13	14	7	6	6
	(3.682)	(6.732)	(5.585)	(14.919)	(8.459)	(9.954)
ICT resources	7	7	8	9	4	9
	(10.770)	(12.241)	(10.039)	(13.548)	(11.439)	(7.468)
Student's perceptions and attitudes						
Fear failure	16	15	15	16	11	11
	(3.328)	(3.668)	(4.162)	(4.160)	(4.972)	(6.783)
Feel awkward	11	8	10	13	15	10
	(6.044)	(11.537)	(8.827)	(7.633)	(4.012)	(7.261)
Feel outsider	9	11	13	10	9	4
	(6.858)	(8.804)	(6.690)	(8.610)	(5.485)	(13.151)
Self confidence	10	12	16	15	7	7
	(6.064)	(7.650)	(3.174)	(6.350)	(7.687)	(9.532)
Sociable	8	14	9	14	13	12
	(7.514)	(6.100)	(9.811)	(7.189)	(4.779)	(6.639)
Enjoyment reading	3	5	4	5	2	2
	(28.026)	(17.202)	(24.107)	(15.801)	(21.917)	(15.872)
Class and school environment						
Teach support	13	10	11	11	16	14
	(5.391)	(9.793)	(7.762)	(8.410)	(3.382)	(4.484)
Discipline language class	6	6	7	8	10	13
	(11.564)	(12.506)	(12.214)	(14.189)	(5.063)	(5.744)
Longest book	2	3	2	2	3	3
	(33.932)	(25.663)	(35.136)	(32.190)	(17.019)	(14.790)
Class size	12	9	3	4	14	16
	(5.899)	(10.854)	(26.130)	(25.284)	(4.318)	(1.614)
Perception cooperation	4	2	5	3	12	15
	(26.196)	(26.195)	(23.954)	(26.604)	(4.866)	(3.713)
% explained variance-RF	23.74	22.78	30.14	29.28	18.36	21.61
PVRE	10.07	7.81	8.52	8.58	0.82	1.55
MSE	0.587	0.597	0.281	0.296	0.303	0.317

Note: The Table reports the position of the student level variables in the ranking provided by Random Forest models and the associated value of $Inc\%MSE$ (in brackets). Note that the higher the variable importance, the higher the position in the ranking. The ranking position covered by the most important variable is 1. The Table presents results for all the six models, by gender (males and females) and by gender gap group, where Group 1 represents countries where males outperform females; Group 2, the countries with no gap; and Group 3, the countries where females outperform males. The bottom part of the Table reports the percentage of variability explained by the random forest in the multilevel model, the PVRE, and the Mean Square Error (MSE) of each model.

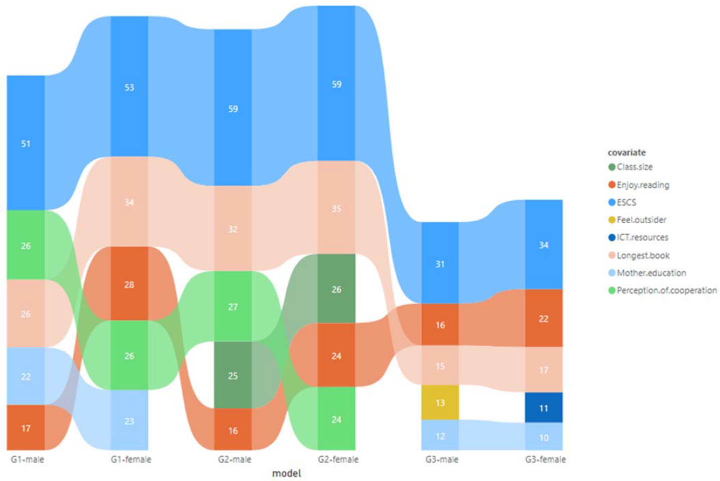
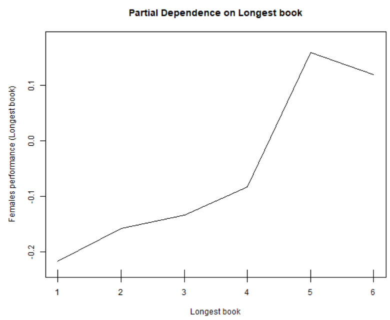


Figure 2: Summary of the five most important variables in the Random Forest, for the six models based on $Inc\%MSE$ – shown in the labels.

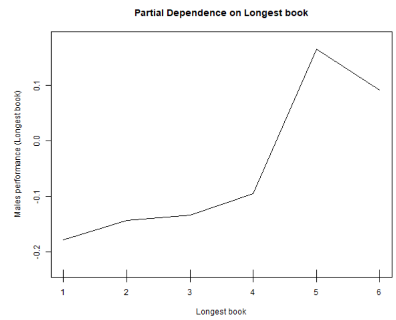
By comparing the results of the six models, it is possible to identify a set of variables that represents the main determinants independently from the group of countries and gender. First, student socioeconomic status (ESCS) is the most important variable in influencing math achievement in all the models. This result is in line with the findings in the literature, which identifies socioeconomic disparities as the main determinant of differences in mathematics performance (Martins and Veiga, 2010). The lower absolute importance associated with ESCS in Group 3 is also not surprising. Indeed, countries such as Finland and Iceland are generally characterised by higher social equality with, as a consequence, a lower influence of family background on educational performance (Martins and Veiga, 2010).

