I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12436

Societal Inequalities Amplify Gender Gaps in Math

Thomas Breda
Elyès Jouini
Clotilde Napp

JUNE 2019

DISCUSSION PAPER SERIES

IZA DP No. 12436

Societal Inequalities Amplify Gender Gaps in Math

Thomas Breda
Paris School of Economics, CNRS and IZA

Elyès Jouini
Université Paris Dauphine and IZA

Clotilde Napp
CNRS and Université Paris Dauphine

JUNE 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Societal Inequalities Amplify Gender Gaps in Math*

While gender gaps in average math performance are close to zero in developed countries, women are still strongly underrepresented among math high performers. Using data from five successive waves of the Programme for International Student Assessment (PISA), we show that this underrepresentation is more severe in more unequal countries. This relationship holds for a wide range of societal inequalities that are not directly related to gender. It is also observed in other parts of the performance distribution and among various sets of countries, including developing countries. Similar relationships are found in science and reading. Such findings highlight how differences in socio-economic and cultural factors can affect gender gaps in performance.

JEL Classification: I24, J16, Z1

Keywords: gender gap in math, culture, societal inequality, income inequality, top math performers

Corresponding author:
Thomas Breda
Paris School of Economics
48 Bd Jourdan
75014 Paris
France
E-mail: thomas.breda@ens.fr

* The paper has been published in Science 16 Mar 2018: Vol. 359, Issue 6381, pp. 1219-1220, DOI: 10.1126/science.aar2307.

Main Text: There is abundant evidence that girls tend to perform on average equally or slightly worse than boys in mathematics (*1*,*2*). For example, the gap in average math performance is around 10% of a standard deviation in the Programme for International Student Assessment (PISA) in the 2000s (Table 1), which can be considered as close to zero (*3*), and it is not statistically significant in most countries. However, if we focus on high levels of performance, a large gap remains: in the top decile of the math performance distribution among countries belonging to the Organisation for Economic Co-operation and Development (OECD), there are on average only seven girls for ten boys. This underrepresentation of girls at high levels of performance is a common feature of all 35 OECD countries (Table S1), and it has remained remarkably stable since 2000 (Table 1). Gender gaps of the same magnitude are also observed in science and reading, the latter one being however in favor of girls (Table 1 and S1).

These substantial gender gaps among high performers at 15 years old are a source of concern because they affect educational choices and contribute to the underrepresentation of women in math and science, especially in higher education, and to their subsequent worse position on the labor market (*4*-*7*).

The reasons behind the gender performance gaps, especially in math, have been debated for more than a century, centered on the roles of nature and nurture (*3*,*8*-*13*), with recent research highlighting the interplay between the two (*14*,*15*). A series of papers in the 1990s and 2000s have related the performance gap in math to measures of countries' cultural attitudes toward women or labor market gender inequalities (*16*-*19*). However, more recent studies have challenged this view by showing that the former relationship is weak, not very robust across time, and not robust to the inclusion of some developing countries in the analysis (*20*-*23*). These studies imply that there is so far no clearly identified robust cross-country relationship between the gender performance gap in math and socioeconomic or cultural factors. They question anew whether differences in such factors across countries could affect this gender gap.

Former research is typically based on the *gender stratification hypothesis* according to which gender differences in opportunities and status shape numerous socialization processes that in turn may affect performance (*16*). We elaborate on this idea, assuming that the processes that transform differences in status into differences in performance depend on the degree of countries' inclusiveness; indeed, inclusive countries are likely to mitigate the impact of status differences in general. We hypothesize that women have a lower status than men in virtually all countries, but that this lower status is more likely to be detrimental to girls' performance in countries that are in general less fair and inclusive: the more unequal a country, the more the status difference between boys and girls should translate into actual differences in school performance.

To test this hypothesis, we analyze data from five successive PISA surveys. PISA is an every-three-year international survey of 15-year-old students, aimed at determining their knowledge and skills in mathematics, reading, and science. We relate the gender gaps in performance to measures of countries' inequalities that are not directly related to gender. Instead, those measures relate either to societies in general (income and other socioeconomic variables) or to their educational system in particular. Regarding the educational system, we consider inequality in both students' performance (e.g., the share of students from low socio economic background among high performers, see SM) and learning opportunities across different schools or types of students, in the spirit of a recent paper (*24*).

PISA scales are divided into proficiency levels from Level 1 (lowest) to Level 6 (highest). We focus primarily on what PISA refers to as "top performers", i.e. on students who perform at Levels 5 or 6 (10.7% of all students in OECD countries in math in 2015). Our main measure of the gender performance gap in math is the ratio of girls to boys among those high performers. We do this for two reasons. First, this ratio is still very unfavorable to girls and it has not narrowed over time. Second, performance at high levels is more closely related to the underrepresentation of women among STEM college graduates.

Among the 35 countries of the OECD observed in 2015, we find that this ratio is negatively correlated with common inequality measures such as the income Gini index (r=-0.63, Fig. 1), the income Palma ratio (r=-0.59) or the variance in the socioeconomic and cultural background of a country's students (r=-0.66), a measure that incorporates several non-financial aspects of inequalities such as parental education or cultural resources available to the family (see SM). Similar correlations are found with countries' poverty rate or infant mortality, and a wide range of inequality measures within the educational system such as the share of students from a low socioeconomic and cultural background among high performers (r=0.57, Fig. 1), or the between-school variation in students' socioeconomic background (r=-0.59). Consistent with our hypothesis of a link between the gender gap in math performance and the way countries perpetuate or reduce initial status differences, we also find a strong correlation between the gender gap in math and intergenerational earnings elasticity (r=-0.57), education mobility measured as parents-child correlations in years of schooling (r=-0.50), or the index of inequality of economic opportunity (r=-0.64), which measures the part of income inequality due to predetermined circumstances beyond individual control (like region of birth or parental background).

We provide a more systematic study of those relationships using linear regression models in which measures of inequalities are introduced one by one. All the measures of inequalities we have considered (which have been normalized to have a standard deviation of 1) are associated with a significantly lower girls-to-boys ratio among high performers in math (Tables 2 and S2, column 1). A one standard deviation increase in inequalities is typically associated with a drop of the girls-to-boys ratio of .05 to .08. About 30% of the cross-country variance in this ratio can be accounted for by a single measure of inequalities, and the explained variance reaches 60% if we consider three measures of inequalities (see SM).

A more comprehensive study of the cross-country relationship between inequalities and gender performance gaps in math is provided in the SM where we make seven main points: (i) the relationship is robust to controlling for countries' GDP and extent of gender stratification (see Table S3), (ii) with a few exceptions, the relationship is maintained when we include non-OECD countries that participated in PISA (mostly developing countries, Table S4) and when we test it using the Trends in International Mathematics and Science Study (TIMSS) instead of PISA (see all details in SM), (iii) the relationship is slightly weaker but still holds when we consider the gender gap in average performance (Table S5) and more generally it is not driven by the way we define the relative performance of girls and boys (Table S6), (iv) these patterns are also observed in 2003, 2006, 2009 and 2012 (Table S7), (v) they are observed again when looking at performance in numeracy among adults instead of students, based on the Programme for the International Assessment of Adult Competencies (PIAAC, see SM), (vi) inequalities are detrimental to both girls and boys, but the associations are systematically larger for girls than for boys (Table S8), (vii) past inequalities (in the 1980s—see Table 2, and to some extent in the 1950s and 1960s—see SM), are also negatively related to girls' relative performance in math in 2015.

We conclude from these analyses that the relationship between the math gender performance gap and several general measures of inequality is larger and much more robust than other relationships already documented with more obvious country characteristics such as gender stratification or economic development. To fully backup this point, we run "horse races" to compare the explanatory power of six possible measures of gender inequality (like the World Economic Forum's Gender Gap Index GGI) and of seven measures of countries' development (like GDP or human development index) to our measures of non-gender related inequalities. We find that when an inequality-related and another explanatory variable are jointly included in regression models, the former systematically remains statistically significant, while the latter usually does not (Table S9). In particular, the effect of the GGI on the gender gap in math becomes statistically non-significant as soon as a general measure of inequalities is introduced as a competing explanatory variable in regression models. We also apply three standard machine learning techniques to select among all explanatory variables those that are providing the "best statistical model". We systematically find that the models with the highest explanatory power include several general inequality-related variables whereas non-inequality-related and gender stratification variables are often dropped from those models (see all details in SM).

There is also some evidence that gender stratification and inequalities have a cumulative effect: the girls-to-boys ratio among high performers in math decreases strongly from .77 among OECD countries that are below the sample averages in terms of both income inequalities (measured with the GINI index) and gender stratification (imperfectly measured as minus the GGI index), to about .71 when only one of those indexes is above average, to .63 when they are both above average (albeit, with only the difference between the first and last groups being statistically significant, $p=0.005$).

We finally analyze the link between societal inequalities and gender performance gaps in science and reading (Tables S2, S4, S5 and S8). We find that societal inequalities are associated with lower girls-to-boys ratios in science (increasing the gender gap, as in math) and higher boys-to-girls ratio in reading (reducing the gender gap). This means that, consistent with our hypothesis, inequalities are detrimental to the performance of girls relative to boys in the three topics math, reading and science.

To better understand the origin of the relationship between non-gender-related inequality measures and gender performance gaps, we offer three different strategies: (i) individual-level regressions, (ii) panel analysis and (iii) instrumental variables.

We first replicate our analysis at the student level, with individual controls for grade repetition, parents' education, and households' economic and cultural resources, as well as the same controls interacted with students' gender. In those analyses, all estimates and their standard errors are corrected for measurement error in individuals' ability, and for sampling error in each country (see SM). We still find that inequalities are more detrimental to girls than to boys (Table 2, column 2, Table S10), suggesting that the main explanation behind our results is not unobserved individual heterogeneity. We can reject that the results are driven by a differential effect of parents' education on their daughters' and sons' math performance, or more generally by a differential allocation of household resources across children's gender (which would for example be detrimental to girls, especially when inequalities are large and resources are scarce).

Even if individual-level regressions allow us to better control for individual heterogeneity, it could still be the case that some unobserved country characteristics are driving the results. We control for countries' time-constant unobserved heterogeneity using country fixed-effects models estimated on an unbalanced panel dataset that includes the 2003, 2006, 2009, 2012, and 2015 PISA surveys. The identification of those fixed-effects models rely on the joint evolution over time of countries' gender performance gaps and general inequality measures. Time variations in our main inequality indicators are almost all significantly related to time variations in the girls-to-boys ratio in math, showing that countries that reduce relatively more (or increase relatively less) socioeconomic inequalities also reduce relatively more the gender performance gap in math (Table 2, column 3, Fig. S1). Except for the GINI index, the estimated effects of societal inequalities obtained from the fixed-effect models are however smaller and statistically less significant than those obtained in cross-section. All results are further confirmed by the estimation of random-effect models that include as regressors both the country-specific average level of each inequality measure over the period and contemporaneous deviations from this average (see Table S11).

To assess further a possible causal link, we finally exploit institutional differences between countries' labor markets as instruments for their extent of income inequality. We argue that such institutional differences of labor markets relate to political choices and are not likely to have a direct effect on the gender performance gap at schooling age (exclusion restriction). We use as instruments indicators such as bargaining coverage or union density, which are unlikely to have a direct impact on gender performance gaps at school, and we systematically find that variations in the Gini index that can solely be explained by those variables affect to the same extent the relative performance of girls in math (Table S12). This result suggests that inequalities driven by institutional factors affect the gender gap in performance at school.

To look at the role of institutions more closely, we study if institutional features of education systems that are known to impact social inequalities at school also affect gender performance gaps. We consider measures of inequalities in learning opportunities across schools or across students' socioeconomic background, measures of vertical stratification at school, such as the extent of grade repetition, and measures of the quality of education. All those measures are known to impact socioeconomic inequalities at school (*25-27*). We show that all those variables are also directly related to gender performance gaps across countries (Tables S2, S4-S7).

It is striking that a large variety of general indicators of inequalities can explain so well the general patterns of gender differences in math, science and reading performance across countries (while other indicators directly related to gender stratification have limited explanatory power). Countries that are generally speaking more egalitarian tend to reduce altogether several forms of inequality, as well as the gender gap in math at 15 years old. In more egalitarian countries, differences in initial status seem less likely to translate into differences in performance (in math, reading or science) and girls are more represented among high performers as are, for example, students from a low socio-economic and cultural background. This suggests that the gender gap in math is a form of social inequality like many others.

This interpretation is consistent with the fact that gender performance gaps at school are linked to countries' institutions that more generally reduce social and economic inequalities. Those institutions may also enhance girls' performance at school. As a consequence, gender equality may not only be a matter of gender norms and stereotypes. More general policies in favor of more inclusive, less vertically stratified, and more standardized education systems may for example have a positive impact on both social inequalities and gender inequalities in performance in math.

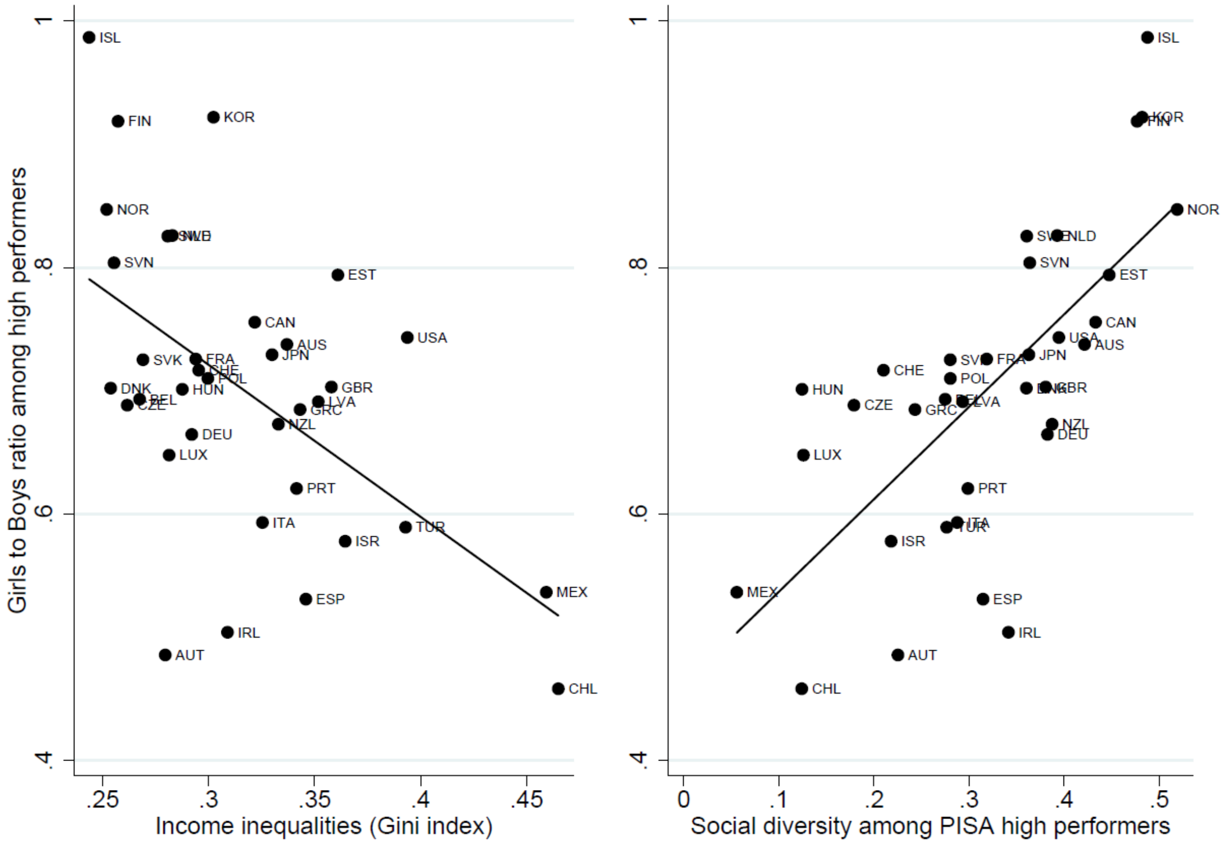
Subheads.

