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

DISCUSSION PAPER SERIES

IZA DP No. 12503

Girls' Comparative Advantage in Reading Can Largely Account for the Gender Gap in Math-Intensive Fields

Thomas Breda
Clotilde Napp

JULY 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12503

Girls' Comparative Advantage in Reading Can Largely Account for the Gender Gap in Math-Intensive Fields

Thomas Breda
Paris School of Economics, CNRS and IZA

Clotilde Napp
CNRS, Université Paris Dauphine and PSL Research University

JULY 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

Girls' Comparative Advantage in Reading Can Largely Account for the Gender Gap in Math-Intensive Fields*

Gender differences in math performance are now small in developed countries and they cannot explain on their own the strong under-representation of women in math-related fields. This latter result is however no longer true once gender differences in reading performance are also taken into account. Using individual-level data on 300,000 15-year-old students in 64 countries, we show that the difference between a student performance in reading and math is 80% of a standard deviation larger for girls than boys, a magnitude considered as very large. When this difference is controlled for, the gender gap in students' intentions to pursue math-intensive studies and careers is reduced by around 75%, while gender gaps in self-concept in math, declared interest for math or attitudes towards math entirely disappear. These latter variables are also much less able to explain the gender gap in intentions to study math than is students' difference in performance between math and reading. These results are in line with choice models in which educational decisions involve intra-individual comparisons of achievement and self-beliefs in different subjects as well as cultural norms regarding gender. To directly show that intra-individual comparisons of achievement impact students' intended careers, we use differences across schools in teaching resources dedicated to math and reading as exogenous variations of students comparative advantage for math. Results confirm that the comparative advantage in math with respect to reading at the time of making educational choices plays a key role in the process leading to women's under-representation in math-intensive fields.

JEL Classification: I24, J16

Keywords: gender gap, math-intensive fields, students' achievement, comparative advantage

Corresponding author:
Thomas Breda
Paris School of Economics
48 Bd Jourdan
75014 Paris
France

E-mail: thomas.breda@ens.fr

* The paper is published in the Proceedings of the National Academy of Science of the United States of America and available at https://doi.org/10.1073/pnas.1905779116.

Introduction

Women are underrepresented in science, technology engineering and mathematics (STEM) university majors and jobs. STEM is however a broad group that includes fields in which women are not under-represented, such as life science or psychology. Scholars have underlined the necessity to focus more narrowly on the STEM fields which are math intensive, such as computer science or engineering (1-3), as the under-representation of women in these fields remains large and has not decreased at all in most developed countries during the two past decades (3-5). For example, over the period 2004–2014, the share of bachelor's degrees awarded to women in engineering and computer science in the U.S. has stagnated around 20 percent, while it has decreased from 46 percent to 43 percent in mathematics and statistics and from 42 percent to 40 percent in physical science (6).

This under-representation of women in math-intensive fields is a source of concern for two main reasons. First, it contributes to gender wage inequality in the labor market as math-intensive jobs pay more (7-9) and are also subject to a smaller gender wage gap (10). Second, it represents a loss of talent that can reduce aggregate productivity (11)–as many talented girls shy away from math-intensive careers–leading to the shortage of workers with math-related skills at a time when the demand for such skills is increasing (12).

Gender differences in math test scores are now very limited in most countries and can only explain a small fraction of this under-representation of women in math-intensive fields (13-14, Table S1). This has pushed scholars to look for other explanations, such as discrimination against women in STEM, or the role of social norms and stereotypes in shaping educational choices. Evidence of direct discrimination is limited (3, 14-15), and many scholars now emphasize the role of gender differences in preferences, self-concept and attitudes towards math, as well as the social processes and institutions possibly shaping these differences (see references in (1)).

We revisit the role of abilities and test scores to explain the gender gap in students' decision to enroll in math-related fields. Our examination is motivated by the idea that students are likely to decide to major in a given field on the basis of their relative (rather than absolute) ability in that field with respect to other fields (16-18). This simple theory is backed-up by studies suggesting that students tend to think in terms of "what they are better at" rather than in terms of "required skills to succeed in a particular career" (19), and that they are encouraged to do so by teachers and their environment (20). Research in social psychology also shows that "people think of themselves as either math persons or verbal persons but not both" (21). Hence, a student that is good at math but even better at reading may favor humanities because she perceives herself as a verbal person. This is despite the fact that her career prospects (which students tend to be unaware of) may be better after math-related studies.

While in most countries, at the age of making irreversible educational choices, girls now perform only slightly worse than boys in math, they however strongly outperform them in reading (18). This gives girls a comparative advantage for disciplines related to reading/literature rather than math. Former studies concluded that this relative advantage could not explain gender differences in STEM choice (*e.g.*, in Sweden in the 1990s (16) and in the U.S. in the 2000s (22)). In contrast, we show that in 2012, it can explain a very large fraction of the gender gap in 15-year-old students' intentions to pursue math-intensive studies and careers in virtually all developed countries and several developing countries.

Comparative advantage and gender gap in intentions to pursue math-related studies and careers

Our main analyses are based on data from the 2012 Programme for International Student Assessment (PISA2012), an every-three-year international assessment of the knowledge and skills of 15-year-old students in mathematics, reading, and science. PISA2012 takes place in the 34 mostly developed countries belonging to the Organisation for Economic Co-operation and Development (OECD) in 2012 and an additional 30 developing countries (see details on all data and analyses in Supplemental Information (SI)). PISA2012 is well-adapted for our purpose for three reasons. First, it allows us to focus directly on math-intensive fields rather than STEM fields. Second, it focuses on a critical age, corresponding to the end of middle school or beginning of high school. In most countries, the majority of students of that age have not yet strongly specialized in a specific field (e.g., in the U.S. this PISA assessment occurs prior to the opportunity to enroll in AP courses such as Calculus BC, Physics C, and Computer Science), so that gender differences in abilities are unlikely to capture anterior specialization. However, 15-year-old students in developed countries are also in the process of choosing high-school courses that will determine their future major and the gender gap in STEM at university (23). A final advantage of PISA is its coverage, as it includes students from 80% of the world economy.

The first column of Table 1 shows that boys outperform girls in math by about 10% of a standard deviation (s.d.). This difference is lower than 25% of a s.d. in most OECD countries and is not statistically significant at the 5% level in four of them (Table S1 completes Table 1 for all countries). In contrast, girls outperform boys by about a third of a s.d. in reading. Together, these observations suggest that girls have a comparative advantage in reading, something that appears more strikingly when we look at the gender gap in the difference between math and reading ability (Table 1, column 3). Worldwide, this gap reaches about 80% of a s.d., or equivalently 24 percentile ranks of the variable. The phenomenon occurs in virtually all countries: the gap is larger than 75% in all OECD countries and larger than 60% everywhere else except in Singapore. Such magnitudes are commonly considered as very large by social scientists.

PISA2012 includes questions related to intentions to pursue math-intensive studies and careers. These intentions are measured through a series of five questions that ask students if they are willing (i) to study harder in math versus English/reading courses (ii) to take additional math versus English/reading courses after school finishes, (iii) to take a math major versus a science major in college, (iv) to take a maximum number of math versus science classes, and (v) to pursue a career that involves math versus science. Our main measure of math intentions is an index constructed from these five questions (for details see SI) and available for more than 300,000 students. It captures the desire to do math versus both reading and other sciences. We complete the analysis with the study of the first two variables that capture more specifically the arbitrage between math and reading.

Column 4 of Table 1 shows that the gender gap in math intentions amounts to 22% of a s.d. worldwide (respectively 26% and 17% among OECD and non-OECD countries). This gap varies across countries. 2 OECD (5 non-OECD) countries have no gap or even a small gap in favor of girls (e.g., Turkey, Malaysia or Thailand, see Table S1). 5 OECD (11 non-OECD) countries have a small-to-medium gap (between 10 and 20% of a s.d.). 17 OECD (13 non-

OECD) countries have large gaps (between 20% and 45% of a s.d.). Finally, 10 OECD (1 non-OECD) countries have gaps larger than 45% of a s.d. (e.g., Australia, Germany, Finland).

The gender gap in intentions cannot be explained by differences in math ability across genders. When one controls for math ability in a linear regression of these intentions on a gender dummy, the estimate for the gender dummy is reduced by less than 10% worldwide and in a majority of the studied countries (column 5 of Tables 1 and S1). Similarly, controlling for reading ability barely affects the gender gap in intentions.

In contrast, the gender gap in intentions to pursue math-intensive studies and careers disappears almost entirely when one controls for individual-level differences in ability between math and reading (MR). Column 7 of Table 1 shows that MR can explain 78% of the gender gap in intentions worldwide (95% confidence interval=[71%,83%], see SI). The corresponding statistics is 81% for OECD countries only, 88% for the U.S., 52% for Germany, 43% for France and 157% for Japan, a country where conditional on MR, girls become more willing to study in math-intensive fields than boys. Similar results are obtained when we measure math intentions with the two questions that capture more specifically the arbitrage between math and reading (Table S2).

MR is more strongly associated with students' intentions to study math than are math or reading abilities taken in isolation (Figure 1 and Table S3). This association is large and very similar for boys and girls, implying that the gender gap in intentions is small and almost constant—only ~5% of a s.d.—along the distribution of MR. In contrast, absolute levels of math or reading abilities leave a large gender gap in intentions unexplained.

The simple difference MR summarizes relatively well the relevant information on abilities that is needed to predict intentions to pursue math-intensive studies and careers. We show in the SI that MR alone captures about 75% of the total capacity of the distributions of math, reading and science abilities to predict intentions to pursue math-related studies and careers. We also show that when we include detailed controls for students' abilities in regression models of these intentions, our results remain qualitatively similar (Table S4).

Our analyses of the relationship between abilities and intentions invite to nuance two ability-based arguments that are sometimes advanced to explain the gender gap in enrolment in STEM: the fact that girls remain under-represented among high math achievers, hence less able to pursue math-related studies, and the fact that they are more often good in both math and reading, hence less constrained than boys in their choice of study (24).

An under-representation of girls among high math-achievers is indeed observed in most countries (25), but taken in isolation, this phenomenon is unlikely to be a good explanation for the gender gap in math-intensive fields. Indeed, this gender gap tends to be larger among high math-achievers (Figure 1, Table S5).

Turning to the second possible explanation, we observe that the gender gap in intentions is not reduced among students that perform above a given threshold in both math and reading (see Table S5 for results based on various thresholds). In contrast, the gender gap in intentions among students that are better in math than in reading (68% of them boys) or better in reading than in math (68% of them girls) is more than twice lower than the average gap.

A possible limitation of the results presented so far is that declared intentions to study math may not capture well actual schooling decisions and gender gaps in enrolments. A first

reassuring element is that sex differences in occupational plans in high school have been found to be a strong predictor of actual gender differences in STEM majors (26, 27). Moreover, we show that cross-country variations in gender gaps in intentions to pursue math-intensive studies and careers measured in PISA are well correlated with objective country-level measures of sex segregation by field of study, like (i) the percentage of women among STEM graduates in tertiary education (rho=-0.52, see SI), (ii) female over-representation in humanities (rho=0.39), or (iii) the female-to-male ratio in computer science (rho=-0.5, see Table S8).

To discuss more directly the effect of MR on actual schooling decisions, we use an auxiliary dataset for France. It includes ability measures in math and reading as well as information on both intentions to study STEM and future enrolment in STEM (see SI for details). We find that the correlation between intentions to specialize in STEM in Grade 11, declared in Grade 10 during the period January-March (corresponding to the same age as that of PISA students), and actual enrolment in Grade 11 is strong but not perfect (78%). However, and crucially to us, MR is a good predictor of both intentions to study STEM and STEM enrolment, and it reduces the gender gaps in these variables to the same extent (46% for intentions and 49% for actual enrolment in STEM, which also corresponds to what we find for France with PISA, see Table S6). From these observations and more detailed analyses presented in the SI, we conclude that our results on students' intentions of study are likely to generalize to their actual course choices in high-school and at university (the latter being strongly related to the former (23)).

Finally, our analysis of the relationship between abilities and intentions highlights a "better self-selection" of boys in math-intensive fields. Indeed, if the relation between MR and math intentions is similar for boys and girls, the relation between math ability and math intentions is larger for boys than for girls (see Figure 1 and SI). Boys take more into account their math ability when they intend to pursue math studies. This leads to a higher gender gap in intentions to study math among students performing above the median in math (Table S5), and to a larger gender gap in math performance among individuals who intend to choose math studies over reading (see SI). These selection patterns are likely to result in an overperformance of boys in math-intensive studies. Similarly, we show in the SI that girls self-select better in humanities and are likely to over-perform boys in university humanity majors even more than they do before specialization occurs. These likely larger gender gaps in performance after specialization, which are generated by gender differences in the self-selection process across fields of study, can feed the stereotype that math is not for girls and humanities not for boys.

Comparative advantage and gender gaps in math self-concept, interest for math, and other math-related attitudes

Gender differences in math self-concept *(i.e.*, how students perceive their math ability and their ability to learn math quickly) is one of the most commonly advanced explanations for the gender gap in math enrolment (1, 28, 29). A series of questions in PISA2012 makes it possible to build an index to measure this concept at the student level (see SI). The gender gap in math self-concept is indeed large (around 30% of a s.d.) but nevertheless three times smaller than the gender gap in MR (Table 2, column 1 for results worldwide and Table S7 for results on a selection of countries/regions). Interestingly, gender differences in the way students perceive their math ability are barely reduced when this ability is controlled for in a linear regression model, while they almost entirely disappear when one controls for MR (Table 2, columns 2 and 3

and Figure 2). We then perform the opposite exercise and show that gender differences in MR cannot be directly explained by gender differences in math self-concept (Table 2, column 4). Such results are fully in line with Marsh's Internal/External (I/E) Frame of Reference model (30) according to which people compare their performances across domains (in particular math versus reading) to reach conclusions about their ability.

Similar results are obtained when we replace math self-concept by other well-known proximate sources of the gender gap in math-related fields, such as gender differences in declared interest for math, instrumental motivation for math, anxiety with respect to math, willingness to get involved in math-related activities, or having a strong "math environment" (i.e., family support for doing math and friends being positive about math). There is a gender gap in the variables that attempt to capture these concepts (Table 2 and SI for details), but (i) these gaps are 3 to 8 times smaller than the gender gap in MR, (ii) they get close to zero when one conditions on MR (except for the involvement in math-related activities), and in contrast (iii) they barely explain the gender gap in MR.

We finally show that the math self-concept and our variables capturing other possible mechanisms are related to students' intentions to study math (Table 2, column 5) but account for a much smaller share of the gender gap in these intentions than does MR (Table 2, columns 6 and 7 for the gender gap in intentions conditional on each variable separately and together with MR). MR is not more strongly associated with intentions than are the other studied variables (Table 2, column 5). This implies that the larger explanatory power of this variable is mostly due to the fact that it is subject to a very large gender gap.

Even if all our analyses so far are only descriptive and not causal, they consistently point to an important role of the comparative advantage of boys in math versus reading for the understanding of women's under-representation in math-intensive fields. This does not rule out the operation of other (perhaps earlier-occurring) factors, of course. Math and reading abilities at 15 years old are likely to be determined by earlier socialization processes that shape preferences and investment in the different fields. These processes are themselves likely to be influenced by countries' socio-economic environment and culture (25, 31) or institutions such as parents and schools which jointly determine future abilities, interests and self-concepts. For example, we observe that the gender gap in MR at 15 years old is larger in countries where the stereotype associating math with men is stronger (Table S8). We also observe that the gender gap in MR at 15 years old is larger in educational systems in which horizontal stratification by field of study is higher or occurs earlier, and in which mandatory standardized tests are less frequent (SI). These observations and more broadly all our analyses are entirely consistent with the choice models developed by Eccles and coauthors in which educational decisions involve intra-individual comparisons of achievement, self-beliefs and motivation in different subjects as well as cultural norms, in particular surrounding gender (32, 33). As such, the present paper provides new supporting evidence for these models.

Instrumental variables and causal inference

While the co-determination of the variables examined here has to be kept in mind, it is not contradictory with our hypothesis that the comparative advantage is an important independent determinant of educational choices, so that exogenous variations in this advantage

(*e.g.*, due to educational policies) can lead students to change their choice of study. We suggest that this is indeed the case by exploiting differences across schools in the availability or shortage of resources to learn math. We show, for example, that in schools that experience a shortage of math teachers but not of reading teachers, both girls' and boys' comparative advantage in math is significantly lower.

The majority of 15-year-old students go to the closest school from where they live and those who do otherwise might struggle to observe shortages in some types of teachers or the quality of math teachers. As a consequence, we assume that quantity and quality of math teachers in their school is to a certain extent exogenous to students' initial intentions to study math. Based on this assumption, we use these school-level variables as instruments for students' comparative advantage in math and show that variations in this comparative advantage that solely arise from differences of "math resources" across schools do affect girls' and boys' intentions to study math (even more than non-instrumented variations, see Table S9).

Our approach would fail to show causality if the students with a large comparative advantage selected into better schools that are likely to have more math resources. For this reason, we include controls for school quality and use as instrumental variables resources devoted to math *relative* to other subjects rather than absolute math resources (which are more directly correlated with school quality, see all details in SI). Finally, we show that the results also hold on the subsample of schools that mostly recruit students based on the geographical location as a self-selection of students in these schools based on their prior comparative advantage appears less likely.

Policy implications

The analysis above suggests that external factors influencing students' comparative advantage are likely to have consequences for their educational choices. As a consequence, any educational policy that could reduce the gender imbalances in comparative advantage is likely to limit the under-representation of women in math-intensive fields. As the gender gap in reading performance is much larger than that in math performance, policy makers may want to focus primarily on the reduction of the former. Systematic tutoring for low reading achievers, who are predominantly males, would be a way for example to improve boys' performance in reading. A limitation of this approach, however, is that it will lower the gender gap in math-intensive fields mostly by pushing more boys in humanities, hence reducing the share of students choosing math. In a context of high and increasing demand for math-intensive skills, improving boys' performance in reading without also improving girls' performance in math can therefore be detrimental for the economy and the latter should also be considered as a valuable option.

The general organization of a country's educational system can also play an important role to limit gender imbalances in comparative advantage. As mentioned above, educational systems with early tracking or specialization are associated with larger gender gaps in comparative advantage, possibly because stereotypes and social norms have a stronger influence on choices at younger age. Delaying the time of making hard-to-reverse educational choices may therefore limit gender gaps in comparative advantage and gender segregation across fields.

