IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10625

Let the Girls Learn! It is not *Only* about Math… It's about Gender Social Norms

Natalia Nollenberger
Núria Rodríguez-Planas

MARCH 2017

DISCUSSION PAPER SERIES

IZA DP No. 10625

Let the Girls Learn! It is not *Only* about Math… It's about Gender Social Norms

Natalia Nollenberger
IE University

Núria Rodríguez-Planas
City University of New York (CUNY), Queens College and IZA

MARCH 2017

ABSTRACT

Let the Girls Learn! It is not *Only* about Math… It's about Gender Social Norms*

Using PISA test scores from 11,527 second-generation immigrants coming from 35 different countries of ancestry and living in 9 host countries, we find that the positive effects of country-of-ancestry gender social norms on girls' math test scores relative to those of boys: (1) expand to other subjects (namely reading and science); (2) are shaped by beliefs on women's political empowerment and economic opportunity; and (3) are driven by parents' influencing their children's (especially their girls') preferences. Our evidence further suggest that these findings are driven by cognitive skills, suggesting that social gender norms affect parent's expectations on girls' academic knowledge relative to that of boys, but not on other attributes for success--such as non-cognitive skills. Taken together, our results highlight the relevance of general (as opposed to math-specific) gender stereotypes on the math gender gap, and suggest that parents' gender social norms shape youth's test scores by transmitting preferences for cognitive skills.

JEL Classification: I21, I24, J16, Z13

Keywords: gender gap in math, reading and science, immigrants, beliefs and preferences, cognitive and non-cognitive skills, culture and institutions

Corresponding author:
Núria Rodríguez-Planas
Economics Department
CUNY Queens College
306-G Powdermaker Hall
65-30 Kissena Blvd.
Queens, NY 11367
USA
E-mail: nuria.rodriguezplanas@qc.cuny.edu

* This paper has benefitted from early discussions with Almudena Sevilla, as well as comments from participants of the 2015 COSME Gender Economic Workshop, the 2014 ESPE conference, the 9th IZA/World Bank Conference on Employment and Development, the 2014 IZA/SOLE Trasatlantic Meeting of Labor Economists, and seminars at San Diego State University, Syracuse University, and the Inter-American Development Bank.

"Maybe it means telling your sons that it's okay to cry, and your daughters that it's okay to be bossy. Maybe it means encouraging your daughters, not just your son, to study math and science and sign up for the football team. And if there isn't a team for girls, maybe it means asking why not.

That's how all of you will begin to break down those old stereotypes and biases. That's how you'll change the way that women and girls are seen. And that's the kind of work that we need to be doing around the world –- the work of changing culture. The work of changing expectations and standards that we have for women and girls."

First Lady of the United States of America, Michelle Obama at the "Let the Girls Learn" conference in Madrid, June 30, 2016

1. Introduction

Using PISA data on close to 12,000 second-generation immigrants from 35 different countries of ancestry and living in 9 host countries, Nollenberger, Rodríguez-Planas and Sevilla (2016) present evidence of the persistence of culture and how it affects the math gender gap—the relative underperformance of girls at math test scores.[1] More precisely, they find that the math gender gap decreases for second-generation immigrants whose parents come from more gender-equal countries.[2] According to their findings, "*a one standard deviation increase in the* (county-of-ancestry) *gender equality index is associated with a reduction of 7.47 score points in the math gender gap* (in the host county)*"*, which represents a 0.29 of the standard deviation in the math gender gap.[3]

The current paper addresses a different but highly critical and policy-relevant question for the cognitive performance of girls. How much of girls' relative underperformance at math test scores is explained by math-specific

[1] Much of the research documenting gender gap in math scores has been based upon US data. The size of the gap reported depends on the test and time-period. Some recent studies suggest that the average gender gap in math scores among teenagers has been narrowing (Hyde and Mertz 2009), while others document persisting large differences in the average performance of girls relative to boys (Fryer and Levitt 2010). There is a wide consensus that substantial differences persist at the top of the distribution (Ellison and Swanson 2010; Hyde and Mertz 2009) and that the fraction of males to females who score in the top 5 percent of the distribution in high-school math has remained constant at two to one over the past 20 years (Xie and Shauman 2003). Ellison and Swanson (2010) document that the gender gap in secondary-school math at high-achievement levels is present in every US high school, although the size of the gap varies between schools. Bedard and Cho (2010) review the existing evidence documenting gender gap in math scores in OECD countries. Guiso *et al.* (2008) document that "*girls' math scores average 10.5 (or 2 percent) lower score points than those of boys",* using PISA data as in this paper.

[2] Second-generation immigrants are defined as individuals born in country they live in to parents (both of them) born in a different country.

[3] Others had previously highlighted the relevance of societal factors in explaining the math gender gap (Guiso et al. 2008; Pope and Sydnor 2010; González de San Román and de la Rica 2016), but they did not disentangle the role of culture defined as social norms or beliefs transmitted "*fairly unchanged from generation to generation*" (Guiso, Sapienza, and Zingales 2006).

gender stereotypes versus general gender stereotypes? Examples of math-specific gender stereotypes include: "math is for boys, reading is for girls", "boys are good at math, girls are good at writing", or "it is always men who work at science, engineering and technical fields". In contrast, examples of general gender stereotypes include: "the best women are stay-at-home moms", "women are supposed to make less money than men", "women are not politicians", "girls have to work hard to learn in school, whereas boys are naturally gifted", or "women are nurses, not doctors".

Using the same data and empirical strategy as in Nollenberger, Rodríguez-Planas, and Sevilla (2016), we first analyze whether cultural beliefs on the role of women in society affect girls' *beliefs* in their own math abilities ("as I am a girl, I am not good at math"); their beliefs in the institutional constraints she may face ("as I am a girl, math will not help my career prospects");[4] their anxiety on performing in math ("as I am a girl, I am told math is not for me, which generates anxiety and reduces my performance in math"); or their *preferences* regarding math ("as I am a girl, I dislike math") relative to those of boys.

We find that girls hold similar beliefs in their *ability* to do math and report similar anxiety when performing math than boys *irrespective* of cultural background. However, we find that girls whose parents come from more gender-equal countries have higher *preferences* for math; to put it differently, they just like math more!

Then, we analyze which country-of-ancestry institutional channels shape the gender cultural attitudes that ultimately improve girls' relative math test performance. We find that country-of-ancestry institutions related to women's political empowerment and economic opportunity (as opposed to education, health and survival) are driving the math gender gap. To the extent, that the transmission of beliefs is related to political empowerment and economic opportunity as opposed to education *per se*, suggests that more general stereotypes (as opposed to only math-gender stereotypes) are at play.

[4] Expected institutional constraints may be driven by actual constraints in the country of ancestry. Note that this is still a story about beliefs, even though beliefs and institutions are closely intertwined.

Our analysis then shifts towards the effect of culture on other subjects, namely reading (where girls outperform boys in our sample) and science (where the gender gap is small and not statistically significantly different from zero in our sample).[5] Evidence that the effects of culture expand beyond math suggests that gender social norms are affecting female academic performance more broadly. As shown by Figure 1, which plots gender gaps in test scores in the country of residence against an index of gender equality in the country of ancestry,[6] we find that second-generation immigrant girls whose parents come from more gender-equal countries gain an absolute advantage over boys on reading and science, (as well as in math), suggesting that beliefs about women's role in society affect girls' relative test performance of different subjects alike.[7] More specifically, we find that a one standard deviation increase in the country-of-ancestry GGI is associated with an *increase* of 0.31 and 0.34 standard deviation in the reading and science gender gaps, respectively. Our results are robust to omitted variable bias, different specification strategies, selective migration, adjustments of standard errors, alternative measures of gender equality, and changes in sample criteria. Most importantly, the effect of gender social norms on the reading and science gender gaps remains even *after* we control for a large set of parental, household, and school characteristics.

We also find that again, as in math, cultural attitudes regarding women's political empowerment and economic opportunity in the country of ancestry are most relevant in determining second-generation immigrants' reading and science gender gaps. Our analysis concludes with evidence that our findings are not driven by non-cognitive skills (such as motivation, agreeableness and ambition),

[5] Even though girls' better reading skills are well known in the literature (Guiso *et al.*, 2008), less is known about gender relative performance and science-test scores. Most of the research on science gender gaps has focused on explaining gaps in science course taking or degree pursuit (see, for example, Ost 2010; Turner and Bowen 1999). Most recently, Quinn and Cooc (2015) find that there is a relatively stable science gender gap in the US between 3rd and 8th grade, which averages -0.19 standard deviations and is slightly larger than the math gender gap (in their sample, -0.12 standard deviations at 8th grade).

[6] The index of gender equality it the gender Gender Gap Index (GGI) from the World Economic Forum (Hausmann, Tyson, and Zahidi 2009), which is the same as the one used by Guiso *et al.* (2008), Fryer and Levitt (2010), Nollenberger, Rodríguez-Planas, and Sevilla (2016), and Rodríguez-Planas and Sans-de-Galdeano (2016).

[7] Figure 1 displays the raw relationship between the gender gap in each test score and the Gender Gap Index (GGI) from the World Economic Forum. This relationship persists even after adjusting the gender gap in each test score by individual characteristics and the GDP of the country of ancestry as shown in Appendix Figure 1.

supporting that social gender norms affect parent's expectations on girls' learning cognitive skills, but not on other attributes for success. Taken together these findings suggest that cultural beliefs on the role of women in society are not specific to math skills, but instead more general as they also apply to reading and science skills, increasing the relevance of the First Lady Michelle Obama's statement: "*And that's the kind of work that we need to be doing around the world – the work of changing culture. The work of changing expectations and standards that we have for women and girls"*.

Our work contributes to findings from Nollenberger, Rodríguez-Planas and Sevilla (2016) in that we show that gender equality affects female performance in math not necessarily through its effects on female math-related gender identity, but instead through its effects on general gender stereotypes, and (possibly) through its direct effects on girls' preferences. Our work also complements earlier findings from Guiso *et al.* (2008) and Pope and Sydnor (2010). Using PISA data from 40 countries, the former find that, in more gender-equal societies, girls close the gender gap by becoming better at both math and reading. In contrast, Pope and Sydnor (2010) find the opposite result by exploiting regional variation in the US. While our findings are closer to Guiso *et al.* (2008), they also use the same data source *and* exploit cross-country variation as opposed to cross-regional variation within one country.[8] Most importantly, our findings reveal that the positive effects of gender social norms on girls' math test scores relative to those of boys: (1) expand to other subjects (namely reading and science), (2) are shaped by beliefs on political empowerment and economic opportunity, and (3) are driven by parents' influencing their children (especially their girls') preferences. To put it differently, our findings suggest that parents' gender social norms shape youth's test scores by transmitting preferences for cognitive skills.

Our wok also exploits findings from a recent literature that views survey and tests as performance tasks (in addition to a measure of knowledge). This literature shows that the rate of decline in performance in tests or item nonresponse rates in survey questionnaires are proxies of agreeableness, motivation and ambition, but

[8] It is important to highlight that Guiso *et al.* (2008) only use 2003 PISA data, focus on a slightly different groups of countries than we do, and include natives as well as first- and second-generation immigrants. Most importantly, their analysis, which relates country-of-residence gender equality measures with the math and reading gender gap in the country of residence, is silent on the direction of the causality or the role of parental transmission of beliefs.

not to cognitive performance, the former (Borghans and Schils 2012); and conscientiousness, the latter (Hitt, Trivitt, and Cheng 2016; Zamarro et al. 2016). Borghans and Schils (2012) and Zamarro, Hitt, and Mendez (2016) find that between one fifth and one third of the between country variation in PISA scores is driven by these non-cognitive skills measures. Our contribution to this literature is to explore whether country-of-ancestry gender social norms are related to gender differences in non-cognitive skills and whether these non-cognitive skills are driving the results that second-generation girls coming from more gender-equal countries of ancestry outperform their male counterparts in math, reading and science. We find no evidence of this.

The remainder of this paper is organized as follows. Sections 2 and 3 describe the empirical strategy, and the data and sample selection, respectively. Section 4 presents results on self-reported beliefs on math performance. Section 5 analyzes which country-of-ancestry institutions affect the math gender gap. Section 6 presents results on the reading and science gender gap. Sections 7 and 8 present heterogeneity analysis and robustness checks, respectively. Section 9 analyses whether the effect of social gender norms on gender test gaps is driven by cognitive or non-cognitive skills, before concluding in section 10.

2. Empirical Strategy

To examine whether culture can explain gender differences in test performance, we focus on second-generation immigrants who are exposed to the same host country's labor market, regulations, laws and institutions, but are also influenced by the different cultural beliefs of their parents.[9] Evidence that gender equality in the immigrant´s country of ancestry can explain test scores of second-generation immigrants living in a particular host country would suggest that the preferences and beliefs of the immigrant's ancestors matter and have been transmitted to them by their parents and/or their ethnic community.[10]

[9] Throughout the paper, we will refer to the country where each individual is born and lives as their "host country". Given that they are second-generation immigrants, the country where they were born and live is actually the host country of their parents.

[10] Using a similar approach, several studies have examined the effect of culture on different socio-economic outcomes, including savings rates (Carroll, Rhee, and Rhee 1994), fertility and female labor force participation (Antecol 2000; Fernández 2007; Fernández and Fogli 2006, 2009), living arrangements (Giuliano 2007), the demand for social insurance (Eugster *et al.* 2011), preferences for a child's sex (Almond, Edlund, and Milligan 2013); the math gender gap (Nollenberger,

We use OLS to estimate the following baseline specification:

$$E_{ijkt} = \alpha_1 female_i + \alpha_2(female_i GE_j) + X'_{ijkt}\beta_1 + X'_{ijkt} female_i \beta_2 + \lambda_j + \lambda_k + \lambda_t + \delta(female_i \lambda_k) + \varepsilon_{ijkt} \tag{1}$$

where E_{ijkt} is the test score of individual *i* who lives in host country *k* at time *t* and is of ancestry *j*. To identify the differences in test scores between girls and boys, the variable $female_i$ is an indicator equal to one if the individual is a girl and zero otherwise. GE_j is a measure of gender equality from the individual *i*'s country of ancestry *j*, such that a higher value is associated with a more gender-equal culture. The vector X_{ijkt}, includes a set of individual characteristics that may affect educational attainment for reasons unrelated to gender equality, and that vary with the specification considered. These individual characteristics are also interacted with the female indicator. λ_j, λ_k, and λ_t are a full set of dummies that control for the country of ancestry *j*, the host country *k*, and the PISA cohort *t*. Country-of-ancestry fixed effects (λ_j) control for the gender equality (GE_j) in the country of ancestry, and for any other factors that affect the test scores of boys and girls in the same way. Year fixed effects (λ_t) account for cohort differences and other time variation. Following Alesina and Giuliano (2010 and 2011), Luttmer and Singhal (2011), and Nollenberger, Rodríguez-Planas and Sevilla (2016) who also look at immigrants living in multiple host countries, we include host-country fixed effects (λ_k) in our specification to account for the host country's characteristics that may be related to test performance. Most importantly, host-country dummy variables (λ_k) are interacted with $female_i$ to account for variation in the host country's test-scores gender gaps that may arise from cross-country differentials in cultural or institutional channels.