References and Notes:

1. OECD, PISA 2015 Results (Volume I): Excellence and Equity in Education, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264266490-en (2016)
2. I.V.S. Mullis, M.O. Martin, P. Foy, M. Hooper, TIMSS 2015 International Results in Mathematics (2016).
3. J.S. Hyde, J.E. Mertz, *Proceedings of the National Academy of Science* USA **106**, 8801 (2009)
4. C. Brown, M. Corcoran, *Journal of Labor Economics*, **15**, 431 (1997)
5. C.J. Weinberger, *Industrial Relations*, **38**, 407 (1999)
6. J. Hunt, J.P. Garant, H. Herman, D.J. Munroe, *Research Policy*, **42**, 831 (2013)
7. R.B. Adams, B.M. Barber, T. Odean WP, available at SSRN, https://ssrn.com/abstract=2933241 (March 14, 2017).
8. H. Ellis, *Man and woman: A Study of Secondary and Tertiary Sexual Characters* (Heinemann, 5. London, 1894)
9. S.J. Ceci, W.M. Williams, *Proc. Nat. Acad. Science*. USA, **108**, 3157 (2011)
10. C.P. Benbow, J.C. Stanley, *Science* **222**, 1029 (1983)
11. S. Machin, T. Pekkarinen, *Science*, **322**, 1331 (2008)
12. J.S. Eccles, T. Adler, J.L. Meece, *Journal of Personality and Social Psychology*, **46**, 26 (1984)
13. A. González de San Román, S. de la Rica, *Estudios de Economia Applicada* **24**, 79 (2016)
14. J. Currie, *American Economic Review*, **101**, 1 (2011)
15. N. Borghol, M. Suderman, W. McArdle, A. Racine, M. Hallett, M. Pembrey, C. Hertzmann, C. Power, M. Szyf, *International Journal of Epidemiology*, **41**, 62 (2012)
16. D.P. Baker, D. Perkins Jones, *Sociology of Education* **66**, 91 (1993)
17. A.M. Penner, *AJS* **114**, 138 (2008)
18. L. Guiso, F. Monte, P. Sapienza, L. Zingales, *Science* **320**, 1164 (2008)
19. N.M. Else-Quest, J.S. Hyde, M.C. Linn, *Psychol Bull*. **136**, 103 (2010)
20. G. Stoet, D.C. Geary, *Intelligence*, **48**, 137-151. (2015)
21. G. Stoet, D.C. Geary, *PLoS ONE* 8(3): e57988 doi:10.1371/journal.pone.0057988
22. J.M. Kane, J.E. Mertz, *Notices of the American Mathematical Society* **59**, 10 (2012)
23. R.G. Jr Fryer, S.D. Levitt, *American Economic Journal: Applied Economics* **2**, 210 (2010)
24. H. Ayalon, I. Litveh, *Social Science Research*, **42**, 432 (2013)
25. W.H. Schmidt, N.A. Burroughs, P. Zoido, R.T. Houang, *Educational Researcher*, **44**, 371 (2015)
26. H.G. Van de Werfhorst, J.J.B. Mijs, *Annu. Rev. Sociol*., **36**, 407 (2010)
27. G. Montt, *Sociology of Education*, **84**, 1 (2011)
28. OECD, PISA 2015 Technical Report. OECD Publishing, Paris. http://www.oecd.org/pisa/data/2015-technical-report/
29. J. M. Wooldridge, *Econometric analysis of cross section and panel data*. MIT press (2010)
30. J. Stock, M. Yogo Testing for Weak Instruments in Linear IV Regression. In: Andrews DWK Identification and Inference for Econometric Models. New York: Cambridge University Press ; pp. 80-108 (2005)

Acknowledgments: We are very grateful to J.-M. Marin for his tremendous help with machine-learning techniques; to F. Avvisati, F. Borgonovi, and F. Keslair at the OECD for their support and help regarding how to handle the PISA data; and to A. Aguilar, G. Piaton, and F. Riva for their suggestions on the manuscript.

Figure 1: Two measures of Inequalities and Girls underrepresentation among high performers in mathematics among OECD countries



Note: Country codes from ISO3166-1 standard. "Social diversity" is measured by the share of students from low socio economic background among PISA high performers.

Table 1: Recent evolutions of the gender gap in mathematics and reading among OECD countries

Year	Mean gap between girls and boys in math	Mean gap between girls and boys in reading	Girls-to-boys ratio among high performers in math	Boys-to-girls ratio among high performers in reading	Number of countries
2003	-0.11s.d.	0.37s.d.	0.7	0.58	30
2009	-0.12s.d.	0.43s.d.	0.7	0.50	30
2015	-0.09s.d.	0.28s.d.	0.71	0.69	30

Notes: Mean gaps in columns 2 and 3 are expressed as a fraction of the standard deviation (s.d.) of the score distribution. Statistics are established on the 30 OECD countries that participated in all three PISA surveys. Similar results are obtained in 2009 and 2015 on the 35 OECD countries that participated in the two surveys.

Table 2: Inequalities and gender performance gaps among high performers in math	(1)	(2)	(3)
Estimation Method:	**Country-level regressions on PISA 2015**	**Individual-level logistic models based on PISA 2015 with individual and country-level controls**	**Country fixed-effect models based on PISA 2003 to 2015**
Dependent variable:	**Girls-to-boys ratio among high math performers**	**Being a high performer in math**	**Girls-to-boys ratio among high math performers**
Explanatory variable (defined at country-level, interacted with students' gender in model (2)):			
a) Income inequalities			
GINI index (OECD)	-0.067***	-0.094***	-0.094***
GINI index in 1985	-0.065***	-0.083*	-
b) Socioeconomic and cultural inequalities			
Variance in the socioeconomic and cultural background of PISA students	-0.081***	-0.169***	-0.041*
c) Social inequalities in school performance			
Share of pupils from low socioeconomic background among high performers*	-0.085***	-0.157***	-0.027**
d) Inequalities in learning opportunities			
Between-school variation in students' socioeconomic background	-0.075***	-0.154***	-0.072

*Note: The Table reports the partial effect of four inequality variables (standardized) on the relative probability of girls and boys to be high math performers in OECD countries. The exact definition of inequality variables is given in the SM. Estimates in model (1) are obtained from linear country-level regression models on 35 countries in PISA 2015. Estimates in model (3) are obtained from linear regression models with country fixed-effects on 35 countries in PISA 2003, 2006, 2009, 2012 and 2015. In models (1) and (3), the performance of girls relative to boys in a given country is measured based on the ratio of girls and boys among high math performers. Model (2) estimates from individual-level logistic regressions the differential effect of inequalities on girls relative to boys. It includes country fixed-effects and individual-level controls for gender, parents' education, grade repetition, household wealth, and household economic, cultural, educational and Information and Communication Technology resources. Controls for the interaction between students' gender and other individual characteristics as well as countries' log GDP per capita are also included. Standard errors in model (2) account for measurement error in individual math proficiency, and country-level sampling error. *** p<0.01, ** p<0.05, * p<0.1. See SM for methodological details, sample sizes, standard errors, and R-squares. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.*

Supplementary Materials for

General Societal Inequalities Linked to Larger Gender Gaps in Math among High Performers

Thomas Breda, Elyès Jouini, Clotilde Napp

Correspondence to: thomas.breda@ens.fr or clotilde.napp@dauphine.fr

This PDF file includes:

Materials and Methods
Supplementary Text
Fig. S1
Tables S1 to S18

Other Supplementary Materials for this manuscript includes the following:

Databases and codes allowing the replication of the results have been deposited on a third-party server and can be accessed at http://doi.org/10.3886/E101800V1.

Table of contents

Materials and Methods

PISA data and gender performance gaps

The Programme for International Student Assessment (PISA) is an every-three-year international survey of 15-year-old students aimed at determining their knowledge and skills in different domains. Students' abilities are assessed in the three curricular domains: mathematics, reading, and science. Students also answer a background questionnaire, seeking information about the students themselves, their homes, and their school and learning experiences. School principals also complete a questionnaire that covers the school system and the learning environment.

The assessment does not just ascertain whether students can reproduce knowledge; it also examines how well students can extrapolate from what they have learned and can apply that knowledge in unfamiliar settings, both in and outside of school.

The PISA target population is made up of all students in any educational institution between the ages of 15 years and 3 months and 16 years and 2 months at the time of the assessment. This specific age has been chosen because it is close to the end of compulsory education in most countries. Efforts have been made to insure the absence of cultural or national biases in the test items and in the evaluation of performance.

We analyse data from the five surveys PISA 2003, PISA 2006, PISA 2009, PISA 2012 and PISA 2015.

The student data set in 2015 contains 540,000 observations, which roughly represent a population of 29 million 15-year-olds attending seventh grade or above in 72 countries, 35 of which belong to the Organisation for Economic Co-operation and Development (OECD).

PISA scores have been scaled during the first PISA survey in 2000 to have a mean of 500 and a standard deviation of 100, when students from all OECD countries are included. PISA scales are divided into proficiency levels from Level 1 (lowest) to Level 6 (highest). We focus primarily on what PISA refers to as "top performers", i.e. on students who perform at Level 5 and above, which corresponds to scores above 607 in mathematics, above 626 in reading and above 633 in science. In 2015, across OECD countries, 10.7% of students are top (or high) performers in math, 8.3% of students are top (or high) performers in reading, and 7.7% are top (or high) performers in science. Our main measure of gender performance gap in math is the ratio between the number of girls that score at or above Level 5 and the number of boys that score at or above Level 5. We consider the similar ratio in science, and the reverse ratio in reading (number of boys over number of girls) as girls outperform boys in that subject. Table S1 provides this ratio in math, reading and science in all OECD countries.

To show that this specific focus on the top decile of the distribution does not drive our results, we also consider as alternative measures girls-to-boys ratios among students performing at or above Level 4, or at Level 6. Among OECD countries in 2015, the share of students that score at Level 4 or above in math, reading and science is (respectively) 29.3%, 29% and 26.7%. The corresponding statistics for level 6 are 2.3% in maths, and 1.1% in reading and in science.

One advantage of those girls-to-boys ratios is to provide gender performance among students that are in the same ability range in all countries as they are beyond a given absolute threshold. Such measure can however become misleading or statistically imprecise due to sampling errors in countries where there are very few students scoring above the chosen threshold (Level 5 in most cases). We have chosen to keep potential outliers in all analyses based on OECD countries only (but mention them in the discussion of the supplementary tables, see below), but to exclude them in analyses that also include partner countries which are more numerous to have very few high performers (see below).

We also build alternative measures of gender performance gaps in the upper part of the score distribution by considering the ratio between a given percentile (90th or 75th) of the girls' and of the boys' score distributions. Such measures are not defined according to a level of performance that is common across countries. However, they do not generate outliers. We use them as robustness checks.

Finally we also consider "effect sizes" which are in each subject the difference between the mean score of boys and the mean score of girls divided by the standard deviation of all scores. These are standard measures of gender performance gaps that do not specifically target high performers.

<u>Measures of Inequality and other variables used in the study</u>

The paper relies on several country-level variables that we have retrieved from various data sources. We have classified those measures in eight groups: "Economic and socio-cultural inequalities", "Other non-educational inequality variables", "Inequalities in performance opportunities", "Inequalities in learning opportunities", "Quality of educative systems", "Measures of gender segmentation", "Countries' development" and "Institutional features of countries' labor markets". Details on variables in each group are given below. Table S12 provides a summary of all variables.

<u>*Economic and socio-cultural inequality*</u>

GINI index measures the extent to which the distribution of income among individuals or households within an economy deviates from a perfectly equal distribution where income is defined as household disposable income in a particular year. It consists of earnings, self-employment and capital income and public cash transfers; income taxes and social security contributions paid by households are deducted. A Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.
Source:
GINI index OECD: OECD (2017), Income inequality (indicator), doi: 10.1787/459aa7f1-en (accessed on April 14, 2017) http://stats.oecd.org/viewhtml.aspx?datasetcode=IDD&lang=en (data from 2013 or 2014 are matched with PISA 2015)
GINI index World Bank: http://data.worldbank.org/indicator/SI.POV.GINI 2015 data refer to the most recent year available
GINI coefficient 2005-2013: United Nations Development Program report 2015 (Data refer to the most recent year available during the period specified)
OECD Gini 85 http://dx.doi.org/10.1787/888932535185
Palma ratio is the share of all income received by the 10% people with highest disposable income divided by the share of all income received by the 40% people with the lowest disposable income.

Source:
Palma Ratio OECD https://stats.oecd.org/Index.aspx?DataSetCode=IDD
Palma ratio 2005-2013: United Nations Development Program report 2015 (Data refer to the most recent year available during the period specified).

Ratio of the average income of richest and poorest 20% of the population

Source:
Human Development Report 2007/2008, UNDP. HDRO calculations based on data from World Bank

Ratio of the average income of richest and poorest 20% of the population 2005-2013: United Nations Development Program report 2015 (Data refer to the most recent year available during the period specified).

Income share held by the poorest decile
Source: World Bank, http://data.worldbank.org/indicator/SI.DST.FRST.10, http://data.worldbank.org/indicator/SI.DST.FRST.20

Top income shares in the 1940s and 1950s
We retrieved from the World Wealth and Income Database website (http://wid.world/) the fiscal income share held by the top percentile of the income distribution. Fiscal income is defined as the sum of all income items reported on income tax returns, before any deduction. It includes labour income, capital income and mixed income. The concept of fiscal income varies with national tax legislations, which can introduce some noise in international comparisons. The population is comprised of individuals over age 20. The base unit is the individual (rather than the household). This is equivalent to assuming no sharing of resources within couples.
We focus on the top 1% income share because it is the indicator for which we have the most complete historical series. We managed to retrieve yearly top 1% income share data from 1945 onwards for 15 OECD countries (Argentina, Australia, Canada, Denmark, France, Germany, Japan, Netherlands, New Zealand, Norway, Singapore, Sweden, Switzerland, the United Kingdom, the United States) and one non-OECD country covered by the PISA survey (Malaysia).
Variance in the socioeconomic and cultural background of PISA students The PISA data set contains one indicator of social status, called Economic, Social and Cultural Status (ESCS). This captures parental education, parental occupation, and home possessions, as reported by the student. This index has been normalized to have mean zero and variance one in the OECD student population.

Other non-educational Inequality Variables

Poverty Rate (Relative income poverty) Share of the population with an income of less than 50% of the respective national median income (income after taxes and transfers, adjusted for differences in household sizes)
Source: OECD (2017), Poverty rate (indicator), doi: 10.1787/0fe1315d-en (accessed on April 14, 2017)
Infant mortality rate per 1,000 live births. Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation (UNICEF, WHO, World Bank, UN DESA Population Division) at childmortality.org. World Bank Data

Adolescent fertility rate Number of births per 1,000 women aged 15-19. There is a well-known link between poverty and high adolescent fertility rates.
Source: World Bank, http://data.worldbank.org/indicator/SP.ADO.TFRT

Index of human inequality Average inequality in three basic dimensions of human development: 1. Inequality in life expectancy: Inequality in distribution of expected length of life based on data from life tables estimated using the Atkinson inequality index. 2. Inequality in education: Inequality in distribution of years of schooling based on data from household surveys estimated using the Atkinson inequality index. 3. Inequality in income: Inequality in income distribution based on data from household surveys estimated using the Atkinson inequality index.

Source: United Nations Development Program Report. See *Technical note 2* at http://hdr.undp.org.

Intergenerational Earnings Elasticity Intergenerational economic mobility is measured as the elasticity between paternal earnings and a son's adult earnings, using data on a cohort of children born, roughly speaking, during the early to mid 1960s and measuring their adult outcomes in the mid to late 1990s. The higher the earnings elasticity, the lower the mobility. The estimates of the intergenerational earnings elasticity are derived from published studies, adjusted for methodological comparability in a way described in the appendix to Corak (2006), updated with a more recent literature review reported in Corak (2013).

Source: Corak, M. (2013) Income Inequality, Equality of Opportunity and Intergenerational Mobility, IZA DP 7520

Index of inequality of economic opportunity measures the part of income inequality due to predetermined circumstances beyond individual control like age, ethnic group, region of birth or of residence, parental background. IEO-L (Level) represents inequality of opportunity in absolute terms, and IEO-R (Ratio) represents inequality of opportunity as percentage of total inequality.

Source: Checchi, D., Peragine, V., and L. Serlenga (2013) Income inequality and opportunity inequality in Europe.

Education (im)mobility Parents child correlations in years of schooling, ages 20-69. Surveyed between 1994 and 2004

Source: Hertz, T., T, Jayasundera, P. Piraino, S. Selcuk, N. Smith and A. Verashchagina (2007) 'The inheritance of educational inequality: International Comparisons and Fifty-Year Trends', The B.E. Journal of Economic Analysis and Policy, Vol 7, Issue 2 (Advances), Article 10.

Inequalities in performance opportunities

Percentage of the variation in performance explained by PISA ESCS index It is measured as the percentage of the variation in performance explained by the PISA index of economic, social and cultural status (ESCS): r-squared*100. The strength of this relationship refers to how well socio-economic status predicts performance.

Share of students from low socio economic background among high performers Ratio of the percentage of low ESCS students (bottom quarter of the ESCS index) scoring at or above Level 5 to the percentage of all students scoring at or above Level 5. As we have kept this variable for panel analysis and it is not available all years, we have recomputed it directly from the individual-level data based on PISA definition for all PISA surveys after 2003.
Increased likelihood of students from low socio economic background to score low Increased likelihood of students from the bottom quarter of the PISA ESCS index to score below Level 2.
Percentage of resilient students A student is classified as resilient if he or she is in the bottom quarter of the PISA index of economic, social and cultural status (ESCS) in the country of assessment and performs internationally in the top quarter of students, after accounting for socio-economic status.
Percentage of students scoring below level 2 Level 2 is considered a baseline level of proficiency that all young adults should be expected to attain in order to take advantage of further learning opportunities and participate in the social economic and civic life.

Ratio of the mean score of low ESCS students/high ESCS students: Ratio of the mean score of students from the bottom quarter of ESCS index /mean score of students from the top quarter of ESCS index

Inequalities in learning opportunities

Between-school variation in students' socioeconomic background (OECD index of social inclusion). The *index of social inclusion* is calculated as 100*(1-rho), where rho stands for the intra-class correlation of socio-economic status, i.e., the variation in the PISA index of social, economic and cultural status of students (ESCS) between schools, divided by the sum of the variation in students' socio-economic status between schools and the variance in students' socio-economic status within schools.