Besides socioeconomic status, the longest book that students have to read for school often covers one of the first positions in the variables importance ranking of Table 3. Reading long books or texts seems particularly important for countries with no gender gap (Group 2) and where boys perform better than girls (Group 1). In particular, when teachers require to read books longer than 100 pages, students tend to achieve higher test scores in math, both for boys and girls (see Panels a and b in Figure 3.[1]). This highlights the importance of teaching behaviours in supporting mathematics learning, even in not-strictly scientific subjects. Indeed, being able to read long texts implies several transversal skills that can support the mathematical competencies of kids, such as the capacity to engage and concentrate over a substantial period of time (Moss and McDonald (2004)), or the ability to interpret questions and texts of a mathematical problem (Jerrim et al., 2020).

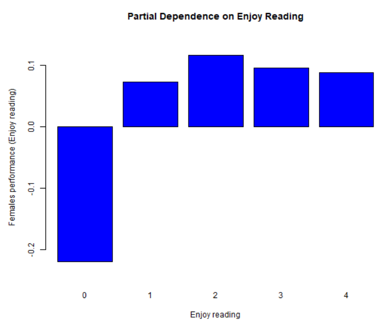
On the same line, the free time spent on reading is also an important covariate, especially in countries where girls have higher math achievement than boys (in Group 3, the variable represents the second position for importance). In this case, the variable does not refer to a teacher's requirement, but it is associated with the personal attitude of the student. The partial plots in Panels c and d of Figure 3 show how math achievement varies depending on the enjoyment of reading of boys and girls in Group 1. The figure reveals that students reporting low enjoyment for reading are associated with significantly lower achievement in math, whereas students reading during their free time perform better-regardless of how much time spent on reading. Comparing the plots across gender, we may notice that the negative effect associated with no enjoyment of reading (category 0) has a considerably larger extent for girls than for boys. This result is in line with the ranking of importance of Table 3, which shows that enjoyment of reading is more important for female students than for males (especially in Group 1). Moreover, this finding could be related to the difference in the number of observations in category 0 between females and males: only 28% of girls reported that they do not read for enjoyment, while for the boys it is 51%. Therefore, it is likely that category 0 of Enjoyment reading is more precise in capturing lower performers among girls than boys.



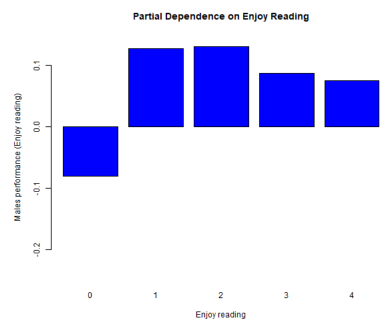
(**a**)



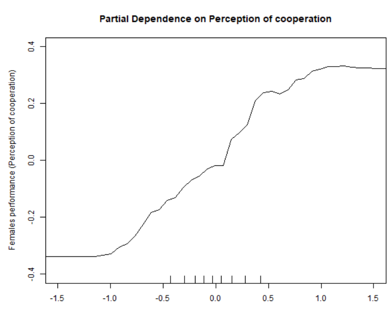
(**b**)



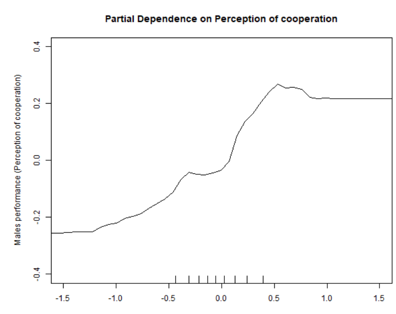
(**c**)



(**d**)



(**e**)



(**f**)

Figure 3: Partial plots of Longest Book, Enjoy Reading, and Perception of Cooperation for Group 1 models. (**a**) Partial plot of Longest Book for females. (**b**) Partial plot of Longest Book for males. (**c**) Partial plot of Enjoy Reading for females. (**d**) Partial plot of Enjoy Reading for males. (**e**) Partial plot of Perception of Cooperation for females. (**f**) Partial plot of Perception of Cooperation for males.

Perception of cooperation is also relevant in explaining math achievement, especially in countries where males perform better or the same as females. Interestingly, the importance of this determinant is higher for boys than for girls. Indeed, while this is the second (Group 1) or third (Group 2) variable in terms of importance for male students, the position in the ranking is lower for females. On the other hand, the partial plots in Panels e and f of Figure 3 reveal that, for students reporting very high values of perception of cooperation, the increase in the math scores associated with this covariate is much higher for girls than for boys. Moreover, the partial plot for males achieves a steady level (thus, an absence of correlation with performance) more quickly than the one for females. The finding seems to imply that girls benefit from a climate of cooperation when this is significantly high, whereas boys can perform well also with lower levels of cooperation.

