Research Article

Psychological Science
1–13

DOI: 10.1177/0956797617741719
www.psychologicalscience.org/PS

The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education

Gijsbert Stoet[1] and David C. Geary[2]
[1]School of Social Sciences, Leeds Beckett University, and [2]Department of Psychological Sciences, University of Missouri

Abstract
The underrepresentation of girls and women in science, technology, engineering, and mathematics (STEM) fields is a continual concern for social scientists and policymakers. Using an international database on adolescent achievement in science, mathematics, and reading (N = 472,242), we showed that girls performed similarly to or better than boys in science in two of every three countries, and in nearly all countries, more girls appeared capable of college-level STEM study than had enrolled. Paradoxically, the sex differences in the magnitude of relative academic strengths and pursuit of STEM degrees rose with increases in national gender equality. The gap between boys' science achievement and girls' reading achievement relative to their mean academic performance was near universal. These sex differences in academic strengths and attitudes toward science correlated with the STEM graduation gap. A mediation analysis suggested that life-quality pressures in less gender-equal countries promote girls' and women's engagement with STEM subjects.

Keywords
cognitive ability, cross-cultural differences, educational psychology, science education, sex differences, open materials

Received 5/11/17; Revision accepted 10/12/17

The underrepresentation of girls and women in science, technology, engineering, and mathematics (STEM) fields is a worldwide phenomenon (Burke & Mattis, 2007; Ceci & Williams, 2011; Ceci, Williams, & Barnett, 2009; Cheryan, Ziegler, Montoya, & Jiang, 2017). Although women are now well represented in the social and life sciences (Ceci, Ginther, Kahn, & Williams, 2014; Su & Rounds, 2016), they continue to be underrepresented in fields that focus on inorganic phenomena (e.g., computer science). Despite considerable efforts toward understanding and changing this pattern, the sex difference in STEM engagement has remained stable for decades (e.g., in the United States; National Science Foundation, 2017). The stability of these differences and the failure of current approaches to change them calls for a new perspective on the issue.

Here, we identified a major contextual factor that appears to influence women's engagement in STEM education and occupations. We found that countries with high levels of gender equality have some of the largest STEM gaps in secondary and tertiary education; we call this the *educational-gender-equality paradox*. For example, Finland excels in gender equality (World Economic Forum, 2015), its adolescent girls outperform boys in science literacy, and it ranks second in European educational performance (OECD, 2016b). With these high levels of educational performance and overall gender equality, Finland is poised to close the STEM gender gap. Yet, paradoxically, Finland has one of the world's largest gender gaps in college degrees in STEM fields, and Norway and Sweden, also leading in gender-equality rankings, are not far behind (fewer than 25% of STEM graduates are women). We will show that this pattern extends throughout the world, whereby the graduation gap in STEM increases with increasing levels of gender equality.

Corresponding Author:
Gijsbert Stoet, Leeds Beckett University, Psychology, Room CC-CL-919, City Campus, Leeds, LS1 3HE, United Kingdom
E-mail: stoet@gmx.com

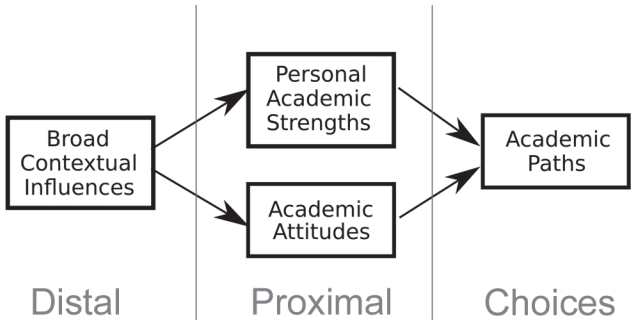

Fig. 1. Schematic illustration of the factors influencing educational and occupational choices. Distal factors, such as relatively poor living conditions, might influence the development of personal academic strengths and attitudes toward different academic fields, which in turn result in choices individuals make in secondary education, tertiary education, and occupations.