Index of equity in resource allocation. Equity in resource allocation (PISA 2015) refers to the difference in the index of shortage of schools' educational resources (see Variability in the index of shortage of educational material) between socio-economically advantaged and disadvantaged schools. It assesses the extent to which the socioeconomic profile of a school is positively or negatively associated with the principal's concern about the lack or inadequacy of educational material or education staff at school. Positive values indicate that principals of disadvantaged schools reported less concern about the material resources at their schools than principals of advantaged schools. Higher values indicate a higher equity in resource allocation. Advantaged (disadvantaged) schools are those in the top (bottom) quarter of the distribution of the school-level PISA index of economic, social and cultural status (ESCS) within each country/economy.

Vertical stratification: variability in grade level Both within and between countries, students can be enrolled in different grades. In PISA, the distribution of 15-year-old students across grade levels is the main measure of vertical stratification.

Variability in the index of shortage of educational material PISA 2015 asked schools principals to report the extent to which their school's capacity to provide instruction was hindered (not at all, very little, to some extent or a lot) by a shortage or inadequacy of educational material such as textbooks, library materials, laboratory equipment, instructional material and computers. The responses were combined to create an *index of shortage of educational material*. The average of the index is zero and the standard deviation is one across OECD countries. Positive values reflect principals' perceptions that the shortage of educational material hinders the capacity to provide instruction to a greater extent than the OECD average; negative values indicate that school principals believe the shortage hinders the capacity to provide instruction to a lesser extent.

Variability in the time devoted to science lessons PISA 2015 asked students to report the average number of minutes per class period, the total number of class periods per week, and the number of class period for science, language of instruction and mathematics.

Quality of education systems

Mean years of schooling Average number of years of education received by people aged 25 and older, converted from education attainment levels using official durations of each level. Data refer to 2014 or the most recent year available. *Source* : Human development report
Staff provided by school to help students with homework For the first time, PISA 2015 asks schools principals if the school provides staff who can help students with homework.

Proportion of 15-years old enrolled at school Coverage (index 3): total population of 15 years old enrolled at grade 7 or above/total population of 15 years old. It therefore also reflects the proportion of 15-year-olds excluded or not at school.

Number of years at preprimary school Attendance at preprimary education (whether and how long students are enrolled in preprimary education) is another aspect of time resources invested in education.

Measures of Gender segmentation

Gender Gap Index (GGI) The Gender Gap Index, from the World Economic Forum, synthesizes the position of women in any given country by taking into account economic opportunities, economic participation, educational attainment, political achievements, and health and well-being. Larger values point to a better position of women in society.

Female economic participation and opportunity Subindex of the GGI.
Gender development index Ratio of female to male HDI values. See *Technical note 3* at http://hdr. undp.org for details on how the Gender Development Index is calculated.

Gender Inequality index A composite measure reflecting inequality in performance between women and men in three dimensions: reproductive health, empowerment and the labour market. See *Technical note 4* at http://hdr.undp.org/en for details on how the Gender Inequality Index is calculated.

Gender wage gap The gender wage gap is unadjusted and is defined as the difference between median earnings of men and women relative to median earnings of men. Data refer to full-time employees and to self-employed.

Countries' development

Human Development Index The Human Development Index (HDI) is a composite statistic of life expectancy, education, and per capita income indicators
Employment rate Total, % of working age population https://data.oecd.org/emp/employment-rate.htm

Life expectancy at birth UNDP 2015 Number of years a newborn infant could expect to live if prevailing patterns of age-specific mortality rates at the time of birth stay the same throughout the infant's life.

Health status OECD, http://stats.oecd.org/BrandedView.aspx?oecd_bv_id=health-data-en&doi=data-00540-en

Institutional features of countries' labor markets

We use annual series for the recent period obtained from OECD and ILO statistics. As most series have gaps and are often not available for recent years (2014 onwards), we have averaged them by country over the decade 2005-2014. Those averages are used as instruments in cross-sectional analyses of the gender gaps in PISA 2015.
Bargaining coverage 2005-2014. Downloaded on May 15th 2017 from ILOSTAT. Share of the workforce whose pay and/or conditions of employment are determined by one or more collective agreements. Averaged over the period 2005-2014.
Union density 2005-2014. Downloaded on May 15th 2017 from ILOSTAT. Share of the workforce that is member of a trade union. Averaged over the period 2005-2014.
Average tax wedge 2005-2014. Extracted on 14 Feb 2017 from OECD.Stat. It is the average tax wedge (including income tax and social security contributions and benefits) for a single person at 100% of average earnings with no child.
Employment Protection index 2005-2014. Extracted on 14 Feb 2017 from OECD.Stat. We use the second version of the OECD index measuring the strictness of employment protection. This is because the third and most recent version is not available before 2008. The index is the weighted sum of sub-indicators concerning the regulations for individual dismissals (weight of 5/7) and additional provisions for collective dismissals (2/7). It incorporates 12 detailed data items. It is averaged over the period 2005-2014.
Min wage (parity of purchasing power) 2005-2014. Extracted on 14 Feb 2017 from OECD.Stat for OECD countries.
Spending in Active Labor Market Policies 2005-2014. Extracted on 14 Feb 2017 from OECD.Stat for OECD countries. Expressed as a share of GDP.

Methods

Country-level estimators of gender gaps

Our country-level measures of gender gaps are based on the readily available aggregated statistics published by PISA. PISA reports (see (1) for 2015) provide the following statistics in math, reading and science in each country: the mean score of boys and the mean score of girls, percentiles of the girls and boys score distributions, the proportion of boys and the proportion of girls performing at or above Level 4, Level 5, Level 6, etc.
When a country-level statistics is not provided, we have computed it directly from the aggregated data. Gender gaps in mean performance, or effect sizes, are estimated by taking the difference between the mean score of boys and the mean score of girls, divided by the standard deviation of all scores. In math and science, the girls-to-boys ratios at or above Level 5 in each country are simply obtained by dividing the share of girls performing at or above Level 5 by the share of boys performing at or above Level 5. The inverse ratio is considered in reading. What we call the gender gap at a given percentile corresponds to the ratio of that percentile of the girls and boys scores distributions.

Details about PISA methodology can be found in PISA Technical reports (see (28) for 2015) and only the main idea is reported here. PISA adopts the Item Response Theory models, and does not provide for each student actual scores but plausible values. These plausible values (5 for PISA surveys until 2012, and 10 in 2015) are random numbers drawn from the distribution of scores that could be reasonably assigned to each individual, given his or her answers - that is, the marginal posterior distribution. Any estimation procedure in PISA (for instance mean score of boys) involves the calculation of the required statistic for each plausible value (appropriately weighting with the reported student weights) and the final estimate is the arithmetic average of the five or ten estimates obtained. Standard errors are calculated with a replication method that takes into account the stratified two-stage sample design for selection of schools and of students within schools.

Note that instead of estimating the ratios of girls-to-boys scoring at or above a given level by taking the ratios of the PISA estimated proportions of girls-to-boys at or above this level, we could estimate it by starting from the individual data provided by PISA, computing the ratio for each plausible value provided for each student (appropriately weighting with the reported student weights) and averaging the values of the ratio for each plausible value. None of these two estimators is theoretically better than the other (both are consistent). We prefer to adopt the former estimator, because the denominator of our quantity represents the probability of scoring above a given level, which might be close to zero in some countries, and starting by averaging this quantity is likely to bring more stability. We have also done the computations with the latter estimator and the results for OECD countries are essentially unchanged.

Statistical analyses of the relationship between inequalities and gender gaps

Most of our analyses are based on the country-level estimators of gender gaps described above and correlate those estimators with various country-level measures of inequality. We use linear regression models with one or a few independent variables that are solved by ordinary least squares. We also use standard fixed-effects and random-effects estimators in empirical specifications based on panel data. Finally, we estimate a few instrumental-variable regressions using a two-stage least squares estimator.

We also provide estimates from individual-level regressions on a sample of about 250,000 students in 35 OECD countries in 2015. This allows us to control for unobserved individual-level heterogeneity and to correct estimates and standard errors for measurement errors in our country-level measures of gender performance gaps (see below). The closest micro-level counterpart to our cross-country regressions of the girls-to-boys ratio among high performers is a logistic regression of the type:

$$\ln\left(\frac{P(Top_{ij}=1)}{1-P(Top_{ij}=1)}\right)=\alpha_j+\beta Female_i+\gamma\,(Female_i*Inequality_j)+\delta X_i+\mu(Female_i*X'_{ij})+\varepsilon_{ij}$$

where Top_{ij} is a dummy variable equal to 1 if the score of student i in country j is above level 5, and to 0 otherwise, α_j are country fixed effects, $Female_i$ a dummy variable for female students, $Inequality_j$ a country-level measure of societal inequalities usually standardized to have a mean of 0 and a standard deviation of 1 over the considered sample of countries, X_i are student-level controls, typically including parental education, the fact to have repeated a grade, and measures of household wealth and economic, cultural, educational or Information and Communication Technology resources, X'_{ij} corresponds to the individual-level controls as well as some country-level controls such as the log of GDP per capita.
The coefficient of interest is γ, which measures the extent to which inequalities affect the likelihood for girls relative to boys to be a high performer. More precisely, β estimates, in log points, the difference between girls and boys in the odds of being a high performer, while γ is an estimate of how this difference varies when cross-country inequalities vary by one standard deviation. However, γ is not directly comparable to the association between inequalities and the girls-to-boys ratio estimated at the country level. First, the odds correspond to the probability of being a high performer (about 12% for boys and 9% for girls) divided by the probability not to be one (about 88% for boys and 91% for girls). The ratio of those latter probabilities for girls and boys is a first small corrective term. The second, more important, difference comes from the fact that γ provides relative variations. The girls-to-boys ratio is typically equal to 0.7. A one standard deviation increase in the Gini index typically decreases it by about 0.07, which corresponds to 0.1 log-points. This is why estimates from micro-level regressions tend to be about 40% larger.

Statistical inference

The statistical inference for the country-level regressions does not correct for measurement error in gender performance gaps and inequality variables. This choice is driven by the fact that our estimates are usually very significant by statistical standards, and that measurement error on the left hand side of our regression models is likely to inflate the standard errors of the estimates rather than reducing them (29). This implies that not doing those corrections in our baseline analyses likely plays against the statistical significance of our results. Another reason for not doing those corrections systematically is to keep the baseline analysis as simple as possible, so that it can be easily replicated.

To confirm nevertheless that measurement error is not an issue, we have run again our main analyses at the individual level, and corrected standard errors for measurement error in the corresponding individual-level regressions (see Table S9 and its discussion). Measurement error can arise for two reasons. First, as explained above there is some uncertainty on the ability measure of each student and PISA provides ten plausible values drawn from a posterior distribution of ability. Second, there is standard sampling error at country-level as performance gaps are not established over the universe of students in a given country. To deal with sampling error, PISA provides 80 alternative sets of individual weights and detailed guideline to use those weights. The computation of corrected standard errors basically relies on bootstrap techniques: one needs to run the regression of interest for each plausible value, weighting it first by the "true" set of individual weights and then by the 80 alternative sets of weights. The correct point estimates is the average of the 10 regressions ran with the "true" set of weights, while the standard errors is computed according to a formula that sums both of the measurement errors described above. See (28) for all details regarding those bootstrap techniques. Note finally that we systematically normalize weights in individual-level regressions so that the sum of individual weights in each country is identical. The use of those "senate" weights makes sure that all countries have the same weight in the analysis instead of contributing according to their total population.

Supplementary Text

This section discusses in more detail the evidence presented in the supplementary tables.

The gender gaps among high performers

Table S1 shows the girls-to-boys ratio at or above Level 5 in each country in math (columns 1 and 4) and science (column 3) and the inverse of this ratio in reading (column 2). The underrepresentation of girls in math at high levels of performance is a common feature of all OECD countries. The most "gender-equal" country in math at high levels of performance is Iceland, with a girls-to boys ratio of 0.99.

As explained on p.9 and recalled in the legend of Table S1, these country-level estimates can be computed in two different ways. By comparing columns 1 and 4, we can see that the two approaches provide very close estimates. We prefer to adopt the first estimator because the denominator of our quantity represents the probability of scoring above a given level, which might be close to zero in some countries, and starting by averaging this quantity is likely to bring more stability.

The relationship between inequalities and the gender performance gap in math: Additional measures of Inequality

The first column of Table 2 in the text shows the relationship between inequalities and the gender gap in math for some inequality measures. Table S2a extends the country-level linear regression analysis to additional inequality measures and illustrates the robustness of the negative relation. We consider different categories of inequalities: income inequalities (like GINI index), other non-educational inequalities (like the index of human inequality), inequalities in performance opportunities (like the share of students from low socioeconomic background among high performers) and inequalities in learning opportunities (like the between-school variation in students' socioeconomic background). A one standard deviation increase in inequalities is typically associated with a decrease in the girls-to-boys ratio (among students scoring at or above level 5 in PISA) of .05 to .08.

The relationship between inequalities and the gender performance gap in reading

Table S2b provides results about the link between inequalities and the gender gap in reading, measured by the boys-to-girls ratio among high performers, i.e., at Level 5 or above. Estimates are obtained for linear country-level regression models on 35 OECD countries in 2015. We consider the same inequality measures as in Table S2a. The relation is inverse to the relation in math: inequalities are always detrimental to girls' relative performance. They increase the gender gap in math but reduce the gender gap in reading.

The link between inequalities and the gender performance gap in reading seems slightly less robust than the link between inequalities and the gender performance gap in math, but this slight lack of robustness is mainly driven by Turkey, a country that has very few high performers and therefore a boys-to-girls ratio which is hard to interpret. When Turkey is removed, the magnitude and statistical significance of the coefficients tend to increase (e.g., Variance in the socioeconomic and cultural background of PISA students, Income share held by the poorest decile, or the Poverty rate become statistically significant at the 5% level).

The relationship between inequalities and the gender performance gap in science

Table S2c provides results about the link between inequalities and the girls-to-boys ratios among high performers, i.e., at Level 5 or above, in science. Estimates are obtained for linear country-level regression models on 35 OECD countries in 2015. We consider the same inequality measures as in Tables S2a and S2b. The relationship between inequalities and the gender gap in science is strong and very similar to the relationship between inequalities and the gender gap in math.

Controls

Table S3 shows that controlling for GDP and gender stratification (the GGI index) does not change the results about the relationship between the gender gap in math and our main measures of inequalities.

The relationship between inequalities and the gender performance gaps in math, reading and science on a larger set of countries

Table S4 shows that the relationship between inequalities and gender gaps in math, reading and science is robust to the inclusion of countries that participated in PISA but are not part of the OECD. In math and science, the relationship between inequalities and the gender performance gap on the larger set of countries is roughly of the same magnitude than on OECD countries. It is however quantitatively weaker in reading.

In this table, estimates are obtained from linear country-level regression models. We have excluded in each subject all countries that have too few high performers (less than 0.5% of girls or boys scoring above level 5, which corresponds to about 10 countries in each subject). This is because the small number of high performers in those countries (less than 50) is likely to imply sampling error and imprecise measures of the girls-to-boys ratio. An alternative option would have been to focus on a measure of the gender performance gap that is less subject to sampling errors when those partner countries are included. When we do so, we do not have to remove outliers and still find strong associations between measures of inequalities and gender performance gaps, as shown in Table S6 (see below).

The relationship between inequalities and the gender gaps in average performance (effect sizes) in math, reading and science

Table S5 shows that the detrimental impact of inequalities on the relative performance of girls in math (Table S5a), reading (Table S5b) and science (Table S5c) is valid not only at high levels of performance but also for average levels of performance. Estimates are obtained from linear country-level regression models. As usual in the literature, we consider effect sizes, i.e., the difference between the boys mean score and the girls mean score divided by the standard deviation of all scores. We retrieve the same qualitative patterns as those exhibited for the girls-to-boys ratios at or above Level 5.

In reading, the relationship between inequalities and effect sizes is roughly of the same strength (in terms of p-value) than the relationship between inequalities and the girls-to-boys ratio among high performers. It is however quantitatively weaker in math and in science. This is especially the case for measures of income inequalities. In math for example, the p-values of the five reported estimates range between .08 and .15. This is consistent with (21), that finds no consistent correlation between the Gini index and the gender gap in average performance in math. All our other measures of inequalities except one (the Index of equity in resource allocation) are more significantly related to effect sizes. In contrast, the GGI, for which we also report an estimated effect, has a higher p-value than almost all our inequality measures (p=.18). Finally, we also tried to include OECD partner countries in the analysis and observed that the relationships between income inequalities and effect sizes in math are maintained, while the relationship between the GGI and effect sizes in math completely disappears.

Using alternative measures of gender performance gaps among high performers in math

Table S6 shows the validity of the negative relationship between gender gaps in math and inequalities for alternative measures of the gender gap.

We consider four alternative measures: a) the girls-to-boys ratio at or above Level 4, b) the girls-to-boys ratio at Level 6, c) the ratio of the 75th percentile of the girls' score distribution to the same percentile of the boys' score distribution and d) the ratio of the 90th percentile of the girls' score distribution to the same percentile of the boys' score distribution. We perform OLS country-level regressions.

For the four alternative measures, we retrieve the same qualitative patterns as those exhibited for the girls-to-boys ratio at or above Level 5. Quantitatively, the relationships with inequalities are a bit weaker when gender gaps are measured using ratios of the 75th or 90th percentiles of girls and boys score distributions than when we compute girls-to-boys ratio at or above Level 4 or at Level 6. They are the strongest when we focus on the girls-to-boys ratio at or above Level 4. Some income inequality measures such as the Gini index from the World Bank or the Palma ratio can explain more than 50% of the variation in the girls-to-boys ratio at or above Level 4.