Our coefficient of interest on the interaction between the GE_j and the female indicator, α_2, captures the role of gender equality in explaining the gender differences in test scores of second-generation immigrant boys and girls. A positive and significant α_2 would suggest that more gender-equal attitudes in the immigrant's country of ancestry are associated with a higher relative test performance of second-generation immigrant girls over boys, and thus a *smaller*

Rodríguez-Planas, and Sevilla 2016); divorce (Furtado, Marcén, and Sevilla 2013); and the gender smoking gap (Rodríguez-Planas and Sans-de-Galdeano 2016).

gender gap if the initial gap is negative (as it is in math), but a *greater* gender gap if the initial gap is non-negative (as is the case in reading and science).

3. Data and Sample

Program for International Student Assessment (PISA) Data

Our main data set uses the 2003, 2006, 2009 and 2012 student-level data from the Program for International Student Assessment (*PISA*), an internationally standardized assessment conducted by the Organization for Economic Cooperation and Development (OECD) and administered to 15-year olds in schools every three years since 2000. PISA assesses a range of relevant skills and competencies in three main domains: mathematics, reading, and science. To do so, PISA randomly distributes the participating students into booklets, which differ (also randomly) in type and order of questions. The PISA test has an average of 60 questions across the three different subjects and is expected to last about 2 hours.

The purpose of PISA is to test whether students have acquired the essential knowledge and skills for full participation in society near the end of compulsory education. These skills include whether they can analyze, reason and communicate effectively. According to the OECD (2003), the PISA math test assesses "*the capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen*". At the same time, the PISA reading test assesses "*the capacity to understand, use and reflect on written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society*", and the PISA science test assesses "*the capacity to use scientific knowledge, to identify scientific questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity*". In addition, students and school principals also answer questionnaires to provide information about the students' background, school and learning experience, as well as the broader school system and learning environment. Appendix Table A.1 presents a detailed description of all PISA variables used in the analysis, as well as basic descriptive statistics.

Our analysis begins in 2003 because questions entering the math scores are not comparable before and after that year. For each subject, PISA tests are paper and pencil tests, lasting up to two hours. Each subject is tested using a broad sample of tasks with differing levels of difficulty to represent a comprehensive indicator of the continuum of students' abilities. The PISA program presents the tests scores in standardized form, whereby they have a mean of 500 test-score points and a standard deviation of 100 test-score points across the OECD countries.

As explained by Guiso *et al.* (2008) in their Supporting Material online: *"PISA assigns a probability distribution to each possible response pattern in each test to describe the ability associated with that pattern. From this distribution, PISA draws a set of five values associated with each student. These values are called plausible values (hereinafter PV) because they represent alternative estimates of the students ability that could have been obtained*." As is standard in this literature and recommended by the OECD, we use PV in all of our analysis that involves test scores. Hence, we estimate one regression for each set of PV and, subsequently, report the arithmetic average of these estimates.

PISA sample is stratified at two stages: first, schools are randomly selected; and second, students at each school are randomly assigned to carry out the test in all three subjects. A minimum participation rate of 65% of schools and 80% of students from the original sample is required for a country to be included in the international database. Following OECD recommendations, we apply the Fay's Balanced Repeated Replicated (BRR) methodology to estimate standard errors that will take into account PISA's stratified, two-stage sample design. Results are robust to clustering standard errors at the country-of-ancestry level.

Gender Equality Measures

To measure gender equality in an immigrant's country of ancestry, we follow Guiso *et al.* (2008), Fryer and Levitt (2010), Nollenberger, Rodríguez-Planas, and Sevilla (2016), and Rodríguez-Planas and Sans-de-Galdeano (2016) and use the Gender Gap Index (GGI hereafter) from the World Economic Forum (Hausmann, Tyson, and Zahidi 2009). The GGI measures the relative position of women in a society taking into account the gap between men and women in economic opportunities, economic participation, educational attainment, political achievements, health and well-being.

To explore which country-of-ancestry institution shape the beliefs that end up mattering the most and test the robustness of our results to alternative measures of gender equality, we also use other measures of gender equality from the World Economic Forum, namely an index of ***economic participation and opportunity*** based upon: (1) female labor force participation over male, (2) wage equality between women and men to similar work, (3) female earned income over male, (4) female legislators, senior officials and managers over male, (5) female professional and technical workers over male; an index on ***educational attainment*** based upon: (1) female literacy rate over male, (2) female net primary level enrolment over male value, (3) female net secondary level enrolment over male, (4) female gross tertiary level enrolment over male value; an index on ***political empowerment*** based upon: (1) the ratio women to men with seats in parliament, (2) the ratio of women to men in ministerial level, and (3) the ratio of the number of years with a women as head of state to the years with a man; and an index on ***health and survival*** based upon: (1) the gap between women and men's healthy life expectancy and, (2) the sex ratio at birth, which aims to capture the phenomenon of "missing women". All these indexes range from 0 to 1, with larger values indicating a better position of women in society.

Information on the GGI is available from 2006 on. In this year, 115 countries were included, in 2007 128, in 2008 130, and in 2009 134. In order to maximize the number of countries in our sample, we focus on the year 2009 as Nollenberger, Rodríguez-Planas, and Sevilla (2016). The use of contemporaneous measures of gender equality rather than those observed at the time parents migrate is a common practice in the literature. First, it is reasonable to expect that countries' aggregated preferences and beliefs about the role of women in society change slowly over time. Second, as Fernández and Fogli (2009) point out, "*one could argue that the values that parents and society transmit are best reflected in what their contemporaneous counterparts are doing in the country of ancestry*".

Sample

Our sample comprises second-generation immigrants who were born and reside in a participating host country but whose parents (both of them) were born in another country. Choosing second-, rather than first-generation immigrants, is preferred by the epidemiological literature as it minimizes the role of institutions in the country of ancestry for immigrant's outcomes, given that the probability to return to the country of ancestry of second-generation immigrants is much lower than the probability of first-generation immigrants. We pool the 2003, 2006, 2009 and 2012 PISA waves to have the larger variation possible in terms of both host countries and countries of ancestry. To determine the students' country of ancestry, we need specific information on their parents' country of birth. This question is not consistently asked among participating countries. For instance, when asking about the country of origin, the US only provided the options "United States of America" and "another country". Consequently, only data from those participating countries providing detailed information about the parents' birth place were used in the analysis.[11]

Based upon Blau *et al.* (2013), who find that the effect of mother's country of origin on second-generation immigrants girls tend to be stronger than the effect of the father's country of origin when parents come from different countries, we assign the mother's country of origin.[12] We restrict our sample to those individuals for whom we observe gender equality measures for both their country of ancestry and their host country, focusing our analysis on host countries with immigrants from at least four countries of ancestry to ensure that we do not compromise the identification in our model, which arises from variation in gender equality in the immigrant's country of ancestry within a given host country.[13] We also drop second-generation immigrants whose country of ancestry has fewer than

[11] These are Australia, Austria, Belgium, Denmark, Finland, Germany, Greece, Latvia, Liechtenstein, Luxembourg, New Zealand, Norway, Portugal, Switzerland and Scotland in 2003, 2006, 2009 and 2012 PISA; Argentina, Czech Republic, Israel, Netherlands and Qatar in 2009 and 2012 PISA; and China, Costa Rica and Turkey in 2012.

[12] In any case, 85% of the second-generation immigrants in our sample have parents who emigrate from the same country.

[13] The lack of gender equality measures for all countries implies losing the following countries of ancestry: Afghanistan, Bosnia and Herzegovina, Cape Verde, Occupied Palestine, Iraq, Lebanon, Liechtenstein, Netherlands Antilles, Somalia, Somoa and Serbia-Montenegro (4,345 observations) and the host country of Liechtenstein (or 135 observations). In any case, most of the countries of ancestry we lose are from conflictive zones, which are commonly excluded from this kind of analysis (see Fernández and Fogli 2009, and Furtado, Marcén, and Sevilla 2013). In addition, by limiting our analysis to host countries with at least four different groups of immigrants we lose 3,983 observations from the following ten host countries (Costa Rica, China, Denmark, Germany, Greece, Latvia, Norway, Portugal, Qatar and Turkey), and seven countries of ancestry (Brazil, Bulgaria, Belarus, Jordan, Egypt, Nicaragua and Yemen).

15 observations in a given host country.[14] In the robustness section, we explore the robustness of our results to changes in sample criteria.

Our final sample has 11,527 second-generation migrants from 35 different countries of ancestry and living in nine host countries (as shown in Appendix Table A.2). Host countries are mainly OECD countries, whereas countries of ancestry are from various continents and levels of development. For instance, the countries of ancestry in our sample cover all continents, with many European (14 countries) and some transition economies (Albania, Poland and Russia), several countries in the Americas (Bolivia, Chile, Paraguay, Suriname, United States and Uruguay), some in Asia (China, India, Korea, Malaysia, Philippines and Vietnam), Africa (Ethiopia, Morocco and South Africa) and Oceania (Australia, Republic of Fiji and New Zealand). Second-generation immigrants whose country of ancestry is Portugal, Turkey or Italy represent 49% of the sample. Host countries with the highest sample of second-generation immigrants are Switzerland, Australia and Luxembourg (immigrants living in these countries represent 71% of the sample). As with other papers using the epidemiological approach with survey data, a possible drawback of our approach is that PISA is not necessarily representative of the second-generation immigrant population. To partly deal with this concern we address issues of selection of immigrants in the robustness section.

Descriptive Statistics

Appendix Table A.3 presents summary statistics for our sample of second-generation immigrants by country of ancestry. The first three columns show the average gap in different test scores of second-generation immigrant girls relative to boys. The gender gap is measured as the average of the girls' minus the average of the boys' scores, whereby a negative gap means that boys over perform girls while a positive gap means that girls over perform boys. The first, second, and

[14] This is a common practice in the literature. For instance, Fernández and Fogli (2009) eliminate those countries of ancestry with fewer than 15 observations. Given that our regressions are ran at the individual level, whether we include these small numbers of observations does not affect our results. With this adjustment, we lose 201 individuals and 11 different countries of ancestry (Argentina, Bangladesh, Colombia, Czech Republic, Denmark, Hungary, Iran, Panama, Slovenia, Sweden and Thailand).

third columns show the average gap in math, reading, and science test scores of second-generation girls relative to boys, respectively.

Countries of ancestry are ordered from the more math gender biased countries to the least. Column 1 shows a large variation in the gender gap in math scores across countries of ancestry. On average, the difference in math score between girls and boys across our sample is -15.70, the equivalent to 4.5 less months of schooling. In contrast, we find that, on average, girls outperform boys in reading test scores (Column 2). The difference in reading score between girls and boys across our sample is +30.16), the equivalent to 9 more months of schooling. Column 3 shows that even though, on average, boys outperform girls in science, the average difference (-6.37) is considerably smaller than in math.

It is important to highlight that these gender gaps in test scores are quite similar to those observed among all second-generation immigrants and natives living in the host countries included in our analysis, and are not too distant from those shown when all countries participating in PISA assessments are considered (see Appendix Table A.4).

Panel A in Appendix Figure A.2 shows that second-generation immigrant girls from a given country of ancestry who perform better in math than their male counterparts also tend to perform relatively better in reading. Panel B also shows that second-generation immigrant girls who have a higher score in math relative to their male counterparts also have a relative higher score in science. Panel C shows a similar relationship between reading and science test scores.

Columns 9 to 12 in Appendix Table A.3 show the value of different gender-equality measures by country of ancestry. Our main variable, GGI, averages 0.69 with a standard deviation of 0.05, varying from 0.58 in Turkey to 0.79 in New Zealand. Further detail on the other indices of gender equality is provided in Section 5 below.

4. Math Test Scores and Self-Beliefs on Math Performance

Replicating Nollenberger, Rodríguez-Planas and Sevilla (2016)

Column 1, Panel A, in Table 1, which only controls for the female indicator and the year, country-of-ancestry and host-country fixed effects, reveals that second-generation immigrant girls underperform boys in math by, on average, 14.77 score points within host country, country of ancestry, and survey year. Column 2, Panel

A, in Table 1 replicates Nollenberger, Rodríguez-Planas and Sevilla's (2016) main result: if a girl's parents, originally from a country with an "average" GGI, had instead come from a country with a GGI one standard deviation above the mean, her math test score in the host country would have increased by 7.47 score points relative to that of a male counterpart, the equivalent of a reduction in the math gender gap of 0.29 standard deviation.[15] To put estimate α_2 into context: if immigrants from Turkish descent, whose country of ancestry has a GGI of 0.58 and who present a gender gap in math scores of -13.77 score points, were characterized by the mean gender equality in our sample (GGI = 0.69), the statistical model would suggest that the mean score performance in mathematics of second-generation Turkish girls relative to boys would increase by 16.45 score points, thus reversing the gender gap.[16]

Culture and Self-Reported Beliefs Regarding Math

What is driving this finding? Girls' relative underperformance in math could be the result of cultural beliefs on the role of women in society affecting girls' *beliefs* in their own math abilities ("as I am a girl, I am not good at math"); their beliefs in the institutional constraints she may face ("as I am a girl, math will not help my career prospects"); their anxiety on performing in math ("as I am a girl, I am told math is not for me, which generates anxiety and reduces my performance in math"); or girls' *preferences* regarding math ("as I am a girl, I dislike math").

To explore this, we estimate equation (1) using as left-hand-side (LHS) variable one of the following five PISA constructed indices on self-reported beliefs or preferences regarding math, available *only* in waves 2003 and 2012 (OECD 2013).[17] The first two indices capture students' beliefs on their math abilities. The *"math self-concept"* captures students' beliefs on their own math's abilities, including whether they believe they are good and fast at *learning* math;[18]

[15] This is calculated as $\frac{(\alpha_2=149.55)*(GGI_{StDev}=0.05)=7.47}{(Gap\ in\ Math_{StDev}=26.04)} = 0.29$

[16] This is calculated as $[GGI_{AVG}\ (0.69) - GGI_{TUR}(-0.58) = 0.11] * \alpha_2(149.55) = 16.45$

[17] The main finding that culture affects the gender gap in math generally holds when estimating the effect using the subsample for whom each of the self-beliefs was reported (see Appendix Table A.5), albeit we lose precision as the sample size is smaller.