We propose that the educational-gender-equality paradox is driven by two different processes, one based on distal social factors and the other on more proximal factors. The latter is student's own rational decision making based on relative academic strengths and weaknesses as well as attitudes that can be influenced by distal factors (Fig. 1).

Our proposal that students' own rational decisions play a key role in explaining the educational-gender-equality paradox is inspired by the *expectancy-value theory* (Eccles, 1983; Wang & Degol, 2013). On the basis of this theory, it is hypothesized that students use their own relative performance (e.g., knowledge of what subjects they are best at) as a basis for decisions about further educational and occupational choices, and this has been demonstrated for STEM-related decision making in the United States (Wang, Eccles, & Kenny, 2013). The basic idea that individuals choose academic paths on the basis of perceived individual strengths is reflected in common practice by school professionals: When students have the opportunity to choose their coursework in secondary education, they are typically recommended to make choices on the basis of their strengths and enjoyment (e.g., Gardner, 2016; Universities and Colleges Admissions Service, 2015).

Wider social factors may influence engagement in STEM fields through students' utility beliefs or the expected long-term value of an academic path (Eccles, 1983; Wang & Degol, 2013). Social factors that might influence STEM engagement are best assessed by comparing countries that vary widely in the associated costs and benefits of a STEM career. One possibility is that contexts with fewer economic opportunities and higher economic risks may make relatively high-paying STEM occupations more attractive relative to contexts with greater opportunities and lower risks. This may contribute to the educational-gender-equality paradox, because economic and general life risks are lower in gender-equal countries, which in turn results in greater opportunity for individual interests and academic strengths to influence investment in one academic path or another, as demonstrated by Wang et al. (2013) for the United States.

In the present article, we report analyses of the academic achievement of almost 475,000 adolescents across 67 nations or economic regions. We found that girls and boys have similar abilities in science literacy in most nations. At the same time, on the basis of a novel approach for examining intraindividual differences in academic strengths and relative weaknesses, we report that science or mathematics is much more likely to be a personal academic strength for boys than for girls. We then report that the relation between the sex differences in academic strengths and college graduation rates in STEM fields is larger in more gender-equal countries. Finally, we conducted a mediation analysis that suggests that the latter is related to overall life satisfaction, which, in turn, is related to income and economic risk in less developed countries (cf. Pittau, Zelli, & Gelman, 2010).

Method

Programme for International Student Assessment (PISA)

PISA (OECD, 2016b) is the world's largest educational survey. PISA assessments in science literacy, reading comprehension, and mathematics are conducted every 3 years, and in each cycle, one of these domains is studied in depth. In 2015, the focus was on science literacy, which included additional questions about science learning and attitudes (see below). We used this most recent data set, in which 519,334 students from 72 nations and regions participated. In order to prevent double-counting of samples, we excluded regions for which we also had national data (Massachusetts and North Carolina, several Spanish regions, and Buenos Aires, because we had data from the United States, Spain, and Argentina as a whole); this exclusion resulted in a sample of 472,242 students in 67 nations or regions (Table S1 in the Supplemental Material available online), which represents 25,141,223 students (i.e., the sum of weights provided by PISA for each student). Our data set included the following regions: Hong Kong, Macao, Chinese Taipei, and the Chinese provinces of Beijing, Shanghai, Jiangsu, and Guangdong (i.e., these four Chinese provinces were combined into one sub-data set by PISA).

The PISA organizers selected a representative sample of schools and students in each participating country or region. Participating students were between 15 years and 3 months and 16 years and 2 months old. All participating students completed a 2-hr PISA test that assessed how well they can apply their knowledge in the domains of reading comprehension, mathematics,

and science literacy. The same (translated) test material was used in each country.

PISA uses a well-developed statistical framework to calculate scores for science literacy, mathematics, reading comprehension, and numerous other variables related to student attitudes and socioeconomic factors (OECD, 2016a). The scores of each student in each academic domain are scaled such that the average of students in Organization for Economic Cooperation and Development (OECD) countries is 500 points and the standard deviation is 100 points.