Overall, the relationship between inequalities and gender performance gaps in math tend to be weaker at average level of performance (effect size, see Table S5) than at higher level of performance (ratios of the 75th or 90th percentiles), suggesting that inequalities are more strongly associated to gender performance gaps in the upper part of the score distribution.

Also note that some items, like the percentage of resilient students, are directly related, in their definition, to the different levels of performance in PISA and have a higher explanatory power with measures also related to the representation of girls and boys at those levels (measures a and b), while some items like the percentage of the variation in performance explained by PISA ESCS index have a higher explanatory power for measures related to the ratios of percentiles (measures c and d).

The relationship between inequalities and the gender performance gaps at different points in time

Table S7 presents OLS regressions showing the robustness across time of the negative relation between inequalities and gender gaps in math. The negative relation is found on the 5 successive PISA surveys from 2003 to 2015. Note that all the measures of inequalities used for 2015 are not always available for the previous surveys. We have therefore added analogous measures of inequalities, for the 4 different categories of inequalities (economic inequalities, other non-educational inequalities, inequalities in performance and inequalities in learning opportunities):

Difference in the index of the quality of infrastructures between socio-economically advantaged and disadvantaged schools The PISA 2009 and PISA 2012 index of the quality of a school's infrastructures was based on the school principals' reports on the extent to which 15-year-olds were hindered in their learning by: poor condition of building; poor heating, cooling and/or lighting systems; and lack of instructional space (e.g., classrooms). Advantaged (disadvantaged) schools are those in the top (bottom) quarter of the distribution of the school-level PISA index of economic, social and cultural status (ESCS) within each country/economy.

Difference in the index of the quality of educational resources between socio-economically advantaged and disadvantaged schools The PISA 2009 and PISA 2012 index of the quality of a school's educational resources was based on the school principals' reports on the extent to which 15-year-olds were hindered in their learning by a lack of resources. Equity in educational resources refers to the difference in the index of the quality of educational resources between socio-economically advantaged and disadvantaged schools. Advantaged (disadvantaged) schools are those in the top (bottom) quarter of the distribution of the school-level PISA index of economic, social and cultural status (ESCS) within each country/economy.

Percentage of PISA students who have repeated a grade Percentage of students who reported that they had repeated a grade at least once over the course of compulsory schooling.

Variability in the index of shortage of educational staff The lack or quality of the human resources in schools is measured by asking principals if the lack or quality of teaching and assisting staff hinders the capacity to provide instruction in the school. Principals' responses were combined to create an index of shortage of education staff. The average on the index is zero and the standard deviation is one across the OECD countries. Positive values reflect principals' perceptions that a shortage of education staff hinders the capacity to provide instruction to a greater extent than the OECD average; negative values indicate that school principals believe a shortage hinders the capacity to provide instruction to a lesser extent.

Variability in school principal's leadership PISA 2009 asked principals to report on their level of involvement in and leadership of several issues, including making sure that teachers' work and development reflects the educational goals of the school, monitoring student performance and classroom activities, and working with teachers to resolve problems. An *index of school principal's leadership* combines their answers to evaluate whether or not principals are active in improving teachers practices and the working environment within the school. This index has a mean of zero and a standard deviation of one for the OECD countries.

Competition between schools -percentage of schools that likely transfer students for behavioral or academic reasons Percentage of students in schools where the principal reported that a student in national modal grade would be "very likely" transferred to another school because of one of the following reasons: low academic performance, behavioral problems or special learning needs

Variability in the index of extracurricular activities at school Schools principals were asked whether the school offers various extracurricular activities to students in the modal grade for 15-year-olds. Some of the principals responses were combined to create an *index of extracurricular activities*.

Variability in the student/teacher ratio PISA 2015 asked the total number of teachers and students in their schools from which the students teacher ratio was computed. PISA 2015 also asked the total number of teachers and students in their schools from which the student-teacher ratio in the school was computed.

Difference in school activities to promote sciences between socio-economically advantaged and disadvantaged schools Advantaged (disadvantaged) schools are those in the top (bottom) quarter of the distribution of the school-level PISA index of economic, social and cultural status (ESCS) within each country/economy.

<u>The relationship between inequalities and the performance of girls and boys separately</u>

We investigate the relationship between inequalities and the performance of each gender separately in Table S8. Estimates are obtained from linear country-level regression models on 35 OECD countries. We find that inequalities are strongly detrimental to both girls and boys, in both math and reading. However, the associations are systematically stronger for girls than for boys, both in math (columns 3 and 4) and reading (columns 5 and 6). The difference between the partial effect of each inequality variable on the log share of girls and the log share of boys among high performers is by construction its partial effect on the log of what we call the girls-to-boys ratio among high performers as this ratio is obtained by dividing the two former variables. Hence, estimates in column (1) of Table S8 are simply the difference between estimates in column (3) and (4), and estimates in column (2) of Table S8 are the difference between estimates in column (5) and (6). Most estimates in columns (1) and (2) are statistically significant, showing that the stronger association between inequalities and performance for girls than for boys is significant from a statistical point of view.

<u>Horse races between gender-related and non-gender-related inequality measures as predictors of the gender performance gap in math high achievers</u>

The first row of Table S9 provides from separate univariate linear regressions the partial effect on the girls-to-boys ratio among high achievers of six possible indicators of gender inequalities: the gender gap index, the gender gap in economic participation and opportunity, the female labor force participation, the gender development index, the gender inequality index, and the gender wage gap. Those variables have been systematically normalized to have a standard deviation of 1, and to correspond to *more* inequalities when they increase. Except the gender wage gap, those variables tend to have significant effect on gender performance gaps, as already shown in the literature. Those effects are however always quantitatively smaller than those of non-gender inequality measures (provided in the first column of Table S9). More importantly, when a gender-related and a non-gender-related inequality measure are jointly included as possible predictors of the gender gap among high achievers, the former almost always becomes statistically non-significant. For the six gender-related measures tested against our four main non-gender-related inequality measures, only one of the 24 tests leads to a significant estimate: this is for the gender wage gap tested against the Gini index. However, the estimated effect of the gender wage gap in that case does not even have the expected sign. The second striking fact is that our four main societal inequality measures remain always statistically significant when tested against any of the six gender-related measures (Table S9).

Focussing on the GGI in particular, we also find that its effect on gender performance gaps is not robust to controlling for GDP in 2015. It is also not robust across time; in 2009 for instance, there is no significant relation, neither among OECD countries (Table S7) nor among all countries that participated in PISA. Finally it does not hold in science nor in reading, unlike the relation between general inequalities and the gender gap in performance (Tables S2b and S2c, last row).

Model selection

A model with three well-chosen explanatory variables, one for each category of inequalities (income inequalities, inequalities of performance opportunities, and inequalities of learning opportunities), can explain more than 60% of the variation of the gender math ratios across countries, with all three coefficients being significant. For example, a linear country-level regression model that includes the GINI index, the share of students from low socioeconomic background among high performers, the variability in grade level, and a constant term has an R-squared of 65%.

To determine the "best model(s)" to explain the variance in the gender gap in math performance across OECD countries and to make sure that inequalities are stronger than other possible explanatory variables, we apply procedures of model selection to a set of variables, including measures of inequality but also other candidates to explain the gender gap, like measures of gender equality, and measures of development. We restrict our attention to variables for which we have observations for all OECD countries.

More precisely, we include in our analysis 6 measures of gender equality (gender gap index, gender gap in economic participation and opportunity, female labor force participation, gender development index, gender inequality index, gender wage gap), 3 measures of socioeconomic inequality (GINI index, variance in the socio economic and cultural background of PISA students, difference between the 95th and the 5th percentile of PISA ESCS index), 6 measures of inequality of opportunities of performance (share of students from low socio economic background among high performers in math, in reading, percentage of the variation in performance explained by PISA ESCS index, percentage of resilient students, ratio of the mean score of students from low socio economic background by the mean score of students from high socio economic background, percentage of students scoring below level 2), 4 measures of inequality of opportunities of learning (rate of grade repetition, variability in grade level of 15 years old, index of equity in resource allocation, variability in the index of shortage of education material) and 7 measures of development (human development index, GDP, average level of PISA ESCS index, health status, employment rates, life expectation, mean years of schooling). We consider three different procedures of model selection: minimization of Bayesian information criterion (BIC), maximization of the adjusted R-squared, and the LASSO procedure (Least Absolute Shrinkage and Selection Operator) that forces the sum of the absolute value of the regression coefficients to be less than a given constant.

The main conclusion is that the three different strategies of model selection always extract the following explanatory variables among the 26 possible variables: GINI index, share of students from low socio economic background among high performers, variability in grade level of 15 years old and variability in the index of shortage of education material. These 4 variables are general inequality variables. This result confirms the importance of inequalities in explaining the gender gap in math performance; in particular, inequalities have a much higher explanatory power than measures of gender equality or general measures of country development.

Let us detail the results of the three procedures. Concerning the BIC procedure, the strongest explanatory variables are the GINI index and the share of students from low socioeconomic background among high performers. The best BIC models keep as additional explanatory variables the variability in grade level and variability in shortage of education material, as well as the gender equality index.

Concerning the R-squared procedure, the strongest explanatory variables are the same variables as in the BIC procedure, i.e., the GINI index and the share of students from low socio economic background among high performers. The best adjusted R-squared model keeps the GINI index and the share of students from low socioeconomic background among high performers, then the variability in grade level, then the gender wage gap, then the variability in the index of shortage of educational material, then life expectation, the percentage of students scoring below level 2 and the rate of grade repetition. The problem with this selection procedure is that due to a high number of possible explanatory variables with respect to the number of observations, several models have an adjusted R-squared between 0.72 and 0.73 and cannot be distinguished.

Concerning the LASSO procedure, the best model extracts the following variables: the share of students from low socio economic background among high performers, the GINI index, the variability in grade level, the variability in shortage of education material, then the difference between the 95th and the 5th percentile of PISA ESCS index.

Results based on historical measures of inequalities

We find that using the GINI index in the early 1980s instead of the contemporaneous Gini has virtually no effect on our estimates (Table 2, Tables S2). Trying to exploit the newly available World Wealth and Income Database, we even find a statistically significant negative relationship between the share of total fiscal income held by the top 1% income earners in the 1950 or 1960 and the gender gap in math (albeit only for the subset of 16 countries for which top income shares at that time are available).

Those additional results based on historical measures of inequalities suggest that we are unlikely to capture reverse causality—an effect of the gender gaps at school on inequalities—nor the effect of contemporaneous confounding factors. They however rely on some methodological choices that we now explain.

To avoid to lose too much information due to some specific years missing in the data, we have averaged the top income shares for each country by decade from 1945 onwards (1945-1954, 1955-1964, etc.). We also kept all available countries including Malaysia and found that a one standard deviation in the top income share in the first two decades following WW2 was associated with a drop in the girls-to-boys ratio in math of about 0.085 ($p<0.01$ for both decades). For more recent decades, the relationship still has the expected sign but often found of a smaller magnitude, and not always statistically significant. For the first two decades following WW2, the relationship also becomes non-significant if we exclude Malaysia and only focus on OECD countries. In total, the evidence of a relationship between income inequality sixty years ago and the gender performance gap today remains suggestive and will need to be completed when data for more countries will be made available. Such evidence together with the results obtained with the income GINI index in the 1980s for a larger set of countries, suggest that the association between inequalities in a society and gender performance gaps in math is historically rooted.

Results based on student-level regressions

Results based on students' level regressions are presented in Table S10. The analysis of Table 2 (column 2) in the main text is extended to a larger set of measures of inequality, and sample sizes, R-squared and p-values are systematically provided for all estimates. Estimates are obtained from individual-level logistic regressions in which estimates' standard errors are corrected for measurement error in gender performance gaps (see p.10 for details about the method). Country fixed-effects and individual-level controls for gender, parents' education, grade repetition, household wealth, and household economic, cultural, educational and Information and Communication Technology resources are included. Controls for the interaction between students' gender and other individual characteristics as well as countries' log GDP per capita are also included.

We still find that inequalities are more detrimental to girls than to boys. For all the considered variables, the negative relation between inequalities and the relative performance of girls in math is very strong. We actually also ran the individual-level regressions without any control variables and obtained estimates and standard errors (not presented) that were very close. This allows us to make two points. First, the main explanation behind the macro-level results is not unobserved heterogeneity in economic or cultural resources available to students, nor a differential effect of those resources on girls and boys. Second, correcting standard errors for measurement error in gender performance gaps has little impact on statistical significance of the results established at the macro level.

Note finally that estimates from micro-level regressions tend to be larger than those from country-level regressions (Table S2a). These two estimates are not directly comparable. As explained in detail on p.10, the estimates from micro-level regressions are related to relative variations in the difference between girls and boys in the odds ratio of being a top performer, and these two distinctive features (odds ratios versus standard probabilities, and relative versus absolute variation) can explain those larger estimates.

Results based on panel data estimators

Main results:
Results based on specifications including country fixed-effects are presented in Table 2. They are established for our main measures of inequalities: one measure of income inequalities (the

GINI index), one general measure of social inequalities (the variance in the socioeconomic and cultural background of PISA students), one measure of inequalities in school performance (the share of pupils from low socioeconomic background among high performers) and one measure of inequalities in learning opportunities (between-school variation in students' socioeconomic background).

For those variables, we also present estimates from random-effects models in Table S11. In those models, both the country-specific mean and the deviation from that mean are included as competing explanatory variables (as well as a set of year dummies). Such an approach makes it possible to grasp quickly the relative importance of within-country and between-country variations in inequalities to explain variations in the girls-to-boys ratio among high performers. Indeed, the effect of the country-specific mean is identified from between-country variation, while the effect of the deviation from the mean is identified from within-country variation. We see that the within and between-country variation in the GINI index imply similar variations in the gender performance gap in math (models 1 to 3): an increase of a one standard deviation in the GINI index, across or within countries, is associated with a drop of the girls-to-boys ratio of about 0.09. Between-country variation in the three other measures of inequality are more strongly related to gender performance gaps than their within country variation. This shows that for those measures the static cross-sectional relationship with the gender performance gap in math is larger than the dynamic equivalent relationship. This is perhaps not surprising, and does not prevent all estimates to be statistically significant at conventional statistical levels.

We also provide a visual representation of the relationship between variations in the four measures of inequalities considered in Table S11 and variations in the girls-to-boys ratio. This is done in Figure S1 that presents scatter plots of the two types of variations over the period 2006-2015.

Technical issues and robustness:
The choice of inequality indicators in those models is driven by their generality, but also by their greater availability at various points in time. The GINI index (retrieved from the OECD website) is however only available from 2000 to 2014 at the latest. The series for each country also has some gaps, especially prior to 2003. We start by presenting estimates based on (country, year) observations in PISA 2003, 2006, 2009, 2012 or 2015 for which we could retrieve a GINI index for the exact same year (Table S11, model 1). In order to increase the number of observations, we have then matched PISA 2015 with GINI indexes for 2013, and PISA 2003 with GINI indexes for 2004 (the closest years for which the GINI indexes were available for most countries). The association between within-country variations in the GINI and the girls-to-boys ratio is virtually unchanged, but becomes more precisely estimated and statistically significant at the 1% level (model 2). We finally retrieve for each (country=c, year=t) observation in PISA the GINI index in country c for the closest year available starting from t-1, then t+1, then t-2, then t+2. Doing so we increase even more the number of observations and the significance of the estimated effect of within-country variations in the GINI, without changing its value. The fact that only the standard errors of the estimated associations are affected by our imputation procedure suggests that this procedure is not generating biases in the estimates.

Note that we have not provided standard errors and sample sizes for the fixed-effects models estimates presented in Table 2. Sample sizes can directly be inferred from sample sizes given in Table S11 for the random-effect counterpart of each fixed-effect estimate. Then, it is worth noting that fixed-effects estimates are usually very close to the estimates of deviations from the mean in Table S11, and their standard errors are also very close. Hence, the statistical significance of estimates from fixed-effect models, in particular when imputations of the GINI index are made, can be deducted directly by looking at Table S11. The only exception concerns the between-school variation in students' socioeconomic background which has a smaller and less statistically significant effect in fixed-effect models (estimate=-0.075, p-value=0.13) than the deviation from the mean in random-effect models (estimate=-0.131, p-value<0.001).

Note that for most of the variables we consider, a within-country relationship between inequalities and gender performance gaps in reading or science cannot be detected with our data. Except the GINI index which is positively related to the boys-to-girls ratio in reading (as expected), other estimates, both in science and reading, are not statistically significant.

<u>Results based on instrumental variables</u>

Results based on instrumented inequalities are presented in Table S12. Estimates are obtained when our main indicator of economic inequalities, the income GINI index, is instrumented by institutional characteristics of countries' labor markets.

Our first specification in Table S12 instruments the standardized GINI index by the extent of bargaining coverage and union density in OECD countries' labor markets. The idea is that income inequalities are largely driven by labor income inequalities, which are themselves known to be related to the extent of collective bargaining in a country. We see that a one standard deviation increase in the GINI index is still associated with a 0.055 decrease in the girls-to-boys ratio among high performers in math (column 1). Instruments have the expected sign in the first stage: a larger bargaining coverage or union density is associated with less income inequalities. The Fisher statistics is 17.3, indicating that the instruments are not too weak (30). A Sargan's J test reveals that the validity of our over-identification restrictions cannot be rejected. In the second specification, we added the average tax wedge (social security contributions and income tax expressed as a fraction of labor cost) as an additional instrument (column 2). This instrument is perhaps a more direct predictor of income inequalities. It also has the expected sign as a higher tax wedge is associated with less disposable income inequalities. The estimated effect of the GINI index is almost identical, while the first stage becomes a bit weak, with a Fisher statistics slightly below 10. In the last specification, we show that the results are not driven by the use of collective bargaining variables and replace those variables by the value of the statutory minimum in parity of purchasing power. Former results are confirmed on a subset of 25 countries, with a stronger first stage, and no indication that over-identification restrictions are violated[1].