Another consideration emerging from the results concerns school-level factors. Variables such as class size and discipline in class are particularly important for countries without a gender gap (i.e., Group 2) but not for the other two groups. In countries where a gender gap is found (Groups 1 and 3), family and personal characteristics of students, such as mother's education, feeling outsider, and ICT resources, play, instead, a more relevant role. This seems to suggest that, in countries without a gender gap, school factors are able to mitigate the influence of family background and student characteristics to a larger extent. In addition, it should be considered that the strong influence of some family characteristics, such as mother's education, may be strengthened by cultural and socioeconomic differences between the groups of countries.

Finally, the results reveal some determinants playing a marginal role across groups of countries. This is particularly interesting when the variables reporting higher differences between boys and girls are examined, such as for ICT resources and fear of failure (see Table 2). Indeed, even if female students reported higher fear of failure and lower use of ICT resources at home (Groups 1 and 2), these gender disparities are unlikely to be translated into a significant gap in math achievement.

In terms of models' performances, the percentage of variability explained by the random forest ranges between 19.34% (relative to the model for females in Group 3) and 30.23% (relative to females in Group 2). Models for Group 2 provide the lowest MSE, while models in Group 1 provide the highest one. Globally, the models for Group 2, where there is no evidence of a gender gap, appear to be the best ones in terms of percentage of explained variability and MSE, both for males and females.

4.2.2 Country Level

The inclusion of a country-level random intercept allows us to estimate the percentage of the unexplained variability in the students' performance that is given to the country level, measured by the PVRE (reported in the bottom part of Table 3). For Groups 1 and 2, the heterogeneity across countries is fairly high, being the PVRE between 8 and 10% for both boys and girls. On the opposite, the PVRE in the two models for Group 3 (where females outperform males) is very low (less than 2%). This is partly due to the fact that in Group 3 we only observe three countries and the heterogeneity across them is relatively small. Figure 5 in Appendix A displays the estimated random intercepts associated with each country, for both males and females. We notice that the countries' effect is typically similar between boys and girls, i.e., countries have the same type of impact, either positive or negative, on both males' and females' performance. On the other hand, the high values of PVRE in Groups 1 and 2 reflect some differences that can be observed across countries within the same group. For instance, in Group 1, other characteristics being equal, Belgium is the country associated with the highest student performance in math, both for males and females, whereas the lowest scores are, by far, in Romania. In Group 2, Lithuania is associated with the lowest student scores, with a random intercept remarkably distant from the rest of the group. Finally, concerning the countries in Group 3, Malta is the country with the highest performance, and Iceland is the one with the lowest achievement in math (again, controlling for the rest of individual, family, and school features).

5 Discussion

Education is one of the most powerful tools to promote equality of opportunity and favour inter-generational mobility, despite any socioeconomic characteristics or disadvantaged background (Torche, 2015). Better education reduces criminality, fosters cooperation, and is associated with higher salaries, health, and life satisfaction. For these reasons, it is of foremost importance to guarantee the benefit of education to all children, especially the ones belonging to disadvantaged backgrounds or minorities (Lee, 2012).

Despite evidence that shows that males and females have similar learning abilities when entering school, in most countries males and females perform significantly differently. This is particularly relevant in mathematics, given the long-run implication on job opportunities and salaries (Frye and Levitt, 2010; Borgonovi et al., 2018). In this paper, we investigate the determinants of student performance by gender, with the purpose of identifying the possibly heterogeneous mechanisms that enhance or hinder pupils' learning. By acknowledging the existence of different learning needs, the educational systems can localise solutions to meet the necessity of every student and to boost the equalising role of schooling.

Our analysis relies on a multilevel random forest estimation. This approach gives us the possibility to combine the advantages of multilevel analysis and that of random forest techniques, i.e., it allows us to account for country-level heterogeneity, while estimating the effects of multiple covariates in a flexible and interpretable way.

In line with previous evidence in the literature, our findings show that around 8–10% of performance variability (within the male and female groups) is explained by country-level variation, especially among the groups of countries where a gender gap is observed (Group 1 and 3), which are the highest portion of countries. Moreover, as the main element of novelty, our results reveal the way through which some key factors influence math achievement can be significantly different between males and females. In terms of heterogeneity in the determinants of performance for boys and girls, four points summarise our findings. First, as previous evidence has highlighted (Broer et al., 2019), the socioeconomic background of the students is the most relevant factor that influences student achievement, especially in countries where a gender gap is observed (Group 1 and 3). However, its influence decreases when no gender gap is observed; thus, more equal countries from a socioeconomic standpoint are also those where gender equality is also more pronounced. In this respect, results suggest the importance of addressing social and educational equality overall. Second, results point out that reading (both in terms of school assignments or extra-curricula activity) is an important determinant of mathematics performance, especially for females. This finding is in line with previous evidence (Breda and Napp, 2019), supporting the idea that closing the gender gap in mathematics is also a matter of reading abilities.