Another option in terms of policy is to better inform students regarding the returns to different fields of study, something that is likely to trigger large effects on educational choices

(34). As labor market opportunities and earnings are significantly higher in math-related careers (11), many (mostly female) students who have a comparative advantage in reading but are nevertheless talented in math would have better career prospects in math-related fields. Hence, adequate information campaigns on future career prospects may be a welfare-improving way (because students can make better informed choices) of reducing the importance of the comparative advantage in students' decision making, and therefore the gender gap in enrolment in math-related fields (35). Similarly, interventions involving teachers or parents targeted at limiting the role of the comparative advantage in educational choices could also be effective. Of course, these options should complement rather than replace interventions directly aimed at limiting the negative effects of gender stereotypes.

Acknowledgments: We warmly thank Francesco Avvisati and Matthias von Davier for their helpful suggestions regarding the optimal way to deal with PISA plausible values in the context of this paper, Stephen Ceci for his support and advice and Fabrice Riva for critical reading. We are also grateful to Julien Grenet, Marion Monnet and Clémentine Van Effenterre for allowing us the access to their data for France which are used in Table S6. The support of the Women and Science Chair (a Dauphine Foundation Chair) and of its sponsors (Fondation L'Oréal, Generali, la Poste, Talan) is gratefully acknowledged by both authors.

References:

1. Kahn, S., & Ginther, D. (2017). *Women and STEM* (No. w23525). National Bureau of Economic Research.
2. Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262-265.
3. Breda, T., & Hillion, M. (2016). Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science*, *353*(6298), 474-478.
4. National Science Foundation, *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2017*, National Science Foundation and National Center for Science and Engineering Statistics. 2017. Special Report NSF 17-310. Arlington, VA.
5. OECD, *Education at a Glance 2016, Paris: OECD Publishing, 2016.*
6. See https://www.nsf.gov/statistics/2017/nsf17310/data.cfm.
7. Brown, C. and M. Corcoran, "Sex-Based Differences in School Content and the Male-Female Wage Gap," *Journal of Labor Economics*, 1997, *15* (3), pp. 431–465.
8. Black, D.A., A.M. Haviland, S.G. Sanders, and L.J. Taylor, "Gender Wage Disparities among the Highly Educated," *Journal of Human Resources*, 2008, *43* (3), 630–650.
9. Blau, F.D. and L.M. Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.
10. Beede, D., T. Julian, D. Langdon, G. McKittrick, B. Khan, and M. Doms. 2011. "Women in STEM: A Gender Gap to Innovation." U.S. Department of Commerce Economics and Statistics Administration Issue Brief, #04-11. Available at: http://www.esa.doc.gov/sites/default/files/womeninstemagaptoinnovation8311.pdf.

11. Weinberger, C.J, "Mathematical College Majors and the Gender Gap in Wages," *Industrial Relations*, 1999, *38* (3), pp. 407–413.
12. Carnevale, A.P., N. Smith, and M. Melton, *STEM: Science, Technology, Engineering, Mathematics*, Washington, DC: Georgetown University Center on Education and the Workforce, 2011.
13. Mann, A., & DiPrete, T. A. (2013). Trends in gender segregation in the choice of science and engineering majors. *Social Science Research*, 42(6), 1519-1541.
14. Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, 15(3), 75-141.
15. Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2: 1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences*, 201418878.
16. Jonsson, J. O. (1999). Explaining sex differences in educational choice an empirical assessment of a rational choice model. *European Sociological Review*, 15(4), 391-404.
17. Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, *29*(4), 581-593.
18. Park, G., Lubinski, D., & Benbow, C. P. (2007). Contrasting intellectual patterns predict creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science*, *18*(11), 948-952.
19. Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1-51.
20. Gardner, A. (2016). How important are GCSE choices when it comes to university? London, UK: Which? Retrieved from http://university.which.co.uk/advice/gcse-choices-university/how-important-are-gcse-choices-whenit-comes-to-university.
21. Marsh H.W. and K.T. Hau (2003). Relations between academic self-concept and achievement in mathematics and language: a cross-cultural generalizability. Paper presented at the joint AARE/NZARE conference Auckland. Available at https://www.aare.edu.au/data/publications/2003/hau03756.pdf
22. Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. *American Educational Research Journal*, *49*(6), 1048-1073.
23. Card, D., & Payne, A. A. (2017). *High school choices and the gender gap in STEM* (No. w23769). National Bureau of Economic Research.
24. Wang, M. T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science*, *24*(5), 770-775.
25. Breda, T., Jouini, E., & Napp, C. (2018). Societal inequalities amplify gender gaps in math. *Science*, *359*(6381), 1219-1220.

26. Morgan, S.L., Gelbgiser, D., Weeden, K.A. (2013). Feeding the pipeline: Gender, occupational plans, and college major selection. *Social Science Research*, *42*(4), 989-1005.
27. Xie, Y., Shauman K.A. (2003). Women in Science: Career Processes and Outcomes. Cambridge, MA, Harvard University Press.
28. Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology*, *106*(6), 1691-1730.
29. Nix, S., Perez-Felkner, L.C, and Thomas, K (2015). Perceived mathematical ability under challenge: a longitudinal perspective on sex segregation among STEM degree fields. Frontiers in Psychology, 6. Doi :10.3389/fpsyg.2015.00530
30. Marsh, H. W. (1986). Verbal and math self-concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal*, *23*, 129-149.
31. Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. S*cience*, *320*(5880), 1164.
32. Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., ... & Spence, J. T. (1983). Achievement and achievement motivation. *Expectancies, values and academic behaviors*, 75-146.
33. Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., & Eccles, J. S. (2008). Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value. In H. M. G. Watt & J. S. Eccles (Eds.), *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 115-143). Washington, DC, US: American Psychological Association.
34. Baker, R., Bettinger, E., Jacob, B., & Marinescu, I. (2018). The effect of labor market information on community college students' major choice. *Economics of Education Review*, 65, 18-30.
35. Barone, C., Schizzerotto, A., Assirelli, G., & Abbiati, G. (2018). Nudging gender desegregation: a field experiment on the causal effect of information barriers on gender inequalities in higher education. *European Societies*, 1-22.

Figure 1: Intentions to pursue math-intensive studies and careers as a function of ability in math, reading and the comparative advantage in math versus reading

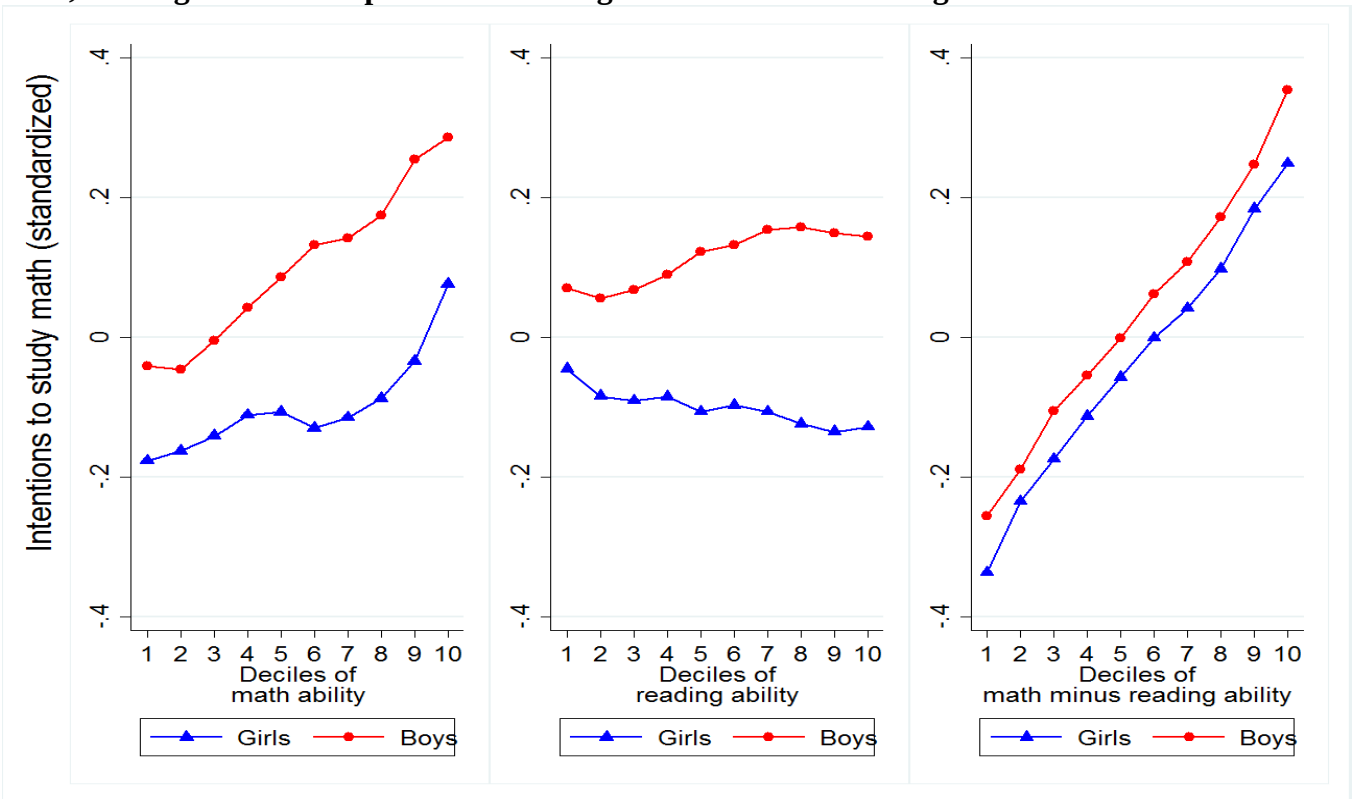


Figure 2: Math self-concept as a function of ability in math, reading and the comparative advantage in math versus reading

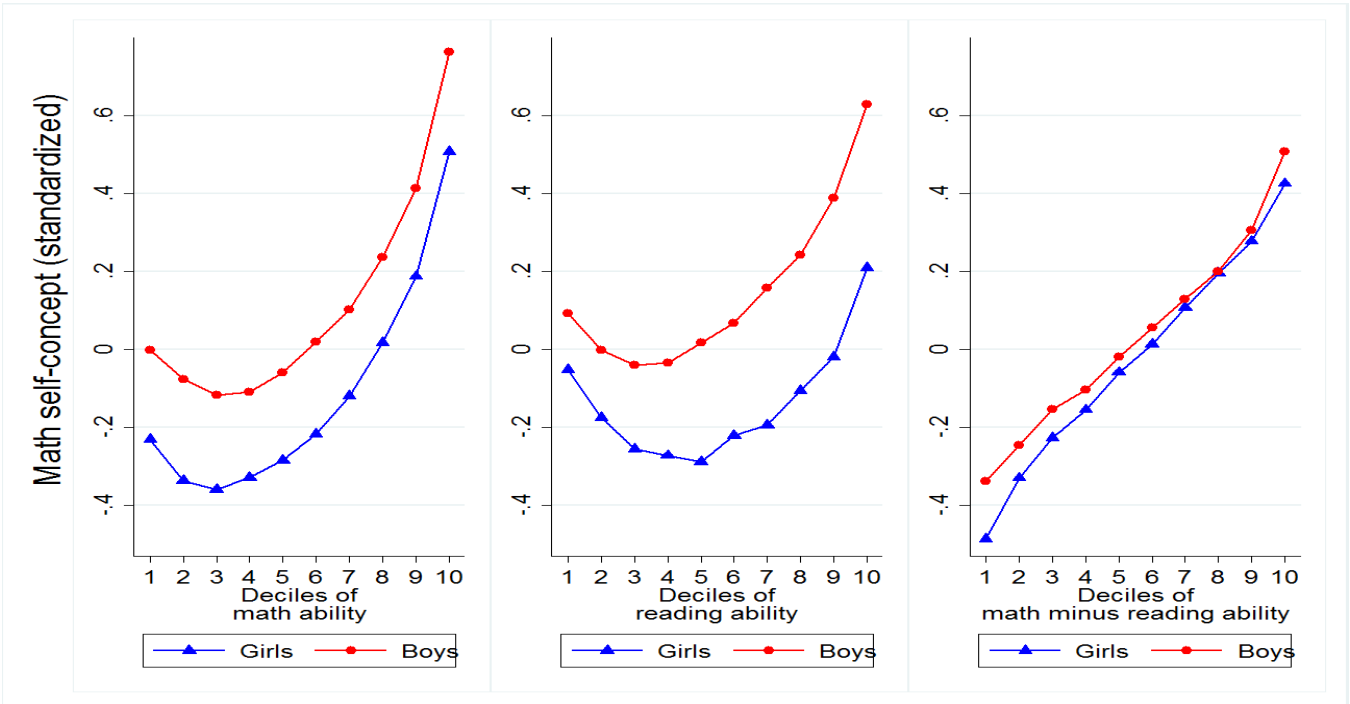


Table 1: Females comparative advantage in reading and the gender gap in intentions to pursue math-intensive studies and careers

	Gender gaps (girls minus boys, as a fraction of variable s.d.)				**Share of the gender gap in intentions to study math explained by ability in…**		
	math	reading	math minus reading	intentions to study math	math	reading	math minus reading
ALL countries	-0.136	0.351	-0.832	-0.218	0.074	-0.036	0.784
OECD countries	-0.159	0.318	-0.883	-0.258	0.064	-0.002	0.810
Non-OECD countries	-0.11	0.389	-0.775	-0.171	0.087	-0.101	0.768
Selected countries (with a gender gap in intentions to study math larger than 0.25 s.d.)							
United States	-0.111	0.276	-0.805	-0.292	0.024	0.069	0.881
United Kingdom	-0.124	0.276	-0.889	-0.250	-0.009	0.146	1.026
Canada	-0.159	0.343	-0.841	-0.485	0.031	0.009	0.348
Germany	-0.186	0.45	-1.238	-0.461	0.016	0.032	0.516
France	-0.186	0.323	-0.958	-0.393	0.027	0.02	0.429
Finland	-0.069	0.569	-1.096	-0.494	0.040	-0.138	0.741
Denmark	-0.202	0.336	-0.967	-0.336	0.046	-0.032	0.27
Brazil	-0.195	0.379	-0.914	-0.276	0.073	-0.002	0.625
Russia	-0.005	0.419	-0.659	-0.398	0.002	-0.060	0.419

Notes: All variables are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country. Intentions to study math is an index built from 5 questions. The total sample includes 301,360 students. See SI for details and statistical significance of each estimate.

Table 2: Comparing the explanatory power of the comparative advantage with that of other possible determinants of the gender gap in math-intensive fields

	Gender gap in each variable (fraction of a s.d.)				Intentions to study math (Standardized)		
	Absolute	Conditional on math ability	Conditional on MR	Gender gap in MR conditional on the variable	Association with each variable	Gender gap conditional on each variable	Gender gap conditional on each variable plus MR
Variable (standardized):						*(raw gap is -0.218 s.d.)*	
Math minus reading ability (**MR**)	-0.832	*n.a.*	*n.a.*	*n.a.*	0.215	-0.047	*n.a.*
Math self-concept	-0.270	-0.231	-0.012	-0.780	0.372	-0.132	-0.033
Declared interest for math	-0.174	-0.160	-0.003	-0.802	0.396	-0.150	-0.046
Instrumental motivation for math	-0.104	-0.088	0.007	-0.820	0.352	-0.182	-0.049
Math anxiety (opposite of)	-0.174	-0.129	-0.020	-0.824	0.228	-0.192	-0.033
Math involvement	-0.293	-0.288	-0.150	-0.789	0.227	-0.155	-0.018
Math environment	-0.096	-0.101	-0.034	-0.827	0.174	-0.202	-0.041

Notes: All variables are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country. See SI for details on the construction of variables and statistical significance of each estimate.

PNAS
www.pnas.org

Supplementary Information for

Girls' Comparative Advantage in Reading can Largely Account for the Gender Gap in Math-Intensive Fields

Thomas Breda, Clotilde Napp

Corresponding author: Thomas Breda
Email: thomas.breda@ens.fr

This PDF file includes:

Supplementary text
Figs. S1 to S3
Tables S1 to S9
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Link to data and programs available on Thomas Breda webpage.

Table of contents

Materials and Methods

The 2012 PISA survey

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 PISA 2012 survey. The student data set in 2012 contains around 510,000 observations, which roughly represent a population of 28 million 15-year-olds attending seventh grade or above in 64 countries, including the 34 countries belonging to the Organisation for Economic Co-operation and Development (OECD) in 2012.

PISA surveys systematically assess students' performance and knowledge in three core subjects: mathematics, reading and science. However, one of the three core subjects is chosen to be covered in greater depth in each survey. In 2012, mathematics literacy is the major subject area, as it was in 2003. This allows us to get more in-depth information on the students' mathematics skills. On top of taking math tests, students fill out a background questionnaire that provides contextual information about themselves, their homes, and their school and learning experiences. The background questionnaire takes 30 minutes to complete and seeks information about students' engagement with and at school in general and engagement with mathematics in particular. It includes questions on students' motivation to succeed in mathematics, the beliefs they hold about themselves as mathematics learners, and their dispositions and behaviors in math-related fields. Of particular interest to us are questions about students' intentions to pursue math-related studies and careers as well as questions about their self-concept, their anxiety in mathematics, their instrumental or intrinsic motivation for math, their math behaviors, and about subjective norms in mathematics (see details below).

The student questionnaire has a rotation design, which means that all students do not answer the same set of questions. The rotated design is such that there are three different forms of the questionnaire, each containing a common part and a rotated part. The common part (which is administered to all students) contains questions about gender, language at home, migrant background, home possessions, parental occupation and education. The rotated part (which is administered to one-third of students) contains questions about attitudinal and other non-cognitive constructs. The rotation design is such that constructs are asked in two out of the

three different forms of the questionnaire to allow joint analyses of these constructs. This results in responses from two thirds of students per construct.

Data restrictions

PISA treats Florida, Connecticut and Massachusetts separately from the rest of the United-States because they have a decentralized (specific) management of the PISA survey. This is also the case for the Perm region of the Russian Federation. We have integrated these states/regions to the country they belong to. PISA also considers separately Chinese provinces and cities that claim their independence from China or have strong cultural specificities: Shanghai, Macao, Taiwan and Hong Kong. We have not grouped these cities/regions in a single "China" sample as such grouping may hide strong cultural differences. This leaves us with 64 countries, 34 of them belonging to the OECD

Our main variable of interest (students' intentions to pursue math-related studies and careers) is obtained through a series of five items in a questionnaire only administered to two thirds of PISA participating students (selected randomly). As a consequence, our main sample includes 301,360 observations (out of the 510,000 initial ones) with available information on intentions to study math as well as information on abilities in math, reading, and science.