As explained in the main text, the objective of those analyses is to isolate economic inequalities that are arguably driven by institutional factors. Our IV estimates are valid providing that the institutional features of the labor market that we use as instruments affect the gender performance gap among 15 y.o. students only through their effect on inequalities. This does not seem implausible as it is hard to imagine a direct effect of the minimum wage or the bargaining coverage on gender performance gaps at school. The fact that over-

[1] The reason why we do not use the average tax wedge as a single instrument on a sample of 34 countries is that this instrument is too weak.

identification restrictions cannot be rejected is also reassuring as it suggests that if one of the instruments we use in each specification is deemed valid, the others are valid as well.

Correlations between the main inequality measures

Table S14 provides a correlation matrix between our main indicators of inequalities. Non-surprisingly, those indicators are positively correlated. The correlations are not however systematically very high. For example, the correlation of the GINI with other measures ranges between 0.27 (for variability in grade level) and 0.94 (Palma ratio).
When we consider separately the three main categories of inequalities that we have looked at in the paper (socioeconomic and cultural inequalities including pure income inequalities, inequalities in school performance, and inequalities in learning opportunities), we find that correlations between variables belonging to the same category are usually very high, sometimes above 0.75, while correlations between variables belonging to different categories are typically much smaller, usually below 0.5. For example, the correlation of the GINI index with a measure of inequalities in school performance (the share of students from low socioeconomic background among high performers) is equal to 0.37 and its correlation with a measure of inequalities in learning opportunities (the variability in the shortage of education resources) is equal to 0.36.

Extension of the results to the PIAAC data

The OECD also has a program to assess adults' proficiency in numeracy, literacy, and problem solving in technology rich environments. Aggregate data from this program (the Programme for the International Assessment of Adult Competencies - PIAAC) is readily available online (https://nces.ed.gov/surveys/piaac/ideuspiaac/dataset.aspx) and is suited to check if our main results also hold among adults of OECD countries. This is indeed the case, as shown in Table S15, which provides the partial effect of various measures of income inequalities and of the GGI on the girls-to-boys ratio among either young adults (aged 16-34, panel a) or all adults (aged 16-65, panel b) high performers. The partial effect of income inequalities is systematically statistically significant and is comparable in magnitude to estimates obtained with PISA. An effect of the GGI is found among young adults, but not when we focus on the entire adult population.

Extension of the results to the TIMSS data

PISA is not the only available data source regarding students' achievement in various countries in the world. The main alternative source is the Trends in International Mathematics and Science Study (TIMSS) which is conducted on a 4 year cycle by the International Association for the Evaluation of International Achievement (IEA) in collaboration with Statistics Canada and the Educational Testing Service. TIMSS provides an international assessment of mathematics and science learning among fourth, eighth, and twelfth graders.

We have checked if our main results could be replicated on the TIMSS data available since the 2000s, namely on TIMSS 2003, 2007, 2011 and 2015. We have focused on eighth graders exclusively (usually between 14 and 15 years old) since they are the most comparable cohort to the 15 years old students observed in PISA. We have restricted the analysis to income inequalities that are typically available in a large set of countries. Most of our other measures of inequalities were coming from PISA. Equivalent measures are not readily available in

TIMSS, and, except in one case (see below), we have not taken the step to construct them from the TIMSS micro data.

We first confirm the existence of gender performance gaps among TIMSS high performers in math (Table S16). However, those gaps are smaller than those observed with PISA, a result that remains if we restrict the analysis of gaps in PISA and TIMSS to the subset of countries that are present in both surveys. This difference is already known (19) and may be explained by the fact TIMSS and PISA have explicitly different goals. PISA aims at assessing mathematics literacy, which is defined as the capacity to identify and understand the role that math plays in the world, whereas TIMSS aims at assessing the attained curriculum, or what students have learned in the classroom. According to the OECD, 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". In light of these differing aims, TIMSS is more curricular-based while PISA is more skill-based. Compared to PISA, TIMSS tends to test pupil's content knowledge rather than their ability to apply it. It emphasizes basic knowledge and routine problem solving and appears to be a less challenging assessment (19).

Turning to the link between income inequalities and gender performance gaps among high achievers, we find that it still holds in TIMSS for years 2003, 2007 and 2015. The estimated relationships for those years tend to be however a bit smaller than those estimated on PISA. Most of them are significant at the 5% statistical level, but the associated p-values tend to be larger than those obtained in PISA. In 2011, the relationships between income inequalities and gender performance gaps on TIMSS is no longer statistically significant at conventional levels. However, all estimated coefficients have the good sign, they remain relatively large in magnitude, and the associated p-values are usually between 0.2 and 0.3.

We started by looking at the link between income inequalities and the girls-to-boys ratio among high achievers. This ratio was not available in the readily-provided macro data for 2015 at the time of this study, and we therefore computed it directly from the TIMSS micro data for that year. For 2003, 2007 and 2011, we used the values provided online at https://nces.ed.gov/timss/idetimss/. It is difficult to take a definition of high performers in TIMSS similar to that used in PISA (typically the top decile worldwide) because several TIMSS countries have very few students in such a group, and the ratio of interest is not provided directly for those countries. We therefore considered a larger group of top performers which corresponds to students performing above TIMSS High International Benchmark (550) and includes around 25% of the sample worldwide (Table S16). Table S17 shows that in all years except 2011, most of our income inequality measures are significantly associated with the girls-to-boys among top performers defined that way. In 2015, we used the micro-data to define a narrower group of top performers that includes the top decile of the worldwide score distribution. The micro-level data allowed us to compute directly the ratio of the share of girls and the share of boys belonging to that group in each country. This variable closely matches our main measure of gender performance gaps at the top in PISA. When we use it in 2015, we confirm the negative relationship between income inequalities and gender performance gaps at the top (Table S17). In 2015, we also recomputed from the microdata a measure of social inequalities in school performance based on the TIMSS family resources index, and found that this measure is also significantly associated with girls-to-boys ratio at the top. Finally, we show that GGI has no effect on gender performance gaps among high achievers in TIMSS.

Table S18 provides similar associations when the girls-to-boys ratio among high achievers is replaced by the ratio of the 90th percentiles of girls and boys score distribution in each country, a definition which is less subject to sampling errors in countries where overall performance is low and that have few students above a high performance cutoff common to all countries. Most results are confirmed, but, here again, the estimated associations tend to be smaller than in PISA, and are not always statistically significant (and never in 2011). The absence of a significant relationship between the GGI and gender performance gaps is also further confirmed (the only significant coefficient has the wrong sign).

Note that in both Tables S17 and S18, we have excluded Botswana for years when it appeared in the sample of countries. This is because Botswana is a clear outlier in terms of income inequalities, with for example a GINI index above 0.6 while in all other countries, the GINI index is below 0.5. Reintroducing Botswana in the analysis does not radically change the flavor of the results, but it weakens a bit the observed associations.

Understanding why the estimates in Tables S17 and S18 are smaller and tend to be less robust than those observed with PISA would require additional analyses. We make four main hypotheses and provide some preliminary evidence to discuss them.

First, the difference may be due to the fact that TIMSS might be less suited to study high achievers due to its curricular-based framing. To investigate this, we have re-run our analysis considering only the set of questions in TIMSS which requires deeper reasoning and is identified as more challenging (the Reasoning subdomain). Doing so, we find that results tend to be reinforced, suggesting that this first explanation may hold at least partly.

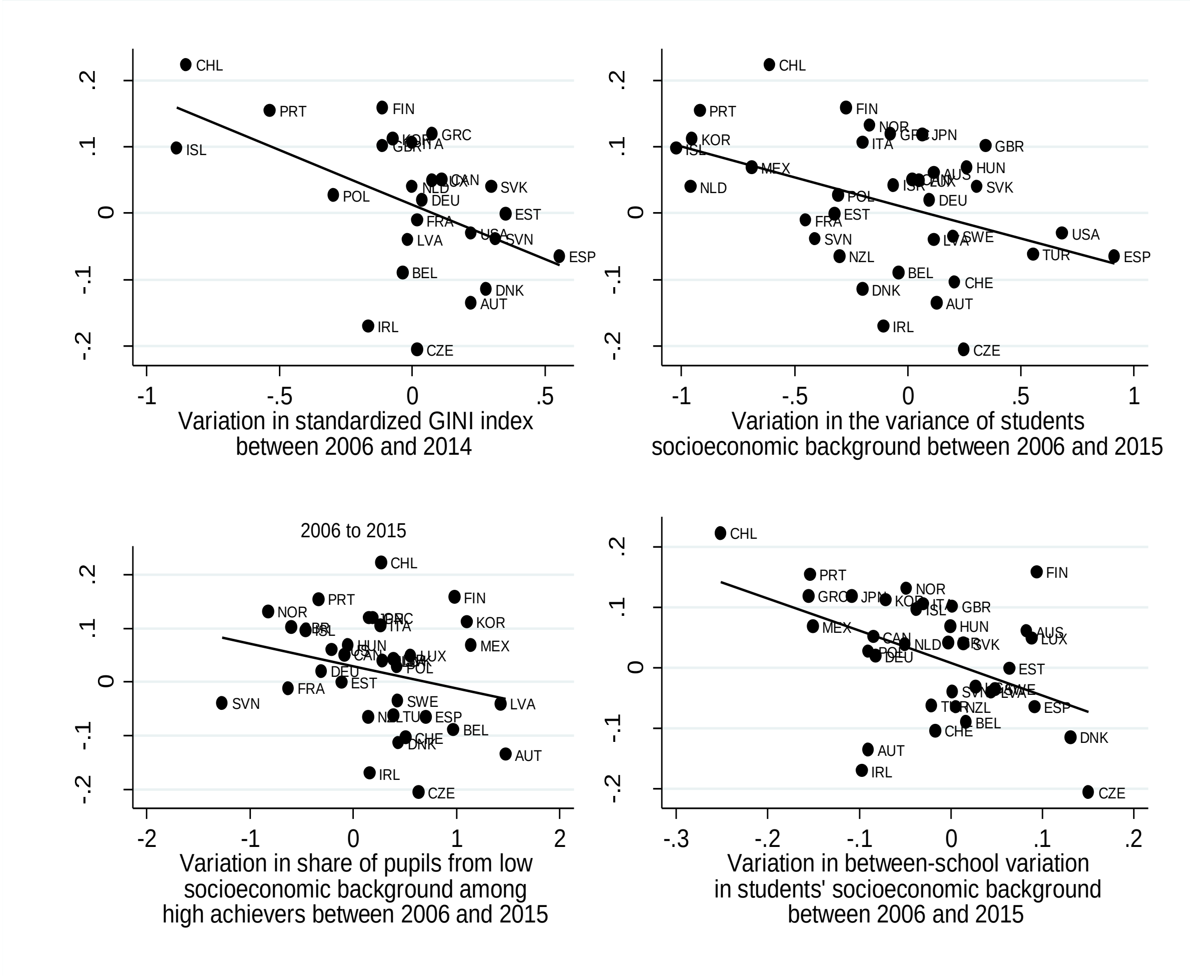
Second, it could be explained by the fact that TIMSS includes many countries for which measures of inequalities and performance gaps may be less reliable. For example, in 2011, the absence of a significant relationship is partly driven by Ukraine, which appears to be the most equal country in the sample. Some observers have claimed that Ukraine appears to have a very low level of income inequality because its black market is large and not included in official statistics. When we remove Ukraine and focus only on the Reasoning subdomain in 2011, we get back the significant relationship between income inequalities and gender performance gaps at the top, as for every other year covered by TIMSS.
A related point is that performance gaps in TIMSS are only established for people that attain the 8th grade. This can lead to important selection biases in countries where only a minority of girls and boys reach that schooling level. As TIMSS includes a lot of developing countries, this selection issue might be more severe in TIMSS.

Fourth, the lower relationship in TIMSS could simply be driven by the fact that TIMSS includes more developing countries than PISA, and that gender issues are still less salient in developing countries than in developed ones, which may limit the scope for an effect of inequalities on gender performance gaps. Testing this explanation further would require a deeper investigation of the results obtained on PISA partner countries which are also mostly developing countries. At first sight, results on PISA partner countries appeared a bit weaker than those on PISA OECD countries.

Differences in results between TIMSS and PISA may finally be explained by more basic differences between the two surveys, such as the fact that TIMSS is based on grade (8th grade) rather than age (15 years 3 month to 16 years 2 month in PISA), or the larger set of countries available in PISA (35 OECD countries and 37 partner countries in PISA versus 39 countries in TIMSS).

Tables and Figures

Figure S1: Country-level relationship between variations in four measures of inequalities and variations in the gender performance gap in math between 2006 and 2015.



Note: OECD countries only. Country codes from ISO3166-1 standard. All inequality variables (x-axis) are standardized to have a mean equal to 0 and a standard deviation equal to 1 on the panel of OECD countries observed over the period 2003-2015. As the GINI index is not systematically available in 2015, we use its value for 2014 instead.

Table S1: Ratio of the shares of girls and boys scoring at or above level 5 in PISA 2015

Country	Girls-to-boys ratio in math[a]	Boys-to-girls ratio in reading	Girls-to-boys ratio in science	Girls-to-boys ratio in math (from microdata)[b]
Australia	0.74	0.69	0.74	0.74
Austria	0.49	0.74	0.52	0.49
Belgium	0.69	0.80	0.65	0.69
Canada	0.76	0.67	0.85	0.76
Chile	0.46	0.79	0.47	0.46
Czech Republic	0.69	0.77	0.60	0.69
Denmark	0.70	0.67	0.67	0.70
Estonia	0.79	0.64	0.80	0.79
Finland	0.92	0.50	1.17	0.92
France	0.73	0.72	0.71	0.73
Germany	0.66	0.72	0.71	0.67
Greece	0.68	0.49	0.78	0.69
Hungary	0.70	0.67	0.74	0.70
Iceland	0.99	0.49	0.81	0.99
Ireland	0.50	1.00	0.55	0.51
Israel	0.58	0.97	0.57	0.58
Italy	0.59	0.77	0.53	0.59
Japan	0.73	0.88	0.69	0.73
Korea	0.92	0.60	0.83	0.92
Latvia	0.69	0.40	0.99	0.69
Luxembourg	0.65	0.79	0.63	0.65
Mexico	0.54	1.14	0.22	0.56
Netherlands	0.83	0.75	0.73	0.83
New Zealand	0.67	0.68	0.74	0.67
Norway	0.85	0.58	0.72	0.85
Poland	0.71	0.67	0.64	0.71
Portugal	0.62	0.89	0.54	0.62
Slovak Republic	0.73	0.56	0.71	0.73
Slovenia	0.80	0.47	0.96	0.80
Spain	0.53	0.75	0.58	0.53
Sweden	0.83	0.59	0.79	0.83
Switzerland	0.72	0.68	0.74	0.72
Turkey	0.59	0.34	0.86	0.61
United Kingdom	0.70	0.68	0.89	0.70
United States	0.74	0.78	0.75	0.75

Notes: The Table shows the ratio of the share of girls and the share of boys scoring at or above Level 5 in math in each country (column 1). The same ratio is shown in science (column 3), and the inverse of that ratio is shown in reading (column 2). Those ratios are country-level estimates that can be computed in two different ways. Column 1 shows the estimator we use in our baseline analyses. It corresponds to the ratio of means over all plausible values provided in PISA of the share of girls and boys scoring at or above Level 5. It can be directly recovered from the aggregated statistics provided by PISA. Column 4 shows an alternative estimator based on PISA micro-data which computes for each plausible value provided in PISA the ratio of the share of girls and boys scoring at or above Level 5 in math and takes the average of those ratios over all plausible values. Both estimators are very close.