The additional science literacy assessments in 2015 focused on attitudes, including *science self-efficacy, broad interest in science*, and *enjoyment of science*. For science self-efficacy,

> PISA 2015 asked students to report on how easy they thought it would be for them to: recognize the science question that underlies a newspaper report on a health issue; explain why earthquakes occur more frequently in some areas than in others; describe the role of antibiotics in the treatment of disease; identify the science question associated with the disposal of garbage; predict how changes to an environment will affect the survival of certain species; interpret the scientific information provided on the labelling of food items; discuss how new evidence can lead them to change their understanding about the possibility of life on Mars; and identify the better of two explanations for the formation of acid rain. For each of these, students could report that they "could do this easily", "could do this with a bit of effort", "would struggle to do this on [their] own", or "couldn't do this". Students' responses were used to create the index of science self-efficacy. (OECD, 2016b, p. 136)

Broad interest in science was assessed as follows:

> Students reported on a five-point Likert scale with the categories "not interested", "hardly interested", "interested", "highly interested", and "I don't know what this is", their interest in the following topics: biosphere (e.g., ecosystem services, sustainability); motion and forces (e.g., velocity, friction, magnetic and gravitational forces); energy and its transformation (e.g., conservation, chemical reactions); the Universe and its history; how science can help us prevent disease. (OECD, 2016b, p. 284)

Enjoyment of science was assessed using the following questions:

> I generally have fun when I am learning <broad science> topics; I like reading about <broad science>; I am happy working on <broad science> topics; I enjoy acquiring new knowledge in <broad science>; and I am interested in learning about <broad science>. (OECD, 2016b, p. 284; different science topics were inserted in <broad science> across questions)

In order to estimate whether a student would, in principle, be capable of study in STEM, we used a proficiency level of at least 4 (of a possible 6) in science, mathematics, and reading comprehension. For science literacy for instance and according to the PISA guidelines,

> At Level 4, students can use more complex or more abstract content knowledge, which is either provided or recalled, to construct explanations of more complex or less familiar events and processes. They can conduct experiments involving two or more independent variables in a constrained context. They are able to justify an experimental design, drawing on elements of procedural and epistemic knowledge. Level 4 students can interpret data drawn from a moderately complex data set or less familiar context, draw appropriate conclusions that go beyond the data and provide justifications for their choices." (OECD, 2016b, p. 60)

We believe that level 4 would be a minimal requirement.

> At Level 3, students can draw upon moderately complex content knowledge to identify or construct explanations of familiar phenomena. In less familiar or more complex situations, they can construct explanations with relevant cueing or support. They can draw on elements of procedural or epistemic knowledge to carry out a simple experiment in a constrained context. Level 3 students are able to distinguish between scientific and non-scientific issues and identify the evidence supporting a scientific claim." (OECD, 2016b, p. 60)

Publications further detailing the PISA framework and methodology are available via http://www.oecd.org/pisa/pisaproducts/.

STEM degrees

The United Nations Educational, Scientific and Cultural Organization (UNESCO) reports national statistics on, among other things, education. We used the UNESCO graduation data (http://data.uis.unesco.org) labeled "Distribution of tertiary graduates" in the years 2012 to 2015 in natural sciences, mathematics, statistics, information and communication technologies, engineering, manufacturing,

and construction (Table S1). The percentage of women among STEM graduates ranged from 12.4% in Macao to 40.7% in Algeria; the median was 25.4%.

Gender equality

The World Economic Forum publishes *The Global Gender Gap Report* annually. We used the 2015 data (World Economic Forum, 2015). For each nation, the Global Gender Gap Index (GGGI) assesses the degree to which girls and women fall behind boys and men on 14 key indicators (e.g., earnings, tertiary enrollment ratio, life expectancy, seats in parliament) on a 0.0 to 1.0 scale, with 1.0 representing complete parity (or men falling behind). For the countries participating in the 2015 PISA, GGGI scores ranged from 0.593 for the United Arab Emirates to 0.881 for Iceland (Table S1).