Variables of interest in PISA 2012

Intentions to study math

PISA 2012 asked students to report their intentions to use mathematics in their future studies and careers. Five items (ST48) used the forced-choice format to measure students' plans regarding mathematics at some stage in the future. The first item had students decide between taking additional courses in mathematics or the language of the test after school finished. The second item asked whether students planned on majoring in a subject requiring mathematics or science at college (or equivalent educational institution in different countries). The third item asked whether students were willing to study harder in either their mathematics classes or in classes teaching the language of the test. For the fourth item, respondents had to indicate whether they were planning on taking as many mathematics or science classes as possible during their education. The fifth and last item of that battery required respondents to choose whether they were planning on pursuing a career that involved a lot of mathematics or science. Students' responses to these five items are used to create the index (MATINTFC), reflecting the extent to which a student intends to pursue math studies and careers.

All items were reversed so that respondents who chose mathematics over either science or the test language were allocated a higher code.

As underlined above, two thirds of students respond to questions in the rotated part of the student background questionnaire, which results in 301,360 observations about intentions to study math.

Difference in performance between math and reading

We study the extent to which the difference between objective measures of proficiency in math and reading can explain the gender gap in intentions to study math. To this aim, we use individual-level PISA scores in these fields. These scores are on a 0-1000 scale. They have been scaled during the first PISA survey in 2000 to have a mean of 500 and a standard deviation of 100. In robustness analyses, we also use students' scores in science that have been scaled similarly. We re-standardize them to get statistics that can be easily interpreted country by country and on the whole PISA 2012 sample (see *methods* section). Importantly, the scores in math, reading and science are imputed based on seven math tests, three science tests, and three reading tests and other students' characteristics. Students only take a subset of 4 tests out of the 13 possible ones. Allocation of students to tests is based on a rotation design by clusters of students. For example, some of these clusters of students do not take any test in reading or in science, in which case their score in these domains are entirely imputed from their individual characteristics and on the scores obtained in reading by other students that have similar characteristics and that have performed similarly to them in non-reading tests. More generally, one should keep in mind that the PISA survey design implies that individual-level measures of performance are very noisy. This does not prevent researchers to make valid statistical inference on the underlying true individual-level abilities of students, but specific techniques need to be used to do so (see *methods* section).

Self-concept

PISA 2012 measures students' self-concept by using students' responses as to whether they strongly agree, agree, disagree or strongly disagree that (i) they are just not good in mathematics; (ii) they get good grades in mathematics; (iii) they learn mathematics quickly; (iv) they have always believed that mathematics is one of their best subjects; (v) they understand even the most complex concepts in mathematics class.

Students' responses are used to create the index of mathematics self-concept which was standardized to have a mean of 0 and a standard deviation of 1 across OECD countries.

Our sample includes 314,607 observations about self-concept, and since self-concept and intentions in mathematics are not in the same question set, the number of students responding to both intentions and self-concept questions is reduced to 148,490.

Other variables

Intrinsic motivation or declared interest for math refers to the drive to perform an activity purely for the joy gained from the activity itself. Students are intrinsically motivated to learn mathematics when they want to do so because they find learning mathematics interesting and enjoyable and because it gives them pleasure, not because of what they will be able to achieve upon mastering mathematical concepts and solving math problems. PISA 2012 measures students' intrinsic motivation to learn mathematics through students' responses as to whether they strongly agree, agree, disagree or strongly disagree that (i) they enjoy reading about mathematics; (ii) they look forward to mathematics lessons; (iii) they do mathematics because they enjoy it and (iv) they are interested in the things they learn in mathematics.

Students' responses are used to create the index of intrinsic motivation for math which was standardized to have a mean of 0 and a standard deviation of 1 across OECD countries.

Our sample includes 316,708 observations about intrinsic motivation, and 299,950 observations about both intrinsic motivation and intentions in math.

Instrumental motivation to learn mathematics refers to the drive to learn mathematics because students perceive it as useful to them and to their future studies and careers. PISA 2012 measures the extent to which students feel that mathematics is relevant to their own lives through students' responses as to whether they strongly agree, agree, disagree or strongly disagree that making an effort in mathematics is worth it because it will help them in the work that they want to do later on ; because learning mathematics can improve their career prospects and chances ; because they need mathematics for what they want to study later on ; and because learning many things in mathematics will help them get a job.

Our sample includes 316,322 observations about instrumental motivation, and 299,778 observations about both instrumental motivation and intentions in math.

Mathematics Anxiety refers to the feeling of helplessness and stress when dealing with mathematics. Students responses about their feelings of stress associated with anticipating math tasks, anticipating their math performance and while attempting to solve math problems were used to identify students' specific level of anxiety toward math and to construct the index of mathematics anxiety. More precisely, PISA 2012 asked students to report whether they agree or strongly agree that they often worry that mathematics classes will be difficult for them; that they get very tense when they have to do mathematics homework; that they get very nervous doing mathematics problems; that they feel helpless when doing a mathematics problem; and that they worry that they will get poor grades in mathematics.

Our sample includes 316,322 observations about math anxiety, and 299,778 observations about both math anxiety and intentions in math.

Students engagement in mathematics activities at and outside school or math involvement. PISA 2012 asks students to report how often they participate in mathematics-related activities at or outside of school. Mathematics-related activities considered in PISA2012 include: (i) talking about mathematics problems with friends; (ii) helping friends with mathematics; (iii) doing mathematics as an extracurricular activity; (iv) taking part in mathematics competitions; (v) doing mathematics for more than two hours a day outside of school; (vi) playing chess; (vii) programming computers; and (viii) participating in a mathematics club. Students can report engaging always or almost always, often, sometimes or never or rarely.

Our sample includes 313,847 observations about math activities, and 300,193 observations about both math activities and intentions in math.

Subjective norms in math or math environment. Pisa 2012 asks students to report on how people who are important to them, such as their parents and friends, view mathematics. Specifically, students are asked to report whether they strongly agree, agree, disagree or strongly disagree that (i) most of their friends do well in mathematics, (ii) most of their friends work hard in mathematics, (iii) their friends enjoy taking mathematics tests, (iv) their parents believe it is important for the student's career, or (v) their parents like mathematics. Students' responses are used to create the index of subjective norms in mathematics (SUBNORM), reflecting the extent to which a student's social environment promotes mathematics and the study of mathematics.

Our sample includes 313,847 observations about subjective norms, and 300,193 observations about both subjective norms and intentions in math.

Methods

Weights and representativeness

PISA provides weights to make surveyed students representative of the 28 million 15-year-old students of the surveyed countries. We use these weights in all country-specific and worldwide analyzes, so all the results we provide are not subject to sample selection and consistent estimates of the underlying parameters at the country and world levels. Note that the sum of PISA weights by country is proportional to countries' total population, so that more populated countries are given more weights in the worldwide statistics.[1]

Standardized variables

To get results that can be interpreted both worldwide and country-by-country, we normalize most variables of interest, so that their weighted mean is zero in each country, while their weighted standard deviation is one in each country. This transformation implies that variables have the same mean and standard deviation in each country and makes it easy to obtain gender gaps in the variables of interest that are directly expressed as a fraction of variables' standard deviation and, as such, directly comparable across countries. The transformation is done separately for each plausible value of math, reading or science ability, for each plausible value of the gap between math and reading ability, and for each variable capturing intentions to study math, self-concept in math or other dimensions of interest. Dummy variables for which gender gaps can be simply expressed in percentage points difference in the rate of positive answers are the only ones that are not systematically standardized.

We also provide results based on variables standardized at the world-level, so that differences in means and standard deviation across countries are conserved. In this latter case, we apply an affine transformation of each variable so that the weighted mean and standard deviation of the new variable are 0 and 1 worldwide. Results at the world level are similar when we adopt this alternative approach.

General estimation procedure in PISA

Details about PISA methodology can be found in PISA Technical reports (see (1) for 2012) 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 2012) 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

[1] An alternative approach is to use "senate weights" which give an equal weight to each country. We have not used such weights.

weighting with the reported students' weights) and the final estimate is the arithmetic average of the five 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.

Statistical inference

Sources of uncertainty in PISA are twofold. First, as explained above, there is some uncertainty on the ability measure of each student and PISA provides five 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 of the five plausible values, weighting it first by the "true" set of individual weights and then by the 80 alternative sets of weights. The correct point estimate is the average of the 5 regressions ran with the "true" set of weights, while the standard error is computed according to a formula that sums both of the measurement errors described above (see (2) for all details regarding those bootstrap techniques).

All our results (from simple means to more complex regressions) involving individual-level measures of abilities are produced using the procedure described above.[2] In particular, results involving measures of the relative advantage for math over reading at the individual level require that we construct plausible values for this relative advantage. We do so by taking the difference between the plausible values in math and reading, from the first to the fifth, hence generating five new plausible values for "math minus reading" ability. This approach works because plausible values offered by PISA had been jointly drawn, so that they capture the intra-individual correlations between abilities in different subjects.[3]

Objects of interest and estimation approach

The main goal of this research is to study gender gaps in intentions to pursue math-related studies and careers conditional on students' comparative advantage for math versus reading, or other measures of students' abilities. This approach is well illustrated in Figure 1, which shows how girls' and boys' intentions evolve with math, reading or math minus reading ability.

Formally we estimate $E[Intentions|Girl, MR] - E[Intentions|Boy, MR]$ through linear regression models of students' intentions on an indicator variable for students' sex and a control for the difference between a student math and reading abilities (MR):

$$Intentions_i = \alpha Girl_i + \beta MR_i + \varepsilon_i \quad (1)$$

[2] The only exception is the IV regressions which do not always use the procedure (see explanations below). We use the Stata command "Repest" which applies this correct procedure to produce estimates and confidence intervals.

[3] We thank Matthias von Davier for suggesting alternative approaches (leading to similar results) and confirming that this one was correct.

MR_i may be replaced by a higher-order polynomial of MR_i to better control for possibly non-linear effects of the comparative advantage. It may also be replaced by a scalar or vector of ability characteristics $Ability_i$. The coefficient of interest is α, and we are interested in comparing how α varies across specifications that control or do not control for MR_i and other measures of ability. We report the estimates α and their standard errors for these different specifications.

Alternatively, we compute the share of the raw gender gap in intentions to pursue math-related studies or careers that can be accounted for by MR_i (or other measures of ability such as math ability or reading ability considered on their own). This is done in Tables 1 and S1. This share is computed as one minus the ratio of α obtained from equation (1) above with a single control for math minus reading ability and the unconditional gender gap. Inference on this ratio is provided for analyses based on the whole PISA2012 sample and on OECD and non-OECD countries (see supplementary text). It is done by standard bootstrap, meaning that we measure the ratio from the two aforementioned regressions (using estimation techniques adapted to PISA, see above) on 120 random subsamples of the initial sample and use the distribution of the ratios computed from these estimates to compute 95% confidence intervals.

Interpretation and causal approach

Our estimates of the share of the gender gap in intentions that can be accounted for by various measures of students' ability should not be interpreted as a causal effect of students' abilities on the gender gap in intentions. These estimates are descriptive. They only allow us to capture the actual gender gap in intentions among students having similar abilities (as shown in Figure 1 for example). However, the gender gaps conditional on abilities that we estimate are not directly informative on the gender gap that would prevail if the distributions of abilities were identical among girls and boys. This is because the factors that explain gender differences in abilities may also explain gender differences in intentions to study math. Changing the distributions of abilities without modifying these factors may therefore leave the gender gap in intentions unchanged.

To discuss the extent to which our results may be interpreted as causal, we first rely on theoretical arguments and the existing literature on the determinants of educational choices. Together they provide and validate mechanisms that make plausible the fact that affecting students' relative abilities in math and reading will directly affect their educational choices. We then turn to an instrumental variable approach to provide a more direct validation. These two (very different) types of validation are presented below in turn.

Theory and literature on educational choices

Abundant literature comforts the idea that students rely heavily on their grades and relative abilities across subjects to build their self-concept (as well as their interest for math, instrumental motivation, etc.) and make educational choices (see main text, first page). According to Eccles and colleagues' (1983) Expectancy/value (EV) model (3), educational and occupational choices are assumed to be influenced by the intraindividual hierarchy of expectations of success and values (interest, intrinsic motivation, etc.). The model predicts that the critical comparisons are not comparisons within domain across individuals but domain comparisons within individuals. Expectancy/value theory moreover suggests that

interest is influenced by the belief that one can succeed in a given field. Educational choices result from individuals' rank orderings for their ability self-concepts across an array of domains.

According to Marsh's (1986) Internal/External (I/E) frame of reference model (4), math self-concept results from both math achievement and the discrepancy between the math and verbal achievement scores. This model permits in particular to explain why math and verbal self-concepts are nearly uncorrelated, although math and verbal achievements are strongly correlated. Marsh and colleagues demonstrate that even bright students may have an average or below average self-concept in their weakest school subject that may seem paradoxical in relation to their good achievement. Although Marsh's model is a model of self-concept formation, the model I/E suggests that high verbal achievement and self-concept may lead to a much lower likelihood of individuals entering science majors, even when they have the requisite math ability to do so.

Nagy and colleagues (2008) integrate Marsh's I/E frame of reference model and Eccles and colleagues' E/V theory. In their model (5), educational and occupational choices result from intraindividual comparisons of achievement in the different domains, leading to comparisons of self-concepts, and comparisons of subjective task values (e.g., interest, instrumental motivation).

In all these models, the effect of the comparative advantage on self-concept, values (like interest) and, in turn, educational or occupational choices is causal. This existing research in psychology on educational choices therefore supports a causal interpretation of our results: if students rely on their relative abilities to make educational choices, changing exogenously these abilities should also affect students' choices. It makes plausible the idea that gender differences in relative abilities in math versus reading can explain in a causal rather than purely statistical sense the gender gap in intentions to study math.

Instrumental variable approach

To assess causality more directly, we show that school environments that are more favorable to learning math than reading or other disciplines increase students' ability gap between math and reading, and in turn strengthen their intentions to study math rather than reading or science.

The idea is to use as instrumental variables school characteristics that are conductive to a better student performance in math than reading (or vice versa). We construct these variables using the PISA school-level questionnaire. This questionnaire provides information on school resources and school quality in general, but also on resources devoted to math in particular (as the focus of PISA2012 was on math). We know for example the existence of a shortage of math or literature teachers in the school, the number of part-time and full-time teachers, and of part-time and full-time math teachers, the share of each type of teachers that is qualified, the average number of students per teacher in the school, the average number of students per math teacher in the school, etc.

From the available information, we build three main instruments: (1) the existence of a shortage of math teachers minus the existence of a shortage of literature teachers, (2) the share of math teachers among the teaching staff, and (3) the share of math teachers that are

qualified minus the share of total teaching staff that is qualified (meaning that has at least a bachelor's degree). Instrument (1) is a simple difference between two direct questions asked in the survey. Instruments (2) and (3) exist separately for part-time and full-time teachers. Our instruments aggregate part-time and full-time teachers by computing the total number of teachers in each relevant group as the sum of full-time teachers and one half times the number of part-time teachers. Alternatively, we have included two instruments for (2) and two instruments for (3) which are built using separately the information on part-time or full-time teachers (leaving us with 5 instruments in total). Results are similar, but some of the instruments tend to be less significant in the first stages, making the results harder to read. Of course, we could have built other instruments based on the same logic, or have constructed the instruments differently. Here again, we have checked that the results are robust to alternative choices (see Supplementary Text).

The logic of our instruments is to capture schooling environments that are more favorable to learning math than reading, while not capturing variations in school quality. We do so because better students usually go into better schools and these better students might be better in math than reading because of their personal characteristics rather than an exogenous variation in their schooling environment. Hence, it is important to hold school quality constant as much as possible. We do so in three ways. First, we systematically construct instruments that capture the available resources of a school for learning math relative to the available resources to learn other subjects. Such instruments are well-suited to explain the difference in ability between math and reading (that does not only depend on math ability by definition) and are also less likely to capture directly school quality. Second we directly control for measures of school quality provided by PISA. Third, we show that the results are robust to restricting the sample to schools that mostly admit students based on their geographic location. Such schools are, to some extent, less likely to have students that strategically selected the school as a response to a shortage of resources to learn math.

We use two stage least square estimators (TSLS) and we systematically provide Montiel-Pfueger robust F statistics to test for the weakness of the instruments[4], and the p-value of a Sargan Khi2 over-identification test.

Finally, we do not use weights in the IV regressions, and we do not account for uncertainty in the individual-level measures of abilities using systematically all plausible values provided by PISA. In practice, we compute a single measure of math minus reading ability as the difference between the first plausible value for math ability and the first plausible value for reading ability. These two choices (not weighting and using one plausible value) are done for both theoretical and practical reasons. On the theoretical ground, the use of weights is of course essential when one wants to provide statistics that are representative of the whole population, which we do in the rest of the paper. Weights are however not required for causal inference (see Solon, Haider and Wooldridge, 2015 (6), for a discussion of the arguments for and against weighting in regressions) and can therefore perfectly be omitted in IV regressions. We do so because the two (indispensable) tests we provide for weak instruments and over-identification cannot easily be implemented on weighting regressions, meaning that we are only able to report the tests for non-weighted estimates. It is therefore more consistent to also report the non-weighted estimates that correspond to the tests. We of course checked that the results were robust to the use of weights. Regarding the use of only one plausible value instead of all of them, this is both to save computational time and because the correction of

[4] This approach for testing the weakness of the instruments follows the advice of James Stock at the 2018 NBER Summer Institute lecture. See http://www.nber.org/econometrics_minicourse_2018/2018si_methods.pdf

standard errors for individual-level uncertainty rely on averaging results over a series of 80 weighted IV regressions for which we cannot provide tests of weak instruments or over-identification. Here again, we have checked that the results are still statistically significant when we do these corrections[5].

Supplementary Text

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

Description of Table S1 and Figure S1 (main results for all countries)

Table S1a extends Table 1 to all countries that participated in PISA. It presents the gender gap in math, in reading and in the comparative advantage MR, as well as the gender gap in intentions to pursue math studies or careers, and the share of the gender gap in intentions explained by ability in math, reading, and MR. For all countries, all OECD countries, all non-OECD countries, and by country.