[18] This index is constructed using student responses to a question over the extent they strongly agree, agree, disagree or strongly disagree with the following statements when asked to think about studying mathematics: "I am just not good at mathematics; I get good in mathematics; I learn

whereas the *"math self-efficacy"* index captures the extent to which students believe in their own ability to *handle mathematical tasks* effectively and overcome difficulties.[19] The higher the value of the index, the higher self-concept or self-efficacy a student has, respectively. The third index, the *"instrumental motivation to learn math"* index, captures students' perception on how useful math may be in their professional future, with a higher value of the index indicating higher instrumental motivation to learn math.[20] The *"math anxiety"* index captures thoughts about doing math, such as feeling of helplessness and stress when dealing with mathematical problems, with a higher index indicating higher anxiety.[21] Finally, the "*intrinsic motivation to learn math"* index includes several questions on enjoyment from doing math. More specifically, the student is asked to strongly agree, agree, disagree or strongly disagree to a series of statements, when asked to think about his or her views on mathematics: "I enjoy reading about mathematics; I look forward to my mathematics class; I do mathematics because I enjoy it; I am interested in the things I learn in mathematics." The higher the value of the index, the more intrinsic motivation the student has.[22]

Columns 3, 5, 7, 9 and 11 in Table 2 explore whether there is a differential gender pattern across these different index variables by estimating a regression with a female indicator, and country-of-ancestry, host-country and year fixed

---

mathematics quickly; I have always believed that mathematics is one of my best subjects; in my mathematics class, I understand even the most difficult work".

[19] This index is calculated based on how confident students report to be at performing the following mathematics tasks: "Using math to work out how long it would take to get from one place to another; calculating how much cheaper a TV would be after a 30% discount; calculating how many square meters of tiles you need to cover a floor; understanding graphs presented in newspapers; solving an equation like 3x+5=17; finding the actual distance between two places on a map with a 1:10,000 scale; solving an equation like 2(x+3)=(x+3)(x-3); calculating the petrol consumption rate of a car".

[20] The index is constructed using students' responses over the extent they strongly agree, agree, disagree or strongly disagree to a series of statements, when asked to think about their views on mathematics: "Making an effort in mathematics is worth because it will help me in the work that I want to do later on; learning mathematics is worthwhile for me because it will improve my career; Mathematics is an important subject for me because I need it for what I want to study later on; I will learn many things in mathematics that will help me get a job".

[21] The index is constructed using student responses to a question over the extent they strongly agree, agree, disagree or strongly disagree with the following statements when asked to think about studying mathematics: "I often worry that it will be difficult for me in mathematics classes; I get very tense when I have to do mathematics homework; I get very nervous doing mathematics problems; I feel helpless when doing a mathematics problem; I worry that I will get poor in mathematics".

[22] In PISA 2003, the index of intrinsic motivation to learn mathematics was named the index of interest and enjoyment in mathematics, but both 2012 and 2003 indices are based on the same questionnaire items.

effects. Columns 4, 6, 8, and 12 in Table 2 re-estimate equation (1) using these alternative LHS variables (instead of the math test score) with the objective of identifying whether country-of-ancestry gender social norms affect these different outcomes differentially for girls than for boys.[23]

Focusing in the odd columns first, we observe that second-generation immigrant girls believe that they are worse at learning math and handling math tasks effectively than their male counterparts (shown in columns 3 and 5, respectively). Second-generation girls are also more likely to report math anxiety than their male counterparts (column 9), and less likely to like math (column 11) and to perceive studying math as useful professionally in the future (column 7) than second-generation boys. All of these estimates are statistically significant at the 1 percent level.

Having girls perceive that their math skills, beliefs or preferences differ from those of boys does *not* necessarily help us better understand the relationship between cultural beliefs on gender roles and the math gender gap found by Nollenberger, Rodríguez-Planas, and Sevilla (2016). For these self-reported skills, beliefs and preferences to be behind the cultural persistence explaining the math gender gap, they must also be related to country-of-ancestry gender social norms. We explore this in the even columns in Table 2. Interestingly, we find that α_2 is positive *and* statistically significant only in the case of *"intrinsic interest in mathematics"* (column 12). The other α_2 estimate that is large (albeit not statistically significant) is the *"instrumental motivation to learn math"* index. When the LHS variable is any of the other indices, the estimates of α_2 are considerably lower in magnitude and not statistically significant.

How much do gender cultural beliefs affect gender differences in math preferences? According to our estimates in Table 2, if a girl's parents, originally from a country with an "average" GGI, had instead come from a country with a GGI one standard deviation above the mean, her *"intrinsic interest in mathematics"* index in the host country would have increased by 1.53, reducing

[23] We use the same covariates as in Nollenberger, Rodríguez-Planas and Sevilla (2016) baseline specification, also shown in column 2, Table 1.

the gender differences in this index by 0.13 standard deviation.[24] This evidence is suggestive that beliefs on gender social norms are transmitted through parents (or parents' social network) from less gender-equal countries instilling to girls lower *preferences* for math relative to boys.

5. Institutional Channels from the Country of Ancestry Shaping Culture

An alternative and complementary exercise is to explore what types of institutional channels in the country of ancestry are shaping the gender cultural attitudes that ultimately affect the math gender gap. Columns 1 to 4 in Table 2 re-estimate our baseline specification replacing the GGI with alternative measures of gender equality that focus on specific areas of society, namely political empowerment (column 1), economic participation and opportunity (column 2), educational attainment (column 3), and health and survival (column 4).

Although these different measures are correlated between them, they capture different aspects of culture, and hence may have independent power to explain the math gender gap.[25] For example, all variables may reflect, in part, the belief as to the appropriate role of women in society, but economic participation and opportunity may also capture some independent cultural preferences for the role of women in the labor market, the education index may also capture some independent cultural beliefs on education opportunities between men and women, and the political empowerment index may also capture some independent cultural beliefs on women's political representation.

Two of the four α_2 estimates shown in Table 2 are positive and statistically significant: the one on political empowerment and the one on economic opportunity, albeit the second one only at the 0.1 level. Column 5 conducts a horse race by estimating a specification that controls for the four estimates of gender equality at the same time, and confirms that these two gender equality indices are the most relevant.[26] According to the estimates in Columns 1 and 2, beliefs transmitted to second-generation immigrants regarding women's political

---

[24] This is calculated as $\frac{1.53 * GGI\ std\ (0.05)}{Index\ Gender\ Gap\ std\ (0.59)} = 0.13$

[25] Correlations between the different measures of gender equality range between 0.23 and 0.77 and are displayed in Appendix Table A.6.

[26] Even though we lose precision due to multicollinearity, we reject the null hypothesis that all four coefficients are jointly equal to zero.

empowerment are those that matter the most, closely followed by those regarding women's economic opportunity. While an increase in the level of the political empowerment index by one standard deviation is associated with a reduction of 0.30 standard deviation in the math gender gap among second-generation immigrants, the reduction is 0.22 standard deviation for the economic-opportunities index. In comparison, an increase in the level of the education (health and survival) index by one standard deviation only reduces the math gender gap by a non-statistically significant 0.09 (0.11) standard deviations. To the extent that the transmission of beliefs is related to political empowerment and economic opportunity as opposed to education *per se,* suggests that more general stereotypes (as opposed to only math-gender stereotypes) are at play.

6. Are Gender-Stereotypes Field Specific?

How much of this improvement of girls' test performance is explained by math-specific gender stereotypes versus general gender stereotypes? The empirical evidence on this matter is mixed. While Guiso *et al.* (2008) find that: "*in countries with a higher GGI, girls close the gender gap by becoming better at both math and reading*", Pope and Sydnor (2010) find the opposite result by exploiting regional variation in the US. More specifically, they find that: "*areas which have smaller gender-disparities in stereotypically-male dominated tests of math and science, also tend to have smaller disparities in stereotypically female-dominated tests of reading."* The authors conclude that: "*variation across states in test scores disparities is not simply a reflection of some states improving the performance of females relative to males. Rather, some states appear to be more gender equal across all tests and adhere less to gender stereotypes in both directions.*"

Column 1 in Table 3, reveals that second-generation immigrant girls outperform boys in reading by, on average, 32.25 score points within host country, country of ancestry, and survey year. Since the average reading test score is 465 among second-generation boys, this implies that second-generation girls' reading test scores are, on average, 7 percent higher than those of boys. Column 3 in Table 3 shows that there is no statistically significant difference in science test scores between second-generation immigrant girls and boys.

Columns 2 and 4 in Table 3 estimate equation (1) using reading and science test scores as the LHS variable. Column 2 shows that second-generation immigrant girls whose country of ancestry is more gender equal also have higher reading scores relative to boys, and hence the girls' reading advantage *widens*. Similarly, Column 4 shows that second-generation girls coming from more gender-equal countries of ancestry outperform their male counterparts also in science. According to these estimates, one standard deviation increase in the GGI is associated with an *increase* of 0.30 (0.36) standard deviation in the reading (science) gender gap, which is very close to the magnitude of effect on the math gender gap (0.29 standard deviation *decrease* of the math gender gap).

Hence, we find that second-generation immigrant girls whose parents come from more gender-equal countries perform better relative to immigrant boys in *both* math, reading and science, suggesting that cultural beliefs on the role of women in society are *not* specific to math skills, but instead more general as they also apply to reading and science skills. While these findings are closer to Guiso *et al.* (2008) than Pope and Sydnor (2010), we use the same data source and exploit cross-country variation as the former, whereas the latter focuses on cross-regional variation in the US, and hence our results and those of Pope and Sydnor (2010) are not necessarily comparable. As explained in footnote 8, Guiso *et al.* (2008) do not use the epidemiological approach but estimate instead correlates between country-of-residence gender equality measures and the math and reading gender gap, being silent on the role of parental transmission of beliefs. Most importantly, our contribution to this literature is that the *transmission of cultural beliefs on the role of women in society* (not societal factors generally defined) affects girls' relative test performance in subjects different from math, namely reading and science. Crucially, this finding added to the earlier results highlight the relevance of general (as opposed to math-specific) gender stereotypes on the math gender gap, and suggest that parents' gender social norms shape youth's test scores by transmitting preferences for cognitive skills.

Table 4 also shows that, as in math, cultural attitudes regarding women's political empowerment and economic opportunity in the country of ancestry matter in determining the reading and science gender gaps of second-generation immigrants in the host country. Except for economic opportunity in the science equation, which is statistically significant at the 0.1 level, the other three coefficients are statistically significant at the 0.05 level. When we do the horse race in Column 5, we observe that cultural attitudes regarding women's political empowerment in the country of ancestry matter the most for both the reading and science gender gaps. The estimates of this index are statistically significant at the 0.1 level and, in both cases, we reject the null hypothesis that all coefficients are jointly equal to zero at the 0.05 level.

The magnitudes from the reading and science estimates in Columns 1 and 2 in Table 4 are similar to those found in math. A one standard increase in the economic opportunity or political empowerment indexes is associated with a 0.21 and 0.32 (0.25 and 0.36) *increase* in the reading (science) gender gap. Hence, beliefs regarding economic opportunity and political empowerment affect girls' test performance relative to boys.[27]

7. Heterogeneity

Panel A in Table 5 estimates the effect of GGI on girls' and boys' test scores separately. The idea here is to explore how much of the culture effect on the test-score gender gaps is explained by the effect of gender social norms on girls' versus that on boys' test scores. To put it differently, do gender social norms improve girls' test scores exclusively? Do they improve girls' test scores more than those of boys? Or do they have a detrimental effect on boys' test scores that could also potentially explain the converging results found earlier.

Interestingly, we observe that the coefficient on the GGI is positive and statistically significant for both boys and girls, suggesting that youth whose parents come from more gender-equal societies perform better in exams regardless of gender or subject type. However, we find that the effect of culture on test scores is more than twice as large for girls than for boys (again regardless of the subject type).[28] Hence, gender social norms seem to be beneficial for all, but more so for

[27] While we would like to perform the same analysis on self-beliefs for reading and science as we did in Table 1 for math, PISA information on reading and science self-beliefs is limited and only available for one of the four waves, reducing the precision of our estimates due to small sample sizes.

[28] Note that for the three test scores, the standard deviation across countries of ancestries is almost the same for boys and for girls (around 64 score points for boys and around 63 score points for girls). Therefore, a one standard deviation increase in the GGI leads to an increase in girls' test score that more than doubles that of boys'.

girls than boys. In all three subjects, we reject the null hypothesis that the effect of culture on girls' and boys' test scores is the same.

Panels B and C in Table 5 explore whether the transmission of cultural beliefs on the role of women in society varies across different types of second-generation immigrants by estimating our preferred specification for two different subgroups of second-generation immigrants. Panel B explores whether culture has a differential effect on girls' test scores relative to boys' according to the concentration of immigrants from the same ethnicity in the school. We calculate the proportion of first- and second-generation immigrants in each school from PISA following Schnepf (2007) (see definition in Appendix Table A.1). Even though we cannot reject that the effect of culture differs across the two groups, the effect of culture on the girls' test scores relative to those of boys is considerably larger and (frequently estimated with greater precision) for second-generation immigrants attending schools with a high concentration of immigrants from the same ethnicity. Fernández and Fogli (2009) and Luttmer and Singhal (2011) also find that the impact of culture is stronger for immigrants who have a greater tendency to cluster with their ethnic community. One possible interpretation is that horizontal transmission of culture through peers may constitute a potential mechanism of the transmission and maintenance of cultural beliefs. As in previous studies, however, to the extent that parents may be sorting into neighborhoods or schools, the stronger cultural effects for this subgroup may be a further consequence of vertical cultural transmission rather than a genuine peer effect.

Panel C re-estimates equation 1 by whether the youth is attending a school with a high or low proportion of girls in the school (as recorded by the school's principal).[29] Although the coefficient on the variable of interest remains positive and tends to be larger in size when youth attend schools with a high proportion of girls, it is only statistically significant for science. Moreover, we cannot reject that the coefficient across the two groups are statistically different from each other.[30]

[29] See Appendix Table A.1 for definition and descriptive statistics of this variable.
[30] We also carried out the same analysis splitting the sample by whether the student attends to a single- or mixed-sex school. We defined single-sex schools as those where the proportion of girls equals one (school for girls) or zero (school for boys). The direct effect of the GGI was again

8. Robustness Checks

Omitted Variable Bias

Appendix Tables A.7 and A.8 present different robustness checks of our reading and science results.[31] While Column 2 presents our baseline specification, Column 1 displays a specification that omits the interaction between country-of-ancestry GDP per capita and the female dummy. The reason for doing so is to explore how sensitive our results are to *only* controlling for the interaction between country-of-ancestry GGI and the female dummy. Although doing so slightly reduces the effect of culture on both the reading and science gender gaps, the effect of culture on the test gender gaps remains large and statistically significant at the 0.05 level or higher, suggesting that this concern is not affecting our main results.