Third, the perception of cooperation is an important variable in countries characterised by better performance of boys (Group 1) or no significant difference in male and female performances (Group 2). On average, this finding holds especially for boys, while girls are positively affected by the perception of cooperative learning in the school when this perception is particularly high. This positive correlation is in line with previous findings on the importance of perceiving schools as a cooperative environment (Ghaith, 2002; OECD, 2019b). Finally, results highlight how school factors are more relevant for students' results in countries where the gender gap is less pronounced. This finding suggests how the more the school and the educational system overall work to achieve a more equitable environment, the more the importance of school factors in influencing performance increases (Lee et al., 1997).

6 Conclusions

Overall, our results indicate that, to boost the equalising role of education and achieve equal opportunity in a globalised world, it is central to reduce cross-country performance variations. Moreover, evidence indicates that schools and teachers can foster students' learning by involving and motivating students in reading activities and by promoting cooperation. Finally, it is also important to note that schools should pay more attention to providing students with the tools they need to culturally emancipate themselves, despite the socioeconomic background of their families. Indeed, our results stress the relevance of individual perceptions and self-beliefs to support student performance.

In terms of future research, it would be interesting to compare results related to European countries with other international contexts, in order to observe how the cultural and educational systems can affect the results.

References

Ajello, Anna Maria, Giulia Accardi, Stefano Aprile, Rosalia Caldarella, Ciriaco Carru, Marcello Ciaccio, Immaculata De Vivo, Caterina Maria Gambino, Mattia Emanuela Ligotti, and Sonya Vasto. 2021. Age and gender-related variations of molecular and phenotypic parameters in a cohort of sicilian population: from young to centenarians. *Aging and disease* 12: 1773.

Ajello, Anna Maria, Elisa Caponera, and Laura Palmerio. 2018. Italian students' results in the pisa mathematics test: does reading competence matter? *European Journal of Psychology of Education* 33: 505–20.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth and Melanie Wasserman, 2019. Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics* 11: 338–81.

Ayalon, Hanna, and Idit Livneh. 2013. Educational standardization and gender differences in mathematics achievement: A comparative study. *Social Science Research* 42: 432–45.

Beede, D. N.; Julian, T. A.; Langdon, D.; McKittrick, G.; Khan, B.; Doms, M. E. Women in STEM: A Gender Gap to Innovation. *SSRN Electron. J.* **2011**. doi:10.2139/ssrn.1964782

Bertocchi, Graziella, and Monica Bozzano. 2020. *Gender Gaps in Education*. Cham: Springer International Publishing.

Borgonovi, Francesca, Álvaro Choi, and Marco Paccagnella. 2021. The evolution of gender gaps in numeracy and literacy between childhood and adulthood. *Econ. Educ. Rev.* doi:10.1007/978-3-319-57365-6_26-1

Bouffard, Suzanne M., and Nancy E. Hill. 2005. Maternal perceptions of competence and children's academic adjustment: Longitudinal relations across early elementary school. *Social Psychology of Education* 8: 441–63.

Breda, Thomas, Elyès Jouini, and Clotilde Napp. 2018. Societal inequalities amplify gender gaps in math. *Science* 359: 1219–20.

Breda, Thomas, and Clotilde Napp. 2019. Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. *Proceedings of the National Academy of Sciences* 116: 15435–40.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45: 5–32.

Brenøe, Anne Ardila, and Shelly Lundberg. 2018. Gender gaps in the effects of childhood family environment: Do they persist into adulthood? *European Economic Review* 109: 42–62.

Broer, Markus, Yifan Bai, and Frank Fonseca. A review of the literature on socioeconomic status and educational achievement. In *Socioeconomic Inequality and Educational Outcomes*. Cham: Springer International Publishing. 2019; vol. 5, pp. 7–17. doi:10.1007/978-3-030-11991-1_2

Buchmann, Claudia, Thomas A. DiPrete, and Anne McDaniel. 2008. Gender inequalities in education. *Annu. Rev. Sociol* 34: 319–37.

Card, David, and A. Abigail Payne. 2021. High school choices and the gender gap in stem. *Economic Inquiry* 59: 9–28.

Cascella, Clelia, Julian Scott Williams, and Maria Pampaka. 2022. Gender differences in mathematics outcomes at different levels of locality to inform policy and practice. *European Educational Research Journal* 21: 705–31 doi:10.1177/1474904121997211

Caviola, Sara, Enrico Toffalini, David Giofrè, Jessica Mercader Ruiz, Dénes Szűcs, and Irene C. Mammarella. 2022. Math performance and academic anxiety forms, from sociodemographic to cognitive aspects: A meta-analysis on 906,311 participants. *Educational Psychology Review* 34: 363–99.

Chaman, Mini J., Kim Beswick, and Rosemary Callingham. 2014. Factors influencing mathematics achievement among secondary school students: A review. *The Future of Educational Research* SensePublishers: Rotterdam, pp. 227–38. doi:10.1007/978-94-6209-512-0_19

Close, Sean, and Gerry Shiel. 2009. Gender and pisa mathematics: Irish results in context. *European Educational Research Journal* 8: 20–33.

Contini, Dalit, Maria Laura Di Tommaso, and Silvia Mendolia. 2017. The gender gap in mathematics achievement: Evidence from italian data. *Economics of Education Review* 58: 32–42.