Each variable is standardized to have a weighted mean equal to 0 and a weighted standard error equal to 1 in each country, as described above. We also present the results of a "basic standardization" which is a more simple linear transformation of the difference between math and reading abilities, so that the resulting variable has a weighted mean equal to 0 and a weighted standard error equal to 1 on the whole sample (only).

Worldwide, the gender gap in reading is three times larger than the gender in math and the gender gap in MR is mostly due to the gender gap in reading. The gender gap in MR is very high (0.83 s.d. worldwide, 0.88 s.d. among OECD countries). This is a common feature of all countries to have a large gender gap in MR: it is larger than 0.75 s.d. in all OECD countries and larger than 0.6 s.d. everywhere, except in Singapore, where it is equal to 0.595 s.d.

The gender gap in math intentions amounts to 22% of a s.d. worldwide. This gap varies across countries. 2 OECD (5 non-OECD) countries have no gap or even a small gap in favor of girls (e.g., Turkey, Malaysia or Thailand, see Table S1). 5 OECD (11 non-OECD) countries have a small-to-medium gap (between 10 and 20% of a s.d.). 17 OECD (13 non-OECD) countries have large gaps (between 20% and 45% of a s.d.). Finally, 10 OECD (1 non-OECD) countries have gaps larger than 45% of a s.d. (e.g., Australia, Germany, Finland).

Gender gaps in intentions to study math (as well as gender gaps in MR) are higher among OECD countries (0.26 s.d.) than among non-OECD countries (0.17 s.d.). Among OECD countries, the smallest gender gap in intentions is in Turkey (the only OECD country with a gap in intentions favoring girls). These observations are consistent with previous literature showing that sex segregation by fields of study is higher in more equal and developed countries (7). Concerning gender equality, Scandinavian countries, although known to be particularly gender equal, have large gender gaps in intentions to study math: Finland (0.5 s.d.), Norway (0.3 s.d.), Sweden (0.3 s.d.), Iceland (0.2 s.d.). This latter finding is consistent

[5] As instrumentation also allows one to get rid of classical measurement error bias, correcting for individual-level uncertainty in IV regressions may be less necessary than for simple OLS.

with the gender equality paradox of (8), according to which more gender equal countries are more segregated.

The absence of a significant gender gap in intentions in the Netherlands is unexpected. It is not in line with the actual segregation by fields of study. The gender gaps in the Netherlands in other measures of math attitudes, like self-concept or interest in math are higher than average (0.41 s.d. and 0.24 s.d., see Table 2 for comparison). Moreover, if we consider another possible measure of intention to pursue math studies that is not related to a specific choice of courses, like item ST29Q07, that measures how much the student considers math as important for future study (see the comments on Table S2), the gender gap in the Netherlands is twice as large as the average gender gap in OECD countries (0.32 s.d. vs 0.16 s.d. for OECD countries and 0.11 s.d. worldwide). This suggests that the result on math intentions may be due to a specificity of the educational system in the Netherlands, e.g., the combinations of certain subjects being compulsory.

When controlling for math ability, the gender gap in intentions is reduced by less than 10% worldwide and in a majority of the studied countries. Similarly, controlling for reading ability barely affects the gender gap in intentions.

In contrast, the gender gap in intentions disappears almost entirely when one controls for MR. MR can explain 78% of the gender gap in intentions worldwide (81% for OECD countries, 77% for non-OECD countries). The corresponding 95% confidence intervals estimated by bootstrap are [0.71,0.83] for the worldwide estimate, [0.72,0.87] for OECD countries only, [0.62,0.86] for non-OECD countries only (see Table S1b for inference results on each specific country).

The share of the gap in intentions explained by MR varies across countries. For instance, among OECD countries, the share explained is equal to 81% on average but varies from 25% for Switzerland to 231% for Korea (we do not take into account Turkey and the Netherlands that encounter no gap or a positive gap in intentions). It is equal to 88% for the U.S., 103% for the United Kingdom, 61% for Norway, 52% for Germany, and 43% for France.

Figures S1 illustrate the latter results and represent for each country participating in PISA 2012, the (unconditional) gender gap in intentions to pursue math studies and careers as well as this gender gap, controlling for math ability, reading ability, and math minus reading ability (comparative advantage). For OECD (Figure S1a) and non-OECD countries (Figure S1b). These figures illustrate the explanatory power of the comparative advantage.

Table S1b provides the standard errors and the 95% confidence intervals for the gender gap in intentions to study math, unconditional as well as conditional on math, reading or MR ability. Figures S1a and S1b provide a visual representation of the same statistics.

Description of Table S2 (intentions to study math captured with alternative variables)

Table S2 shows that our main analysis remains valid when considering three alternative measures of intentions. Results are presented for all countries, for OECD countries, for non-OECD countries, as well as for three specific countries (United States, United Kingdom and France).

Tables S2A and S2B deal with measures of intentions capturing more specifically the arbitrage between math and reading. They rely on two of the five items considered in the index of math intentions (MATINTFC). Table S2A relies on item ST48Q01, asking students if they are willing to take additional math versus reading courses after school finishes and Table S2b relies on item ST48Q03, asking students if they are willing to study harder in math versus reading courses. We measure the gender gap in intentions by the difference between the percentage of boys and the percentage of girls willing to take additional math courses/willing to study harder in math. There are for instance on average 8% fewer girls than boys willing to take additional math courses after school finishes. When standardizing by country, this gap amounts to 0.16 s.d.

For these two measures of intentions to favor math relative to reading in future study, Tables S2A and S2B provide not only the gender gap, but also the share of this gap explained by math, reading and MR ability. We retrieve the same patterns as in Tables 1 and S1. More precisely, the gender gap in intentions for both measures is barely affected by math or reading ability alone, but it almost entirely disappears when one controls for the comparative advantage MR. The reduction of the gender gap when controlling for MR is equal to 84% for the first measure and 103% for the second. These reductions are slightly stronger than for the index, which is understandable since these measures are specifically related to the arbitrage between math and reading, hence more likely to be impacted by the comparative advantage in math relative to reading.

As already observed with the measure of intentions based on the index (see the comments on Table S1), the gender gap in intentions is higher for OECD countries than for developing countries.

We consider in Table S2C an alternative measure of the intention to study math that is not related to the ipsative choice of courses, but relies on student's answers as to whether they strongly agree, agree, disagree or strongly disagree that math is important for their future study (item ST29Q07). We consider the gender difference in the standardized item. We retrieve the same patterns as with the previously studied measures of intentions, i.e., a gender gap in intentions (0.1 s.d. worldwide), that is higher among OECD than non-OECD countries (0.16 s.d. vs. 0.05), that is barely affected by math ability (reduction of 12%) or reading ability (no reduction) but disappears or is strongly reduced when controlled for comparative advantage (reduction of 107%).

Description of Table S3 (relationships between abilities and intentions to study math for all countries)

Table S3 presents the partial effect of math ability, reading ability and the comparative advantage MR on intentions to pursue math studies and careers, for all students.
The effect of MR is about twice stronger than the effect of math ability (0.25 vs. 0.11, for OECD countries), and reading ability has almost no effect.

If we perform the same regressions by gender, we obtain that the partial effect of the comparative advantage is the same for boys and for girls (0.205 for girls and 0.206 for boys), as illustrated by Figure 1.

In contrast, the link between math ability and intentions to pursue math studies is higher for boys than for girls (0.132 vs. 0.106), as can also be seen on Figure 1. The same result holds for the measure of math intentions related to the specific question asking students if they are willing to take additional math versus reading courses after school finishes. This better self-selection of boys in math-related fields implies that the gender gap in math intentions is on average larger in top half of the math ability distribution than in the bottom half, as shown in Table S5 and illustrated in Figure 1. This pattern is also likely to lead to a larger gender gap in math performance among students who choose math-related fields of study. To show this more directly, we compute the gender gap in math performance among students that are willing and not willing to study math further. Consistent with the stronger selection of boys in math-related fields based on their math performance, we find that this gender gap is significantly higher among students that are willing to take additional math courses than among those that are willing to take additional reading courses after school finishes (0.15 vs. 0.05). Moreover, girls' representation among high achieving students is lower for students that are willing to take additional math courses than in the whole sample of students. For instance, the percentage of girls among the students scoring at or above Level 4 in math is 46% in the whole sample but only 42% among those that are willing to take additional math courses.

On the contrary, girls self-select better in « humanities ». Indeed, the gender gap in reading performance is higher among the students who 'choose humanities' (i.e. who are willing to take additional reading rather than additional math courses after school finishes) than among the whole sample of students (0.42 vs. 0.33) and girls' representation among high achieving students in reading is higher after self-selection than before. The percentage of girls among the students scoring at or above Level 4 in reading is 59% before selection but amounts to 69% after selection.

The better self-selection of boys in math and of girls in reading, and their consequences in terms of gender gaps in performance after specialization has occurred can support the survival of the stereotype that math is not for girls and humanities are not for boys.

For the same reasons put forward in the comments of Table S1, the case of the Netherlands appears to be quite specific. To complete our analysis, we can consider, as above, the measure of intentions based on item ST29Q07 ('math is important for future study', see the comments of Table S2). For this measure of intentions, the partial effect of MR for the Netherlands is then comparable to the average effect among OECD countries (0.18 for the Netherlands and 0.17 for OECD countries) and it is stronger than the effect of math ability. This suggests that, as in the other countries, comparative advantage, unlike math ability alone, is likely to have an important explanatory power for understanding sex segregation by field of study/career in the Netherlands.

Description of Table S4 (intentions explained with various measures of ability)

Table S4 shows how the gender gap in intentions evolves (i) when we control for students' family resources and social background as well as country fixed effects and (ii) when we control for additional measures of abilities on top of MR.

The inclusion of individual-level controls and country fixed-effects barely affects the gender gap in intentions once MR is already controlled for (Table S4, column 2). Our individual-

level controls are usually indexes built and provided by PISA. They capture students' family wealth (WEALTH), their family cultural possessions (CULTPOS) and their economic, social and cultural status (ESCS). We also control for the type of study programme a student is enrolled in (ISCEDD) and two measures of misbehavior/absences (st08q01 and st09q01).

Similarly, controlling more flexibly by MR (5^{th} order polynomial) to allow for non-linearities in the effect of MR does not change its effect on the gender gap in intentions (column 3). When we control very flexibly for math and reading abilities on top of MR (column 4), the gender gap in intentions moves up slightly, from an initial -0.047 s.d. when only MR is controlled for, to -0.057 s.d.[6] When we remove controls for math and reading abilities and include instead a single control for science ability on top of MR, the gender gap in intentions is not modified (column 5). Finally, in columns (6) and (7), we try to fully control for abilities by including 5^{th} order polynomials in MR, math, reading and science abilities (column 6) as well as interactions between these variables (column 7). We see that the gender gap in intentions moves up to -0.083 s.d., corresponding to 38% of the raw gender gap in intentions.

Description of Table S5 (share of girls and gender gaps in intentions among subgroups of students)

This table presents the proportion of girls as well as the gender gap in intentions to pursue math studies and careers among different subgroups of performance.

To address the question of the possible explanation of the underrepresentation of girls in math studies by their overrepresentation among students performing well in both math and reading (9), we consider the subgroups of students performing at the top 10%, top 25% and top half in both math and reading. Women are indeed slightly overrepresented in these performance groups (about 53%) but contrary to predictions, the gender gap in intentions is higher than average in the last two groups of performance and only slightly below average in the first group (0.18 vs. 0.22). The fact that girls are less constrained in their choices than boys because of their ability in various domains cannot explain the gender gap in intentions to study math.

To address the question of the possible explanation of the underrepresentation of girls in math studies by their underrepresentation among top performers in math, we consider the subgroups of students performing at the top 10%, top 25% and top half in math. Women are indeed underrepresented in these performance groups (40 to 48% of all students) but the gender gap in intentions is around -0.23 s.d. which is slightly larger than the gap among all students. This is illustrated by Figure 1. The fact that girls are less represented among the students who are the most able to pursue math studies cannot explain the gender gap in intentions to study math.

In contrast, we observe that women are strongly underrepresented in the top 10%, top 25% and top half of MR (from 16% to 32%). It is worth noting that there are only 16% of girls in the top 10% of MR, which is a very small proportion. Moreover, the gender gap in intentions in these three groups is at least twice lower than the average gender gap in intentions.

[6] Math and reading abilities are not perfectly collinear with MR as they have been standardized separately. We include all these variables as controls to saturate the model as much as possible. However, not surprisingly, using only any two of these three variables leads to similar conclusions.

Similarly, women are strongly overrepresented in the bottom 10%, 25%, and bottom half of MR (from 68% to 85%) and the gender gap in intentions is also at least twice lower than average in these performance groups. These results are fully in line with our main analysis showing the power of the comparative advantage to explain the gender gap in intentions to study math.

Description of Table S6 (Effect of MR on both intentions and actual school enrolment in France)

Table S6 is based on an auxiliary data set that includes information on around 12,000 10^{th} grade students enrolled in French high schools of the Paris region. The data has been collected for another research project and detailed information on it can be found in (17). The data set includes ability measures in both reading and math, in the form of class grades as well as in the form of national exam scores, taking place at the end of 9^{th} grade. It also includes intentions to specialize in a scientific track (Première S, emphasizing math, physics and natural science) in 11^{th} grade, declared in 10^{th} grade, between January and March, as well as actual enrolments. Approximately 40% of the students enroll in the scientific track. The gender gaps in intentions and in enrolments are of the same magnitude. Correlation between intentions and actual enrolment is not perfect, but strong (78%). We have excluded repeaters as intentions to repeat are not asked for, implying that repeaters' intentions cannot be equal to their actual enrolment. This leaves us with a final sample of 11,659 students.[7]

The gender gaps in intentions and in actual enrolment are barely affected by the control for math ability, and controlling for reading ability alone increases the gender gaps. In contrast, the comparative advantage in math relative to reading (measured by the exam scores) reduces the gender gap in intentions by 46% and the gender gap in actual enrolment by 49%. The reductions associated with the comparative advantage measured by class grades are slightly higher. These figures are similar to those obtained in the study of PISA results for France.

The data set also includes measures of the gender gap in the intention to study science after high school finishes. Controlling for math or reading ability alone does not reduce the gender gap. The comparative advantage reduces this gender gap, but much less than for the intentions to enroll in the scientific track or for the actual enrolment in the scientific track.

Description of Table S7 (self-concept, attitudes towards math and math environment in various regions of the world) and Figures S2 and S3

Tables S7abcd complete and extend Table 2. They compare the explanatory power of the comparative advantage with that of other often considered determinants of the gender gap in math (self-concept in math, interest in math, instrumental motivation for math, math anxiety, math involvement and math environment) in various regions of the world: worldwide (Table S7a), among OECD countries (Table S7b), among non-OECD countries (Table S7c) and in three specific countries (United States, United Kingdom and France, Table S7d). Tables S7 also provide standard errors. The definition of the different variables as well as how they are measured are provided in the *Methods* section.

[7] Keeping repeaters does not change significantly the results.

Results are very consistent across regions.
The gender gaps in the various variables are quite large and of comparable size, those in self-concept and math involvement being in general slightly larger than the others (first column). The gaps are higher in OECD than in non-OECD countries. In particular, the gender gap in instrumental motivation is three times smaller in non-OECD countries than in OECD countries (but remains significant). This feature is consistent with previous literature underlining that the gender gaps in math attitudes are higher in more developed, gender equal and equal countries (10, 11). As seen above (see the comments on Table S1), sex segregation has also been shown to increase with countries' level of development and equality.

The six studied variables have no impact on the gender gap in MR (column 4). This is illustrated by Figure S3 that represents the comparative advantage in math versus reading as a function of math self-concept or math interest.
On the contrary, conditional on MR, the gender gaps in these variables almost always disappear. Worldwide, the gender gap in self-concept is reduced by 96%, interest for math by 98%, instrumental motivation by 107%. This is less true of the gender gap in math involvement, that is however reduced by at least 50% when controlled for MR.
As can be seen in column 2, the gender gaps in the six variables are reduced by around 10% only when controlling for math ability.

Figure S2 represents the interest in math as a function of ability in math, in reading and of the comparative advantage in math versus reading. It illustrates the fact that controlling for math or reading ability leaves a large gender gap whereas the gender gap in interest for math disappears when controlling for the comparative advantage.
It is interesting to note that interest for math first decreases then increases as a function of math or reading ability. A higher score in math for instance is not necessarily associated with a higher level of interest. In contrast, interest for math increases with MR, for boys and for girls, for all levels of MR. Moreover, the relation between interest for math and MR is the same for boys and for girls.

Of the six considered variables, declared interest for math has the highest association with the intentions to study math (for boys and girls) but self-concept has the highest explanatory power for the gender gap in intentions (due to its higher gender gap). This latter result is in line with previous literature emphasizing the role of the gender gap in self-concept to explain sex segregation by field of study.

MR has a much higher explanatory power than any of these six variables. Worldwide, it reduces the gender gap in intentions by 79%, compared to 39% for self-concept, 31% for interest for math, 17% for instrumental motivation, 12% for math anxiety, 29% for math involvement and 7% for math environment.
Controlling for the combination of MR and math involvement reduces even more the gender gap in intentions (by 92% worldwide). In non-OECD countries, it even becomes positive.

We can consider more closely the case of the United States. It illustrates all the just mentioned patterns. Gender gaps in the six variables lie between 0.08 s.d. (for instrumental motivation) and 0.25 s.d. (for math involvement). None of these gaps remains significant, when controlled for MR. None of the six variable has an impact on the gender gap in MR. The association with the intentions to study math varies between 0.21 for math involvement and 0.45 for declared interest for math. The largest reduction in the gender gap in intentions to

study math is obtained by controlling by the self-concept (28%). It is much smaller that the reduction due to MR alone (88%), or MR and math involvement (92%).