Table S2a: The link between several measures of inequalities and the girls-to-boys ratio among high PISA performers in math in 2015

Explanatory variable	*Estimate*	*p value*	*Adjusted r2*	*Number of observations*
a) Economic and socio-cultural inequalities				
GINI index (OECD)	-0.067	0.001	0.280	35
GINI index in 1985	-0.065	0.005	0.374	17
Palma ratio 2005-2013	-0.073	0.000	0.329	31
Ratio of the average income of richest and poorest 20% of the population	-0.062	0.001	0.272	33
Variance in the socioeconomic and cultural background of PISA students	-0.081	0.000	0.421	35
Income share held by the poorest decile*	-0.073	0.000	0.355	33
b) Other non-educational inequalities				
Poverty rate	-0.049	0.016	0.137	35
Infant mortality rate	-0.050	0.016	0.138	35
Adolescent fertility rate	-0.057	0.005	0.190	35
Index of human inequality	-0.054	0.010	0.166	34
Intergenerational earnings elasticity	-0.072	0.008	0.358	16
Index of inequality of economic opportunity -absolute	-0.075	0.000	0.458	23
Index of inequality of economic opportunity -relative	-0.070	0.001	0.387	23
Education (im)mobility	-0.060	0.017	0.248	19
c) Inequalities in performance opportunities				
Percentage of the variation in performance explained by PISA ESCS index	-0.044	0.033	0.104	35
Share of students from low socio economic background among high performers*	-0.085	0.000	0.469	35
Percentage of students scoring below level 2	-0.061	0.002	0.227	35
Increased likelihood of students from low socio economic background to score low	-0.046	0.027	0.113	35
d) Inequalities in learning opportunities				
Between-school variation in students' socioeconomic background	-0.075	0.000	0.348	34
Index of equity in resource allocation: material*	-0.056	0.006	0.182	35
Vertical stratification: variability in grade level	-0.075	0.000	0.355	35
Variability in the index of shortage of educational material	-0.062	0.002	0.237	35
Variability in the time devoted to science lessons	-0.061	0.002	0.225	35
Variability in school size	-0.049	0.016	0.143	34
e) Quality of education				
Mean years of schooling	0.043	0.037	0.099	35
Proportion of 15-years old enrolled at school	0.057	0.005	0.194	35
Staff provided by school to help students with homework	0.044	0.035	0.101	35
Percentage of teachers fully certified by appropriate authority	0.053	0.015	0.151	33
Number of years at preprimary school	0.047	0.025	0.120	34
e) Gender inequalities measured by the Gender Gap Index (GGI)*	-0.052	0.012	0.153	35

*Notes: The Table shows the marginal effect of the measures of inequalities described in each row on the girls-to-boys ratio among students scoring at or above Level 5 in math in PISA 2015. Results are obtained from OLS regression with a single explanatory variable (and the constant). Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.*

Table S2b: The link between several measures of inequalities and the boys-to-girls ratio among high PISA performers in reading in 2015

Explanatory variable	*Estimate*	*p value*	*Adjusted r2*	*Number of observations*
a) Economic and socio-cultural inequalities				
GINI index (OECD)	0.056	0.049	0.086	35
GINI index in 1985	0.123	0.000	0.635	17
Palma ratio 2005-2013	0.070	0.022	0.139	31
Ratio of the average income of richest and poorest 20% of the population	0.070	0.015	0.151	33
Variance in the socioeconomic and cultural background of PISA students	0.039	0.174	0.027	35
Income share held by the poorest decile*	0.048	0.112	0.050	33
b) Other non-educational inequalities				
Poverty rate	0.037	0.193	0.022	35
Infant mortality rate	0.001	0.961	-0.030	35
Adolescent fertility rate	0.051	0.071	0.068	35
Index of human inequality	0.057	0.049	0.088	34
Intergenerational earnings elasticity	0.055	0.013	0.324	16
Index of inequality of economic opportunity -absolute	0.072	0.012	0.229	23
Index of inequality of economic opportunity -relative	0.075	0.009	0.251	23
Education (im)mobility	0.033	0.271	0.016	19
c) Inequalities in performance opportunities				
Percentage of the variation in performance explained by PISA ESCS index	0.044	0.123	0.042	35
Share of students from low socio economic background among high performers*	0.067	0.016	0.138	35
Percentage of students scoring below level 2	0.023	0.422	-0.010	35
Increased likelihood of students from low socio economic background to score low	0.061	0.029	0.109	35
d) Inequalities in learning opportunities				
Between-school variation in students' socioeconomic background	0.057	0.051	0.086	34
Index of equity in resource allocation: material*	0.073	0.008	0.170	35
Vertical stratification: variability in grade level	0.067	0.016	0.139	35
Variability in the index of shortage of educational material	0.043	0.133	0.039	35
Variability in the time devoted to science lessons	0.050	0.079	0.063	35
Variability in school size	0.089	0.001	0.259	34
e) Gender inequalities measured by the Gender Gap Index (GGI)*	0.021	0.477	-0.014	35

*Notes: The Table shows the marginal effect of the measures of inequalities described in each row on the boys-to-girls ratio among students scoring at or above Level 5 in reading in PISA 2015. Results are obtained from OLS regression with a single explanatory variable (and the constant). Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.*

Table S2c: The link between several measures of inequalities and the boys-to-girls ratio among high PISA performers in science in 2015

Explanatory variable	*Estimate*	*p value*	*Adjusted r2*	*Number of observations*
a) Economic and socio-cultural inequalities				
GINI index	-0.062	0.028	0.112	35
GINI index in 1985	-0.113	0.012	0.311	17
Palma ratio	-0.080	0.009	0.188	31
Ratio of the average income of richest and poorest 20% of the population	-0.070	0.015	0.151	33
Variance in the socioeconomic and cultural background of PISA students	-0.078	0.005	0.195	35
Income share held by the poorest decile*	-0.067	0.025	0.125	33
b) Other non-educational inequalities				
Poverty rate	-0.037	0.197	0.021	35
Infant mortality rate	-0.037	0.197	0.021	35
Adolescent fertility rate	-0.077	0.005	0.189	35
Index of human inequality	-0.072	0.012	0.156	34
Intergenerational earnings elasticity	-0.073	0.073	0.155	16
Index of inequality of economic opportunity -absolute	-0.062	0.070	0.107	23
Index of inequality of economic opportunity -relative	-0.062	0.068	0.110	23
Education (im)mobility	-0.065	0.079	0.122	19
c) Inequalities in performance opportunities				
Percentage of the variation in performance explained by PISA ESCS index	-0.052	0.069	0.070	35
Share of pupils from low socio economic background among high performers*	-0.101	0.000	0.346	35
Percentage of students scoring below level 2	-0.072	0.010	0.162	35
Increased likelihood of students from low socio economic background to score low	-0.062	0.029	0.110	35
d) Inequalities in learning opportunities				
Between-school variation in students' socioeconomic background	-0.088	0.002	0.247	34
Index of equity in resource allocation: material*	-0.096	0.000	0.310	35
Vertical stratification: variability in grade level	-0.080	0.004	0.202	35
Variability in the index of shortage of educational material	-0.055	0.054	0.081	35
Variability in the time devoted to science lessons	-0.064	0.023	0.121	35
Variability in school size	-0.076	0.006	0.185	34
Gender Gap Index*	-0.049	0.091	0.056	35

*Notes: The Table shows the marginal effect of the measures of inequalities described in each row on the boys-to-girls ratio among students scoring at or above Level 5 in science in PISA 2015. Results are obtained from OLS regression with a single explanatory variable (and the constant). Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.*

Table S3: The link between several measures of inequalities and the girls-to-boys ratio among high PISA performers in math in 2015, controlling for GDP and gender stratification

Explanatory variable	*Estimate*	*p value*	*Adjusted r2*	*Number of observations*
GINI index (OECD)	-0.062	0.007	0.289	35
GINI index in 1985	-0.032	0.235	0.467	17
Ratio of the average income of richest and poorest 20% of the population	-0.060	0.004	0.270	33
Variance in the socioeconomic and cultural background of PISA students	-0.073	0.000	0.423	35
Share of students from low socio economic background among high performers*	-0.080	0.000	0.441	35
Percentage of students scoring below level 2	-0.053	0.023	0.241	35
Between-school variation in students' socioeconomic background	-0.074	0.005	0.312	34
Index of equity in resource allocation: material*	-0.046	0.028	0.231	35
Vertical stratification: variability in grade level	-0.066	0.001	0.390	35

*Notes: The Table shows the marginal effect of the measures of inequalities described in each row on the girls-to-boys ratio among students scoring at or above Level 5 in math in PISA 2015. Results are obtained from OLS regressions that include three explanatory variables: the measure of inequality described in each given row, countries' GDP, and countries extent of gender stratification measured using the GGI index. A constant term is also included. Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.*

Table S4: The link between several measures of inequalities and the girls-to-boys ratio among high PISA performers in 2015, including OECD partner countries*

Explanatory variable	*Estimate*	*p value*	*Adjusted r2*	*Number of observations*
a) in math				
GINI index (World Bank)	-0.063	0.001	0.198	52
Ratio of the average income of richest and poorest 20% of the population	-0.042	0.054	0.055	51
Palma ratio	-0.068	0.001	0.216	48
Income share held by the poorest decile*	-0.078	0.000	0.308	51
Variance in the socioeconomic and cultural background of PISA students	-0.058	0.006	0.100	64
Difference between the 95th percentile and the 5th percentile of PISA ESCS index	-0.050	0.021	0.069	64
Adolescent fertility rate	-0.019	0.392	-0.004	61
Index of human inequality	-0.021	0.354	-0.002	54
Share of pupils from low socio economic background among high performers*	-0.062	0.004	0.114	63
Percentage of students scoring below level 2	-0.039	0.072	0.036	64
Percentage of the variation in performance explained by PISA ESCS index	-0.053	0.014	0.080	63
Percentage of resilient students (in science)*	-0.058	0.007	0.098	64
Increased likelihood of students from low socio economic background to score low	-0.062	0.004	0.112	63
Between-school variation in students' socioeconomic background	-0.049	0.025	0.064	63
Index of equity in resource allocation: material*	-0.064	0.002	0.125	64
Vertical stratification: variability in grade level	-0.042	0.055	0.043	64
Variability in the index of shortage of educational material	-0.036	0.097	0.028	64
Variability in science competition offered as a school activity	-0.060	0.005	0.109	64
b) in reading				
GINI index (World Bank)	0.043	0.070	0.047	50
Ratio of the average income of richest and poorest 20% of the population	0.040	0.107	0.034	49
Palma ratio	0.044	0.077	0.048	46
Income share held by the poorest decile*	0.037	0.130	0.028	49
Variance in the socioeconomic and cultural background of PISA students	0.029	0.157	0.017	62
Difference between the 95th percentile and the 5th percentile of PISA ESCS index	0.031	0.137	0.020	62
Adolescent fertility rate	0.012	0.584	-0.012	59
Index of human inequality	0.034	0.157	0.020	53
Share of pupils from low socio economic background among high performers*	0.029	0.168	0.015	62
Percentage of students scoring below level 2	-0.007	0.724	-0.015	62
Percentage of the variation in performance explained by PISA ESCS index	0.033	0.111	0.026	62
Percentage of resilient students (in science)*	-0.001	0.977	-0.017	62
Increased likelihood of students from low socio economic background to score low	0.046	0.025	0.066	62
Between-school variation in students' socioeconomic background	0.040	0.056	0.045	61
Index of equity in resource allocation: material*	0.058	0.004	0.115	62
Vertical stratification: variability in grade level	0.050	0.014	0.081	62
Variability in the index of shortage of educational material	0.029	0.155	0.017	62

Variability in science competition offered as a school activity	0.041	0.047	0.049	62
c) in science				
GINI index (World Bank)	-0.070	0.005	0.145	46
Ratio of the average income of richest and poorest 20% of the population	-0.060	0.039	0.073	46
Palma ratio	-0.073	0.007	0.149	42
Income share held by the poorest decile*	-0.063	0.014	0.113	45
Variance in the socioeconomic and cultural background of PISA students	-0.062	0.014	0.088	57
Difference between the 95th percentile and the 5th percentile of PISA ESCS index	-0.058	0.021	0.077	57
Adolescent fertility rate	-0.027	0.297	0.002	54
Index of human inequality	-0.035	0.229	0.010	48
Share of pupils from low socio economic background among high performers*	-0.047	0.065	0.044	57
Percentage of students scoring below level 2	-0.011	0.657	-0.014	57
Percentage of the variation in performance explained by PISA ESCS index	-0.059	0.019	0.080	57
Percentage of resilient students (in science)*	-0.030	0.245	0.007	57
Increased likelihood of students from low socio economic background to score low	-0.068	0.006	0.112	57
Between-school variation in students' socioeconomic background	-0.076	0.002	0.143	56
Index of equity in resource allocation: material*	-0.092	0.000	0.219	57
Vertical stratification: variability in grade level	-0.053	0.034	0.062	57
Variability in the index of shortage of educational material	-0.047	0.062	0.045	57
Variability in science competition offered as a school activity	-0.042	0.097	0.032	57

*Notes: The Table shows the marginal effect of the measures of inequalities described in each row on the girls-to-boys ratio among students scoring at or above Level 5 in PISA 2015 in math, reading and science. OECD as well as non-OECD countries who participated in PISA are included in the analysis. *In each subject, countries that have less than 0.5% of girls or boys scoring above level 5 have been excluded as the girls-to-boys ratio for those countries is not precise enough due to sampling error. Results are obtained from OLS regression with a single explanatory variable (and the constant). Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.*

Table S5: The link between several measures of inequalities and normalized average performance gaps between boys and girls (boys minus girls) in math, science and reading in PISA in 2015

Explanatory variable	*Estimate*	*p value*	*Adjusted r2*	*Number of obs.*
a) Link between inequalities and the average gender performance gap in math				
GINI index (OECD)	0.021	0.142	0.036	35
GINI index in 1985	0.032	0.102	0.113	17
Ratio of the average income of richest and poorest 20% of the population	0.023	0.126	0.044	33
Palma ratio 2005-2013	0.024	0.119	0.050	31
Income share held by the poorest decile*	0.025	0.082	0.065	33
Variance in the socioeconomic and cultural background of PISA students	0.032	0.023	0.120	35
Intergenerational earnings elasticity	0.055	0.010	0.347	16
Index of inequality of economic opportunity -absolute	0.055	0.002	0.331	23
Index of inequality of economic opportunity -relative	0.054	0.003	0.325	23
Education (im)mobility	0.043	0.021	0.231	19
Share of students from low socio economic background among high performers*	0.037	0.007	0.173	35
Percentage of the variation in performance explained by PISA ESCS index	0.030	0.034	0.103	35
Increased likelihood of students from low socio economic background to score low	0.033	0.018	0.133	35
Between-school variation in students' socioeconomic background	0.036	0.013	0.151	34
Index of equity in resource allocation: material*	0.017	0.234	0.014	35
Vertical stratification: variability in grade level	0.038	0.007	0.176	35
Variability in the time devoted to science lessons	0.036	0.009	0.164	35
Variability in science competition offered as a school activity	0.026	0.069	0.069	35
Variability in school size	0.025	0.060	0.079	34
Gender inequalities measured by the GGI*	0.019	0.183	0.024	35
b) Link between inequalities and the average gender performance gap in reading				
GINI index (OECD)	0.035	0.043	0.092	35
GINI index in 1985	0.053	0.016	0.284	17
Ratio of the average income of richest and poorest 20% of the population	0.035	0.050	0.090	33
Palma ratio 2005-2013	0.042	0.023	0.137	31
Income share held by the poorest decile*	0.038	0.037	0.105	33
Variance in the socioeconomic and cultural background of PISA students	0.038	0.028	0.112	35
Intergenerational earnings elasticity	0.067	0.003	0.438	16
Index of inequality of economic opportunity -absolute	0.066	0.002	0.345	23
Index of inequality of economic opportunity -relative	0.072	0.001	0.417	23
Education (im)mobility	0.035	0.165	0.058	19
Share of students from low socio economic background among high performers*	0.047	0.006	0.183	35
Percentage of the variation in performance explained by PISA ESCS index	0.041	0.016	0.137	35
Increased likelihood of students from low socio economic background to score low	0.048	0.005	0.193	35
Between-school variation in students' socioeconomic background	0.046	0.009	0.170	34
Index of equity in resource allocation: material*	0.031	0.080	0.063	35
Vertical stratification: variability in grade level	0.051	0.002	0.225	35

Variability in the time devoted to science lessons	0.047	0.006	0.185	35
Variability in science competition offered as a school activity	0.042	0.015	0.142	35
Variability in school size	0.051	0.003	0.216	34
Gender inequalities measured by the GGI*	0.029	0.101	0.051	35
c) Link between inequalities and the average gender performance gap in science				
GINI index (OECD)	0.018	0.255	0.010	35
GINI index in 1985	0.036	0.090	0.125	17
Ratio of the average income of richest and poorest 20% of the population	0.027	0.093	0.059	33
Palma ratio 2005-2013	0.023	0.161	0.034	31
Income share held by the poorest decile*	0.019	0.227	0.016	33
Variance in the socioeconomic and cultural background of PISA students	0.023	0.141	0.036	35
Intergenerational earnings elasticity	0.048	0.034	0.231	16
Index of inequality of economic opportunity -absolute	0.038	0.061	0.117	23
Index of inequality of economic opportunity -relative	0.038	0.057	0.122	23
Education (im)mobility	0.039	0.052	0.157	19
Share of students from low socio economic background among high performers*	0.036	0.017	0.136	35
Percentage of the variation in performance explained by PISA ESCS index	0.031	0.041	0.093	35
Increased likelihood of students from low socio economic background to score low	0.037	0.014	0.145	35
Between-school variation in students' socioeconomic background	0.034	0.028	0.115	34
Index of equity in resource allocation: material*	0.025	0.106	0.049	35
Vertical stratification: variability in grade level	0.035	0.022	0.123	35
Variability in the time devoted to science lessons	0.038	0.011	0.154	35
Variability in science competition offered as a school activity	0.020	0.192	0.022	35
Variability in school size	0.034	0.020	0.131	34
Gender inequalities measured by the GGI*	0.013	0.397	-0.008	35

*Notes: The Table shows the marginal effect of the measures of inequalities described in each row on the gender standardized gap in math (mean score of boys minus mean score of girls divided by the standard deviation of all scores), reading and science in PISA 2015. Results are obtained from OLS regression with a single explanatory variable (and the constant). Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.*

Table S6: The link between several measures of inequalities and alternative measures of the gender performance gap among high performers in math in PISA in 2015