Overall life satisfaction (OLS)

We took the OLS score from the United Nations Development Programme (2016, pp. 250–253). The OLS question was formulated as follows:

> Please imagine a ladder, with steps numbered from zero at the bottom to ten at the top. Suppose we say that the top of the ladder represents the best possible life for you, and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time, assuming that the higher the step the better you feel about your life, and the lower the step the worse you feel about it? Which step comes closest to the way you feel?

This score was expressed on a scale from 0 (*least satisfied*) to 10 (*most satisfied; M* = 6.2, *SD* = 0.9, ranging from 4.1 in Georgia to 7.6 in Switzerland and Norway).

Analyses

For each participating student, the PISA data set provides scores for mathematics, science literacy, and reading comprehension. We used these given scores to calculate each student's highest performing subject (i.e., personal strength), second highest, and lowest. To do so, we needed to calculate each student's average score in these three subjects and then compare each subject score to the calculated average score. In order to make such calculations possible, we standardized data first. In other words, we scaled the data into a common format, namely *z* scores, which have a mean of 0 and a standard deviation of 1.

We calculated each students' relative strengths in mathematics, science literacy, and reading comprehension using the following steps:

1. We standardized the mathematics, science, and reading scores on a nation-by-nation basis. We call these new standardized scores *z*Math, *z*Reading, and *z*Science, respectively.
2. We calculated for each student the standardized average score of the new *z* scores. We call this *z*General.
3. Then, we calculated each student's intraindividual strengths by subtracting *z*General as follows: relative science strength = *z*Science – *z*General, relative math strength = *z*Math – *z*General, relative reading strength = *z*Reading – *z*General.
4. Finally, using these new intraindividual (relative) scores, we calculated for each country the averages for boys and girls and subtracted those scores to calculate the gender gaps in relative academic strengths.

To illustrate, one U.S. student had the following three PISA scores for science, mathematics, and reading: 364, 411, and 344, respectively. After standardization (Step 1), these scores were *z*Science = −1.39, *z*Math = −0.69, and *z*Reading = −1.61. The student's *z*General score was −1.27 (Step 2). His relative strengths were calculated by subtracting *z*General from the standardized scores and then again standardizing the difference scores (because they are by definition not standardized). Using this calculation, we obtained the following relative scores for this student: relative science strength = −0.71, relative math strength = 2.23, and relative reading strength = −1.34 (Step 3). Note that although this student's scores in all three subjects are below the standardized national mean (i.e., 0), his personal strength in mathematics deviates more than 2 standard deviations from the national mean of relative mathematics strengths. In other words, the gap between his mathematics score and his overall mean score is much larger (> 2 *SD*s) than is typical for U.S. students. Using these types of scores, we could calculate the intraindividual sex differences for science, mathematics, and reading for the United States (and similarly for all other nations and regions).

Further, we calculated for each student the difference between actual science performance and science self-efficacy (i.e., self-perceived ability). For this, we used the same method as reported elsewhere (Stoet, Bailey, Moore, & Geary, 2016, p. 10): For each participating nation, we first standardized science performance and science self-efficacy scores. Then, we subtracted these two variables for each student and then once more

standardized the difference for the students of each country separately. The resulting score is a measure of the degree to which science self-efficacy is unrepresentative of actual performance (i.e., underestimation of own ability or exaggeration of own ability).

For correlations, we typically applied Spearman's ρ (correlation coefficient abbreviated as r_s), because not all variables were normally distributed. Throughout all analyses, we used an alpha criterion of .05.