Description of Table S8 (cross-country correlations)

Table S8A presents the cross-country correlations between gender gaps in intentions to pursue math related studies and careers and actual gender gaps in fields of study. We consider the following indicators of actual gender segregation:

- *Sex segregation in humanities and social science* is taken from (7). (7) contrasts the female-to-male ratio in a respective field to that in the average field of study, and define the sex segregation in field j by $\ln(\mathrm{Fj}/\mathrm{Mj}) - \sum_1^J \ln(\mathrm{Fj}/\mathrm{Mj})$ where Fj and Mj are the number of women and men graduates and J is the total number of fields. (7) relies on UNESCO data on graduates in higher education.
- *Sex segregation in mathematics and natural science* is also taken from (7).
- *The percentage of women among STEM graduates in tertiary education*, taken from (8). (8) uses UNESCO graduation data from 2012 to 2015 in natural sciences, mathematics, statistics, information and communication technologies, engineering manufacturing and construction.
- *The gender gap in the expectation of a job involving math*, taken from the 2015 Trends in International Mathematics and Science Study (TIMSS), an every-four-year international assessment of the mathematics and science knowledge of 4th and 8th grade students. Our measure represents the gender gap in 8th-grade students answers in the TIMSS 2015 survey to the question: do you agree a lot/agree a little/disagree a little/disagree a lot with the following assertion 'I would like a job that involves using mathematics'?
- *The female-to-male ratio in tertiary education new entrants in computer science*, taken from (12).
- *The percentage of tertiary education graduates in engineering, sciences, technology, and math*, taken from UIS Unesco 2014-2015 (data.uis.unesco.org).

These various indicators differ by the specific field they consider. Some focus on a narrow STEM field like computer science or math only, others associate math with additional fields (natural science, or the whole STEM field) or consider "reading-related" fields (humanities or social sciences). They also differ by the way they measure sex segregation: the first two use "relative" female-to-male ratios, contrasting with the other measures using "absolute" ratios. Finally, the TIMSS indicator does not rely on actual segregation but on expectations from 8th grade students about their future job. These differences may explain why the coefficients in the matrix of correlations are not always significant.

Although different, these indicators however all measure some type of sex segregation in math- or reading- related fields. Importantly for our analysis, we find that the gender gap in intentions to pursue math studies or careers, as measured in PISA, is significantly correlated with all these measures of actual sex segregation. This is reassuring as for the generalizability of our results to actual students' school and career decisions.

Table S8B presents the cross-country correlations between gender gaps in comparative advantage to pursue math studies and the following indicators of gender norms:

- *Gender gap index*, published by the WEF, taken from the Global Gender Gap Report 2012.
- *Index of country subjective norms* (PISA), taken from (13, III.4.7.a). It is given by the effect size (i.e., the gender gap divided by the standard deviation) of the PISA index of subjective norms (SUBNORM) defined above.
- *Gender-science IAT*, data taken from (14). The Gender-science *Implicit Association Test* (IAT) is a measure of implicit gender-science stereotype relying on the Implicit Association Test (15). This measure relies on the differential ease to categorize items depending on whether they associate male and science (and female and liberal arts) or female and science (and male and liberal arts). Most people are able to categorize the items faster and more accurately in the former condition (male=science) compared with the latter (female=science), which is taken to reflect stronger associations of science with male than female and an implicit gender science stereotype. Considering the mean of IAT results for all participants in a given country provides a measure of science-gender stereotype at the country level. Note that the IAT measure relies on self-selected samples of individuals taking the test online. Following Miller, Eagly and Linn, 2015 (14), who worked in detail with these country measures of IAT, we only consider countries for which more than 50 observations are available and we omit Romania.

The index of subjective norms and the IAT-measure of the gender-science stereotype are strongly correlated (rho=0.52)[8], which tends to indicate that they both reflect gender norms concerning math or science. The Gender Gap Index is not related to these measures of gender-math stereotypes.

Importantly for our concerns, the gender gap in comparative advantage MR is significantly correlated with both measures of gender norms (rho=0.5 for the index of subjective norms and rho=0.27 for the gender-science IAT[9]). This suggests that cultural norms associating math with boys and humanities with girls may have a link with the gender gap in comparative advantage. This is in line with the expectancy-value model of Eccles and coauthors in which cultural milieu, and cultural norms and stereotypes surrounding gender have an impact on boys' and girls' involvements and achievements in the various domains.

Table S8C presents the cross-country correlations between gender gaps in comparative advantage and the following characteristics of educational systems:

- The *index of between-school horizontal stratification* is provided by PISA 2012. It is based on five interrelated indicators of horizontal stratification between schools : number of educational tracks, prevalence of vocational and pre-vocational programs, early selection, academic selectivity and school transfer rates.

- The use of *mandatory standardized tests*. PISA 2015 indicates the percentage of students who are assessed at least once a year with mandatory standardized tests.

The gender gap in MR at 15 years old tends to be larger in educational systems in which horizontal stratification is higher (especially among OECD countries only) or in which students are less assessed through mandatory standardized tests.

[8] rho=0.46 if we re-integrate Romania.
[9] The corresponding statistics is 0.23 if we include Romania.

Description of Table S9 (Instrumental variables)

Table S9 shows a large impact of MR on intentions to study math. This is true for both girls and boys, and robust to the inclusion of controls for students background and school quality and resources. More specifically, a one s.d. increase in MR increases girls' (boys') intentions to study math by 70% (90 to 100%) of a s.d (columns 4 to 7). When girls and boys are pulled together, the estimated impact of MR on intentions is found to be a bit lower (around 60% of a s.d.). We can observe that this impact is precisely estimated and significantly larger than the OLS association between standardized MR and standardized intentions to study math (which is 0.215 worldwide - see Table S3 - and 0.23 on the exact same sample as that for column 1 of Table S9).

The inclusion of controls for students' characteristics (social background, parental wealth) and school characteristics (class size, percentage of girls in the school, index of school responsibility for resource allocation, index of school responsibility for curriculum and assessment, number of computers per student, share of computers connected to the web, shortage of teaching staff) has little effect on the estimated impact of MR.

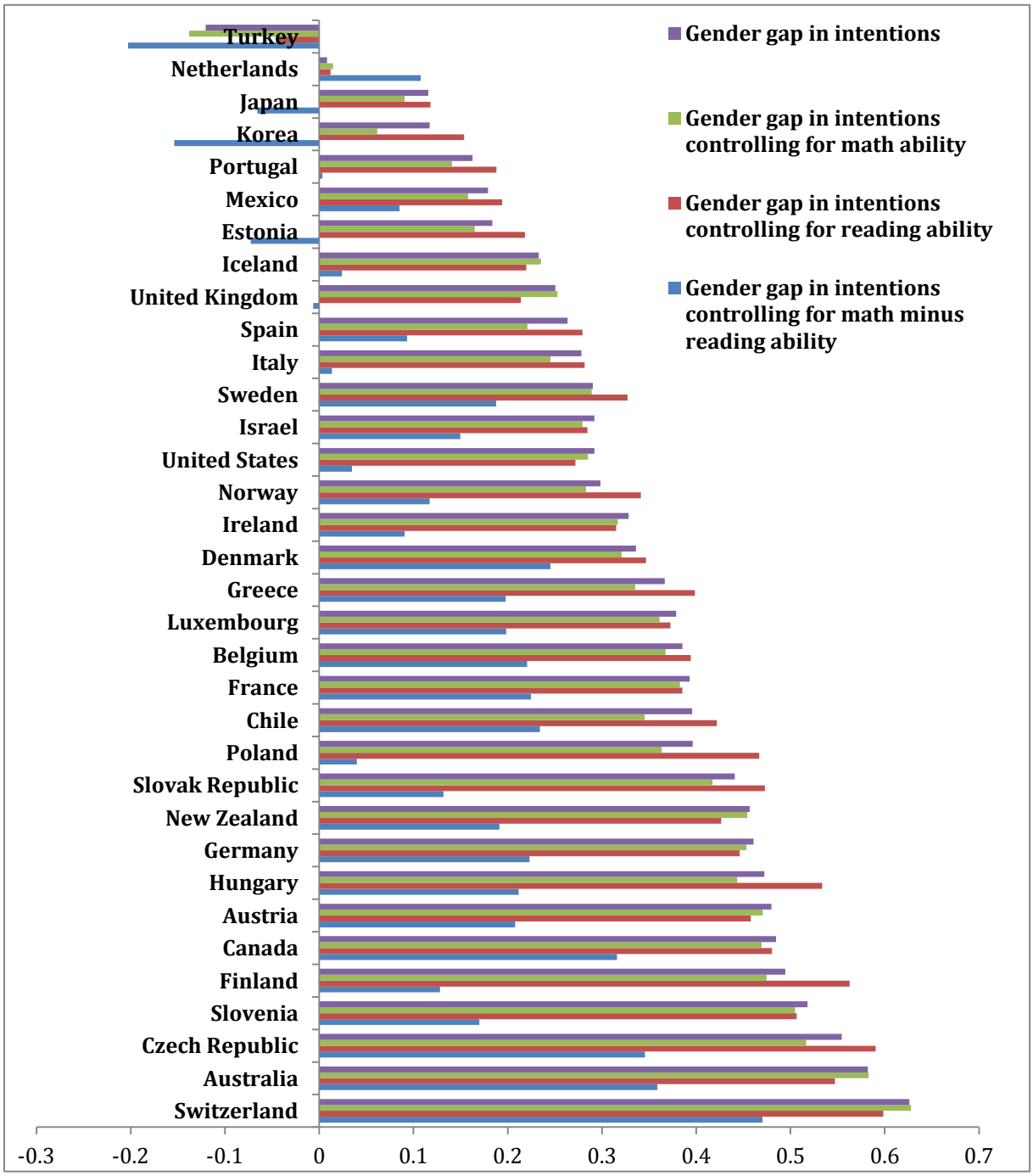
When we focus on girls and boys jointly (columns 1 to 3), all three instruments have a significant effect on MR of the expected sign: a larger share of math teachers or of qualified teachers among math teachers in a school increases students' MR, while a shortage of math teachers decreases it. The first stage is very strong, the Montiel Olea-Pflueger (2013) robust F statistics (16) being well above critical values. The Sargan Khi^2 test reveals that we cannot reject the validity of over-identification restrictions (an interpretation of which is that if we assume that one instrument is exogenous, we cannot reject that the others are as well).
However, when we focus only on girls or boys (especially girls), the instruments tend to become a bit weak (columns 4 to 7). The Montiel Olea-Pflueger statistics in these specifications are such that we can never reject that there is less than a 5% bias between the estimated impact and the true one. The amount of bias that we can reject is typically around 20% to 30% (Table S9 reports only the 20% cut-off, and while some F statistics are above, other are below, meaning that we cannot reject a smaller bias than 20%). What is reassuring however is that the estimates obtained for girls and boys separately are close to those obtained on the whole sample (and not statistically different). This suggests that the slightly weak instruments in these subsample is not too problematic.

Finally, once we restrict the analysis to schools that primarily rely on residence for admissions, we still find a significant impact of MR on intentions, albeit smaller (but still twice larger than the OLS estimate).

In total, results in Table S9 suggest that differences in training or learning conductive to differences in MR can have an impact on choices of study. We have checked the robustness of the results to the use of additional or alternative instruments built with the same logic, the standardization of the instruments or the non-standardization of MR or intentions, the use of weights, and the use of other plausible values: such modifications do not alter substantially the results.

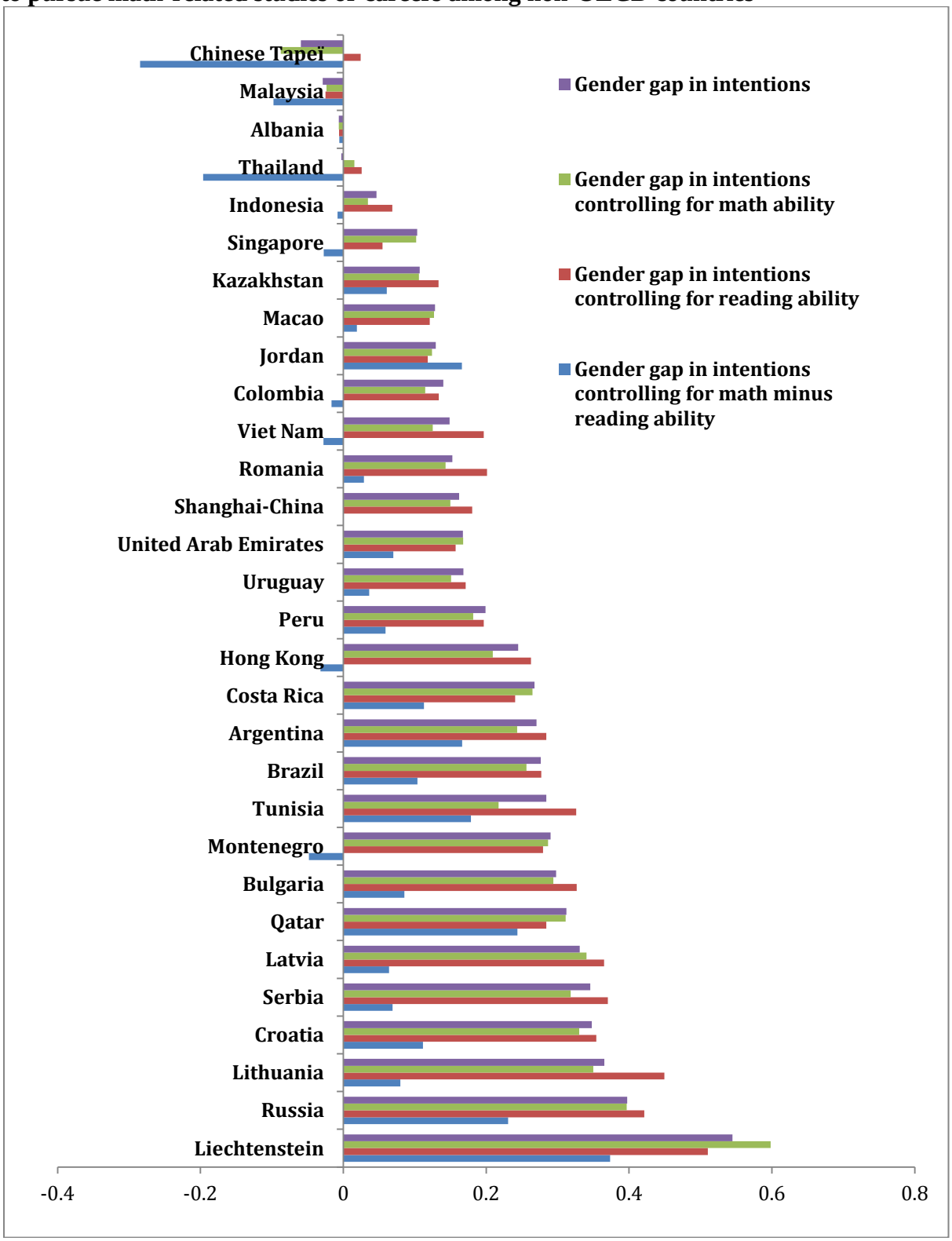
Tables and Figures

Figure S1a: Unconditional and conditional gender gaps (boys minus girls) in intentions to pursue math-related studies or careers among OECD countries



Note: Standard errors and confidence intervals of each estimate are provided in Table S1b.

Figure S1b: Unconditional and conditional gender gaps (boys minus girls) in intentions to pursue math-related studies or careers among non-OECD countries



Note: Standard errors and confidence intervals of each estimate are provided in Table S1b.

Figure S2: Interest for math as a function of ability in math, reading and the comparative advantage in math versus reading

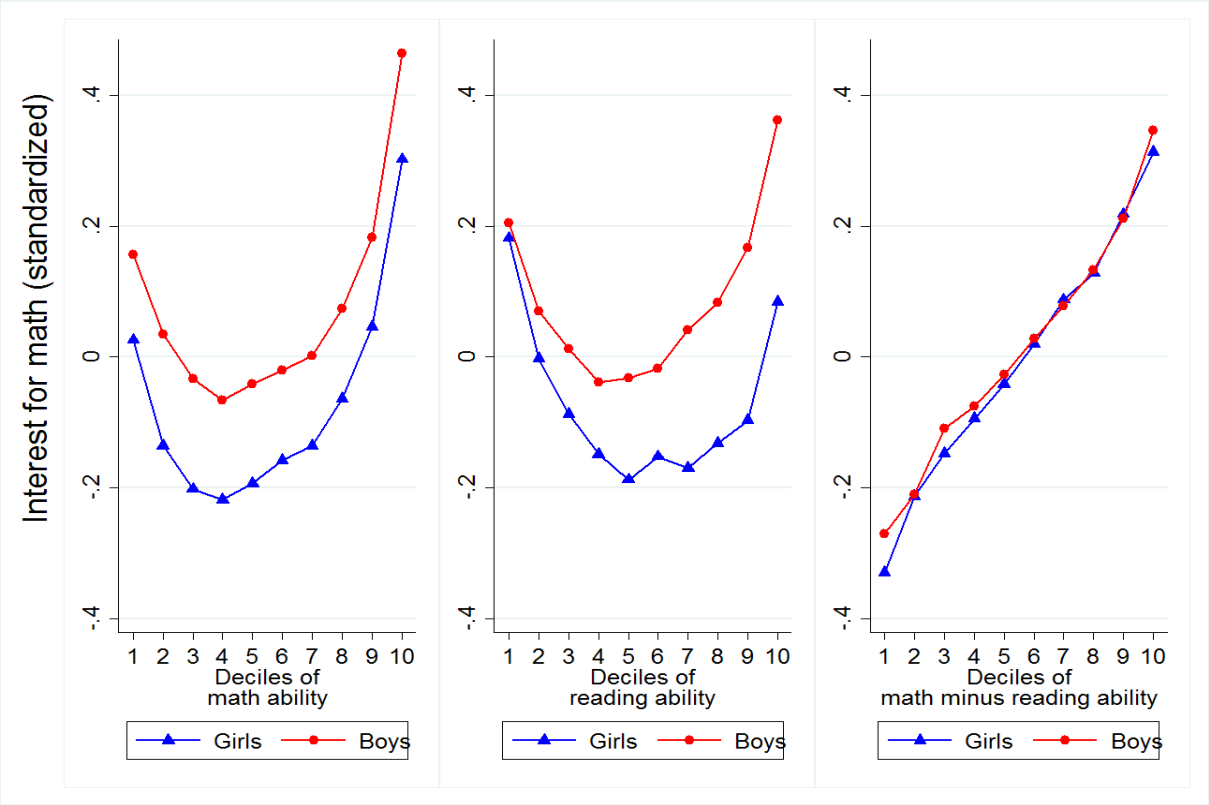


Figure S3: Comparative advantage in math versus reading as a function of math self-concept and math interest