Explanatory variable	*Estimate*	*p value*	*Adjusted r2*	*Number of observations*
a) girls-to-boys ratio among medium and high performers (at or above Level 4)				
GINI index	-0.072	0.000	0.420	35
GINI index in 1985	-0.094	0.000	0.642	17
GINI index (World Bank)	-0.077	0.000	0.531	34
Ratio of the average income of richest and poorest 20% of the population	-0.075	0.000	0.463	33
Palma ratio	-0.079	0.000	0.506	31
Income share held by the poorest decile*	-0.067	0.000	0.381	33
Variance in the socioeconomic and cultural background of PISA students	-0.071	0.000	0.412	35
Difference between the 95th percentile and the 5th percentile of PISA ESCS index	-0.074	0.000	0.445	35
Index of human inequality	-0.064	0.000	0.316	34
Share of pupils from low socio economic background among high performers*	-0.076	0.000	0.476	35
Percentage of students scoring below level 2	-0.066	0.000	0.347	35
Percentage of the variation in performance explained by PISA ESCS index	-0.030	0.103	0.051	35
Percentage of resilient students (in science)*	0.043	0.017	0.135	35
Increased likelihood of students from low socio economic background to score low	-0.034	0.062	0.074	35
Between-school variation in students' socioeconomic background	-0.074	0.000	0.439	34
Index of equity in resource allocation: material*	-0.065	0.000	0.336	35
Vertical stratification: variability in grade level	-0.059	0.001	0.280	35
Variability in the index of shortage of educational material	-0.046	0.012	0.153	35
Variability in science competition offered as a school activity	-0.039	0.036	0.101	35
Gender inequalities measured by the GGI*	-0.041	0.025	0.118	35
b) girls-to-boys ratio among very high performers (at Level 6)				
GINI index	-0.080	0.001	0.278	35
GINI index in 1985	-0.073	0.008	0.342	17
GINI index (World Bank)	-0.088	0.000	0.358	34
Ratio of the average income of richest and poorest 20% of the population	-0.061	0.013	0.158	33
Palma ratio	-0.085	0.001	0.303	31
Income share held by the poorest decile*	-0.087	0.000	0.333	33
Variance in the socioeconomic and cultural background of PISA students	-0.101	0.000	0.461	35
Difference between the 95th percentile and the 5th percentile of PISA ESCS index	-0.100	0.000	0.454	35
Index of human inequality	-0.074	0.002	0.230	34
Share of pupils from low socio economic background among high performers*	-0.089	0.000	0.352	35
Percentage of students scoring below level 2	-0.087	0.000	0.334	35
Percentage of the variation in performance explained by PISA ESCS index	-0.027	0.284	0.005	35
Percentage of resilient students (in science)*	0.050	0.045	0.090	35
Increased likelihood of students from low socio economic	-0.026	0.314	0.001	35

background to score low				
Between-school variation in students' socioeconomic background	-0.069	0.005	0.194	34
Index of equity in resource allocation: material*	-0.066	0.006	0.183	35
Vertical stratification: variability in grade level	-0.082	0.000	0.296	35
Variability in the index of shortage of educational material	-0.084	0.000	0.311	35
Variability in science competition offered as a school activity	-0.058	0.017	0.134	35
Gender inequalities measured by the GGI*	-0.069	0.004	0.202	35
c) ratio of the 75th percentile of the girls and boys score distributions				
GINI index	-0.005	0.045	0.089	35
GINI index in 1985	-0.006	0.021	0.259	17
GINI index (World Bank)	-0.006	0.013	0.154	34
Ratio of the average income of richest and poorest 20% of the population	-0.005	0.015	0.149	33
Palma ratio	-0.005	0.025	0.132	31
Income share held by the poorest decile*	-0.005	0.019	0.137	33
Variance in the socioeconomic and cultural background of PISA students	-0.005	0.029	0.111	35
Difference between the 95th percentile and the 5th percentile of PISA ESCS index	-0.005	0.021	0.125	35
Index of human inequality	-0.003	0.141	0.037	34
Share of pupils from low socio economic background among high performers*	-0.007	0.001	0.259	35
Percentage of students scoring below level 2	-0.002	0.334	-0.001	35
Percentage of the variation in performance explained by PISA ESCS index	-0.006	0.013	0.148	35
Percentage of resilient students (in science)*	0.002	0.419	-0.010	35
Increased likelihood of students from low socio economic background to score low	-0.006	0.005	0.194	35
Between-school variation in students' socioeconomic background	-0.006	0.006	0.185	34
Index of equity in resource allocation: material*	-0.003	0.163	0.030	35
Vertical stratification: variability in grade level	-0.007	0.002	0.229	35
Variability in the index of shortage of educational material	-0.003	0.169	0.028	35
Variability in science competition offered as a school activity	-0.004	0.067	0.071	35
Gender inequalities measured by the GGI*	-0.003	0.146	0.034	35
d) ratio of the 90h percentile of the girls and boys score distributions				
GINI index	-0.003	0.112	0.047	35
GINI index in 1985	-0.004	0.045	0.191	17
GINI index (World Bank)	-0.004	0.032	0.109	34
Ratio of the average income of richest and poorest 20% of the population	-0.004	0.043	0.097	33
Palma ratio	-0.004	0.068	0.079	31
Income share held by the poorest decile*	-0.004	0.033	0.110	33
Variance in the socioeconomic and cultural background of PISA students	-0.004	0.040	0.095	35
Difference between the 95th percentile and the 5th percentile of PISA ESCS index	-0.004	0.039	0.096	35
Index of human inequality	-0.002	0.229	0.015	34
Share of pupils from low socio economic background among high performers*	-0.006	0.001	0.248	35

Percentage of students scoring below level 2	-0.001	0.617	-0.022	35
Percentage of the variation in performance explained by PISA ESCS index	-0.005	0.009	0.165	35
Percentage of resilient students (in science)*	0.001	0.714	-0.026	35
Increased likelihood of students from low socio economic background to score low	-0.005	0.003	0.207	35
Between-school variation in students' socioeconomic background	-0.005	0.014	0.148	34
Index of equity in resource allocation: material*	-0.002	0.227	0.015	35
Vertical stratification: variability in grade level	-0.006	0.002	0.244	35
Variability in the index of shortage of educational material	-0.003	0.093	0.055	35
Variability in science competition offered as a school activity	-0.002	0.236	0.013	35
Gender inequalities measured by the GGI*	-0.003	0.119	0.044	35

*Notes: The Table shows the marginal effect of the measures of inequalities described in each row on various measures of the gender performance gap among high performers in math in PISA 2015. Results are obtained from OLS regression with a single explanatory variable (and the constant). Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.*

Table S7: The link between several measures of inequalities and the girls-to-boys ratio among high PISA performers in math in 2003, 2006, 2009 and 2012

Explanatory variable	*Estimate*	*p value*	*Adjusted r2*	*Number of observations*
a) 2012				
GINI index (World Bank)	-0.106	0.000	0.496	33
Ratio of the average income of richest and poorest 20% of the population	-0.085	0.000	0.354	33
Income share held by the poorest decile*	0.104	0.000	0.484	33
Variance in the socioeconomic and cultural background of PISA students	-0.087	0.000	0.348	35
Infant mortality rate	-0.060	0.012	0.151	35
Adolescent fertility rate	-0.073	0.002	0.236	35
Percentage of students in the lowest 15% of the international ESCS distribution	-0.073	0.002	0.238	35
Intergenerational earnings elasticity	-0.072	0.035	0.228	16
Index of human inequality	-0.091	0.000	0.371	34
Index of inequality of economic opportunity -absolute	-0.064	0.008	0.258	23
Index of inequality of economic opportunity -relative	-0.070	0.003	0.316	23
Education (im)mobility	-0.067	0.020	0.236	19
Percentage of the variation in performance explained by PISA ESCS index	-0.047	0.054	0.081	35
Percentage of resilient students (in maths)*	-0.061	0.011	0.157	35
Ratio of the mean score of low ESCS students/high ESCS students*	-0.064	0.007	0.175	35
Increased likelihood of students from low ESCS to score low (in maths)	-0.047	0.053	0.082	35
Percentage of students scoring below level 2	-0.098	0.000	0.451	35
Between-school variation in students' socioeconomic background	-0.103	0.000	0.484	34
Index of equity in resource allocation*	-0.063	0.009	0.166	35

Difference in the index of the quality of infrastructures between socio-economically advantaged and disadvantaged schools	-0.069	0.004	0.207	35
Difference in the index of the quality of educational resources between socio-economically advantaged and disadvantaged schools	-0.067	0.005	0.195	35
Percentage of PISA students who have repeated a grade	-0.046	0.060	0.076	35
Vertical stratification: variability in grade level	-0.067	0.005	0.193	34
Variability in the index of shortage of educational staff	-0.067	0.004	0.197	35
Variability in the time devoted to science lessons	-0.042	0.087	0.058	35
Gender inequalities measured by the GGI*	-0.078	0.001	0.276	35
b) 2009				
GINI index (World Bank)	-0.087	0.000	0.451	33
Ratio of the average income of richest and poorest 20% of the population	-0.081	0.000	0.391	33
Palma ratio	-0.090	0.000	0.477	31
Quintile ratio 2005-2013	-0.094	0.000	0.520	31
Gini coefficient 2005-2013	-0.096	0.000	0.551	31
Income share held by the poorest decile*	-0.086	0.000	0.436	33
Variance in the socioeconomic and cultural background of PISA students	-0.077	0.000	0.344	35
Adolescent fertility rate	-0.057	0.006	0.180	35
Infant mortality rate	-0.034	0.119	0.044	35
Percentage of students in the lowest 15% of the international ESCS distribution	-0.061	0.003	0.208	35
Percentage loss between HDI and Inequality-adjusted HDI	-0.056	0.012	0.159	33
Intergenerational earnings elasticity	-0.086	0.001	0.528	16
Index of inequality of economic opportunity -absolute	-0.084	0.000	0.446	23
Index of inequality of economic opportunity -relative	-0.080	0.001	0.402	23
Education (im)mobility	-0.048	0.083	0.118	19
Percentage of the variation in performance (in reading) explained by the PISA ESCS index	-0.040	0.067	0.071	35
Percentage of resilient students (in reading)*	-0.035	0.112	0.047	35
Ratio of the mean score of low ESCS students/high ESCS students*	-0.064	0.002	0.231	35
Increased likelihood of students from low ESCS to score low (in reading)	-0.045	0.035	0.101	35
Percentage of students scoring below level 2	-0.084	0.000	0.417	35
Percentage of students scoring below level 2 (in science)	-0.087	0.000	0.457	35
Between-school variation in students' socio economic background	-0.069	0.001	0.271	34
Percentage of PISA students who have repeated a grade	-0.062	0.003	0.215	35
Difference in the index of the quality of educational resources between socio-economically advantaged and disadvantaged schools	-0.065	0.002	0.235	34
Variability in the index of shortage of educational staff	-0.056	0.010	0.164	34
Variability in the index of the quality of educational resources	-0.084	0.000	0.418	34
Variability in the time devote to science lessons	-0.035	0.112	0.047	35
Variability in the index of school principal's leadership	-0.048	0.029	0.113	34
Competition between schools -percentage of schools that likely transfer students for behavioral or academic reasons	-0.055	0.010	0.162	34
Variability in the index of index of extracurricular activities at school	-0.083	0.000	0.405	34
Gender inequalities measured by the GGI*	-0.037	0.093	0.055	35
c) 2006				

GINI index (World Bank)	-0.105	0.000	0.524	32
Ratio of the average income of richest and poorest 20% of the population	-0.099	0.000	0.503	33
Income share held by the poorest decile*	-0.088	0.000	0.361	32
Variance in the socioeconomic and cultural background of PISA students	-0.089	0.000	0.391	35
Adolescent fertility rate	-0.069	0.003	0.222	35
Infant mortality rate	-0.037	0.125	0.042	35
Percentage of students in the lowest 15% of the international ESCS distribution	-0.084	0.000	0.348	35
Percentage loss between HDI and Inequality-adjusted HDI (2010)	-0.079	0.001	0.289	33
Intergenerational earnings elasticity	-0.072	0.070	0.172	15
Index of inequality of economic opportunity -absolute	-0.061	0.010	0.278	20
Index of inequality of economic opportunity -relative	-0.038	0.132	0.073	20
Education (im)mobility	-0.089	0.010	0.322	17
Percentage of the variation in performance (in science) explained by the PISA ESCS index	-0.043	0.070	0.069	35
Percentage of the variation in performance (in reading) explained by the PISA ESCS index	-0.048	0.049	0.088	34
Percentage of resilient students (in science)	-0.044	0.067	0.071	35
Ratio of the mean score in science of low ESCS students/high ESCS students*	-0.053	0.024	0.118	35
Ratio of the mean score in math of low ESCS students/high ESCS students*	-0.067	0.003	0.209	35
Ratio of the mean score in reading of low ESCS students/high ESCS students*	-0.075	0.001	0.266	34
Percentage of students scoring below level 2	-0.099	0.000	0.489	35
Between-school variation in students' socio economic background	-0.089	0.000	0.385	34
Percentage of PISA students who have repeated a grade in secondary education	-0.057	0.017	0.139	34
Variability in the student/teacher ratio	-0.059	0.013	0.152	34
Difference in the index of the index of shortage of educational staff between socio-economically advantaged and disadvantaged schools	-0.058	0.044	0.123	26
Difference in school activities to promote sciences between socio-economically advantaged and disadvantaged schools	-0.063	0.011	0.176	31
Gender inequalities measured by the GGI*	-0.054	0.020	0.127	35
d) 2003				
GINI index (World Bank)	-0.075	0.005	0.243	27
Ratio of the average income of richest and poorest 20% of the population	-0.059	0.007	0.209	29
Income share held by the poorest decile*	-0.040	0.078	0.078	29
Variance in the socioeconomic and cultural background of PISA students	-0.086	0.000	0.369	30
Adolescent fertility rate	-0.041	0.114	0.056	29
Percentage of students in the lowest 15% of the international ESCS distribution	-0.085	0.000	0.359	30
Intergenerational earnings elasticity	-0.023	0.176	0.083	13
Percentage of resilient students (in math)*	-0.062	0.013	0.173	30
Ratio of the mean score (in math) of low ESCS students/high ESCS students*	-0.047	0.066	0.084	30
Percentage of students scoring below level 2 (in maths)	-0.088	0.000	0.391	30
Between-school variation in students' socio economic background	-0.056	0.031	0.129	29

Percentage of PISA students who have repeated a grade	-0.035	0.172	0.032	30
Difference in the index of the quality of educational resources between socio-economically advantaged and disadvantaged schools	-0.028	0.302	0.004	29

*Notes: The Table shows the marginal effect of the measures of inequalities described in each row on various measures of the gender performance gap among high performers in math in PISA 2012, PISA 2009, PISA 2006 and PISA 2003. Results are obtained from OLS regression with a single explanatory variable (and the constant). Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.*

Table S8: Inequalities and performance of girls and boys in math and reading

	Dependent variable					
	Girls-to-boys ratio among high math performers (log)	**Boys-to-Girls ratio among high reading performers (log)**	**Share of girls that are higher performers in mathematics (log)**	**Share of boys that are higher performers in mathematics (log)**	**Share of girls that are higher performers in reading (log)**	**Share of boys that are higher performers in reading (log)**
Explanatory variable:						
a) Economic inequalities						
GINI index (OECD)	-0.099***	0.062	-0.665***	-0.565***	-0.501***	-0.438***
GINI index in 1985	-0.094***	0.153***	-0.691***	-0.597***	-0.675***	-0.522***
Ratio of the average income of richest and poorest 20% of the population	-0.097***	0.092**	-0.630***	-0.533***	-0.436***	-0.343**
b) Economic and sociocultural inequalities						
Variance in the socioeconomic and cultural background of PISA students	-0.117***	0.039	-0.648***	-0.531***	-0.600***	-0.561***
c) Social inequalities in school performance						
Share of pupils from low socioeconomic background among high performers*	-0.122***	0.087**	-0.520***	-0.398***	-0.532***	-0.445***
d) Inequalities in learning opportunities						
Between-school variation in students' socioeconomic background	-0.110***	0.076*	-0.561***	-0.451***	-0.532***	-0.457***
Inequalities in resource allocation according to socioeconomic background	-0.085***	0.084**	-0.685***	-0.601***	-0.625***	-0.540***
Vertical stratification: variability of grade levels of 15 year old students	-0.107***	0.094**	-0.322**	-0.214	-0.309**	-0.215

*The Table reports the partial effect of various standardized inequalities indicators on measures of girls and boys performance in math and reading. Estimates are obtained from linear regression models on 35 countries in PISA 2015. The log of the girls-to-boys ratio in math and of the boys-to-girls ratio in reading are considered. The partial effect of an inequality indicator on those ratio is the difference between its partial effect on log of the share of girls and the log of the share of boys among high performers. See Table S12 for the exact definition of each variable. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.*

Table S9: "Horse races" to compare the partial effect of societal inequality measures and gender-related inequality measures when they are jointly included as predictors of the girls-to-boys ratio in math in 2015

Societal inequality measures / Gender-related inequality measures	None		GGI*		Gender gap in econ. particip. and opport.*		Female labor force particip*		Gender dev. Index*		Gender equality index		Gender wage gap
None			-0.052**		-0.055***		-0.0553***		-0.035*		-0.043**		0.023
			(0.019)		(0.019)		(0.0190)		(0.020)		(0.020)		(0.021)
Gini index (OECD)	-0.067***	-0.055***	-0.026	-0.053**	-0.027	-0.053**	-0.0283	-0.063***	-0.022	-0.096***	0.035	-0.079***	0.044**
	(0.018)	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)	(0.0205)	(0.018)	(0.018)	(0.031)	(0.031)	(0.017)	(0.017)
Variance in the socioeconomic and cultural background of PISA students	-0.081***	-0.072***	-0.025	-0.070***	-0.027	-0.072***	-0.0177	-0.077***	-0.016	-0.081***	-0.000	-0.082***	-0.002
	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)	(0.019)	(0.0188)	(0.016)	(0.016)	(0.019)	(0.019)	(0.017)	(0.017)
Share of pupils from low socioeconomic background among high performers*	-0.085***	-0.080***	-0.010	-0.076***	-0.018	-0.076***	-0.0184	-0.082***	-0.016	-0.080***	-0.013	-0.085***	0.001
	(0.015)	(0.018)	(0.018)	(0.017)	(0.017)	(0.017)	(0.0174)	(0.016)	(0.016)	(0.017)	(0.017)	(0.016)	(0.016)
Between-school variation in students' socioeconomic background	-0.075***	-0.072***	-0.0055	-0.066***	-0.015	-0.065***	-0.0185	-0.071***	-0.019	-0.075***	0.000	-0.074***	0.018
	(0.017)	(0.023)	(0.023)	(0.022)	(0.022)	(0.022)	(0.0212)	(0.018)	(0.018)	(0.022)	(0.021)	(0.018)	(0.017)

Notes: The table shows the estimated partial effects of societal inequality measures and gender-related inequality measures when they are included "two-by-two" as predictors in a linear regression of the girls-to-boys ratio in math. The first row and the first column of estimates provide the partial effect of those measures when they are considered as the single predictor in a univariate linear regression. An asterix indicates that the opposite of the variable has been considered, so that all estimates are expected to be negative. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S10: Inequalities and the difference between girls and boys in the odds of being a high performer. Based on individual-level logistic regressions from PISA 2015.