As our baseline specification includes country-of-ancestry fixed effects, it precludes us from observing the direct effect of country-of-ancestry GGI or GDP per capita on second-generation immigrants reading and science test scores. Column 3 presents a specification that replaces country-of-ancestry fixed effects with country-of-ancestry GGI and GDP per capita. It shows that more gender equality in the country-of-ancestry is associated with higher reading and science test scores among second-generation immigrants and that higher GDP per capita in the country of ancestry is also associated with higher science test scores (but has no effect on reading). Note, however, that this alternative specification leaves our main estimates of culture practically unchanged.

Columns 4 to 6 take a closer look at the relationship between gender social norms and the reading and science gender gaps by sequentially adding covariates. The aim here is to observe how our coefficients of interests vary with the inclusion of additional covariates and to shed some light on the mechanisms through which the relationship between gender social norms and the gender reading and science gaps operates. Most importantly, doing so enables us to assess the relevance of various potential sources of omitted variable bias and how they may affect our

---

larger in size for those attending single-sex schools for the three subjects, but only statistically significant at 10%. Moreover, we could not reject that the coefficients across the two groups were statistically different from each other.

[31] Results are also robust to using math test scores as the LHS variable as shown in Table 1 in Nollenberger, Rodríguez-Planas and Sevilla (2016).

conclusions. Note, however, that some of the additional characteristics that we will sequentially include (such as, for instance, parental education and work status as well as school type) may well be affected by culture. Therefore, by including some of the controls we will introduce below, we are limiting the avenues through which culture is allowed to operate, and estimate the direct effect of culture beyond the indirect ways in which these additional variables could affect such gender gaps through these variables. This is arguably a very demanding test of the relevance of culture. Note also that, by comparing outcomes across second-generation immigrants whose parents came to the host country from different countries of origin, the epidemiological approach is prone to underestimating the true effect of culture for two additional motives. First, cultural transmission is restricted to parents (or parents' social networks). Second, assimilation to the host country's culture is likely to weaken the impact of the country of ancestry's culture.

Column 4 adds to the baseline specification mother's and father's highest education level attained and their interaction with the female dummy. Doing so has little effect on the estimate of culture on the reading gender gap, and slightly increases the estimate of culture on the science gender gap. Not surprisingly, having more educated parents increases reading and science test scores for both girls and boys.

Column 5 adds to the specification in Column 4 controls for mother's and father's work status, as well as a variable measuring the household's possessions, and these variables' interaction with the female dummy. Having parents' work or more household possessions is positively associated with higher reading and science test scores for both boys and girls. While having more home possessions seems to have a larger effect on girls' science test scores than on boys', the opposite is true for having a working mother (father) on reading and science (science) tests scores. Crucially, adding these controls increases the estimate of culture, which remains positive and statistically significant at the 0.01 level.

In addition to the covariates in Column 5, Column 6 adds school controls and their interaction with the female dummy. As discussed earlier, to the extent that parents choose which schools (or neighborhoods) their children enroll (or live in), these variables are endogenous. Including them reduces the size of the coefficient of culture on reading by about 10% and that of science by about 5%.

Nonetheless, both coefficients remain large, positive, and statistically significant at the 0.01 level. As observed in the heterogeneity section, estimates from Column 6 indicate that attending schools with a higher proportion of girls improves girls' science and reading test scores relative to those of boys. In contrast, attending schools in metropolitan areas is more beneficial for boys than for girls.

Additional Robustness Checks

One concern is that our results may capture educational differences in the country of ancestry rather than differences in gender equality. If those more egalitarian countries also have more advanced educational systems, the effect of gender equality measures on math, reading, and science gender gaps would be upward biased.[32] We check this possibility in Panel B of Appendix Table A.9 by adding the interaction between the female dummy and the Human Development Index (HDI), which in addition to income (as GDP per capita), includes measures of life expectancy and education (such as mean years of schooling and expected years of schooling). As can be seen, adding this variable slightly increases the size of all three culture coefficients.

A common concern within the epidemiological approach is that immigrants may "self-select" in some areas in a given country. While most of epidemiological papers focus the analysis in one country, our analysis looks at immigrants not only coming from different countries of ancestries, but also going to multiple destination countries. As the form of selection is likely to differ across different destination countries, this approach limits the scope for selection bias (see Alesina and Giuliano 2011; and Luttmer and Singhal 2011). In addition, gender-equality based selection whereby parents who care more about their girls' success choose to move from ancestry countries with low gender-equality culture to host countries with high-gender equality is likely to attenuate our coefficients, biasing the culture estimates downward. To address this concern, Panel C in Appendix Table A.9 controls for local geographic variation in markets and institutions *within* our host countries by including regional fixed-effects (instead of the host-country dummies) and their interaction with the female indicator.

[32] We also present the results for the math gender gap in Appendix Tables A.9 and A.10 for completeness sake as these robustness checks were not performed in Nollenberger, Rodríguez-Planas, and Sevilla (2016).

Doing so accounts for variation in the host-country region's educational gender gaps that may arise from cross-regional differentials in cultural or institutional channels as a result from immigrants self-selecting in particular areas of the host country. Again, the effect of culture on the three test gender gaps remains robust to this specification change.

Panel D in Appendix Table A.9 also shows that our results remain practically unchanged when we adopt a more flexible specification where each year fixed-effect is interacted by the female indicator to allow different gender gaps by the cohort assessed in different PISA waves. Finally, Panel E shows that our results are robust to clustering the standard errors at the host country level, as opposed to using Fay's BRR methodology to account for the double stratification of the sampling design employed by PISA as explained in the Data Section.

Changes in Sample Criteria

Appendix Table A.10 shows that our results are not driven by the main group of immigrants (the Portuguese) or the host country with the largest sample of immigrants (Switzerland)--shown in panels B and C, respectively. Panel D also shows that the effect remains when only one host country is used (although the coefficient is no longer statistically significant in the case of Switzerland). Panel E shows that the gender gap results also holds when we drop countries that send migrants to only one host country.

9. Culture and Non-Cognitive Skills

A growing literature has found that non-cognitive factors, such as an individual's motivation, eagerness to succeed, agreeableness, or ambition, affect human capital accumulation (Cunha, Heckman, and Schennach 2010) as well as labor market outcomes, engagement in risky behaviors and health outcomes (Heckman, Pinto, and Savelyev 2013; Heckman and Rubinstein 2001). Most relevant to our analysis, Borghans and Schils (2012) show that the rate of decline in performance over the course of the 2006 PISA test's administration is related to non-cognitive factors such as agreeableness, motivation and ambition, and is a good predictor of final levels of educational attainment. They also show that this decline in performance is not related to cognitive performance. Using 2009 PISA, Zamarro, Hitt, and Mendez (2016) expand the methods used by Borghans and Schils (2012), and find that the decline in test performance is a good predictor of international variation in test scores.

At the same time, other researchers have found that individuals' item non-response in survey questionnaires is also related with their performance in school or in the labor market. Using six nationally-representative longitudinal datasets of American secondary school students, Hitt, Trivitt, and Cheng (2016) show that item-nonresponse rates predict students' educational attainment and employment outcomes as adults (even after controlling for cognitive ability measures), concluding that item non-response is a good proxy to measure character skills related to conscientiousness. Similarly, Zamarro et al. (2016) find that careless answering patterns in a nationally representative US survey is related to educational attainment, employment income, a greater likelihood of being employed in a high-skilled job, and self-reported measures of conscientiousness, even after controlling for cognitive ability. Most recently, Zamarro, Hitt and Mendez (2016) use 2009 PISA students' survey questionnaire to build proxies of conscientiousness and diligence by measuring the amount of effort students put forward on the survey that accompanies the PISA test. Consistent with their findings on the decline in test performance, they find that survey item nonresponse is a strong predictor of international variation in test scores.

Below, we first explore whether our main finding, namely that country-of-ancestry gender social norms affect girls' test performance relative to those of boys, is driven by cognitive or non-cognitive factors. Second, we explore whether country-of-ancestry gender social norms are related to gender differences in non-cognitive skills.

Finding that country-of-ancestry social gender norms affect the gender gap in test scores through both cognitive and non-cognitive skills would suggest that parents from less gender-equal societies care less about their daughters' success in life in general than their sons' success. Instead, evidence that country-of-ancestry social gender norms *only* affect the gender gap through cognitive skills would suggest that parents' gender stereotypes do not shape girls' non-cognitive skills relative to boys. To put it differently, believing that "the best women are stay-at-home moms", "women are supposed to make less money than men", "women are not politicians", "girls have to work hard to learn in school, whereas boys are naturally gifted", or "women are nurses, not doctors" may well affect how much parents push their daughters to learn relative to boys, without affecting parents' expectations on their daughters' motivation, ambition or agreeableness relative to that of their sons.

To do so, we follow Borghans and Schils (2012) and use the information on the decline of students' performance in the PISA achievement tests to disentangle cognitive and non-cognitive skills. These authors exploit the randomization in the order of PISA questions to identify the cognitive versus the non-cognitive factors behind the PISA achievement test.

Applying their methodology to our analysis, we use OLS to estimate the following specification:

$$Y_{iqjkt} = \alpha_0 + \alpha_1 female_i + \alpha_2(female_i GE_j) + \gamma_1 order_{iq} + \gamma_2(female_i order_{iq}) + \gamma_3(order_{iq} GE_j) + \gamma_4(female_i order_{iq} GE_j) + X'_{ijkt}\beta_1 + X'_{ijkt} female_i \beta_2 + \lambda_j + \lambda_k + \lambda_t + \lambda_q + \delta(female_i \lambda_k) + \varepsilon_{ijkt} \qquad (2)$$

Where Y_{iqjkt} is 0 if the answer of participant i who lives in host country k at time t and of ancestry j on question q is wrong, 0.5 if the answer is partially right and 1 if the answer is right.[33] The variable $order_{iq}$ indicates the sequence number of the test question q for individual i, rescaled such that the first question is numbered as 0 and the last question as 1. Question fixed effects, λ_q, control for unobserved characteristics of the question such as clarity, difficulty, type of question (math, reading, or science), and nature (multiple choice versus an open question). See equation 1 for an explanation of the other covariates.[34]

The coefficient γ_1 shows the pattern of the test performance drop, that is the variable of interest in Borghans and Schils (2012). A significant and negative γ_1 coefficient would reveal a decline in performance from the first to the last question of the test. The interaction between $order_{iq}$ and the female indicator, γ_2, captures whether there is a gender differential decline in performance along the test. The interaction between $order_{iq}$ and the GE_j, γ_3, captures whether there is a

[33] Following Borghans and Schils (2012), questions that have not been reached by the student are classified as missing, and those that have been skipped are classified as wrong.

[34] To identify the effect of cognitive skills, we assume that non-cognitive skills do not affect the answer on a test in the beginning of the test. This is a normalization that defines cognitive skills as the performance at the first question. Borghans and Schils (2012) show that there is no strong correlation between the decline in performance and the performance on the first question.

differential decline in performance along the test among those second-generation immigrants whose parents come from more (or less) gender equal countries.

Our two coefficients of interest are: (1) the interaction between the GE_j and the female indicator, α_2; and (2) the interaction between $order_q$, the GE_j and the female indicator, γ_4. In equation 2, α_2 captures whether gender equality in the country of ancestry affects gender differences in test scores of second-generation immigrant boys and girls in the host country via the cognitive component. In contrast, γ_4 captures whether gender equality in the country of ancestry affects test scores via a non-cognitive component (such as agreeableness, motivation or ambition). A positive and significant α_2 would suggest that more gender-equal attitudes in the immigrant's country of ancestry are associated with a higher relative test performance of second-generation immigrant girls over boys because of its effect on cognitive skills. In contrast, a positive and significant γ_4 would suggest that more gender-equal attitudes in the immigrant's country of ancestry are associated with a higher relative performance of second-generation immigrant girls over boys via non-cognitive skills.

Column 1 in Table 6 explores whether there is a differential gender pattern in test performance drop by estimating a regression similar to equation 2 with only a female indicator, the $order_{iq}$ variable, their interaction, and country-of-ancestry, host-country, question, and year fixed effects. It is noteworthy that our estimate of the decline in test performance, γ_1, is very similar to that of Borghans and Schils (2012).[35] Moreover, we do not observe a gender differential in the performance decline as α_2 is zero and not statistically significant. The positive and statistically significant coefficient on the gender dummy, α_1, reflects that on average, second-generation girls perform better in the first question of the test.

Column 2 in Table 6 estimates equation 2. Interestingly, we find that α_2 is positive and statistically significant, while γ_4 is also positive, but half the size and not statistically significant. These findings suggest that the evidence that second-generation girls whose parents come from more gender-equal countries

[35] Borghans and Schils (2012) estimate an γ_1 coefficient ranging between 0.07 and 0.09 (shown in Table 1). The intercept, which represent the average student's performance on the first question is also close to Borghans and Schils (2012) estimate, which ranges between 0.46 and 0.50.

outperform their male counterparts in reading and science, as well as in math, is driven by cognitive factors.

Columns 4 and 6 in Table 6 re-estimate equation 2 using more flexible specifications. Column 4 estimates equation 2 using an outcome variable that captures the deviation of each second-generation immigrant's response to each test question to the average response in his or her host country. More specifically, for each individual and each question, we estimated the difference between his or her actual score and the predicted value of his or her score in his or her host country. This predicted value was estimated separately for each host country and PISA wave using all individuals (including natives) but excluding the second-generation immigrant for whom we are predicting the answer. The covariates in the predicted model are $order_{iq}$ and question fixed effects. Concerns that specification 2 does not interact the host-country, year, order and question fixed effects are addressed in this more flexible specification. Results in column 4 resemble those of column 2. Column 6 follows Zamarro, Hitt and Mendez (2016) in that it introduces a random intercept and a random slope to the model of Borghans and Schils (2012).[36] Doing so, allows for students to deviate from the average performance in the first question as well as from the average decline in performance (Zamarro, Hitt and Mendez, 2016). Again, results are similar to those in columns 2 and 4.