Cuevas-Ruiz, Pilar, Cristina Borra, and Almudena Sevilla. 2020. Educación y salud al nacer. *Papeles de Economía Española* 166: 185–202. Available online: https://www.funcas.es/wp-content/uploads/2021/01/Pilar-Cuevas-Cristina-Borra-Almudena-Sevilla.pdf (accessed on 21 June 2022)

Cvencek, Dario, Andrew N. Meltzoff, and Manu Kapur. 2014. Cognitive consistency and math–Gender stereotypes in singaporean children. *Journal of Experimental Child Psychology* 117: 73–91.

De Witte, Kristof, and Mika Kortelainen. 2013. What explains the performance of students in a heterogeneous environment? conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics* 45: 2401–12.

Dossi, Gaia, David Figlio, Paola Giuliano, and Paola Sapienza. 2021. Born in the family: preferences for boys and the gender gap in math. *Journal of Economic Behavior & Organization* 183: 175–88.

Dunnzlaff, Lina, Dirk Neumann, Judith Niehues, and Andreas Peichl. 2011. In *Equality of Opportunity and Redistribution in Europe*. Bingley: Emerald Group Publishing Limited.

Education, European, Culture Executive Agency, and Eurydice. 2012. *Gender Differences in Educational Outcomes: Study on the Measures Taken and the Current Situation in Europe*. Publications Office. Available online: `https://data.europa.eu/doi/10.2797/3598` (accessed on 21 June 2022) .

Else-Quest, Nicole M., Janet Shibley Hyde, and Marcia C. Linn. 2010. Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin* 136: 103.

Evans, David K., Maryam Akmal, and Pamela Jakiela. 2020. *Gender Gaps in Education: The Long View*. Technical Report. Washington, DC: Center for Global Development.

Flore, Paulette C., and Jelte M Wicherts. 2015. Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology* 53: 25–44.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Springer series in statistics. New York: Springer, vol. 1.

Fryer, Roland G., Jr., and Steven D. Levitt. 2010. An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics* 2: 210–40.

Gevrek, Z. Eylem, Deniz Gevrek, and Christian Neumeier. 2020. Explaining the gender gaps in mathematics achievement and attitudes: The role of societal gender equality. *Economics of Education Review* 76: 101978.

Ghaith, Ghazi M. 2002. The relationship between cooperative learning, perception of social support, and academic achievement. *System* 30: 263–73.

Gibbs, Benjamin G. 2010. Reversing fortunes or content change? gender gaps in math-related skill throughout childhood. *Social Science Research* 39: 540–69.

Giuliano, Paola. 2020. Gender and culture. *Oxford Review of Economic Policy* 36: 944–61.

González de San Román, Ainara, and Sara De La Rica. 2012. Gender gaps in pisa test scores: The impact of social norms and the mother's transmission of role attitudes. IZA DP No. 6338. Available online: `https://docs.iza.org/dp6338.pdf` (accessed on 21 June 2022)

Halpern, Diane F., and Simay Ikier. 2002. in *Biology, Society, and Behavior: The Development of Sex Differences in Cognition*. Westport: Ablex Publishing, pp. 3–19.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Berlin/Heidelberg: Springer, vol. 112.

Jaremus, Felicia, Jennifer Gore, Elena Prieto-Rodriguez, and Leanne Fray. 2020. Girls are still being 'counted out': Teacher expectations of high-level mathematics students. *Educational Studies in Mathematics* 105: 219–36.

Jerrim, John, Luis Alejandro Lopez-Agudo, and Oscar D. Marcenaro-Gutierrez. 2020. Does it matter what children read? new evidence using longitudinal census data from spain. *Oxford Review of Education* 46: 515–33.

Lee, Jaekyung. 2012. Educational equity and adequacy for disadvantaged minority students: School and teacher resource gaps toward national mathematics proficiency standard. *The Journal of Educational Research* 105: 64–75.

Lee, Valerie E., Julia B. Smith, and Robert G. Croninger. 1997. How high school organization influences the equitable distribution of learning in mathematics and science. *Sociology of Education* 70 (2), pp. 128–50.

Levine, Susan C., Marina Vasilyeva, Stella F. Lourenco, Nora S. Newcombe, and Janellen Huttenlocher. 2005. Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science* 16: 841–45.

Lewis, Roger J. 2000. An introduction to classification and regression tree (cart) analysis. Paper presented at Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California, San Francisco, CA, USA. May 22–25.

Machin, Stephen, and Patrick A Puhani. 2003. Subject of degree and the gender wage differential: evidence from the uk and germany. *Economics Letters* 79: 393–400.

Marsh, Herbert W., and Andrew J. Martin. 2011. Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology* 81: 59–77.

Martins, Lurdes, and Paula Veiga. 2010. Do inequalities in parents' education play an important role in pisa students' mathematics achievement test score disparities? *Economics of Education Review* 29: 1016–33.

Masci, Chiara, Geraint Johnes, and Tommaso Agasisti. 2018. Student and school performance across countries: A machine learning approach. *European Journal of Operational Research* 269: 1072–85.