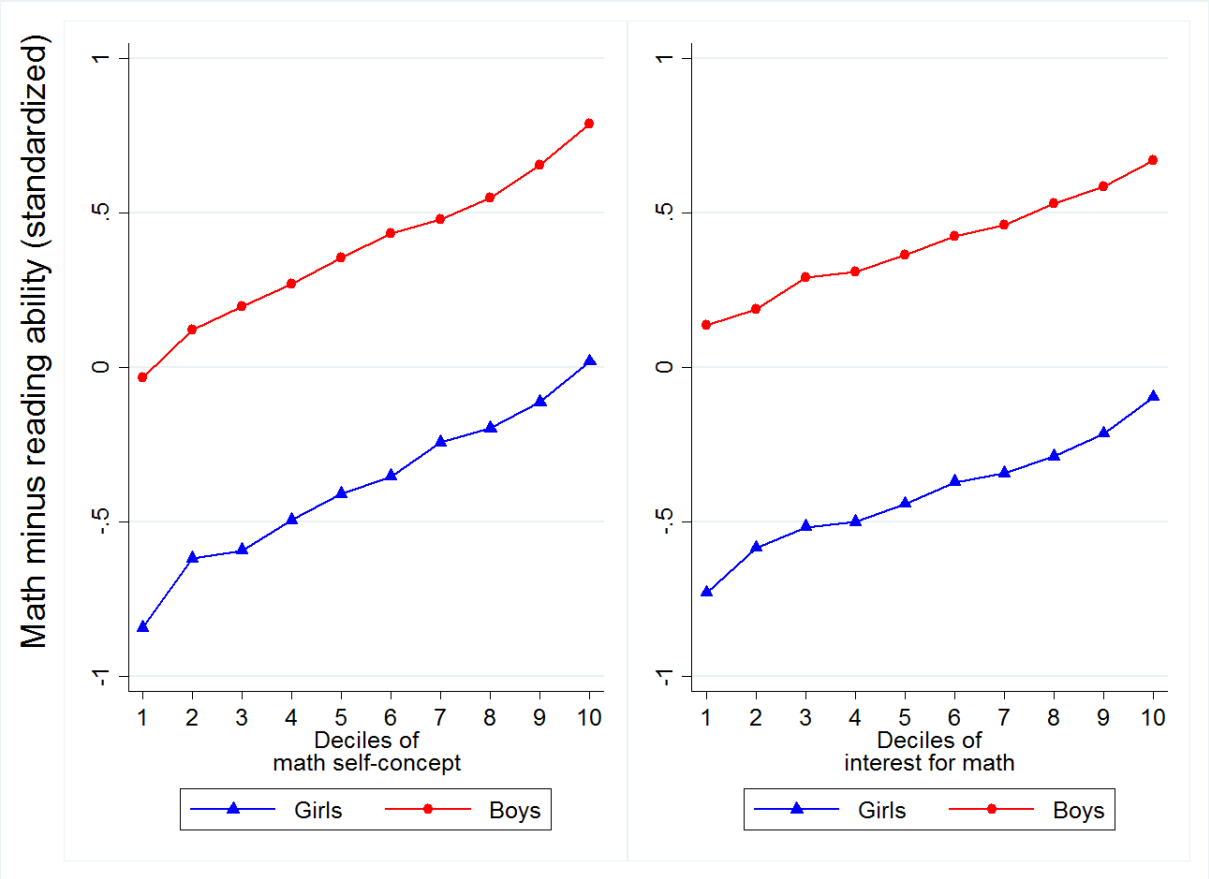


Table S1a: Girls comparative advantage in reading and the gender gap in intentions to study math

	Gender gaps (girls minus boys, as a fraction of variable s.d.)				**Share of the gap in intentions to study math explained by ability in…**		
	math	reading	math minus reading	intentions to study math	math	reading	math minus reading
All PISA2012 Countries	-0.136***	0.351***	-0.832***	-0.218***	0.074	-0.036	0.784
All PISA2012 Countries (basic standardization)	-0.130***	0.286***	-0.803***	-0.213***	0.052	-0.008	0.686
All OECD countries	-0.159***	0.318***	-0.883***	-0.258***	0.064	-0.002	0.810
All Non-OECD countries	-0.110***	0.389***	-0.775***	-0.171***	0.087	-0.101	0.768
OECD countries (ordered by decreasing order of gender gap in intentions to pursue math-related studies or careers							
Switzerland	-0.177***	0.366***	-1.028***	-0.626***	-0.003	0.044	0.249
Australia	-0.164***	0.327***	-0.965***	-0.582***	-0.002	0.060	0.384
Czech Republic	-0.204***	0.365***	-0.990***	-0.554***	0.068	-0.065	0.377
Slovenia	-0.113**	0.553***	-1.131***	-0.518***	0.025	0.022	0.672
Finland	-0.069*	0.569***	-1.096***	-0.494***	0.040	-0.138	0.741
Canada	-0.159***	0.343***	-0.841***	-0.485***	0.031	0.009	0.348
Austria	-0.293***	0.339***	-1.139***	-0.480***	0.019	0.045	0.566
Hungary	-0.125**	0.421***	-1.057***	-0.472***	0.061	-0.130	0.552
Germany	-0.186***	0.450***	-1.238***	-0.461***	0.016	0.032	0.516
New Zealand	-0.217***	0.263***	-0.911***	-0.457***	0.006	0.066	0.581
Slovak Republic	-0.112**	0.383***	-0.916***	-0.441***	0.054	-0.073	0.701
Poland	-0.101**	0.444***	-0.931***	-0.396***	0.083	-0.178	0.899
Chile	-0.302***	0.307***	-0.967***	-0.396***	0.127	-0.066	0.408
France	-0.186***	0.323***	-0.958***	-0.393***	0.027	0.020	0.429
Belgium	-0.148***	0.253***	-0.807***	-0.385***	0.047	-0.024	0.427
Luxembourg	-0.321***	0.249***	-1.009***	-0.379***	0.047	0.016	0.476
Greece	-0.159***	0.440***	-0.870***	-0.367***	0.086	-0.087	0.461
Denmark	-0.202***	0.336***	-0.967***	-0.336***	0.046	-0.032	0.270
Ireland	-0.227***	0.298***	-1.018***	-0.328***	0.035	0.040	0.724
Norway	-0.076**	0.427***	-0.905***	-0.298***	0.051	-0.144	0.607
United States	-0.111***	0.276***	-0.805***	-0.292***	0.024	0.069	0.881
Israel	-0.239***	0.266***	-0.881***	-0.292***	0.044	0.025	0.487
Sweden	-0.008	0.446***	-0.871***	-0.290***	0.004	-0.127	0.353
Italy	-0.249***	0.355***	-1.013***	-0.278***	0.118	-0.013	0.951
Spain	-0.255***	0.245***	-0.845***	-0.263***	0.162	-0.060	0.646
United Kingdom	-0.124**	0.276***	-0.889***	-0.250***	-0.009	0.146	1.026
Iceland	0.025	0.485***	-0.877***	-0.233***	-0.010	0.057	0.896
Estonia	-0.112***	0.503***	-1.058***	-0.184***	0.103	-0.187	1.396
Mexico	-0.169***	0.307***	-0.750***	-0.179***	0.118	-0.084	0.525
Portugal	-0.162***	0.381***	-0.949***	-0.163***	0.135	-0.156	0.979
Korea	-0.209***	0.227***	-0.850***	-0.117***	0.475	-0.311	2.312
Japan	-0.227***	0.208***	-0.830***	-0.116***	0.217	-0.021	1.568

Netherlands	-0.162***	0.233***	-0.800***	-0.008	-0.784	-0.456	-12.008
Turkey	-0.110*	0.538***	-1.014***	0.120**	-0.145	0.623	-0.683
Non-OECD countries (ordered by decreasing order of gender gap in intentions to pursue math-related studies or careers							
Liechtenstein	-0.417***	0.167	-1.113***	-0.545***	-0.098	0.063	0.314
Russia	-0.005	0.419***	-0.659***	-0.398***	0.002	-0.060	0.419
Lithuania	-0.069*	0.578***	-1.143***	-0.365***	0.042	-0.230	0.782
Croatia	-0.192***	0.497***	-1.168***	-0.348***	0.050	-0.018	0.679
Serbia	-0.151***	0.467***	-1.025***	-0.346***	0.079	-0.072	0.801
Latvia	0.057	0.665***	-1.029***	-0.331***	-0.029	-0.103	0.807
Qatar	0.074***	0.541***	-0.891***	-0.312***	0.004	0.090	0.220
Bulgaria	-0.023	0.543***	-1.007***	-0.298***	0.013	-0.096	0.713
Montenegro	-0.024	0.641***	-1.084***	-0.290***	0.012	0.036	1.166
Tunisia	-0.256***	0.279***	-0.738***	-0.284***	0.235	-0.147	0.371
Brazil	-0.195***	0.379***	-0.914***	-0.276***	0.073	-0.002	0.625
Argentina	-0.222***	0.353***	-0.875***	-0.271***	0.101	-0.051	0.385
Costa Rica	-0.368***	0.335***	-0.974***	-0.268***	0.011	0.102	0.579
Hong Kong	-0.185***	0.278***	-0.854***	-0.245***	0.146	-0.073	1.130
Peru	-0.211***	0.254***	-0.810***	-0.199***	0.086	0.012	0.703
Uruguay	-0.168***	0.318***	-0.790***	-0.168***	0.103	-0.019	0.785
United Arab Emirates	0.017	0.540***	-0.984***	-0.168***	-0.002	0.062	0.583
Shanghai-China	-0.071*	0.268***	-0.611***	-0.162***	0.074	-0.114	0.999
Romania	-0.047	0.454***	-0.768***	-0.153***	0.064	-0.318	0.812
Viet Nam	-0.116***	0.436***	-0.817***	-0.149***	0.159	-0.322	1.187
Colombia	-0.352***	0.221***	-0.894***	-0.140***	0.180	0.045	1.117
Jordan	0.135	0.665***	-0.913***	-0.129***	0.040	0.088	-0.283
Macao	-0.025	0.434***	-0.693***	-0.128***	0.012	0.059	0.852
Kazakhstan	-0.011	0.480***	-0.656***	-0.107***	0.011	-0.248	0.430
Singapore	0.020	0.300***	-0.595***	-0.103***	0.011	0.468	1.267
Indonesia	-0.103*	0.356***	-0.660***	-0.046	0.259	-0.477	1.171
Thailand	0.107**	0.644***	-0.792***	0.003	6.620	10.298	-70.535
Albania	0.005	0.095**	-0.146***	0.006	0.000	0.027	0.114
Malaysia	0.068	0.449***	-0.608***	0.029	0.179	0.138	-2.399
Chinese Tapeï	-0.088	0.318***	-0.736***	0.059*	-0.477	1.411	-3.795

*Notes: *** p<0.01, ** p<0.05, * p<0.1. All variables are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country. Intentions to study math is an index built from 5 questions. Estimates and standard errors involving measures of ability are based on plausible values and account for measurement error in these abilities on top of standard sampling error.*

Table S1b: Unconditional and conditional gender gaps (boys minus girls) in intentions to pursue math-related studies or careers

	Gender gap in (standardized) intentions to study math			
	Unconditional	Conditional on math ability	Conditional on reading ability	Conditional on math minus reading ability
All PISA2012 Countries	-0.218	-0.202	-0.226	-0.047
(standard errors)	(0.009)	(0.009)	(0.009)	(0.010)
[95% CI]	[-0.235, -0.201]	[-0.219, -0.184]	[-0.244, -0.208]	[-0.067, -0.027]
All countries (basic standardization)	-0.213	-0.202	-0.214	-0.067
	(0.009)	(0.009)	(0.009)	(0.010)
	[-0.230, -0.196]	[-0.219, -0.185]	[-0.232, -0.197]	[-0.087, -0.047]
OECD countries	-0.258	-0.242	-0.259	-0.049
	(0.012)	(0.011)	(0.012)	(0.013)
	[-0.281, -0.235]	[-0.264, -0.219]	[-0.283, -0.235]	[-0.075, -0.023]
Non-OECD countries	-0.171	-0.156	-0.189	-0.040
	(0.013)	(0.013)	(0.014)	(0.014)
	[-0.197, -0.146]	[-0.182, -0.131]	[-0.217, -0.161]	[-0.068, -0.012]
OECD countries (ordered by decreasing order of gender gap in intentions to pursue math-related studies or careers)				
Switzerland	-0.626	-0.628	-0.598	-0.470
	(0.045)	(0.045)	(0.046)	(0.052)
	[-0.715, -0.537]	[-0.717, -0.539]	[-0.688, -0.509]	[-0.573, -0.368]
Australia	-0.582	-0.583	-0.547	-0.359
	(0.023)	(0.022)	(0.024)	(0.029)
	[-0.626, -0.538]	[-0.627, -0.539]	[-0.594, -0.500]	[-0.415, -0.302]
Czech Republic	-0.554	-0.517	-0.591	-0.346
	(0.045)	(0.045)	(0.046)	(0.055)
	[-0.643, -0.466]	[-0.605, -0.428]	[-0.681, -0.500]	[-0.454, -0.237]
Slovenia	-0.518	-0.505	-0.507	-0.170
	(0.042)	(0.042)	(0.043)	(0.060)
	[-0.600, -0.436]	[-0.587, -0.423]	[-0.590, -0.423]	[-0.287, -0.052]
Finland	-0.494	-0.475	-0.563	-0.128
	(0.033)	(0.032)	(0.037)	(0.043)
	[-0.559, -0.430]	[-0.538, -0.411]	[-0.634, -0.491]	[-0.213, -0.043]
Canada	-0.485	-0.469	-0.480	-0.316
	(0.026)	(0.026)	(0.028)	(0.031)
	[-0.536, -0.433]	[-0.520, -0.419]	[-0.535, -0.426]	[-0.376, -0.255]
Austria	-0.480	-0.471	-0.458	-0.208
	(0.043)	(0.042)	(0.045)	(0.055)
	[-0.564, -0.396]	[-0.553, -0.388]	[-0.546, -0.370]	[-0.316, -0.100]
Hungary	-0.472	-0.444	-0.534	-0.212
	(0.044)	(0.044)	(0.046)	(0.052)
	[-0.559, -0.386]	[-0.530, -0.357]	[-0.624, -0.443]	[-0.313, -0.111]
Germany	-0.461	-0.453	-0.446	-0.223
	(0.042)	(0.043)	(0.042)	(0.051)
	[-0.543, -0.378]	[-0.537, -0.370]	[-0.529, -0.363]	[-0.324, -0.122]
New Zealand	-0.457	-0.454	-0.426	-0.191
	(0.038)	(0.037)	(0.039)	(0.042)
	[-0.531, -0.383]	[-0.526, -0.382]	[-0.503, -0.350]	[-0.273, -0.109]
Slovak Republic	-0.441	-0.417	-0.473	-0.132

	(0.037)	(0.036)	(0.037)	(0.047)
	[-0.513, -0.369]	[-0.488, -0.347]	[-0.545, -0.401]	[-0.224, -0.040]
Poland	-0.396	-0.363	-0.467	-0.040
	(0.039)	(0.035)	(0.039)	(0.042)
	[-0.472, -0.320]	[-0.432, -0.295]	[-0.544, -0.390]	[-0.121, 0.041]
Chile	-0.396	-0.345	-0.422	-0.234
	(0.038)	(0.039)	(0.038)	(0.047)
	[-0.471, -0.321]	[-0.421, -0.269]	[-0.496, -0.348]	[-0.327, -0.141]
France	-0.393	-0.383	-0.385	-0.225
	(0.043)	(0.043)	(0.043)	(0.048)
	[-0.477, -0.310]	[-0.467, -0.298]	[-0.469, -0.302]	[-0.319, -0.130]
Belgium	-0.385	-0.367	-0.394	-0.221
	(0.034)	(0.034)	(0.035)	(0.036)
	[-0.452, -0.318]	[-0.435, -0.300]	[-0.464, -0.325]	[-0.291, -0.150]
Luxembourg	-0.379	-0.361	-0.373	-0.198
	(0.031)	(0.031)	(0.032)	(0.042)
	[-0.439, -0.318]	[-0.422, -0.300]	[-0.434, -0.311]	[-0.280, -0.117]
Greece	-0.367	-0.335	-0.399	-0.198
	(0.040)	(0.039)	(0.041)	(0.045)
	[-0.446, -0.288]	[-0.412, -0.258]	[-0.479, -0.318]	[-0.285, -0.110]
Denmark	-0.336	-0.321	-0.347	-0.245
	(0.035)	(0.036)	(0.036)	(0.038)
	[-0.405, -0.267]	[-0.391, -0.251]	[-0.417, -0.276]	[-0.320, -0.171]
Ireland	-0.328	-0.317	-0.315	-0.091
	(0.037)	(0.038)	(0.036)	(0.042)
	[-0.400, -0.256]	[-0.392, -0.242]	[-0.386, -0.245]	[-0.173, -0.008]
Norway	-0.298	-0.283	-0.341	-0.117
	(0.042)	(0.040)	(0.043)	(0.048)
	[-0.380, -0.216]	[-0.361, -0.205]	[-0.426, -0.256]	[-0.212, -0.023]
United States	-0.292	-0.285	-0.272	-0.035
	(0.033)	(0.033)	(0.035)	(0.039)
	[-0.357, -0.227]	[-0.350, -0.221]	[-0.340, -0.204]	[-0.110, 0.041]
Israel	-0.292	-0.279	-0.285	-0.150
	(0.044)	(0.045)	(0.045)	(0.053)
	[-0.378, -0.206]	[-0.367, -0.191]	[-0.372, -0.197]	[-0.255, -0.045]
Sweden	-0.290	-0.289	-0.327	-0.188
	(0.044)	(0.043)	(0.046)	(0.052)
	[-0.377, -0.204]	[-0.374, -0.205]	[-0.418, -0.237]	[-0.290, -0.086]
Italy	-0.278	-0.245	-0.282	-0.014
	(0.018)	(0.018)	(0.018)	(0.023)
	[-0.314, -0.242]	[-0.281, -0.209]	[-0.317, -0.246]	[-0.058, 0.031]
Spain	-0.263	-0.221	-0.279	-0.093
	(0.027)	(0.026)	(0.027)	(0.031)
	[-0.316, -0.211]	[-0.271, -0.171]	[-0.332, -0.226]	[-0.154, -0.032]
United Kingdom	-0.250	-0.253	-0.214	0.006
	(0.029)	(0.029)	(0.028)	(0.036)
	[-0.307, -0.194]	[-0.309, -0.196]	[-0.269, -0.159]	[-0.063, 0.076]
Iceland	-0.233	-0.235	-0.220	-0.024
	(0.049)	(0.049)	(0.049)	(0.053)
	[-0.329, -0.137]	[-0.331, -0.139]	[-0.315, -0.124]	[-0.129, 0.080]
Estonia	-0.184	-0.165	-0.218	0.073