Results

Sex differences in science literacy

For each of the 67 countries and regions participating in the 2015 PISA, we first tested for sex differences in science literacy (i.e., average score of boys – average score of girls, by country; Fig. 2a). We found that girls outperformed boys in 19 (28.4%) countries, boys outperformed girls in 22 (32.8%) countries, and there was no statistically significant difference in the remaining 26 (38.8%) countries. The mean national effect size (Cohen's *d*) was −0.01 (*SD* = 0.13, 95% confidence interval, or CI = [−0.04, 0.02]), ranging between −0.46 (95% CI = [−0.50, −0.41]) in favor of girls (in Jordan) and 0.26 (95% CI = [0.21, 0.31]) in favor of boys (in Costa Rica). The relation between the effect size of the absolute science gap and gender equality (GGGI) was not statistically significant (r_s = .23, 95% CI = [−.18, .46], p = .069, n = 62).

Sex differences in academic strengths

As we previously reported for reading and mathematics (Stoet & Geary, 2015), there were consistent sex differences in intraindividual academic strengths across reading and science. In all countries except for Lebanon and Romania (97% of countries), boys' intraindividual strength in science was (significantly) larger than that of girls (Fig. 2b). Further, in all countries, girls' intraindividual strength in reading was larger than that of boys, while boys' intraindividual strength in mathematics was larger than that of girls. In other words, the sex differences in intraindividual academic strengths were near universal. The most important and novel finding here is that the sex difference in intraindividual strength in science was higher and more favorable to boys in more gender-equal countries, r_s = .42, 95% CI = [.19, .61], p < .001, n = 62 (Fig. 3a), as was the sex difference in intraindividual strength in reading, which favored girls in more gender-equal countries, r_s = −.30, 95% CI = [−.51, −.06], p = .017, n = 62.

Another way of calculating these patterns is to examine the percentage of students who have individual strengths in science, mathematics, and reading, respectively. To do so, we first determined students' individual strength. Next, we calculated the percentage of boys and girls who had science, mathematics, or reading as their personal academic strength; this contrasts with the above analysis that focused on the overall magnitude of these strengths independently of whether they were the students' personal strength. We found that on average (across nations), 24% of girls had science as their strength, 25% of girls had mathematics as their strength, and 51% had reading. The corresponding values for boys were 38% science, 42% mathematics, and 20% reading.

Thus, despite national averages that indicate that boys' performance was consistently higher in science than that of girls relative to their personal mean across academic areas, there were substantial numbers of girls within nations who performed relatively better in science than in other areas. Within Finland and Norway, two countries with large overall sex differences in the intraindividual science gap and very high GGGI scores, there were 24% and 18% of girls, respectively, who had science as their personal academic strength, relative to 37% and 46% of boys.

Finally, it should also be noted that the difference between the percentage of girls with a strength in science or mathematics was always equally large or larger than the percentage of women graduating in STEM fields; importantly, this difference was again larger in more gender-equal countries (r_s = .41, 95% CI = [.15, .62], n = 50, p = .003). In other words, more gender-equal countries were more likely than less gender-equal countries to lose those girls from an academic STEM track who were most likely to choose it on the basis of personal academic strengths.

The above analyses show that most boys scored relatively higher in science than their all-subjects average, and most girls scored relatively higher in reading than their all-subjects average. Thus, even when girls outperformed boys in science, as was the case in Finland, girls generally performed even better in reading, which means that their individual strength was, unlike boys' strength, reading. The relevant finding here is that the intraindividual sex differences in relative strengths in science and reading rose with increases in gender equality (GGGI). In accordance with expectancy-value theory, this pattern should result in far more boys than girls pursuing a STEM career in more gender-equal nations, and this was the case (r_s = −.47, 95% CI = [−.66, −.22], p < .001, n = 50; Fig. 3b). And, similarly, girls will be more likely than boys to choose options in which they can gain the most benefit from their relative strength in reading.