An alternative and complementary question is whether second-generation girls put more or less effort in answering survey questionnaires than boys and whether country-of-ancestry gender social norms affect differentially non-cognitive skills of second-generation girls and boys. Estimates in columns 7 and 8 address these two questions by estimating equation 1 using as left-hand-side variable a measure of non-cognitive skills, namely item nonresponse on PISA students' surveys questionnaires. In a second specification (columns 9 and 10) we also control for cognitive abilities (math, reading and science test scores). Estimates from column 7 shows that second-generation girls' item non-response are lower than boys, suggesting that they are more conscientious when answering

[36] While Zamarro, Hitt and Mendez (2016) estimate each equation separately for each country, we pool our 9 host countries together and add a host-country fixed effect as Borghans and Schils (2012).

the students' questionnaires than similar boys. This difference however disappears after controlling for cognitive abilities (column 9). Estimates from column 8 and 10 provide no evidence that country-of-ancestry gender social norms affects this measure of non-cognitive skills, consistent with the results in columns 2, 4, and 6. While Zamarro, Hitt and Mendez (2016) find that about one third of the between country variation in 2009 PISA scores is driven by similar measures of conscientiousness, we find no evidence that countries' gender social norms are driving these results.

10. Conclusion

Merging data from PISA and the World Economic Forum, this papers presents evidence that second-generation girls whose parents come from more gender-equal countries outperform their male counterparts in reading and science, as well as in math, suggesting that cultural beliefs on the role of women in society are not specific to math skills, but instead more general as they also apply to reading and science skills. Interestingly, we find that it is the persistence of beliefs on women's political empowerment (and economic opportunities to a lesser extent) that drive these results. Furthermore, evidence on students' self-reported beliefs on math suggests that parents' gender social norms shape youth's test scores by transmitting preferences for cognitive skills. Our results are robust to a battery of sensitivity checks. These results are driven by a relatively larger effect of cultural beliefs about the role of women in society on test scores on girls than boys. We also find suggestive evidence for horizontal transmission of culture, as well as vertical transmission from parents to children. Our findings do not appear to be driven by non-cognitive skills (such as motivation, agreeableness and ambition), supporting that social gender norms affect parent's expectations on girls' learning cognitive skills, but not on other attributes for success.

References

Alesina, Alberto, and Paola Giuliano. 2010. "The Power of the Family." *Journal of Economic Growth* 15(2): 93–125.

———. 2011. "Family Ties and Political Participation." *Journal of the European Economic Association* 9(5): 817–39..

Almond, Douglas, Lena Edlund, and Kevin Milligan. 2013. "Son Preference and the Persistence of Culture: Evidence from South and East Asian Immigrants to Canada." *Population and Development Review* 39(1): 75–95.

Antecol, Heather. 2000. "An Examination of Cross-Country Differences in the Gender Gap in Labor Force Participation Rates." *Labour Economics* 7(4): 409–26.

Bedard, Kelly, and Insook Cho. 2010. "Early Gender Test Score Gaps across OECD Countries." *Economics of Education Review* 29(3): 348–63.

Blau, Francine D., Lawrence M. Kahn, AlbertYung-Hsu Liu, and Kerry L. Papps. 2013. "The Transmission of Women's Fertility, Human Capital, and Work Orientation across Immigrant Generations." *Journal of Population Economics* 26(2): 405–35.

Borghans, Lex, and Trudie Schils. 2012. "The Leaning Tower of Pisa. Decomposing Achievement Test Scores into Cogntive and Noncognitive Components" http://www.sole-jole.org/13260.pdf.

Carroll, Christopher D., Byung-Kun Rhee, and Changyong Rhee. 1994. "Are There Cultural Effects on Saving? Some Cross-Sectional Evidence." *The Quarterly Journal of Economics* 109(3): 685–99.

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. "Estimating The Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78(3): 883–931.

Ellison, Glenn, and Ashley Swanson. 2010. "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions." *Journal of Economic Perspectives* 24(2): 109–28.

Eugster, Beatrix, Rafael Lalive, Andreas Steinhauer, and Josef Zweimüller. 2011. "The Demand for Social Insurance: Does Culture Matter?" *The Economic Journal* 121: 413–48.

Fernández, Raquel. 2007. "Women, Work and Culture." *Journal of the European Economic Association* 24(4): 329–30.

Fernández, Raquel, and Alessandra Fogli. 2006. "Fertility: The Role of Culture and Family Experience." *Journal of the European Economic Association* 4(2–3): 552–61.

———. 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics* 1(1): 146–77.

Fryer, Ronald, and Steven Levitt. 2010. "An Empirical Analysis of the Gender Gap in Mathematics." *American Economic Journal: Applied Economics* 2(2): 210–40.

Furtado, Delia, Miriam Marcén, and Almudena Sevilla. 2013. "Does Culture

Affect Divorce? Evidence from European Immigrants in the United States." *Demography* 50(3): 1013–38.

Giuliano, Paola. 2007. "Living Arrangements in Western Europe: Does Cultural Origin Matter?" *Journal of the European Economic Association* 5(September): 927–52.

González de San Román, Ainara, and Sara de la Rica. 2016. "Gender Gaps in PISA Test Scores: The Impact of Social Norms and the Mother's Transmission of Role Attitudes." *Estudios de Economía Aplicada* 34(1): 79–108.

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales. 2008. "Culture, Gender, and Math." *Science (New York, N.Y.)* 320: 1164–65.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *The Journal of Economic Perspectives* 20(2): 23–48.

Hausmann, Ricardo, LDA Tyson, and Saadia Zahidi. 2009. *The Global Gender Gap Report 2008*. World Economic Forum.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103(6): 2052–86.

Heckman, James, and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *American Economic Review* 91(2): 145–49.

Hitt, Collin, Julie Trivitt, and Albert Cheng. 2016. "When You Say Nothing at All: The Predictive Power of Student Effort on Surveys." *Economics of Education Review* 52: 105–19.

Hyde, Janet S., and Janet E. Mertz. 2009. "Gender, Culture, and Mathematics Performance." *Proceedings of the National Academy of Sciences of the United States of America* 106(22): 8801–7.

Luttmer, Erzo F. P, and Monica Singhal. 2011. "Culture, Context, and the Taste for Redistribution." *American Economic Journal: Economic Policy* 3(1): 157–79.

Nollenberger, Natalia, Núria Rodríguez-Planas, and Almudena Sevilla. 2016. "The Math Gender Gap: The Role of Culture." *American Economic Review* 106(5): 257–61.

OECD. 2003. *The PISA 2003 Assessment Framework - Mathematics, Reading, Science and Problem Solving Knowledge and Skills.* París.

Ost, Ben. 2010. "The Role of Peers and Grades in Determining Major Persistence in the Sciences." *Economics of Education Review* 29(6): 923–34.

Pope, Devin G, and Justin R Sydnor. 2010. "Geographic Variation in the Gender Differences in Test Scores." *Journal of Economic Perspectives* 24(2): 95–108.

Quinn, D.M., and N. Cooc. 2015. "Science Achievement Gaps by Gender and Race/Ethnicity in Elementary and Middle School: Trends and Predictors." *Educational Researcher* 44(6).

Rodríguez-Planas, Núria, and Anna Sans-de-Galdeano. 2016. "Social Norms and Teenage Smoking : The Dark Side of Gender Equality" IZA Discussion Paper No. 10134.

Schnepf, Sylke Viola. 2007. "Immigrants' Educational Disadvantage: An Examination across Ten Countries and Three Surveys." *Journal of Population Economics* 20(3): 527–45.

Turner, Sarah E., and William G. Bowen. 1999. "Choice Of Major: The Changing (Unchanging) Gender Gap" *Industrial & Labor Relations Review* 52(2): 289–313.

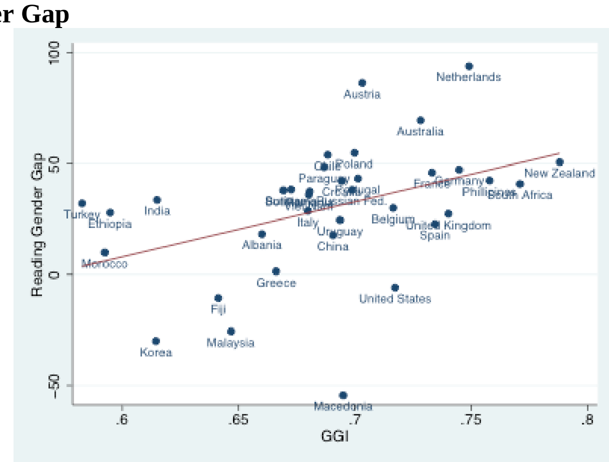
Xie, Yu, and Kimberlee A. Shauman. 2003. *Women in Science: Career Processes and Outcomes*. Harvard University Press.

Zamarro, Gema, Albert Cheng, Collin Hitt, and Collin Hitt. 2016. "Comparing and Validating Measures of Character Skills: Findings from a Nationally Representative Sample" (May 9, 2016). EDRE Working Paper No. 2016-08.

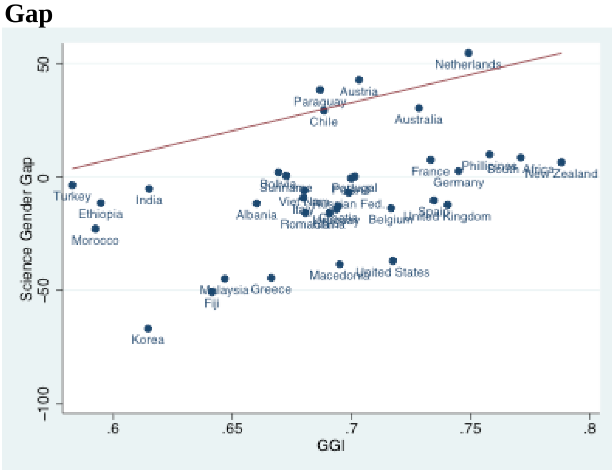
Zamarro, Gema, Collin Hitt, and Ildefonso Mendez. 2016. "When Students Don't Care: Reexamining International Differences in Achievement and Non-Cognitive Skills" (October 2016). EDRE Working Paper No. 2016-18.

Figure 1. Gender Gap in Test Scores of Second-generation Immigrants and Gender Equality in Countries of Ancestry

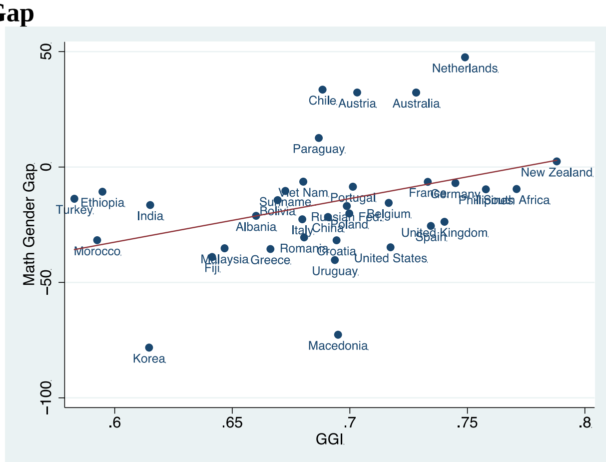
Panel A. Reading Gender Gap



Panel B. Science Gender Gap



Panel C. Math Gender Gap



Notes: These figures display the correlation between the raw average test scores gender gap among second-generation immigrants and the Gender Gap Index (GGI) in the country of ancestry. Panel A and B present the figures for reading and science test scores, respectively, whereas Panel C replicates Figure 1 presented in Nollenberger, Rodríguez-Planas, and Sevilla (2016) for math test scores. The test scores gender gap were obtained from estimating a linear regression using the plausible values provided by the PISA data sets as LHS variable and a female indicator as RHS variable. We estimated one regression for each PV for each country and present the average of the 5 coefficients estimated. We use individuals whose both parents were born in a foreign country from the 2003, 2006, 2009 and 2012 PISA datasets. The index of gender equality is the Gender Gap Index (GGI) from the 2009 World Economic Forum.

Table 1. Culture and Gender Gaps in Math, Math Preferences, Beliefs, and Anxiety

	Math Test Scores		Math self-concept		Math self-efficacy		Math improve career prospects		Anxiety doing math		Math interest	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
Female	-14.77***	-177.15	-0.33***	-1.58	-0.35***	-2.50	-0.25***	-0.74	0.30***	0.19	-0.28***	-3.43
	[2.74]	[298.18]	[0.04]	[2.53]	[0.04]	[2.09]	[0.05]	[2.64]	[0.05]	[2.67]	[0.03]	[2.75]
GGI×Female		**149.55****		**0.31**		**0.31**		**1.48**		**0.60**		**1.53****
		[62.62]		**[0.99]**		**[1.11]**		**[1.04]**		**[1.12]**		**[0.63]**
Age of student		7.90		0.05		0.13		-0.14		-0.10		-0.07
		[6.71]		[0.11]		[0.10]		[0.10]		[0.13]		[0.09]
Age×Female		6.07		0.11		0.16		0.06		-0.03		0.16
		[9.54]		[0.15]		[0.13]		[0.16]		[0.17]		[0.17]
Diff. grade		-13.82***		-0.01		0.06		-0.22***		0.02		-0.08*
		[4.69]		[0.06]		[0.06]		[0.07]		[0.05]		[0.05]
Diff. grade×Female		-5.64		-0.07		-0.10		0.17		0.08		0.01
		[6.30]		[0.10]		[0.07]		[0.10]		[0.09]		[0.07]
GDP×Female		-3.94		-0.09		-0.06		-0.20***		0.01		-0.06
		[3.30]		[0.05]		[0.05]		[0.06]		[0.06]		[0.06]
Constant	372.32***	243.53**	0.24	-0.83	-0.23	-2.44	0.39**	2.76*	0.43	1.88	0.40	1.72
	[33.33]	[117.25]	[0.41]	[1.71]	[0.14]	[1.54]	[0.16]	[1.61]	[0.27]	[2.02]	[0.50]	[1.34]
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country of ancestry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country FE× Female	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	11,527	11,527	4,396	4,396	4,507	4,507	4,514	4,514	4,399	4,399	4,521	4,521
R^2	0.34	0.35	0.09	0.10	0.13	0.14	0.09	0.11	0.09	0.10	0.09	0.10