Mejias, Paulina Perez, Dora Elias McAllister, Karina G. Diaz, and Javiera Ravest. 2021. A longitudinal study of the gender gap in mathematics achievement: evidence from chile. *Educational Studies in Mathematics* 107: 583–605.

Mizala, Alejandra, Francisco Martínez, and Salomé Martínez. 2015. Pre-service elementary school teachers' expectations about student performance: How their beliefs are affected by their mathematics anxiety and student's gender. *Teaching and Teacher Education* 50: 70–78.

Moss, Gemma, and John W. McDonald. 2004. The borrowers: Library records as unobtrusive measures of children's reading preferences. *Journal of Research in Reading* 27: 401–12.

Munir, Farzana, and Rudolf Winter-Ebmer. 2018. Decomposing international gender test score differences. *Journal for Labour Market Research* 52: 1–17.

OECD. 2016. *PISA 2015 Results (Volume I).* Paris: OECD. https://doi.org/10.1787/9789264266490-en.

OECD. 2019a. *PISA 2018 Results (Volume II).* Paris: OECD. https://doi.org/10.1787/b5fd1b8f-en.

OECD. 2019b. *PISA 2018 Results (Volume III): What School Life Means for Students' Lives.* Paris: OECD.

Pajares, Frank. 2005. *Gender Differences in Mathematics Self-Efficacy Beliefs.* Cambridge: Cambridge University Press.

Pellagatti, Massimo, Chiara Masci, Francesca Ieva, and Anna M. Paganoni. 2021. Generalized mixed-effects random forest: A flexible approach to predict university student dropout. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 14: 241–57.

Piazzalunga, Daniela. 2018. The gender wage gap among college graduates in italy. *Italian Economic Journal* 4: 33–90.

Rainey, Katherine, Melissa Dancy, Roslyn Mickelson, Elizabeth Stearns, and Stephanie Moller. 2019. A descriptive study of race and gender differences in how instructional style and perceived professor care influence decisions to major in stem. *International Journal of STEM Education* 6: 1–13.

Schiltz, Fritz, Chiara Masci, Tommaso Agasisti, and Dániel Horn. 2017. Using Machine Learning to Model Interaction Effects in Education: A Graphical Approach. MOX-Report No. 31/2017. Available online: `https://www.mate.polimi.it/biblioteca/add/qmox/31-2017.pdf` (accessed on 21 June 2022).

Schleicher, Andreas. 2019. *Pisa 2018: Insights and Interpretations.* Paris: OECD Publishing.

Sikora, Joanna, and David GW Pitt. 2019. Does advanced mathematics help students enter university more than basic mathematics? gender and returns to year 12 mathematics in australia. *Mathematics Education Research Journal* 31: 197–218.

Starr, Christine R., and Sandra D. Simpkins. 2021. High school students' math and science gender stereotypes: relations with their stem outcomes and socializers' stereotypes. *Social Psychology of Education* 24: 273–98.

Steinthorsdottir, Olof Bjorg, and Bharath Sriraman. 2008. Exploring gender factors related to pisa 2003 results in iceland: a youth interview study. *ZDM* 40: 591–600.

Tomassini, Cecilia. 2021. Gender gaps in science: Systematic review of the main explanations and research agenda. *Education in the Knowledge Society (EKS)* 22: 25.

Torche, Florencia. 2015. Analyses of intergenerational mobility: An interdisciplinary review. *The ANNALS of the American Academy of Political and Social Science* 657: 37–62.

UNICEF. 2020. Towards an equal future: Reimagining girls' education through stem. Available online: https://www.unicef.org/media/84046/file/Reimagining-girls-education-through-stem-2020.pdf (accessed on 25 June 2022)

van Hek, Margriet, Claudia Buchmann, and Gerbert Kraaykamp. 2019. Educational systems and gender differences in reading: A comparative multilevel analysis. *European Sociological Review* 35: 169–86.

Wang, Yurou, Mengya Xia, Wenjing Guo, Fangjie Xu, and Yadan Zhao. 2022. Academic performance under covid-19: The role of online learning readiness and emotional competence. *Current Psychology* doi:10.1007/s12144-022-02699-7 .

Zhou, Yisu, Xitao Fan, Xiaoxin Wei, and Robert H. Tai. 2017. Gender gap among high achievers in math and implications for stem pipeline. *The Asia-Pacific Education Researcher* 26: 259–69.

A

Table 4: Details of gender differences among countries by PISA math scores.