	(0.037)	(0.037)	(0.038)	(0.051)
	[-0.256, -0.111]	[-0.238, -0.092]	[-0.293, -0.144]	[-0.027, 0.173]
Mexico	-0.179	-0.158	-0.194	-0.085
	(0.015)	(0.014)	(0.016)	(0.018)
	[-0.209, -0.150]	[-0.186, -0.130]	[-0.225, -0.163]	[-0.121, -0.049]
Portugal	-0.163	-0.141	-0.188	-0.003
	(0.032)	(0.032)	(0.034)	(0.039)
	[-0.225, -0.100]	[-0.204, -0.078]	[-0.254, -0.122]	[-0.079, 0.072]
Korea	-0.117	-0.062	-0.154	0.154
	(0.039)	(0.037)	(0.038)	(0.044)
	[-0.193, -0.042]	[-0.134, 0.011]	[-0.229, -0.079]	[0.067, 0.241]
Japan	-0.116	-0.091	-0.118	0.066
	(0.033)	(0.032)	(0.033)	(0.035)
	[-0.180, -0.052]	[-0.153, -0.028]	[-0.182, -0.054]	[-0.002, 0.134]
Netherlands	-0.008	-0.015	-0.012	-0.108
	(0.037)	(0.037)	(0.037)	(0.049)
	[-0.080, 0.063]	[-0.087, 0.058]	[-0.085, 0.061]	[-0.204, -0.011]
Turkey	0.120	0.138	0.045	0.203
	(0.047)	(0.046)	(0.047)	(0.054)
	[0.028, 0.213]	[0.048, 0.228]	[-0.047, 0.138]	[0.097, 0.309]
Non-OECD countries (ordered by decreasing order of gender gap in intentions to pursue math-related studies or careers)				
Liechtenstein	-0.545	-0.598	-0.510	-0.374
	(0.132)	(0.138)	(0.131)	(0.155)
	[-0.802, -0.287]	[-0.869, -0.327]	[-0.768, -0.253]	[-0.678, -0.069]
Russia	-0.398	-0.397	-0.421	-0.231
	(0.034)	(0.033)	(0.039)	(0.040)
	[-0.465, -0.330]	[-0.461, -0.333]	[-0.498, -0.344]	[-0.309, -0.153]
Lithuania	-0.365	-0.350	-0.450	-0.080
	(0.041)	(0.039)	(0.045)	(0.051)
	[-0.446, -0.285]	[-0.426, -0.274]	[-0.538, -0.361]	[-0.180, 0.020]
Croatia	-0.348	-0.330	-0.354	-0.112
	(0.040)	(0.039)	(0.042)	(0.052)
	[-0.426, -0.269]	[-0.407, -0.253]	[-0.436, -0.273]	[-0.214, -0.009]
Serbia	-0.346	-0.318	-0.370	-0.069
	(0.040)	(0.039)	(0.040)	(0.046)
	[-0.424, -0.267]	[-0.395, -0.242]	[-0.449, -0.291]	[-0.158, 0.021]
Latvia	-0.331	-0.341	-0.365	-0.064
	(0.042)	(0.042)	(0.044)	(0.054)
	[-0.414, -0.248]	[-0.422, -0.259]	[-0.452, -0.278]	[-0.170, 0.042]
Qatar	-0.312	-0.311	-0.284	-0.244
	(0.027)	(0.027)	(0.028)	(0.030)
	[-0.365, -0.260]	[-0.363, -0.259]	[-0.338, -0.230]	[-0.303, -0.185]
Bulgaria	-0.298	-0.294	-0.327	-0.086
	(0.037)	(0.038)	(0.039)	(0.045)
	[-0.370, -0.226]	[-0.368, -0.220]	[-0.403, -0.250]	[-0.174, 0.003]
Montenegro	-0.290	-0.287	-0.280	0.048
	(0.037)	(0.037)	(0.040)	(0.058)
	[-0.362, -0.218]	[-0.358, -0.215]	[-0.358, -0.201]	[-0.066, 0.162]
Tunisia	-0.284	-0.217	-0.326	-0.179
	(0.042)	(0.043)	(0.043)	(0.046)

	[-0.367, -0.201]	[-0.301, -0.134]	[-0.410, -0.243]	[-0.270, -0.088]
Brazil	-0.276	-0.256	-0.277	-0.104
	(0.026)	(0.025)	(0.027)	(0.030)
	[-0.326, -0.226]	[-0.305, -0.207]	[-0.330, -0.224]	[-0.162, -0.046]
Argentina	-0.271	-0.243	-0.284	-0.166
	(0.046)	(0.046)	(0.047)	(0.054)
	[-0.360, -0.181]	[-0.333, -0.154]	[-0.376, -0.192]	[-0.273, -0.060]
Costa Rica	-0.268	-0.265	-0.241	-0.113
	(0.042)	(0.043)	(0.043)	(0.054)
	[-0.350, -0.186]	[-0.349, -0.181]	[-0.324, -0.157]	[-0.219, -0.007]
Hong Kong	-0.245	-0.209	-0.263	0.032
	(0.032)	(0.031)	(0.032)	(0.036)
	[-0.307, -0.182]	[-0.270, -0.148]	[-0.326, -0.199]	[-0.039, 0.103]
Peru	-0.199	-0.182	-0.197	-0.059
	(0.035)	(0.037)	(0.036)	(0.040)
	[-0.268, -0.130]	[-0.254, -0.109]	[-0.266, -0.127]	[-0.137, 0.018]
Uruguay	-0.168	-0.151	-0.171	-0.036
	(0.037)	(0.038)	(0.038)	(0.042)
	[-0.241, -0.096]	[-0.225, -0.077]	[-0.246, -0.097]	[-0.118, 0.046]
United Arab Emirates	-0.168	-0.168	-0.157	-0.070
	(0.030)	(0.030)	(0.033)	(0.037)
	[-0.226, -0.109]	[-0.226, -0.110]	[-0.221, -0.093]	[-0.142, 0.002]
Shanghai-China	-0.162	-0.150	-0.180	-0.000
	(0.044)	(0.043)	(0.045)	(0.044)
	[-0.249, -0.075]	[-0.235, -0.065]	[-0.268, -0.093]	[-0.086, 0.086]
Romania	-0.153	-0.143	-0.201	-0.029
	(0.044)	(0.043)	(0.045)	(0.047)
	[-0.240, -0.066]	[-0.228, -0.058]	[-0.290, -0.112]	[-0.121, 0.064]
Viet Nam	-0.149	-0.125	-0.196	0.028
	(0.037)	(0.038)	(0.035)	(0.043)
	[-0.221, -0.076]	[-0.199, -0.051]	[-0.265, -0.127]	[-0.056, 0.112]
Colombia	-0.140	-0.115	-0.134	0.016
	(0.041)	(0.042)	(0.041)	(0.050)
	[-0.220, -0.060]	[-0.197, -0.033]	[-0.215, -0.053]	[-0.082, 0.115]
Jordan	-0.129	-0.124	-0.118	-0.166
	(0.043)	(0.045)	(0.051)	(0.045)
	[-0.214, -0.045]	[-0.212, -0.036]	[-0.217, -0.019]	[-0.254, -0.078]
Macao	-0.128	-0.127	-0.121	-0.019
	(0.034)	(0.034)	(0.035)	(0.038)
	[-0.195, -0.062]	[-0.194, -0.060]	[-0.190, -0.052]	[-0.093, 0.055]
Kazakhstan	-0.107	-0.106	-0.133	-0.061
	(0.037)	(0.037)	(0.040)	(0.040)
	[-0.179, -0.035]	[-0.177, -0.034]	[-0.211, -0.056]	[-0.138, 0.017]
Singapore	-0.103	-0.102	-0.055	0.028
	(0.031)	(0.030)	(0.031)	(0.032)
	[-0.163, -0.043]	[-0.162, -0.043]	[-0.115, 0.005]	[-0.036, 0.091]
Indonesia	-0.046	-0.034	-0.069	0.008
	(0.047)	(0.045)	(0.048)	(0.048)
	[-0.138, 0.045]	[-0.122, 0.053]	[-0.163, 0.026]	[-0.087, 0.103]
Thailand	0.003	-0.015	-0.026	0.196
	(0.036)	(0.035)	(0.037)	(0.038)

	[-0.067, 0.073]	[-0.084, 0.053]	[-0.099, 0.048]	[0.121, 0.271]
Albania	0.006	0.006	0.006	0.006
	(0.049)	(0.049)	(0.050)	(0.051)
	[-0.091, 0.103]	[-0.091, 0.103]	[-0.092, 0.104]	[-0.094, 0.105]
Malaysia	0.029	0.024	0.025	0.098
	(0.035)	(0.035)	(0.037)	(0.041)
	[-0.039, 0.097]	[-0.046, 0.093]	[-0.048, 0.098]	[0.017, 0.179]
Chinese Tapeï	0.059	0.088	-0.024	0.285
	(0.035)	(0.039)	(0.036)	(0.038)
	[-0.009, 0.127]	[0.011, 0.164]	[-0.095, 0.046]	[0.209, 0.360]

Notes: Standard errors in parenthesis and 95% confidence intervals in brackets. All variables are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country. Intentions to study math is an index built from 5 questions. Estimates and standard errors involving measures of ability are based on plausible values and account for measurement error in these abilities on top of standard sampling error.

Table S2: Girls comparative advantage in reading and the gender gap in intentions to pursue math-related studies and careers. Alternative measures of intentions

	Gender gap in variable	Share of the gap in variable explained by ability in…		
		math	reading	math minus reading
Panel A: intentions to take more courses in math than reading *(dummy variable, not standardized)*				
All PISA2012 Countries	-0.077***	0.163	-0.265	0.841
All PISA2012 Countries (basic standardization)	-0.077***	0.126	-0.182	0.704
All OECD countries	-0.093***	0.164	-0.205	0.837
All Non-OECD countries	-0.059***	0.164	-0.370	0.876
Selected countries:				
US	-0.123***	0.094	-0.120	0.832
UK	-0.136***	0.087	-0.075	0.836
FRA	-0.110***	0.161	-0.232	0.449
Panel B: intentions to study harder in math than reading *(dummy variable, not standardized)*				
All PISA2012 Countries	-0.056***	0.187	-0.291	1.027
All PISA2012 Countries (basic standardization)	-0.056***	0.103	-0.107	0.794
All OECD countries	-0.088***	0.131	-0.133	0.851
All Non-OECD countries	-0.022***	0.419	-1.026	1.849
Selected countries:				
US	-0.071***	0.052	0.018	1.037
UK	-0.077***	0.062	-0.008	0.964
FRA	-0.180***	0.108	-0.083	0.543
Panel C: Consider that math is important in future study (standardized)				
All PISA2012 Countries	-0.106***	0.118	-0.155	1.067
All PISA2012 Countries (basic standardization)	-0.106***	0.100	-0.111	0.958
All OECD countries	-0.156***	0.150	-0.203	0.966
All Non-OECD countries	-0.050***	0.066	0.099	1.536
Selected countries:				
US	-0.104***	0.140	-0.178	1.480
UK	-0.201***	0.036	-0.007	0.765
FRA	-0.251***	0.087	-0.086	0.909

*Notes: *** p<0.01, ** p<0.05, * p<0.1. Ability measures are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country. The dependent variables in panels A and B are dummy variables built from students' decision between taking additional courses in mathematics or the language of the test after school finished (panel A) and students' willingness to study harder in either their mathematics classes or in classes teaching the language of the test (panel B). In Panel C, we use students' answers as to whether they strongly agree, agree, disagree or strongly disagree that math is important for their future study. Estimates and standard errors involving measures of ability are based on plausible values and account for measurement error in these abilities on top of standard sampling error.*

Table S3: Results from univariate linear regressions showing the partial effect of math ability, reading ability and math minus reading ability on intentions to study math

	Dependent variable is intentions to pursue math-related studies or careers		
	Partial effect of math ability	Partial effect of reading ability	Partial effect of math minus reading ability
All PISA2012 Countries	0.126***	0.002	0.215***
All countries (basic standardization)	0.092***	-0.010**	0.195***
OECD countries	0.113***	-0.020***	0.248***
Non-OECD countries	0.141***	0.026***	0.178***
OECD countries			
Australia	0.019	-0.154***	0.317***
Austria	0.068***	-0.106***	0.299***
Belgium	0.136***	0.008	0.250***
Canada	0.115***	-0.056***	0.267***
Switzerland	0.018	-0.133***	0.273***
Chile	0.193***	0.052**	0.223***
Czech Republic	0.215***	0.039	0.300***
Germany	0.063***	-0.090***	0.258***
Denmark	0.093***	-0.001	0.152***
Spain	0.182***	0.046***	0.221***
Estonia	0.173***	0.039*	0.224***
Finland	0.295***	0.030	0.369***
France	0.075***	-0.059**	0.229***
United Kingdom	-0.010	-0.149***	0.288***
Greece	0.211***	0.026	0.237***
Hungary	0.244***	0.088***	0.301***
Ireland	0.070***	-0.070***	0.256***
Iceland	0.089***	-0.058**	0.244***
Israel	0.071***	-0.049**	0.195***
Italy	0.147***	-0.017	0.265***
Japan	0.116***	0.005	0.205***
Korea	0.270***	0.152***	0.286***
Luxembourg	0.084***	-0.049***	0.229***
Mexico	0.131***	0.034***	0.142***
Netherlands	-0.039*	0.015	-0.102***
Norway	0.206***	0.059***	0.227***
New Zealand	0.038	-0.144***	0.335***
Poland	0.337***	0.105***	0.391***
Portugal	0.141***	0.048**	0.169***
Slovak Republic	0.223***	0.033	0.371***
Slovenia	0.131***	-0.098***	0.355***
Sweden	0.142***	0.042*	0.160***
Turkey	0.156***	0.146***	0.030
United States	0.070***	-0.093***	0.327***
Non-OECD countries			
Albania	-0.001	0.003	-0.005
United Arab Emirates	0.022*	-0.042***	0.117***
Argentina	0.137***	0.012	0.156***
Bulgaria	0.177***	0.005	0.233***
Brazil	0.116***	-0.025	0.213***
Colombia	0.082***	-0.036	0.172***

Costa Rica	0.031	-0.101***	0.186***
Hong Kong	0.203***	0.046**	0.317***
Croatia	0.106***	-0.034	0.235***
Indonesia	0.118***	0.056*	0.082***
Jordan	-0.043**	-0.042**	0.001
Kazakhstan	0.101***	0.039	0.080***
Liechtenstein	-0.068	-0.229***	0.251***
Lithuania	0.228***	0.075***	0.272***
Latvia	0.164***	-0.008	0.275***
Macao	0.064***	-0.030*	0.161***
Montenegro	0.146***	-0.065***	0.299***
Malaysia	0.076***	0.012	0.098***
Peru	0.091***	-0.023	0.185***
Qatar	-0.021*	-0.092***	0.132***
Shanghai-China	0.172***	0.056***	0.265***
Romania	0.209***	0.083***	0.167***
Russia	0.207***	0.011	0.292***
Singapore	-0.058***	-0.165***	0.216***
Serbia	0.193***	0.007	0.287***
Chinese Tapeï	0.319***	0.261***	0.254***
Thailand	0.170***	0.040**	0.206***
Tunisia	0.275***	0.126***	0.177***
Uruguay	0.110***	-0.005	0.174***
Viet Nam	0.207***	0.088***	0.210***

*Notes: *** p<0.01, ** p<0.05, * p<0.1. All variables are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country. Intentions to study math is an index built from 5 questions. Estimates and standard errors involving measures of ability are based on plausible values and account for measurement error in these abilities on top of standard sampling error.*

Table S4: Gender gap in intentions to pursue math-related studies or careers controlling for various measures of abilities and individual characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Girl	-0.0471***	-0.0517***	-0.0453***	-0.0567***	-0.0454***	-0.0830***	-0.0833***
	(0.0100)	(0.0104)	(0.00999)	(0.0102)	(0.0100)	(0.0102)	(0.0102)
Math minus reading ability	0.205***	0.207***			0.205***		
	(0.00611)	(0.00635)			(0.00611)		
Science ability					0.0143***		
					(0.00424)		
Observations	301,360	291,336	301,360	301,360	301,360	301,360	301,360
R-squared	0.048	0.052	0.048	0.055	0.048	0.067	0.067
Controls:							
Country fixed-effects and students' social background	No	Yes	No	No	No	No	No
5th order polynomial in MR	No	No	Yes	Yes	Yes	Yes	Yes
5th order polynomials in math and reading abilities	No	No	No	Yes	No	No	Yes
5th order polynomial in science ability	No	No	No	No	No	Yes	Yes
Interactions between math, reading and science abilities	No	No	No	No	No	No	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Intentions and ability measures are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country. Controls for students' social background include indexes capturing family wealth, family cultural possessions and economic, social and cultural status. We also control for the type of study program a student is enrolled in, and two measures of misbehavior/absences. Intentions to study math is an index built from 5 questions. Estimates and standard errors involving measures of ability are based on plausible values and account for measurement error in these abilities on top of standard sampling error.