Notes: Results from estimating equation 1 using as LHS variable the PISA indices displayed in each column (refer to the main text for a definition of each index and Appendix Table A.2 for descriptive statistics). Specification in odd columns include a female dummy, year and countries fixed-effects (for both ancestry and host countries). Specification in even columns add our variable of interest (GGI× Female) and control for individuals' age and dummies for any students who are in a grade different from the modal one in the country and its interactions with the female indicator, and the GDP per capita from the country of ancestry interacted by the female indicator. These indices are only available in 2003 and 2012 PISA waves, which are focused in math. In Appendix Table A.5, we show the results from estimating the same specification using as LHS variable math scores over this reduced sample. Standard Errors are adjusted following the Fay's BRR methodology using the 80 alternative weights provided by the PISA datasets.
* p<0.1, ** p<0.05, *** p<0.01

Table 2. Effect of Gender Equality in the Country of Ancestry on the Math Gender Gaps, by Measure of Gender Equality

	Math Scores				
	(1)	(2)	(3)	(4)	(5)
Female	-100.90	-135.25	-139.99	-392.91	-140.25
	[154.09]	[155.37]	[158.73]	[344.62]	[365.51]
GGI Pol. Emp.×Female	71.72**				52.72
	[33.53]				[37.18]
GGI Ec. Opp.×Female		56.62*			61.61
		[29.58]			[42.82]
GGI Educ.×Female			38.83		-63.77
			[63.78]		[78.47]
GGI Health×Female				295.37	53.76
				[338.44]	[358.82]
Constant	242.38**	242.57**	242.70**	240.96**	241.36**
	[118.94]	[118.18]	[117.41]	[119.79]	[120.43]
Year FE	Yes	Yes	Yes	Yes	Yes
Country of ancestry FE	Yes	Yes	Yes	Yes	Yes
Host country FE	Yes	Yes	Yes	Yes	Yes
Host country FE×Female	Yes	Yes	Yes	Yes	Yes
N	11,527	11,527	11,527	11,527	11,527
R^2	0.35	0.35	0.35	0.35	0.35
H_0: All coefficients are jointly equal to zero (Prob>χ^2)					0.051

Notes: Results from estimating our baseline specification (specification in Column 2, Table 1) using alternative measures of Gender Equality (see Appendix Table A.1 and Table A.2 for definitions and descriptive statistics of each measure). In all cases, we use the five plausible values of math test scores provided by PISA datasets and report the average coefficient (Stata command *pv*). Standard errors are adjusted following the Fay's BRR methodology using the 80 alternative weights provided by the PISA datasets.
p<0.1, ** p<0.05, *** p<0.01

Table 3. Culture and Gender Gaps in Reading and Science Test Scores

	A. Reading Test Scores		B. Science Test Scores	
	(1)	(2)	(3)	(4)
Female	32.25***	-339.29	-4.73	-343.57
	[3.18]	[517.76]	[3.19]	[519.63]
GGI×Female		**179.27*** **		**186.90*** **
		[68.25]		**[65.67]**
Age of student		0.61		4.24
		[6.69]		[6.96]
Age×Female		17.86*		16.22
		[9.80]		[9.99]
Diff. grade		-13.79***		-14.00***
		[5.00]		[4.60]
Diff. grade×Female		-9.12		-6.79
		[7.07]		[6.73]
GDP×Female		-3.26		-5.02
		[4.01]		[3.74]
Constant	373.40***	360.49***	383.99***	306.93***
	[53.22]	[110.32]	[48.31]	[111.61]
Year FE	Yes	Yes	Yes	Yes
Country of ancestry FE	Yes	Yes	Yes	Yes
Host country FE	Yes	Yes	Yes	Yes
Host country FE×Female	No	Yes	No	Yes
N	11,527	11,527	11,527	11,527
R^2	0.34	0.35	0.32	0.33

Notes: Results from estimating equation 1 on individuals' tests scores. In all cases we use the five plausible values of math test scores provided by PISA datasets and report the average coefficient (Stata command *pv*). Specification in columns (1) include a female dummy, year and countries fixed-effects (for both ancestry and host countries). Specification in columns (2) add our variable of interest (GGI× Female) and control for individuals' age and dummies for any students who are in a grade different from the modal one in the country and its interactions with the female indicator, and the GDP per capita from the country of ancestry interacted by the female indicator. Standard Errors are adjusted following the Fay's BRR methodology using the 80 alternative weights provided by the PISA datasets.
* p<0.1, ** p<0.05, *** p<0.01

Table 4. Effect of Gender Equality in the Country of Ancestry on the Reading and Science Gender Gaps, by Measure of Gender Equality

	A. Reading Scores					B. Science Scores				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Female	-246.77	-286.86*	-302.19*	-706.31**	-476.54	-248.28	-288.71*	-314.61**	-668.18*	-457.09
	[154.76]	[155.81]	[158.22]	[322.23]	[339.75]	[153.16]	[154.63]	[156.30]	[357.80]	[393.36]
GGI Pol. Emp.×Female	86.67**				68.33*	85.52**				68.94*
	[34.06]				[37.81]	[35.03]				[39.02]
GGI Ec. Opp.×Female		62.64*			43.80		65.34**			36.59
		[32.52]			[43.91]		[30.16]			[46.73]
GGI Educ.×Female			63.95		-18.32			87.32		16.88
			[73.02]		[87.99]			[69.61]		[93.95]
GGI Health×Female				472.93	242.05				430.18	203.32
				[337.17]	[353.08]				[358.58]	[390.03]
Constant	359.12***	359.26***	359.95***	357.13***	358.56***	305.45***	305.65***	306.91***	303.57***	305.82***
	[108.14]	[109.39]	[110.81]	[108.17]	[107.42]	[112.61]	[113.34]	[114.36]	[112.83]	[112.43]
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country of ancestry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country FE×Female	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	11,527	11,527	11,527	11,527	11,527	11,527	11,527	11,527	11,527	11,527
R^2	0.34	0.34	0.34	0.34	0.35	0.33	0.33	0.33	0.33	0.33
H_0: All coefficients are jointly equal to zero (P-value)					0.047					0.046

Notes: Results from estimating our baseline specification (specification in column 2, Table 1) using alternative measures of Gender Equality (see Table 1 and Table A.2 for definitions and descriptive statistics of each measure). In all cases, we use the five plausible values of math test scores provided by PISA datasets and report the average coefficient (Stata command *pv*). Standard errors are adjusted following the Fay's BRR methodology using the 80 alternative weights provided by the PISA datasets.
p<0.1, ** p<0.05, *** p<0.01

Table 5. Subgroup Analysis

A. By gender	Math scores	Reading scores	Science scores
Boys			
GGI	96.13**	149.37***	138.65***
	[45.50]	[45.49]	[47.11]
Girls			
GGI	240.49***	327.38***	326.37***
	[39.61]	[39.65]	[42.02]
H_0: Equal GGI across samples (P-value)	0.02	0.00	0.00

B. By proportion of immigrants of same origin at school	Math scores	Reading scores	Science scores
Below median			
GGI×Female	88.20	115.48*	150.87**
	[69.21]	[69.42]	[72.46]
Above median			
GGI×Female	294.50***	235.16**	273.85**
	[105.29]	[106.83]	[112.78]
H_0: Equal GGIxFemale across samples (P-value)	0.10	0.35	0.36

C. By proportion of girls at school	Math scores	Reading scores	Science scores
Below median			
GGI×Female	93.21	103.68	87.30
	[87.73]	[87.16]	[92.90]
Above median			
GGI×Female	112.98	136.69	178.31**
	[82.99]	[83.93]	[87.52]
H_0: Equal GGIxFemale samples (P-value)	0.87	0.78	0.48

Notes: Results from estimating our preferred specification (specification in column 2 of Table 2) with different samples. Note that in Panel A, the GGI interacted by gender is not included in the specification. In all cases, we use the five plausible values of math, reading and science test scores provided by PISA datasets and report the average coefficient (Stata command *pv*). Standard Errors are adjusted following the Fay's BRR methodology using the 80 alternative weights provided by the PISA datasets. We report the p-value of a test about equality of coefficients (GGI or GGIxfemale) across different samples (we use the Stata command SUEST) .
* p<0.1, ** p<0.05, *** p<0.01

Table 6. Culture and Gender Gaps in non-cognitive outcomes

Dep. variable:	*Score by question*		*Actual – Predicted Score by question*		*Score by question*		*Item nonresponse rate to survey questionnaire*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female	0.03***	-0.64***	-0.00	-0.68***	0.01	-0.11	-0.01***	0.04	-0.00	-0.08
	[0.01]	[0.10]	[0.00]	[0.12]	[0.01]	[0.09]	[0.00]	[0.19]	[0.01]	[0.18]
Order	-0.08***	-0.05	-0.11***	-0.08	-0.10***	-0.05				
	[0.00]	[0.04]	[0.01]	[0.05]	[0.00]	[0.03]				
Order×Female	0.00	-0.08	0.00	-0.05	0.02***	0.04				
	[0.01]	[0.06]	[0.01]	[0.08]	[0.00]	[0.05]				
GGI×Female		0.21***		0.17**		0.00		-0.02		0.04
		[0.06]		[0.07]		[0.06]		[0.07]		[0.07]
GGI×Order× Female		-0.04		-0.04		-0.06				
		[0.07]		[0.08]		[0.05]				
GGI×Order		0.13		0.08		-0.03				
		[0.09]		[0.11]		[0.07]				
GDP×Female		-0.01***		-0.01**		0.01*		-0.00		-0.00
		[0.00]		[0.00]		[0.00]		[0.00]		[0.00]
Age of student		-0.00		-0.00		0.02***		-0.01		-0.01
		[0.00]		[0.01]		[0.00]		[0.01]		[0.01]
Age×Female		0.04***		0.04***		0.00		-0.00		0.00
		[0.01]		[0.01]		[0.01]		[0.01]		[0.01]
Diff. grade		-0.02***		-0.02***		0.00		0.02***		0.02***
		[0.00]		[0.00]		[0.00]		[0.01]		[0.01]
Diff. grade×Female		-0.01***		-0.02***		- 0.01***		-0.01		-0.01
		[0.00]		[0.00]		[0.00]		[0.01]		[0.01]
Math score									0.00	0.00
									[0.00]	[0.00]
Reading score									-0.00**	-0.00**
									[0.00]	[0.00]
Science score									-0.00**	-0.00**
									[0.00]	[0.00]
Constant	0.30***	0.32***	-0.02	-0.01	0.30***	0.02	0.08***	0.20	0.21***	0.33***
	[0.03]	[0.07]	[0.02]	[0.08]	[0.01]	[0.06]	[0.03]	[0.13]	[0.03]	[0.13]
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country of ancestry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Host country FE x female	Yes	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Question FE	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
N	731767	731767	731648	731648	731648	731648	11527	11527	11527	11527
R^2	0.20	0.20	0.03	0.03	-.-	-.-	0.67	0.68	0.70	0.70

Notes: Columns 1 and 2 display the results from estimating equation (2) over the score the student achieved on each question of the test. Following Borghans and Schils (2012), we include the (random) question order to measure the decline in performance during the test, and its interaction with the female dummy (column 1), as well as the triple interaction with the GGI and the female dummy (column 2). In columns 3 and 4, we estimate the same model but using as dependent variable the difference between the actual score on each question and the score the student would achieve if his/her decline in performance during the test were the same than the average student in the host country where his/her lives. The predicted score comes from estimating the same model as in Borghans and Schils (2012) for each host country and PISA wave, excluding the student for whom we are predicting the answer. In columns 5 and 6, we follow Zamarro, Hitt and Mendez (2016) and estimate equation 2 introducing a random intercept and a random slope. In columns 7 and 8, we estimate the equation 1 using as left-hand side variable the item nonresponse rate on PISA students' surveys questionnaires. In columns 9 and 10 we include also test scores to control for students' cognitive skills.

Appendix

Appendix Table A.1. Individual-level variables: Definition and Descriptive Statistics

Name	Definition	Mean	St. Dev. across countries of ancestry
A. Individual Characteristics			
Female	Dummy variable equal to 1 if the individual is a girl	0.52	0081
Age	Years and months	15.77	0.06
Different grade	Dummy equal to 1 if the current individual's grade is different from the modal grade at the children age in the host country and 0 otherwise.	0.35	0.17
B. Family characteristics			
Mother highest level of education (MISCED)	Index constructed by the PISA program based upon the highest education level of each parent. It has the following categories: (0) None; (1) ISCED 1 (primary education); (2) ISCED 2 (lower secondary); (3) ISCED Level 3B or 3C (vocational/pre-vocational upper-secondary); (4) ISCED 3A (upper-secondary) and/or ISCED 4 (non-tertiary post-secondary); (5) ISCED 5B (vocational tertiary); and (6) ISCED 5A, 6 (theoretically-oriented tertiary and post-graduate).	3.66	1.04
Father highest level of education (FISCED)		3.85	0.85
Mother works	Dummy equal to one if the mother (father) works, and zero otherwise. Due to the direct question about parents' labor status is not included in all PISA waves, we use students' responses about what is the mother (father) main work. The dummy takes the value of zero when the answer is housewife, student or social beneficiary (unemployed, retired, sickness, etc.) and one otherwise.	0.82	0.14
Father works		0.93	0.05
Index of home possessions (homeposs)	The index of home possessions comprises all items on the indices of wealth, cultural possessions and home educational resources, as well as books in the home recoded into a four-level categorical variable (0-10 books, 11-25 or 26-100 books,101-200 or 201-500 books, more than 500 books). The index of wealth is based on the students' responses on whether they had a room of their own, a link to the Internet, a dishwasher, a DVD player, and three other country-specific items; and their responses on the number of cellular phones, televisions, computers, cars and the rooms with a bath or shower. The index of cultural possessions is based on the students' responses to whether they had the following at home: classic literature, books of poetry and works of art. The index of home educational resources is based on the items measuring the existence of educational resources at home including a desk and a quiet place to study, a computer, educational software, books to help with students' school work, technical reference books and a dictionary.	-0.04	0.53
C. ***School characteristics***			
Percentage of girls	PISA index of the proportion of girls enrolled in each school derived from school principals' responses regarding the number of girls divided by the total of girls and boys at a school.	0.49	0.04
Private school	Dummy equal to 1 if school is private and 0 otherwise.	0.24	0.18
School location	Dummy equal to 1 if the school is in a metropolis or city and 0 if the school is in a town or village.	0.29	0.27
Percentage of immigrants from the same ethnicity	Number of immigrants from the same ethnicity (either first or second-generation) divided the total individuals by school. Own calculation based upon PISA samples by year, weighted by student final weight.	0.11	0.06

Appendix Table A.1 (cont.) Individual-level variables: Definition and Descriptive Statistics

Name	**Definition**	**Mean**	**St. Dev. across countries of ancestry**
D. Math-specific variables			
Mathematics self-concept	See main text	-0.36	0.55
Mathematics self-efficacy	See main text	-0.48	0.61
Math improve career prospects	See main text	-0.37	0.64
Mathematics anxiety	See main text	0.33	0.61
Math interest	See main text	-0.35	0.59

Appendix Table A.2. Sample Size by Country of Ancestry and Destiny

		ARG	AUS	AUT	BEL	CHE	ISR	LUX	NLD	NZL	**Total**
1	Albania					132					**132**
2	Australia									36	**36**
3	Austria					46					**46**
4	Belgium							159			**159**
5	Bolivia	131									**131**
6	Chile	24									**24**
7	China		410						27	130	**567**
8	Croatia			77							**77**
9	Ethiopia						151				**151**
10	Fiji									35	**35**
11	France				102	203	67	242			**614**
12	Germany		21	38	41	176		116			**392**
13	Greece		46								**46**
14	India		158								**158**
15	Italy		88			739		256			**1,083**
16	Korea		31							15	**46**
17	Malaysia		34								**34**
18	Morocco								192		**192**
19	Netherlands				50						**50**
20	New Zealand		376								**376**
21	Paraguay	63									**63**
22	Philippines		240								**240**
23	Poland			47							**47**
24	Portugal					777		2,069			**2,846**
25	Romania			58							**58**
26	Russian Fed.						491				**491**
27	Viet Nam		291								**291**
28	South Africa		60								**60**
29	Spain					246					**246**
30	Suriname								107		**107**
31	Turkey			509	440	591			222		**1,762**
32	Macedonia			20							**20**
33	United Kingdom		651							168	**819**
34	United States		29				82				**111**
35	Uruguay	17									**17**
	Total	**235**	**2,435**	**749**	**633**	**2,910**	**791**	**2,842**	**548**	**384**	**11,527**

Notes: Final sample of second-generation immigrants from 2003, 2006, 2009 and 2012 PISA datasets. ARG=Argentina, AUS=Australia, AUT=Austria, BEL=Belgium, CHE=Switzerland, ISR=Israel, LUX=Luxembourg, NLD=Netherlands, NZL= New Zealand.