Explanatory variable (interacted with students' gender)	*Estimate*	*p value*	*Adjusted r2*	*Number of observations*
a) Economic and social-cultural inequalities				
GINI index	-0.094	0.004	0.141	235574
GINI index in 1985	-0.083	0.051	0.127	130157
Palma ratio	-0.114	0.007	0.139	213926
Ratio of the average income of richest and poorest 20% of the population	-0.124	0.001	0.141	227331
Variance in the socioeconomic and cultural background of PISA students	-0.169	0.000	0.141	235574
b) Inequalities in performance opportunities				
Share of pupils from low socio economic background among high performers*	-0.157	0.000	0.141	235574
Percentage of students scoring below level 2	-0.111	0.006	0.141	235574
c) Inequalities in learning opportunities				
Between school variation in students' socioeconomic background	-0.154	0.000	0.141	229793
Index of equity in resource allocation: material	-0.107	0.004	0.141	235574
Vertical stratification: variability in grade level	-0.125	0.000	0.141	235574

*Notes: The Table reports the differential effect of inequality variables (standardized) on the likelihood of girls relative to boys to score at or above Level 5 in math. Estimates are obtained from separate individual-level logistic regressions on 35 countries in PISA 2015. They include country fixed-effects and individual-level controls for gender, parents' education, grade repetition, household wealth, and household economic, cultural, educational and Information and Communication Technology resources. Controls for the interaction between students' gender and other individual characteristics as well as countries' log GDP per capita are also included. Standard errors account for measurement error in individual math proficiency, and country-level sampling error. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities*

Table S11: Estimates of the link between inequalities and the girls-to-boys ratio among high PISA performers in math from country random-effects models estimated on PISA 2003, 2006, 2009, 2012 and 2015

	(1)	(2)	(3)	(4)	(5)	(6)
Gini index:						
Country-specific mean	-0.0801***					
	(0.0145)					
Deviation from the mean	-0.0924*					
	(0.0479)					
GINI index (PISA 2003 matched with GINI for 2004 and PISA 2015 matched with GINI for 2013)						
Country-specific mean		-0.0788***				
		(0.0156)				
Deviation from the mean		-0.0933***				
		(0.0316)				
GINI index (missing observations matched with the closest past or next available years)						
Country-specific mean			-0.0822***			
			(0.0157)			
Deviation from the mean			-0.0936***			
			(0.0296)			
Variance in the socioeconomic and cultural background of PISA students						
Country-specific mean				-0.0865***		
				(0.0152)		
Deviation from the mean				-0.0437**		
				(0.0205)		
Share of pupils from low socioeconomic background among high performers						
Country-specific mean					-0.0835***	
					(0.0181)	
Deviation from the mean					-0.0277**	
					(0.0125)	
Between-school variation in students' socioeconomic background						
Country-specific mean						-0.264***
						(0.0610)
Deviation from the mean						-0.131***
						(0.0445)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	101	145	167	170	169	169
Number of Countries	35	35	35	35	35	35
Within R-squared	0.087	0.085	0.087	0.043	0.048	0.028
Between R-squared	0.496	0.465	0.469	0.489	0.39	0.4

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. For each measure of inequalities, the Table shows from random-effect models the effect of the country-specific mean inequality over the period considered, as well as the effect of the deviation from that mean in a given year. This corresponds to the Mundlak approach. Estimates of deviations from the mean are identical to estimates of inequalities from fixed-effects models. For each variable, comparisons between the estimated effects of the country-specific mean and the deviation of the mean informs the relative importance of between-countrries and within-country variations in inequalities to explain variations in the gender performance gap in math.

Table S12: 2SLS estimates of the effect of Income inequalities instrumented with institutional variables of the labor market on the Girls to Boys ratio among math high performers

Panel A: Second stage estimates (2SLS IV)			
	dependent variable: girls to boys ratio among math high performers		
GINI index (standardized)	-0.0548*	-0.0573**	-0.0567***
	(0.0327)	(0.0284)	(0.0184)
R-squared	0.291	0.296	0.296
Panel B: First stage estimates			
	dependent variable: standardized GINI index		
Bargaining coverage 2005-2014	-0.0152***	-0.0049	
	(0.0054)	(0.006)	
Union density 2005-2014	-0.0129	-0.0177***	
	(0.0042)	(0.005)	
Average tax wedge 2005-2014		-0.0370*	-0.048***
		(0.019)	(0.012)
Log of min wage (parity of purchasing power) 2005-2014			-0.667***
			(0.216)
Weak identification F stat	17.3	9.5	24.0
Test of overidentifying restrictions (Sargan J-test p-value)	0.103	0.21	0.77
Observations	35	34	25
R-squared	0.425	0.512	0.540

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S13: Summary of the variables used in the study

Variable name*	Source
Economic and socio-cultural inequalities	
GINI index in the 2000s	OECD
GINI index in the 2000s	World Bank
GINI index in 1985	OECD
Palma ratio	OECD
Ratio of the average income of richest and poorest 20% of the population	UNDP
Variance in the socioeconomic and cultural background of PISA students	PISA
Income share held by the poorest decile*	World Bank
Top 1% income share in the 1940s and 1950s	World Top Income Database
Other non-educational inequalities	
Poverty rate	OECD
Infant mortality rate	World Bank
Adolescent fertility rate	World Bank
Index of human inequality	UNDP
Intergenerational earnings elasticity	Corak (2013)
Index of inequality of economic opportunity –absolute	Checchi et al. (2013)
Index of inequality of economic opportunity –relative	Checchi et al. (2013)
Education (im)mobility (parents child correlation in years of schooling)	Hertz et al. (2007)
Inequalities in performance opportunities	
Percentage of the variation in performance explained by PISA ESCS index	PISA
Share of students from low socio economic background among high performers*	PISA
Percentage of students scoring below level 2	PISA
Percentage of resilient students*	PISA
Increased likelihood of students from low socio economic background to score low	PISA
Inequalities in learning opportunities	
Between-school variation in students' socioeconomic background (OECD index of social inclusion*)	OECD, PISA
Index of equity in resource allocation: material *	OECD, PISA
Vertical stratification: variability in grade level	OECD, PISA
Variability in the index of shortage of educational material	OECD, PISA
Variability in the time devoted to science lessons	OECD, PISA
Variability in school size	OECD, PISA
Variability in science competition offered as a school activity	OECD, PISA
Quality of education	
Mean years of schooling	UNDP
Proportion of 15-years old enrolled at school	PISA
Staff provided by school to help students with homework	PISA
Percentage of teachers fully certified by appropriate authority	PISA
Number of years at preprimary school	PISA
Gender related inequalities	

Gender Gap Index (GGI)*	World Economic Forum
Gender gap in economic participation and opportunity	World Economic Forum
Female labor force participation between 35 and 54 y.o.	ILO
Gender development index	UNDP
Gender equality index	UNDP
Gender wage gap	OECD
Countries' development	
Gross Domestic Product	PISA
Human Development Index	UNDP
Employment rate	OECD
Average level of PISA ESCS index	PISA
Health status	OECD
Life expectancy at birth	UNDP 2015
Mean years of schooling	UNDP 2015
Countries' labor markets institutional characteristics	
Average Tax wedge 2005-2014	OECD
Employment Protection index 2005-2014	OECD
Min wage (parity of purchasing power) 2005-2014	OECD
Bargaining coverage 2005-2014	ILO
Union density 2005-2014	ILO

*Notes: variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities. ILO is the International Labor Organization.*

Table S14: correlation matrix between main inequality measures

	GINI index (OECD)	GINI index in 1985	Palma ratio 2005-2013	Variance in the socioeconomic and cultural background of PISA students	Poverty rate	Index of human inequality	Intergenerational earnings elasticity	Share of students from low socio economic background among high performers*	Percentage of students scoring below level 2	Between-school variation in students' socioeconomic background	Index of equity in resource allocation: material*	Vertical stratification: variability in grade level	Gender inequalities measured by the Gender Gap Index (GGI)*
GINI index (OECD)	1.00												
GINI index in 1985	0.96	1.00											
Palma ratio 2005-2013	0.92	0.93	1.00										
Variance in the socioeconomic and cultural background of PISA students	0.56	0.58	0.64	1.00									
Poverty rate	0.84	0.81	0.72	0.43	1.00								
Index of human inequality	0.87	0.91	0.91	0.51	0.82	1.00							
Intergenerational earnings elasticity	0.70	0.74	0.65	0.59	0.55	0.66	1.00						
Share of students from low socio economic background among high performers*	0.37	0.47	0.52	0.65	0.21	0.36	0.72	1.00					
Percentage of students scoring below level 2	0.70	0.78	0.78	0.71	0.48	0.71	0.61	0.62	1.00				
Between-school variation in students' socioeconomic background	0.56	0.76	0.60	0.67	0.45	0.53	0.67	0.73	0.66	1.00			
Index of equity in resource allocation: material*	0.61	0.68	0.61	0.60	0.43	0.65	0.18	0.44	0.69	0.56	1.00		
Vertical stratification: variability in grade level	0.27	0.28	0.31	0.65	0.13	0.19	0.39	0.55	0.34	0.62	0.30	1.00	
Gender inequalities measured by the Gender Gap Index (GGI)*	0.46	0.61	0.41	0.38	0.59	0.57	0.51	0.52	0.46	0.65	0.31	0.29	1.00

Note: A star after a variable name indicates that we have considered the opposite of the original variable.

Table S15: Link between income inequalities and gender performance gaps in numeracy among adults' high performers

a) Girls-to-boys ratio at or above level 4 in numeracy in PIAAC 2012-2014, age 16-34				
GINI index	-0.007	0.032	0.209	18
Ratio of the average income of richest and poorest 20% of the population	-0.007	0.021	0.245	18
Palma ratio	-0.007	0.039	0.191	18
Income share held by the poorest decile*	-0.008	0.012	0.292	18
Ratio of the average income of richest and poorest 10% of the population	-0.008	0.008	0.306	19
Gender inequalities measured by the GGI* (GGI 2012)	-0.007	0.03	0.204	19
b) Girls-to-boys ratio at or above level 4 in numeracy in PIAAC 2012-2014, age 16-65				
GINI index	-0.064	0.008	0.32	18
Ratio of the average income of richest and poorest 20% of the population	-0.063	0.01	0.308	18
Palma ratio	-0.056	0.027	0.224	18
Income share held by the poorest decile*	-0.061	0.014	0.279	18
Ratio of the average income of richest and poorest 10% of the population	-0.049	0.039	0.181	19
Gender inequalities measured by the GGI* (GGI 2012)	0.008	0.754	-0.052	19

Notes: The Table shows the marginal effect of the measures of inequalities described in each row on the boys-to-girls ratio among students scoring at or above Level 4 in numeracy in PIAAC 2012-2014. Results are obtained from OLS regressions with a single explanatory variable (and the constant). Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.

Table S16: Gender performance gaps in math in TIMSS

Year	2003	2007	2011	2015
Number of countries	46	50	42	37
Effect size	-0.01	-0.045	-0.038	-0.026
Share of girls above 550 (a)	0.229	0.187	0.228	0.279
Share of boys above 550 (b)	0.235	0.188	0.234	0.291
Ratio a/b	0.908	0.96	0.959	0.915
Ratio of 90th percentiles of girls and boys score distributions	0.991	0.996	0.994	0.992

Table S17: Income inequalities and gender gaps in math among top performers in TIMSS

	Dependent variable: girls-to-boys ratio among high achievers			
Explanatory variable	*Estimate*	*p value*	*Adjusted r2*	*Number of observations*
a) TIMSS 2015 (girls-to-boys ratios not available in macro data, hence computed from micro data, Botswana and South Africa excluded)				
Ratio of the average income of richest and poorest 20% of the population	-0.037	0.067	0.115	22
Palma index	-0.042	0.034	0.167	22
Gini index (UNDP)	-0.039	0.047	0.142	22
Income share held by the poorest decile*	-0.037	0.101	0.074	25
share of students with low family resources among top performers* (*computed from micro data*)	-0.067	0.006	0.174	37
GGI* (GGI2015)	-0.019	0.470	0.014	35
b) TIMSS 2011				
Ratio of the average income of richest and poorest 20% of the population	-0,040	0,200	0,024	31
Palma index	-0,037	0,237	0,015	31
Gini index (UNDP)	-0,035	0,263	0,010	31
Income share held by the poorest decile*	-0,035	0,255	0,012	30
GGI* (GGI2011)	-0,003	0,937	-0,027	38
c) TIMSS 2007 (Botswana excluded)				
ratio top bottom 20%	-0.126	0.005	0.190	35
Palma index	-0.122	0.007	0.179	35
Gini index (UNDP)	-0.104	0.023	0.122	35
Income share held by the poorest decile*	-0,103	0,005	0,184	36
GGI* (GGI2007)	+ 0.030	0.471	-0,012	42
d) TIMSS 2003 (Botswana excluded)				
Ratio of the average income of richest and poorest 20% of the population	-0.079	0.033	0.114	32
Palma index	-0.077	0.040	0.104	32
Gini index (UNDP)	-0.073	0.051	0.092	32
Income share held by the poorest decile*	-0.075	0,051	0.088	33
GGI* (GGI2006)	-0.066	0.115	0.041	38

Notes: The Table shows the marginal effect of the measures of inequalities described in each row on the boys-to-girls ratio among students scoring at or above 550 in TIMSS 2003, 2007, 2011 and 2015. Results are obtained from OLS regressions with a single explanatory variable (and the constant). Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.

Table S18: Income inequalities and the ratio of the 90h percentile of the girls and boys score distributions in TIMSS

	Dependent variable: ratio of the 90h percentile of the girls and boys score distributions			
Explanatory variable	*Estimate*	*p value*	*Adjusted r2*	*Number of observations*
a) TIMSS 2015 (Botswana and South Africa not available)				
Ratio of the average income of richest and poorest 20% of the population	-0,005	0.117	0.074	22
Palma index	-0,004	0.149	0.056	22
Gini index (UNDP)	-0,004	0.234	0.023	22
Income share held by the poorest decile*	-0.006	0.04	0.135	25
GGI* (GGI2015)	+0.004	0.114	0.046	35
b) TIMSS 2011				
Ratio of the average income of richest and poorest 20% of the population	-0.004	0.223	0.018	31
Palma index	-0,003	0,384	-0,007	31
Gini index (UNDP)	-0,002	0,463	-0,015	31
Income share held by the poorest decile*	-0,004	0,234	0,016	30
GGI* (GGI2011)	+0,002	0,471	-0,013	38
c) TIMSS 2007 (Botswana excluded)				
Ratio of the average income of richest and poorest 20% of the population	-0.011	0.002	0.238	35
Palma index	-0.011	0.004	0.205	35
Gini index (UNDP)	-0.009	0.017	0.136	35
Income share held by the poorest decile*	-0.009	0.002	0.224	36
GGI* (GGI2007)	+0.008	0.042	0.076	42
d) TIMSS 2003 (Botswana excluded)				
Ratio of the average income of richest and poorest 20% of the population	-0.006	0.048	0.095	32
Palma index	-0.005	0.095	0.060	32
Gini index(UNDP)	-0.005	0.143	0.039	32
Income share held by the poorest decile*	-0,005	0,110	0,051	33
GGI* (GGI2006)	0.000	0.842	-0.027	38

Notes: The Table shows the marginal effect of the measures of inequalities described in each row on ratio of the 90th percentile of the girls' and boys' score distribution in each country in TIMSS 2003, 2007, 2011 and 2015. Results are obtained from OLS regressions with a single explanatory variable (and the constant). Measures of inequalities have been normalized to have a standard deviation equal to 1, so that results are comparable across rows. Variable names followed by * are variables for which we took the opposite in the statistical analysis, so that an increase in the variable corresponds to more inequalities.