Country	*p*.Value	Mean Score for Females	Mean Score for Males	# obs.	Group
AUT	0	496.979	507.410	6802	better male
BEL	0	541.043	560.719	4888	better male
BGR	0.004	464.893	475.523	2739	better male
HRV	0	508.091	541.658	2094	better male
CZE	0.546	533.560	531.863	4764	no gap
DNK	0.00004	493.128	501.081	7656	better male
EST	0.005	519.645	525.908	5315	better male
FIN	0.005	510.455	504.353	5648	better female
FRA	0.00002	499.445	510.525	4992	better male
DEU	0.002	499.226	507.447	5305	better male
GRC	0.0003	462.327	470.351	5599	better male
HUN	0	494.471	511.977	4294	better male
ISL	0.008	498.026	489.664	3296	better female
IRL	0.005	498.221	504.029	5536	better male
ITA	0	508.668	545.193	5744	better male
LVA	0.001	490.650	498.148	5259	better male
LTU	0.480	481.024	482.566	6758	no gap
LUX	0.0004	489.381	500.135	4123	better male
MLT	0.0003	480.116	467.605	3363	better female
NLD	0.0001	547.000	557.371	3379	better male
POL	0.204	515.407	518.377	5616	no gap
PRT	0	496.962	510.137	4965	better male
ROU	0	435.553	449.897	4437	better male
SVK	0.775	495.028	494.051	3900	no gap
SVN	0.00005	545.966	560.069	2221	better male
ESP	0	488.831	495.895	35,599	better male
CHE	0.389	513.383	515.740	4841	no gap
GBR	0	491.066	498.559	13,762	better male

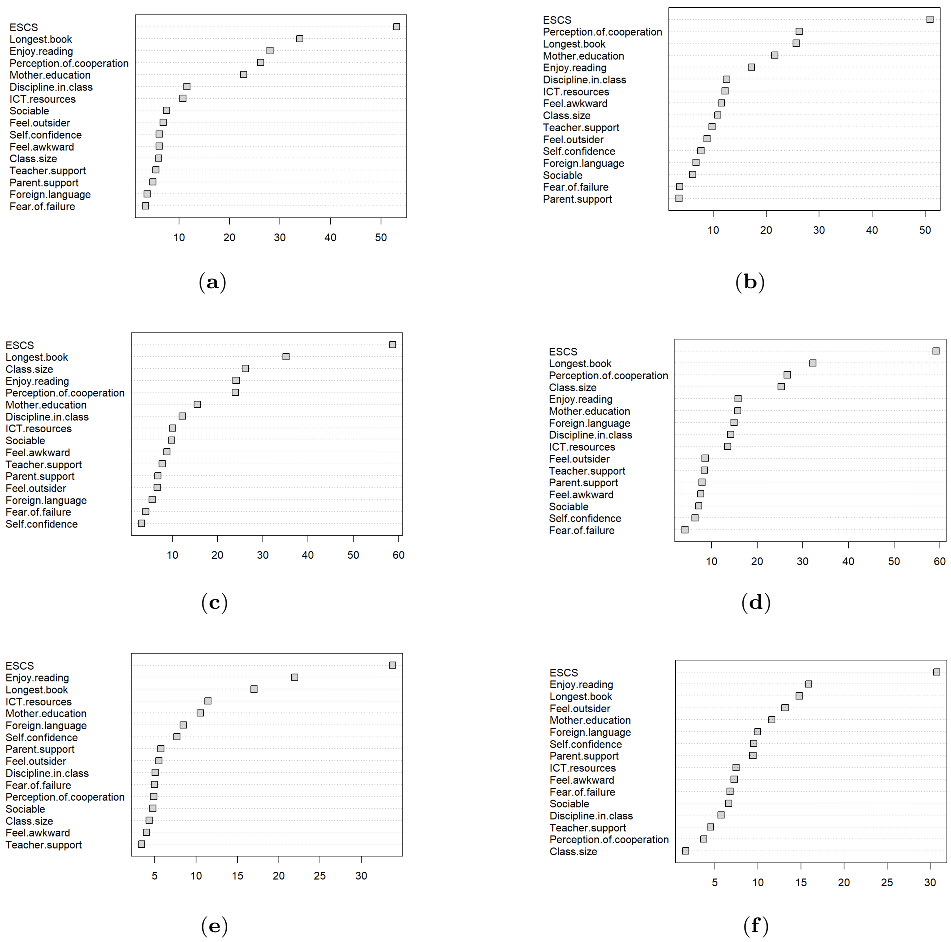


Figure 4: Variable importance plots. (**a**) Females – Countries where males perform better than females. (**b**) Males – Countries where males perform better than females. (**c**) Females – Countries with no gap. (**d**) Males – Countries with no gap. (**e**) Females – Countries where females perform better than males. (**f**) Males – Countries where females perform better than males.