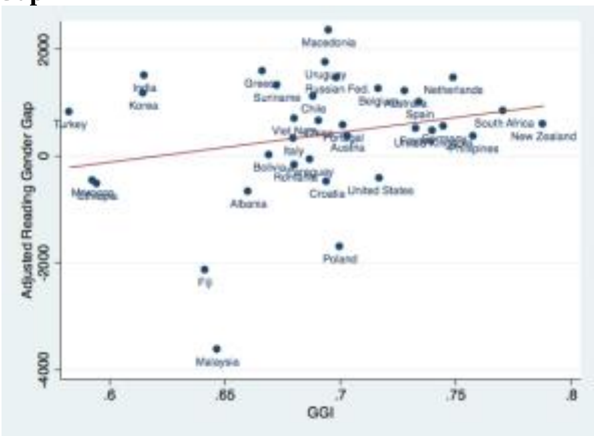
Appendix Table A.3. Gender Gap in Tests Scores and Gender Equality by Country of Ancestry

Country of ancestry	Math gender gap	Reading gender gap	Science gender gap	Math Self-concept gap	Math Self-effic. gap	Math career gap	Math anxiety gap	Math pref. gap	GGI	GGI Ec. Opp.	GGI Educ.	GGI Pol.	GGI Health
Korea	-78.24	-30.06	-66.90	0.06	-0.34	-0.47	0.59	-0.29	0.61	0.52	0.89	0.07	0.97
Macedonia	-72.64	-54.49	-38.52	-0.07	0.74	-0.54	-0.52	0.68	0.69	0.67	0.99	0.16	0.96
Uruguay	-40.31	24.55	-14.18	0.76	-0.47	0.67	-0.73	0.29	0.69	0.65	1.00	0.14	0.98
Fiji	-38.99	-10.64	-50.81	-0.61	-1.03	-0.44	0.17	-0.59	0.64	0.53	0.99	0.06	0.98
Greece	-35.53	1.44	-44.53	-0.16	0.07	-0.76	0.17	-0.49	0.67	0.61	0.99	0.09	0.98
Malaysia	-35.19	-25.68	-44.90	0.12	-0.71	-0.22	0.08	0.34	0.65	0.57	0.99	0.06	0.97
United States	-34.75	-5.90	-37.08	-0.09	-0.81	-0.20	0.11	0.49	0.72	0.75	1.00	0.14	0.98
Croatia	-31.74	42.24	-12.92	-0.13	0.34	0.67	0.77	0.12	0.69	0.65	0.99	0.16	0.98
Morocco	-31.70	9.92	-22.88	-0.35	0.20	0.00	0.35	-0.28	0.59	0.45	0.86	0.10	0.97
Romania	-30.52	37.49	-15.86	-1.08	-0.85	-0.76	0.70	-0.95	0.68	0.71	0.99	0.04	0.98
Spain	-25.55	22.78	-10.36	-0.33	-0.36	-0.26	0.42	-0.22	0.73	0.60	0.99	0.37	0.97
UK	-23.73	27.37	-12.32	-0.51	-0.45	-0.42	0.44	-0.28	0.74	0.71	1.00	0.28	0.97
Italy	-22.65	28.70	-9.18	-0.33	-0.34	-0.47	0.17	-0.29	0.68	0.59	1.00	0.16	0.97
China	-21.69	17.75	-15.95	-0.24	-0.57	0.01	0.20	-0.11	0.69	0.70	0.98	0.14	0.95
Albania	-21.16	18.23	-11.73	-0.23	-0.16	-0.46	-0.68	-0.39	0.66	0.65	0.99	0.04	0.96
Poland	-20.11	54.87	-0.59	-0.06	-1.55	-1.34	-0.79	-1.13	0.70	0.64	1.00	0.18	0.98
Russian Fed.	-16.88	38.20	-6.90	-0.45	-0.34	-0.06	0.44	0.02	0.70	0.74	1.00	0.08	0.98
India	-16.45	33.60	-5.31	-0.23	-0.64	0.25	0.57	-0.13	0.62	0.41	0.84	0.27	0.93
Belgium	-15.56	30.01	-13.81	-0.06	-0.77	-0.70	-0.22	-0.75	0.72	0.65	0.99	0.24	0.98
Bolivia	-14.36	37.98	2.02	-0.14	-0.33	-0.29	0.61	-0.54	0.67	0.59	0.97	0.15	0.97
Turkey	-13.77	32.04	-3.64	-0.35	-0.25	-0.31	0.05	-0.36	0.58	0.40	0.89	0.07	0.97
Ethiopia	-10.69	27.84	-11.48	-0.47	-0.57	0.06	0.52	0.11	0.59	0.60	0.70	0.11	0.97
Suriname	-10.39	38.32	0.43	0.02	-0.37	0.18	0.09	-0.15	0.67	0.57	0.99	0.16	0.97
Philippines	-9.66	42.40	9.93	-0.04	-0.08	-0.08	-0.02	-0.06	0.76	0.76	1.00	0.29	0.98
South Africa	-9.56	40.86	8.48	.	.	.	.	.	0.77	0.66	1.00	0.45	0.98
Portugal	-8.53	43.30	0.18	-0.57	-0.31	-0.59	0.39	-0.50	0.70	0.68	0.99	0.16	0.97
Germany	-6.96	47.27	2.59	-0.64	-0.58	-0.54	0.30	-0.77	0.74	0.70	1.00	0.31	0.98
France	-6.43	46.00	7.47	-0.69	-0.57	-0.28	0.73	0.06	0.73	0.66	1.00	0.29	0.98
Viet Nam	-6.34	35.92	-6.03	-0.28	-0.30	-0.08	0.43	-0.36	0.68	0.73	0.90	0.12	0.97
New Zealand	2.42	50.69	6.49	-0.46	-0.56	-0.28	0.48	-0.15	0.79	0.78	1.00	0.39	0.97
Paraguay	12.61	48.43	38.39	0.77	0.35	0.86	-0.01	-0.08	0.69	0.67	1.00	0.10	0.98
Australia	32.26	69.63	30.31	-1.07	-1.55	-0.48	1.03	-0.79	0.73	0.75	1.00	0.19	0.97
Austria	32.29	86.48	42.77	-2.24	-2.68	-1.55	2.18	-1.87	0.70	0.57	0.99	0.27	0.98
Chile	33.52	53.97	29.21	-1.57	-0.50	-2.69	1.94	-2.29	0.69	0.52	1.00	0.26	0.98
Netherlands	47.53	94.01	54.63	-0.39	-0.08	-0.83	0.45	-0.18	0.75	0.69	0.99	0.34	0.97
Mean	-15.70	30.16	-6.37	-0.36	-0.48	-0.37	0.33	-0.35	0.69	0.63	0.97	0.18	0.97
St. Dev.	26.04	30.03	26.10	0.55	0.61	0.64	0.61	0.59	0.05	0.10	0.06	0.11	0.01

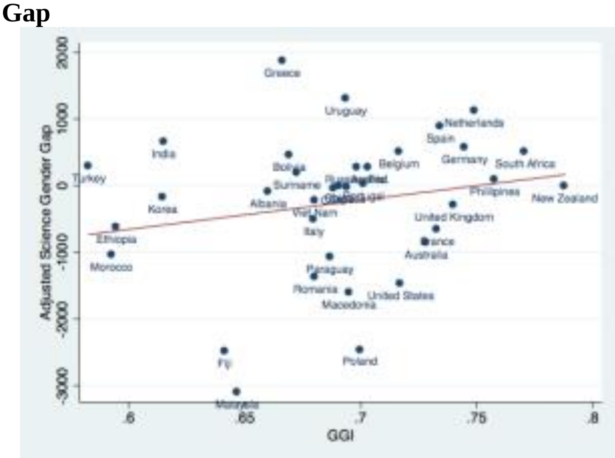
Notes: This table displays the means of test scores gender gaps, math indices and gender equality measures by country of ancestry estimated using our sample of second-generation immigrants from 2003, 2006, 2009 and 2012 PISA. Countries are ordered by the gender gap in math scores. The gap in test scores was obtained from estimating a linear regression using the plausible values provided by the PISA data sets as LHS variable and a female indicator as RHS (we estimated one regression for each PV and present the average of the 5 coefficients estimated). See Appendix Table A.2 for details about gender equality measures. The last two rows display the mean and cross-country standard deviation.

Appendix Figure 1. Adjusted Gender Gap in Test Scores of Second-generation Immigrants and Gender Equality in Countries of Ancestry

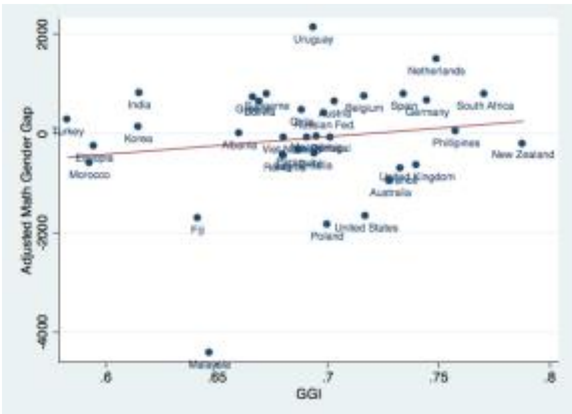
Panel A. Reading Gender Gap



Panel B. Science Gender Gap



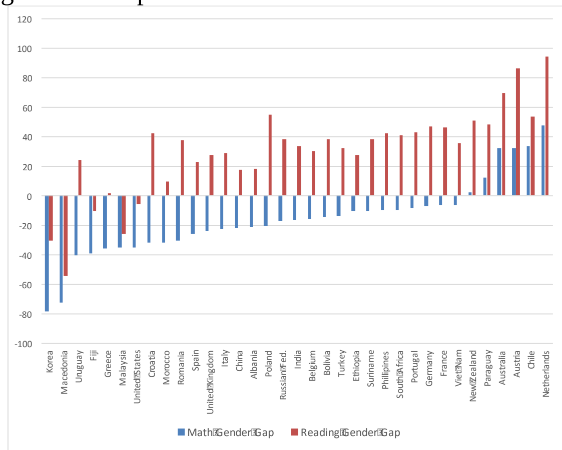
Panel C. Math Gender Gap



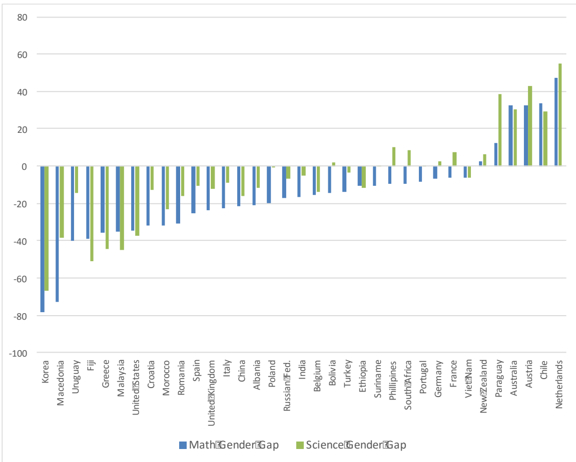
Notes: These figures display the correlation between the average test scores gender gap among second-generation immigrants and the GGI in the country of ancestry after adjusting the test scores gender gap by individual characteristics (age and dummies for being in a grade different from the modal one in the host country) and the GDP per capita of the country of ancestry. More specifically, we first estimate a linear regression using the individual plausible values provided by the PISA data sets as LHS variable and a female indicator, individual's controls and country of ancestry fixed effects as RHS variable. We then take the math gender gap of each country of ancestry resulting from the previous exercise and regress these coefficients on the GDP per capita of the country of ancestry and gender differences in the country of ancestry.

Appendix Figure 2. Math, Reading and Science Test Scores of Second-generation Immigrants, by Country of Ancestry

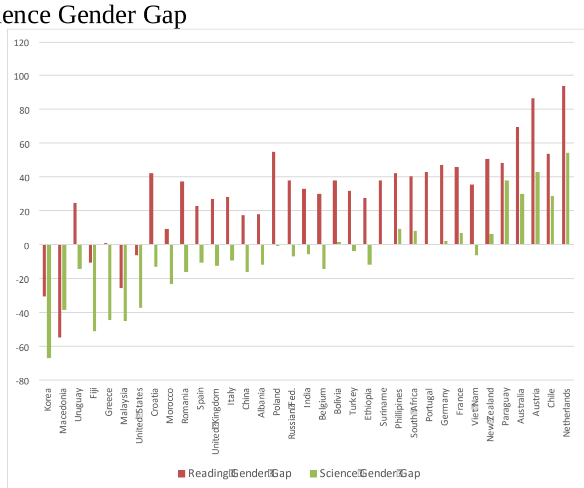
Panel. A Math and Reading Gender Gap



Panel. B Math and Science Gender Gap



Panel C. Reading and Science Gender Gap



Note: Panel A presents the relationship between the raw math and reading gender gaps among second generation immigrants, by country of ancestry. Panel B and C do the same for math and science and for reading and science, respectively. The test scores gender gap were obtained from estimating a linear regression using the plausible values provided by the PISA data sets as LHS variable and a female indicator as RHS variable. We estimated one regression for each PV for each country and present the average of the 5 coefficients estimated. We use individuals whose both parents were born in a foreign country from the 2003, 2006, 2009 and 2012 PISA datasets.