Table S5: Proportion of girls and gender gaps in intentions to pursue math-related studies and careers among various ability groups

	Proportion of girls			Gender gap in intentions		
	All	OECD	Non-OECD	All	OECD	Non-OECD
Top students in math, in math and reading, or in terms of math comparative advantage						
Top 10% of MR	0.165	0.137	0.197	-0.103	-0.071	-0.109
Top 25% of MR	0.232	0.206	0.261	-0.091	-0.075	-0.093
Better in math than reading	0.322	0.302	0.345	-0.110	-0.114	-0.100
Top 10% in math	0.424	0.403	0.448	-0.225	-0.227	-0.225
Top 25% in math	0.454	0.436	0.474	-0.244	-0.275	-0.210
Top half in math	0.481	0.470	0.494	-0.238	-0.285	-0.186
Top 10% in reading and math	0.523	0.505	0.544	-0.180	-0.174	-0.195
Top 25% in reading and math	0.529	0.512	0.551	-0.220	-0.239	-0.201
Top half in reading and math	0.530	0.515	0.549	-0.229	-0.270	-0.182
Bottom students in math, in math and reading, or in terms of math comparative advantage						
Bottom 10% of MR	0.848	0.859	0.835	-0.054	-0.026	-0.074
Bottom 25% of MR	0.776	0.787	0.764	-0.071	-0.062	-0.075
Better in reading than math	0.680	0.684	0.675	-0.103	-0.129	-0.073
Bottom 10% in math	0.512	0.498	0.528	-0.152	-0.170	-0.126
Bottom 25% in math	0.520	0.512	0.530	-0.166	-0.189	-0.139
Bottom half in math	0.524	0.519	0.530	-0.173	-0.207	-0.135
Bottom 10% in reading and math	0.374	0.368	0.382	-0.083	-0.060	-0.114
Bottom 25% in reading and math	0.420	0.417	0.425	-0.121	-0.126	-0.115
Bottom half in reading and math	0.465	0.463	0.468	-0.150	-0.175	-0.119

Notes: All variables are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country. Intentions to pursue math-related studies and careers is an index built from 5 questions. Estimates and standard errors involving measures of ability are based on plausible values and account for measurement error in these abilities on top of standard sampling error.

Table S6: Intentions to study science after Grade 10 and actual enrolment in science tracks in Grade 11 in France

	Gender gap in Variable					Share of gender gap in variable explained by…			
	Raw gap	Conditional on math ability (national exam)	Conditional on reading ability (national exam)	Conditional on math minus reading (national exam)	Conditional on math minus reading (in class grades)	Math (national exam)	Reading (national exam)	Math minus reading (national exam)	Math minus reading (in class grades)
Outcomes are Intentions in Grade 10 (measured between January and March 2016) or actual enrolment in Grade 11 in 2016-2017									
Intentions to go in a General scientific Grade 11	-0.112***	-0.103***	-0.162***	-0.060***	-0.049***	0.081	-0.448	0.459	0.563
	(0.009)	*(0.008)*	*(0.009)*	*(0.009)*	*(0.009)*				
Enrolled in a General scientific Grade 11 in 2016-2017	-0.096***	-0.085***	-0.155***	-0.049***	-0.043***	0.115	-0.610	0.489	0.554
	(0.009)	*(0.007)*	*(0.008)*	*(0.009)*	*(0.009)*				
Consider to study science later on (general question)	-0.180***	-0.172***	-0.219***	-0.128***	-0.114***	0.041	-0.221	0.289	0.367
	(0.009)	*(0.008)*	*(0.009)*	*(0.009)*	*(0.009)*				

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Abilities in math and reading are measured either using either grades at the *Brevet National des Collèges*, which is a national exam graded externally and anonymously taking place at the end of middle school (Grade 10), or using grades given by teachers non-anonymously in class in Grade 10. "General scientific Grade 11" correspond to *"Première S"* in the French system.

Table S7a: Comparing the explanatory power of the comparative advantage with that of other possible determinants of the gender gap in math-intensive fields (whole sample, standard errors included)

	Gender gap in each variable (fraction of a s.d.)				Intentions to study math (Standardized)		
	Absolute	Conditional on math ability	Conditional on MR	Gender gap in MR conditional on the variable	Association with each variable	Gender gap conditional on each variable	Gender gap conditional on each variable plus MR
Variable (standardized):						*(row gap is -0.218 s.d.)*	
Math self-concept	-0.270***	-0.231***	-0.012	-0.780***	0.372***	-0.132***	-0.033**
	(0.005)	(0.011)	(0.013)	(0.013)	(0.002)	(0.005)	(0.014)
Declared interest for math	-0.174***	-0.160***	-0.003	-0.802***	0.396***	-0.150***	-0.046***
	(0.004)	(0.009)	(0.010)	(0.012)	(0.002)	(0.003)	(0.009)
Instrumental motivation for math	-0.104***	-0.088***	0.007	-0.820***	0.352***	-0.182***	-0.049***
	(0.004)	(0.009)	(0.011)	(0.012)	(0.002)	(0.003)	(0.009)
Math anxiety	0.174***	0.129***	0.020	-0.824***	-0.228***	-0.192***	-0.033**
	(0.005)	(0.011)	(0.013)	(0.013)	(0.003)	(0.005)	(0.014)
Math involvement	-0.293***	-0.288***	-0.150***	-0.789***	0.227***	-0.155***	-0.018*
	(0.004)	(0.008)	(0.010)	(0.012)	(0.002)	(0.004)	(0.010)
Math environment	-0.096***	-0.101***	-0.034***	-0.827***	0.174***	-0.202***	-0.041***
	(0.004)	(0.010)	(0.011)	(0.012)	(0.002)	(0.004)	(0.010)

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All variables are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country.

Table S7b: Comparing the explanatory power of the comparative advantage with that of other possible determinants of the gender gap in math-intensive fields (OECD countries only)

	Gender gap in each variable (fraction of a s.d.)				Intentions to study math (Standardized)		
	Absolute	Conditional on math ability	Conditional on MR	Gender gap in MR conditional on the variable	Association with each variable	Gender gap conditional on each variable	Gender gap conditional on each variable plus MR
Variable (standardized):						*(row gap is -0.258 s.d.)*	
Math self-concept	-0.302***	-0.236***	0.014	-0.806***	0.384***	-0.156***	-0.043**
	(0.006)	(0.017)	(0.019)	(0.019)	(0.003)	(0.006)	(0.021)
Declared interest for math	-0.204***	-0.174***	0.013	-0.842***	0.405***	-0.179***	-0.054***
	(0.005)	(0.013)	(0.014)	(0.014)	(0.002)	(0.004)	(0.013)
Instrumental motivation for math	-0.154***	-0.126***	-0.003	-0.861***	0.373***	-0.203***	-0.048***
	(0.005)	(0.013)	(0.015)	(0.014)	(0.002)	(0.004)	(0.012)
Math anxiety	0.216***	0.157***	0.003	-0.853***	-0.253***	-0.217***	-0.038*
	(0.006)	(0.017)	(0.021)	(0.019)	(0.003)	(0.006)	(0.022)
Math involvement	-0.271***	-0.255***	-0.110***	-0.841***	0.213***	-0.204***	-0.030**
	(0.005)	(0.011)	(0.013)	(0.014)	(0.002)	(0.005)	(0.013)
Math environment	-0.111***	-0.106***	-0.038**	-0.875***	0.191***	-0.237***	-0.042***
	(0.005)	(0.013)	(0.016)	(0.014)	(0.002)	(0.005)	(0.014)

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All variables are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country.

Table S7c: Comparing the explanatory power of the comparative advantage with that of other possible determinants of the gender gap in math-intensive fields (Non-OECD countries only)

	Gender gap in each variable (fraction of a s.d.)				Intentions to study math (Standardized)		
	Absolute	Conditional on math ability	Conditional on MR	Gender gap in MR conditional on the variable	Association with each variable	Gender gap conditional on each variable	Gender gap conditional on each variable plus MR
Variable (standardized):						*(row gap is -0.172 s.d.)*	
Math self-concept	-0.236***	-0.219***	-0.036**	-0.747***	0.358***	-0.103***	-0.020
	(0.008)	(0.014)	(0.017)	(0.022)	(0.004)	(0.008)	(0.020)
Declared interest for math	-0.140***	-0.138***	-0.014	-0.755***	0.387***	-0.118***	-0.035**
	(0.006)	(0.011)	(0.014)	(0.018)	(0.003)	(0.005)	(0.014)
Instrumental motivation for math	-0.046***	-0.041***	0.024	-0.771***	0.327***	-0.156***	-0.047***
	(0.006)	(0.014)	(0.016)	(0.017)	(0.003)	(0.006)	(0.013)
Math anxiety	0.125***	0.095***	0.033*	-0.785***	-0.200***	-0.161***	-0.026
	(0.008)	(0.016)	(0.019)	(0.022)	(0.004)	(0.008)	(0.020)
Math involvement	-0.318***	-0.321***	-0.191***	-0.729***	0.243***	-0.097***	0.001
	(0.006)	(0.013)	(0.015)	(0.017)	(0.003)	(0.006)	(0.014)
Math environment	-0.078***	-0.091***	-0.028*	-0.771***	0.155***	-0.160***	-0.036***
	(0.006)	(0.012)	(0.015)	(0.017)	(0.003)	(0.006)	(0.014)

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All variables are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country.

Table S7d: Comparing the explanatory power of the comparative advantage with that of other possible determinants of the gender gap in math-intensive fields (United-States, United Kingdom and France)

	Gender gap in each variable (fraction of a s.d.)				Intentions to study math (Standardized)		
	Absolute	Conditional on math ability	Conditional on MR	Gender gap in MR cond. on the variable	Association with each variable	Gender gap conditional on each variable	Gender gap conditional on each variable plus MR
				United-States		*(raw gap is -0.292)*	
Variable (standardized):							
Math self-concept	-0.180***	-0.130***	0.072	-0.723***	0.408***	-0.210***	-0.052
Declared interest for math	-0.163***	-0.147***	0.045	-0.766***	0.451***	-0.221***	-0.053
Instrumental motivation for math	-0.079***	-0.061	0.065	-0.789***	0.384***	-0.263***	-0.056
Math anxiety	0.177***	0.125**	-0.008	-0.740***	-0.300***	-0.231***	-0.029
Math involvement	-0.246***	-0.237***	-0.059	-0.754***	0.209***	-0.241***	-0.024
Math environment	-0.100***	-0.094**	-0.004	-0.793***	0.233***	-0.267***	-0.032
				United-Kingdom		*(raw gap is -0.250)*	
Math self-concept	-0.404***	-0.344***	0.061	-0.782***	0.360***	-0.144***	0.029
Declared interest for math	-0.130***	-0.101***	0.138***	-0.860***	0.382***	-0.202***	-0.038
Instrumental motivation for math	-0.188***	-0.173***	-0.031	-0.864***	0.357***	-0.186***	0.017
Math anxiety	0.353***	0.299***	0.092**	-0.870***	-0.239***	-0.206***	0.057
Math involvement	-0.143***	-0.130***	0.004	-0.872***	0.198***	-0.228***	0.002
Math environment	-0.048**	-0.043	0.060	-0.887***	0.231***	-0.240***	-0.005
				France		*(raw gap is -0.393)*	
Math self-concept	-0.449***	-0.364***	-0.068	-0.843***	0.379***	-0.338***	-0.275***
Declared interest for math	-0.248***	-0.208***	0.018	-0.903***	0.394***	-0.299***	-0.230***
Instrumental motivation for math	-0.248***	-0.218***	-0.026	-0.912***	0.394***	-0.300***	-0.215***
Math anxiety	0.390***	0.328***	0.156***	-0.915***	-0.160***	-0.446***	-0.285***
Math involvement	-0.361***	-0.349***	-0.193***	-0.907***	0.201***	-0.339***	-0.203***
Math environment	-0.165***	-0.171***	-0.133***	-0.952***	0.195***	-0.362***	-0.200***

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All variables are standardized to have a weighted mean equal to 0 and a weighted standard deviation equal to 1 in each country.

Table S8: Cross-country pairwise correlations

Panel A: Pairwise correlations between countries gender gaps in intentions to study math (PISA 2012) and various measures of country-level gender segregation across fields of studies or jobs

	1	***2***	***3***	***4***	***5***	***6***	***7***
1: Gender gap in intentions to pursue math-related studies or careers	1						
2: Female over-representation in humanities (Bradley-Charles)	-0.39	1					
3: Female representation in math and biology (Bradley-Charles)	0.32	-0.22	1				
4: Percentage of women among stem graduates in tertiary education (Stoet and Geary)	0.52	-0.69	0.38	1			
5: Expectation of a job involving math (TIMSS)	0.51	-0.16	0.34	-0.02	1		
6: Female-to male ratio in computer science	0.5	-0.33	0.21	0.51	0.67	1	
7: Percentage of women among graduates in STEM (Unesco)	0.3	-0.39	0.27	0.76	0.17	0.12	1

Panel B: Pairwise correlations between countries gender gaps in MR (PISA2012) and countries' extent of gender stereotypes/subjective norms related to gender

	1	***2***	***3***	***4***
1: Gender gap in math minus reading ability	1.00			
2: Gender Gap Index	-0.24	1.00		
3: Index of country subjective norms associating girls to math (computed by OECD from PISA2012)	-0.50	-0.01	1.00	
4: Implicit association of girls to humanities and boys to math (from implicit association tests, keeping countries with at least 50 observations)	-0.27	0.10	0.52	1.00

Panel C: Pairwise correlations between countries gender gaps in MR (PISA2012) and characteristics of countries' educational systems

	1	***2***	***3***
1: Gender gap in math minus reading ability	1.00		
2: Index of between-school horizontal stratification	0.40	1.00	
3: Use of mandatory standardized tests	-0.26	-0.37	1.00

Notes: All correlations are established on a sample of at least 26 countries. In panel A, the number of observations for the correlations between 1 and 2, 1 and 3, 1 and 4, 1 and 5, 1 and 6 and 1 and 7 are 39, 39, 45, 26, 31 and 49. In panel B, the number of observations for the correlations between 1 and 2, 1 and 3, 1 and 4 are 57, 64 and 50. Romania is excluded from the correlation between 1 and 4 in panel B because it is flagged as an outlier by the producers of the implicit association data.

Table S9: IV estimates of the effect of the comparative advantage in math versus reading on intentions to pursue math-related studies or careers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent variable is Intentions to study math								
Panel A: Second stage estimates									
Math minus reading ability	0.638***	0.607***	0.818***	0.691***	0.697**	0.926***	1.013***	0.453***	0.436**
	(0.0945)	(0.0921)	(0.184)	(0.247)	(0.342)	(0.221)	(0.282)	(0.106)	(0.177)
	Dependent variable is math minus reading ability								
Panel B: Second stage estimates									
Shortage of math teachers minus shortage of reading teachers	-0.00888**	-0.00901**	-0.000331	-0.000635	0.00335	-0.00276	-0.000606	-0.00750	0.00247
	(0.00363)	(0.00370)	(0.00386)	(0.00449)	(0.00479)	(0.00477)	(0.00514)	(0.00641)	(0.00670)
Share of school teachers that are math teachers	0.167***	0.169***	0.101***	0.159***	0.109***	0.0983***	0.0812***	0.161***	0.101***
	(0.0162)	(0.0164)	(0.0164)	(0.0349)	(0.0365)	(0.0174)	(0.0178)	(0.0188)	(0.0187)
Share of math teachers that are qualified minus Share of all teachers that are qualified	0.0422***	0.0443***	0.0203**	0.0198*	0.0234*	0.0118	0.0153	0.0507***	0.0218
	(0.00900)	(0.00924)	(0.00956)	(0.0111)	(0.0120)	(0.0114)	(0.0123)	(0.0158)	(0.0166)
R-squared	0.003	0.010	0.043	0.011	0.016	0.012	0.016	0.009	0.038
Montiel-Pflueger (MP) robust F statistics	41.48	41.98	12.8	5.74	4.07	11.19	7.47	26.46	9.38
MP weak IV critical value (TSLS 5% threshold)	14.55	14.43	14.58	133.99	13.6	13.96	14.08	14.38	14.93
MP weak IV critical value (TSLS 20% threshold)	6.32	6.28	6.33	6.15	6.02	6.13	6.17	6.27	6.45
p-value of Sargan Khi2 over-identification test	0.839	0.885	0.554	0.539	0.663	0.456	0.352	0.823	0.758
Sample	All students	All students	All students	Girls	Girls	Boys	Boys	School adm. based on residence	School adm. based on residence
Controls for individual characteristics	No	Yes	Yes	No	Yes	No	Yes	No	Yes
Controls for school quality	No	No	Yes	No	Yes	No	Yes	No	Yes
Observations	180,589	174,303	162,643	92,569	83,528	88,020	79,115	60,565	56,732

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Country fixed-effects are included in all models. Math minus reading ability is measured using only the first student plausible values for math and for reading ability (see justification in the supplementary text). Controls for individual characteristics are those of Table S4. Controls for school quality include class size, share of girls in the school, shortage of teaching staff, number of computer per student and share connected to the web, two indexes of school responsibility for resource allocation and for curriculum and assessment.

References:

1. OECD, PISA 2012 Results (Volume I): *Excellence and Equity in Education*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264266490-en (2016)
2. OECD, PISA 2012 Technical Report. OECD Publishing, Paris. http://www.oecd.org/pisa/data/2015-technical-report/
3. Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., ... & Spence, J. T. (1983). Achievement and achievement motivation. *Expectancies, values and academic behaviors*, 75-146.
4. Marsh, H. W. (1986). Verbal and math self-concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal, 23*, 129-149.
5. Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., & Eccles, J. S. (2008). Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value. In H. M. G. Watt & J. S. Eccles (Eds.), *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 115-143). Washington, DC, US: American Psychological Association.
6. Solon, G., S.J. Haider & J.M. Wooldridge, 2015. "What Are We Weighting For?" *Journal of Human Resources*, 50, 301-316.
7. Charles, M. & K. Bradley (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*, 114(4), 924-976.
8. Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, *29*(4), 581-593.
9. Wang, M. T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science*, *24*(5), 770-775.
10. Charles, M., Harr, B., Cech, E. & Hendley, A. (2014). Who likes math where? Gender differences in eighth-graders' attitudes around the world. *International Studies in Sociology of Education*, 24 (1), 85–112.
11. Stoet, G., Bailey, D.H., Moore, A., & Geary, D. C. (2016) Countries with Higher Levels of Gender Equality Show Larger National Sex Differences in Mathematics Anxiety and Relatively Lower Parental Mathematics Valuation for Girls *PLoS One*, 11 (4): e0153857 DOI: 10.1371/journal.pone.0153857
12. OECD (2015), The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence, PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264229945-
13. OECD (2012) PISA 2012 Results: Ready to Learn (Volume III) Students' Engagement, Drive and Self-Beliefs, OECD Publishing, Paris. DOI:https://dx.doi.org/10.1787/9789264201170-en
14. Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology*, *107*(3), 631-644. https://doi.org/10.1037/edu0000005
15. Nosek, B.A. et al. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106 (26) 10593-10597; https://doi.org/10.1073/pnas.0809921106
16. Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, *31*(3), 358-369.
17. Breda, T., Grenet, J., Monnet, M., & Van Effenterre, C. (2018). Can female role models reduce the gender gap in science? Evidence from classroom interventions in French high schools. Paris School of Economics Working Papers n°2018-06. 2018