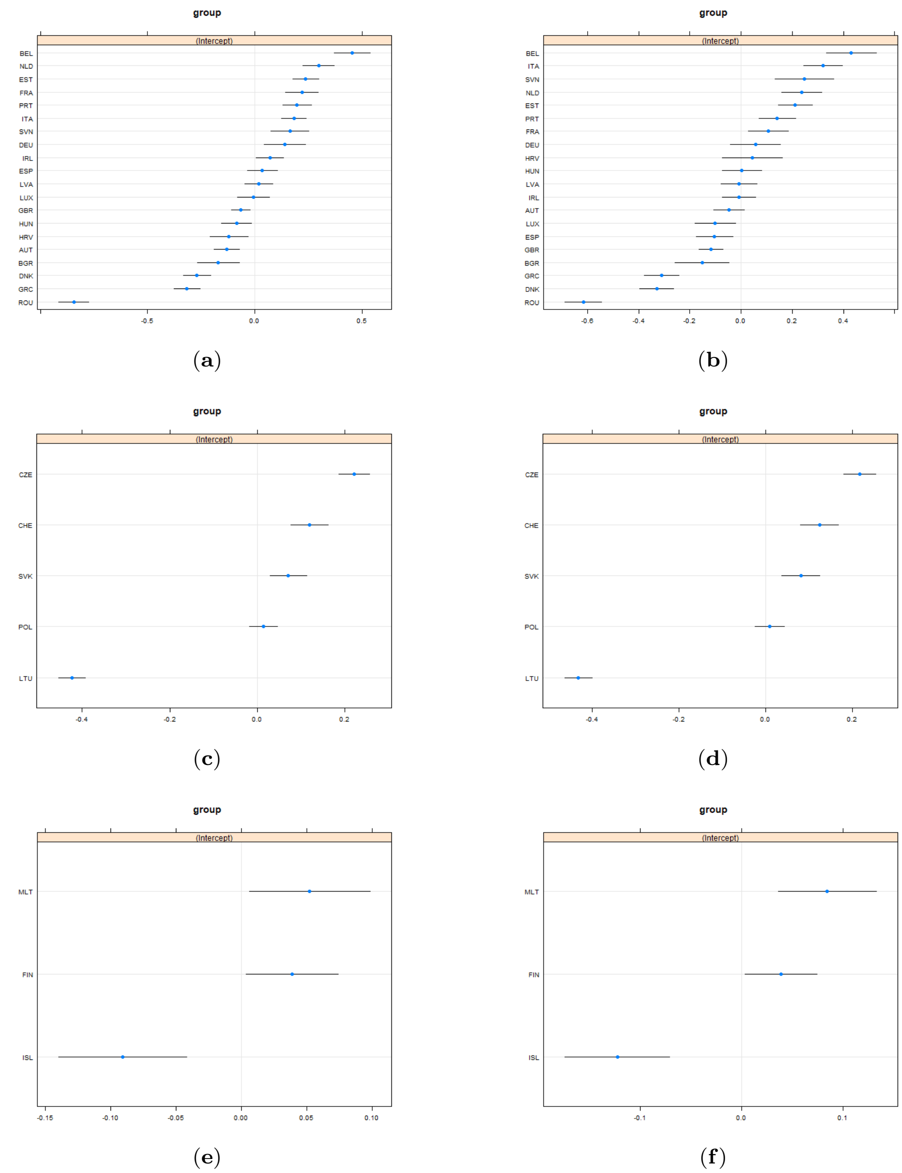


Figure 5: Dotplot of country effect. (**a**) Females−Countries where males perform better than females.(**b**) Males−Countries where males perform better than females. (**c**) Females−Countries with no gap. (**d**) Males −Countries with no gap. (**e**) Females−Countries where females perform better than males. (**f**) Males−Countries where females perform better than males.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

22/2023 Su, Y.; Riccobelli, D.; Chen, Y; Chen, W; Ciarletta, P
Designing dielectric-elastomer balloons with tunable shapes in response to electro-mechanical stimuli

21/2023 Cavinato, L.; Sollini, M.; Ragni, A.; Bartoli, F.; Zanca, R.; Pasqualetti, F.; Marciano, A.; Ieva, F.; and Erba, A.P.
Radiomics-based Inter-lesion Relation Network to Describe [18F]FMCH PET/CT Imaging Phenotypes in Prostate Cancer

17/2023 Savin, M.S.; Cavinato, L.; Costa, G.; Fiz, F.; Torzilli, G.; Vigano', L.; Ieva, F.
Distant supervision for imaging-based cancer sub-typing in Intrahepatic Cholangiocarcinoma

19/2023 Marcinno', F.; Vergara, C.; Giovannacci, L.; Quarteroni, A.; Prouse, G.
Computational fluid-structure interaction analysis of the end-to-side radio-cephalic arteriovenous fistula

20/2023 Ciaramella, G.; Nobile, F.; Vanzan, T.
A multigrid method for PDE-constrained optimization with uncertain inputs

15/2023 Ragni, A.; Masci, C.; Ieva, F.; Paganoni, A. M.
Clustering Hierarchies via a Semi-Parametric Generalized Linear Mixed Model: a statistical significance-based approach

13/2023 Masci, C.; Cannistrà, M.; Mussida, P.
Modelling time-to-dropout via Shared Frailty Cox Models. A trade-off between accurate and early predictions

11/2023 Gatti, F.; Perotto, S.; de Falco, C.; Formaggia, L.
A positivity-preserving well-balanced numerical scheme for the simulation of fast landslides with efficient time stepping

10/2023 Corti, M.; Antonietti, P.F.; Bonizzoni, F.; Dede', L., Quarteroni, A.
Discontinuous Galerkin Methods for Fisher-Kolmogorov Equation with Application to Alpha-Synuclein Spreading in Parkinson's Disease

09/2023 Buchwald, S.; Ciaramella, G.; Salomon, J.
Gauss-Newton oriented greedy algorithms for the reconstruction of operators in nonlinear dynamics