Appendix Table A.4. Gender Gap in Test Scores

	All Countries participating in PISA				Countries included in our sample					
	All individuals		Second-generation immigrants		All individuals		Second-generation immigrants		Second-generation immigrants (final sample)	
Math Scores	(1)		(2)		(3)		(4)		(5)	
Boys	460.13	[105.15]	470.02	[94.64]	488.95	[108.85]	494.46	[104.11]	493.51	[107.78]
Girls	447.70	[100.38]	459.79	[92.77]	473.21	[104.06]	476.00	[98.08]	477.81	[99.52]
Gender Gap	**-12.43**		**-10.23**		**-15.74**		**-18.46**		**-15.70**	
Reading scores										
Boys	441.18	[103.22]	453.67	[100.55]	460.94	[110.27]	465.93	[106.96]	464.69	[110.99]
Girls	472.29	[97.01]	487.64	[94.67]	494.40	[102.74]	495.82	[100.42]	494.84	[103.49]
Gender Gap	**31.11**		**33.97**		**33.46**		**29.89**		**30.16**	
Science scores										
Boys	465.41	[104.89]	469.74	[98.03]	486.28	[111.21]	484.21	[109.20]	483.79	[112.97]
Girls	461.75	[98.75]	466.39	[94.14]	483.32	[103.47]	476.53	[101.92]	477,42	[103.54]
Gender Gap	**-3.66**		**-3.35**		**-2.96**		**-7.67**		**-6.37**	
N	*1,676,363*		*84,426*		*222,082*		*22,910*		*11,527*	

Notes: Author's calculations based upon 2003, 2006, 2009 and 2012 PISA datasets. Mean and standard deviation in brackets. The nine (host) countries included in our sample are: Argentina, Australia, Austria, Belgium, Switzerland, Israel, Luxembourg, Netherlands, New Zealand.

Table A.5. Math Gender Gap and Culture with PISA Math Indexes Sub-Samples

Dep. Variable: Math scores	Sub-sample: 2003 and 2012 PISA	Sub-sample: no missing responses to each PISA math index below				
		Intrinsic motivation to learn mathematics	Instrumental motivation to learn mathematics	Mathematics self-efficacy	Mathematics self-concept	Mathematics Anxiety
GGI×Female	175.33*	204.01*	208.63**	205.23*	118.35	122.91
	[94.14]	[105.56]	[105.77]	[106.20]	[98.83]	[97.84]
N	5,850	4,521	4,514	4,507	4,396	4,399
% of missing		22.7%	22.8%	23.0%	24.9%	24.8%
R^2	0.36	0.33	0.33	0.32	0.34	0.35
Year FE		Yes	Yes	Yes	Yes	Yes
Country of ancestry FE		Yes	Yes	Yes	Yes	Yes
Host country FE		Yes	Yes	Yes	Yes	Yes
Host country FE x female		Yes	Yes	Yes	Yes	Yes

Notes: Results from estimating our preferred specification (specification in column 2 of Table 2) on the samples of respondents to the math indexes reported in each column. In all cases, we use the five plausible values of math test scores provided by PISA datasets and report the average coefficient (Stata command *pv*). Standard errors are adjusted following the Fay's BRR methodology using the 80 alternative weights provided by the PISA datasets.
p<0.1, ** p<0.05, *** p<0.01

Table A.6. Correlations Between Gender Equality Measures

	GGI	GGI Ec. Opp.	GGI Educ	GGI Pol. Emp.	GGI
GGI	1				
GGI Ec. Opp.	0.77†	1			
GGI Educ.	0.69†	0.48†	1		
GGI Pol. Emp.	0.73†	0.23	0.24	1	
GGI Health	0.36†	0.32†	0.39†	0.06	1

Notes: Table A.6 displays Pearson correlations between variables. † Indicates a correlation statistically significant at 5 percent.

Appendix Table A.7. Reading Scores and the Gender Gap Index (GGI)

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-344.68**	-332.49**	-260.55	-352.55**	-311.83*	-317.96**
	[157.53]	[157.31]	[160.68]	[158.03]	[159.35]	[160.43]
GGI ×Female	**146.99*****	**179.27*****	**174.26****	**172.93*****	**192.23*****	**173.91*****
	[55.14]	**[68.25]**	**[67.91]**	**[63.46]**	**[64.31]**	**[64.81]**
Age of student	0.51	0.61	1.92	0.13	2.05	1.46
	[6.71]	[6.69]	[6.59]	[6.75]	[6.62]	[6.71]
Age×Female	17.98*	17.86*	13.92	19.24*	18.27*	17.75*
	[9.82]	[9.80]	[9.95]	[9.92]	[9.83]	[9.91]
Diff. grade	-13.68***	-13.79***	-16.83***	-13.65***	-11.77**	-11.36**
	[4.99]	[5.00]	[5.23]	[5.20]	[5.22]	[5.17]
Diff. grade× Female	-9.37	-9.12	-9.44	-6.99	-7.35	-7.45
	[7.03]	[7.07]	[7.37]	[7.16]	[7.09]	[6.89]
GDP×Female		-3.26	-3.81	-3.08	-3.62	-3.39
		[4.01]	[4.04]	[4.06]	[4.16]	[4.14]
Dad educ.				6.19***	5.10***	4.83***
				[1.45]	[1.48]	[1.47]
Dad educ.× Female				-1.05	-1.44	-1.38
				[2.30]	[2.40]	[2.36]
Mom educ.				3.80**	2.80*	2.55
				[1.59]	[1.61]	[1.59]
Mom educ.× Female				0.92	0.89	1.03
				[2.20]	[2.25]	[2.23]
Dad work					31.13***	30.75***
					[8.77]	[8.58]
Dad work× Female					-16.41	-16.15
					[10.92]	[10.95]
Mom work					20.68***	19.39***
					[5.42]	[5.42]
Mom work× Female					-18.86**	-16.63**
					[7.95]	[7.78]
Home possessions					9.21***	9.08***
					[2.66]	[2.65]
Home possessions ×Female					5.76	5.69
					[3.92]	[3.86]
Proportion of girls at school						-15.44
						[14.02]
Prop. girls× female						44.31**
						[19.43]
Private school						12.39
						[7.89]
Private school× female						-0.78
						[9.13]
School is in a Metropolis						23.40***
						[5.87]
School is in a Metro×Female						-19.98**
						[8.44]
GGI			152.58***			
			[56.42]			
GDP			3.02			
			[3.12]			
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Country of ancestry FE	Yes	Yes	No	Yes	Yes	Yes
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes
Host countryFE x female	Yes	Yes	Yes	Yes	Yes	Yes
N	11,527	11,527	11,527	11,527	11,527	11,527
R^2	0.35	0.35	0.30	0.38	0.39	0.40

Notes: Results from estimating equation 1 on individuals' reading scores. In all cases we use the five plausible values of math test scores provided by PISA datasets and report the average coefficient (Stata command *pv*). Following OECD recommendations, standard Errors are adjusted following the Fay's BRR methodology using the 80 alternative weights provided by the PISA datasets.
* p<0.1, ** p<0.05, *** p<0.01

Appendix Table A.8. Science Scores and the Gender Gap Index (GGI)

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-355.04**	-336.27**	-263.05	-361.59**	-324.46**	-339.57**
	[159.05]	[156.69]	[162.53]	[156.16]	[156.03]	[155.91]
GGI ×Female	**137.19****	**186.90*****	**183.14*****	**194.71*****	**212.52*****	**200.47*****
	[53.81]	**[65.67]**	**[67.03]**	**[61.61]**	**[61.48]**	**[61.46]**
Age of student	4.08	4.24	5.45	3.89	5.07	4.21
	[6.98]	[6.96]	[6.87]	[6.98]	[6.86]	[6.96]
Age×Female	16.40	16.22	12.33	17.50*	17.05*	16.83*
	[10.04]	[9.99]	[10.08]	[9.87]	[9.85]	[9.91]
Diff. grade	-13.84***	-14.00***	-16.94***	-13.77***	-12.51***	-12.11***
	[4.59]	[4.60]	[4.88]	[4.67]	[4.66]	[4.66]
Diff. grade×	-7.17	-6.79	-7.24	-4.88	-4.54	-4.42
Female	[6.72]	[6.73]	[7.14]	[6.78]	[6.73]	[6.48]
GDP×Female		-5.02	-5.85	-4.72	-5.77	-5.37
		[3.74]	[3.80]	[3.78]	[3.84]	[3.77]
Dad educ.				6.47***	5.29***	5.05***
				[1.45]	[1.47]	[1.46]
Dad educ.×				-0.32	-0.87	-0.79
Female				[2.26]	[2.30]	[2.27]
Mom educ.				5.69***	4.49***	4.28***
				[1.48]	[1.49]	[1.49]
Mom educ.×				-1.29	-1.45	-1.32
Female				[2.21]	[2.26]	[2.23]
Dad work					25.32***	25.00***
					[7.41]	[7.34]
Dad work×					-17.75*	-17.31*
Female					[10.26]	[10.38]
Mom work					19.01***	17.86***
					[5.22]	[5.11]
Mom work×					-14.76*	-12.12
Female					[7.97]	[7.86]
Home possessions					10.08***	9.90***
					[2.87]	[2.84]
Home possessions					8.24**	7.76*
×Female					[4.12]	[4.09]
Proportion of girls						-8.34
at school						[15.23]
Prop. girls×						38.76**
female						[18.76]
Private school						16.72**
						[7.73]
Private school×						0.42
female						[8.55]
School is in a						15.45**
Metropolis						[6.10]
School is in a						-17.02**
Metro×Female						[7.48]
GGI			143.11**			
			[56.93]			
GDP			5.35*			
			[3.02]			
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Ancestry country FE	Yes	Yes	No	Yes	Yes	Yes
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes
Host country FE×Fem.	Yes	Yes	Yes	Yes	Yes	Yes
N	11,527	11,527	11,527	11,527	11,527	11,527
R^2	0.33	0.33	0.28	0.36	0.38	0.39

Notes: Results from estimating equation 1 on individuals' science scores. In all cases we use the five plausible values of math test scores provided by PISA datasets and report the average coefficient (Stata command *pv*). Standard Errors are adjusted following the Fay's BRR methodology using the 80 alternative weights provided by the PISA datasets.

* p<0.1, ** p<0.05, *** p<0.01.

Table A.9. Robustness Checks

	Math scores	Reading scores	Science scores
A. Baseline			
GGI×Female	149.55**	179.27***	186.90***
	[62.62]	[68.25]	[65.67]
N	11,527	11,527	11,527
R^2	0.35	0.35	0.33
B. Controlling for ancestry-country HDI and its interaction with female			
GGI×Female	158.79**	192.45***	203.17***
	[66.52]	[73.99]	[72.94]
N	11,527	11,527	11,527
R^2	0.35	0.35	0.33
C. Host-country regional FE			
GGI×Female	133.98**	166.16**	169.53**
	[62.69]	[69.46]	[66.60]
N	11,527	11,527	11,527
R^2	0.36	0.36	0.34
D. Adding Year FE × Female			
GGI×Female	150.13**	179.38***	187.37***
	[64.12]	[68.80]	[67.79]
N	11,527	11,527	11,527
R^2	0.35	0.35	0.33
E. Cluster SE at country-of-ancestry level			
GGI×Female	149.55***	179.27***	186.90***
	[45.98]	[49.70]	[43.99]
N	11,527	11527	11527
R^2	0.37	0.37	0.35

Notes: Results from estimating equation 1 using alternative specifications. In panel B, we control for the more comprehensive Human Development Index (interacted by female) instead of for the GDP. In panel C, we control for host-country regional FE instead of countries FE. In Panel D, we add the interaction between year FE and the female dummy. Panel E presents estimates with standard errors clustered at the country of ancestry level. In all cases we use the five plausible values of math test scores provided by PISA datasets and report the average coefficient (Stata command *pv*). Except for Panel E, standard errors are adjusted following the Fay's BRR methodology using the 80 alternative weights provided by the PISA datasets.

* p<0.1, ** p<0.05, *** p<0.01

Table A.10. Sensitivity to Sample Selection

	Math scores	Reading scores	Science scores
A. Baseline			
GGI×Female	149.55**	179.27***	186.90***
	[62.62]	[68.25]	[65.67]
N	11,527	11,527	11,527
R^2	0.35	0.35	0.33
B. Dropping the most important country of ancestry (Portugal)			
GGI×Female	144.52**	173.54**	184.05***
	[65.15]	[70.81]	[67.56]
N	8,681	8,681	8,681
R^2	0.36	0.35	0.34
C. Dropping the most important host country (Switzerland)			
GGI×Female	148.77**	199.87**	185.84**
	[74.20]	[80.35]	[77.67]
N	8,617	8,617	8,617
R^2	0.38	0.37	0.36
D. Keeping only one host country			
Switzerland	163.12	85.42	184.09
	[136.34]	[137.45]	[149.98]
N	2910	2910	2910
R^2	0.12	0.16	0.14
Australia	199.01**	245.60***	235.03**
	[91.00]	[91.15]	[99.97]
N	2,450	2,450	2,450
R^2	0.16	0.12	0.11
E. Dropping those countries that send immigrants to only one host country			
GGI×Female	228.01**	154.40	194.15*
	[101.93]	[105.10]	[115.99]
N	8,240	8,240	8,240
R^2	0.29	0.32	0.32

Notes: Results from estimating our preferred specification (Baseline) with different samples. In panel B we drop those second-generation immigrants whose ancestries come from Portugal (the country of origin with more observations in our sample). In panel C, we drop the host country with more observations in our sample (Switzerland). In panel D, we replicate our analysis using only one host country (Switzerland or Australia). In panel E, we drop those countries that send immigrants to only one host country. In all cases we use the five plausible values of math test scores provided by PISA datasets and report the average coefficient (Stata command *pv*). Standard Errors are adjusted following the Fay's BRR methodology using the 80 alternative weights provided by the PISA datasets.
* p<0.1, ** p<0.05, *** p<0.01