Science attitudes and gender equality

Next, we considered sex differences in science attitudes, namely science self-efficacy, broad interest in

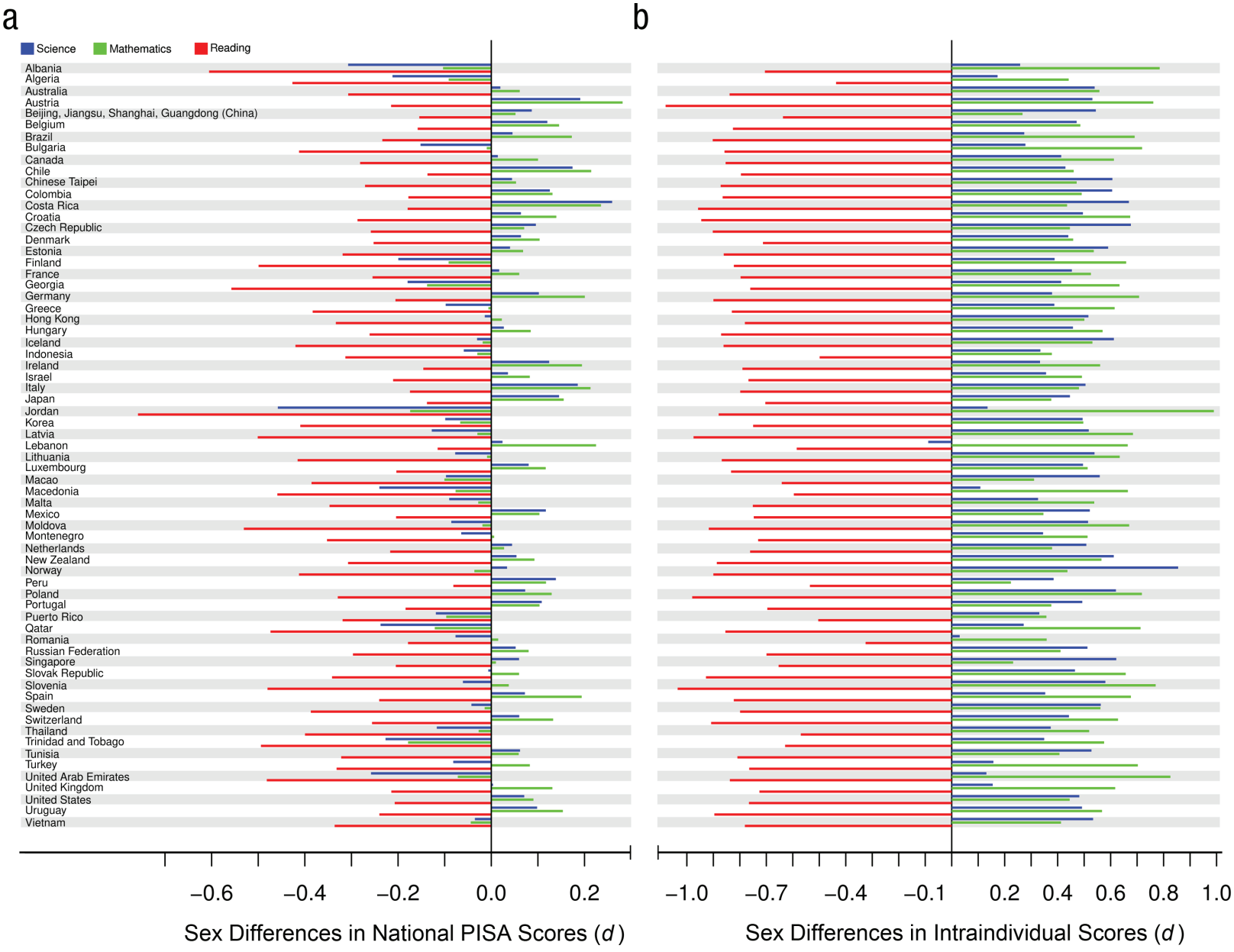

Fig. 2. Sex differences in Programme for International Student Assessment (PISA) science, mathematics, and reading scores expressed as Cohen's *d*s (see Table S2 in the Supplemental Material for confidence intervals). Sex differences were calculated as the scores of boys minus the scores of girls. Thus, negative values indicate an advantage for girls, and positive values indicate an advantage for boys. Results are shown separately for (a) sex differences in absolute PISA scores and (b) sex differences in intraindividual scores.

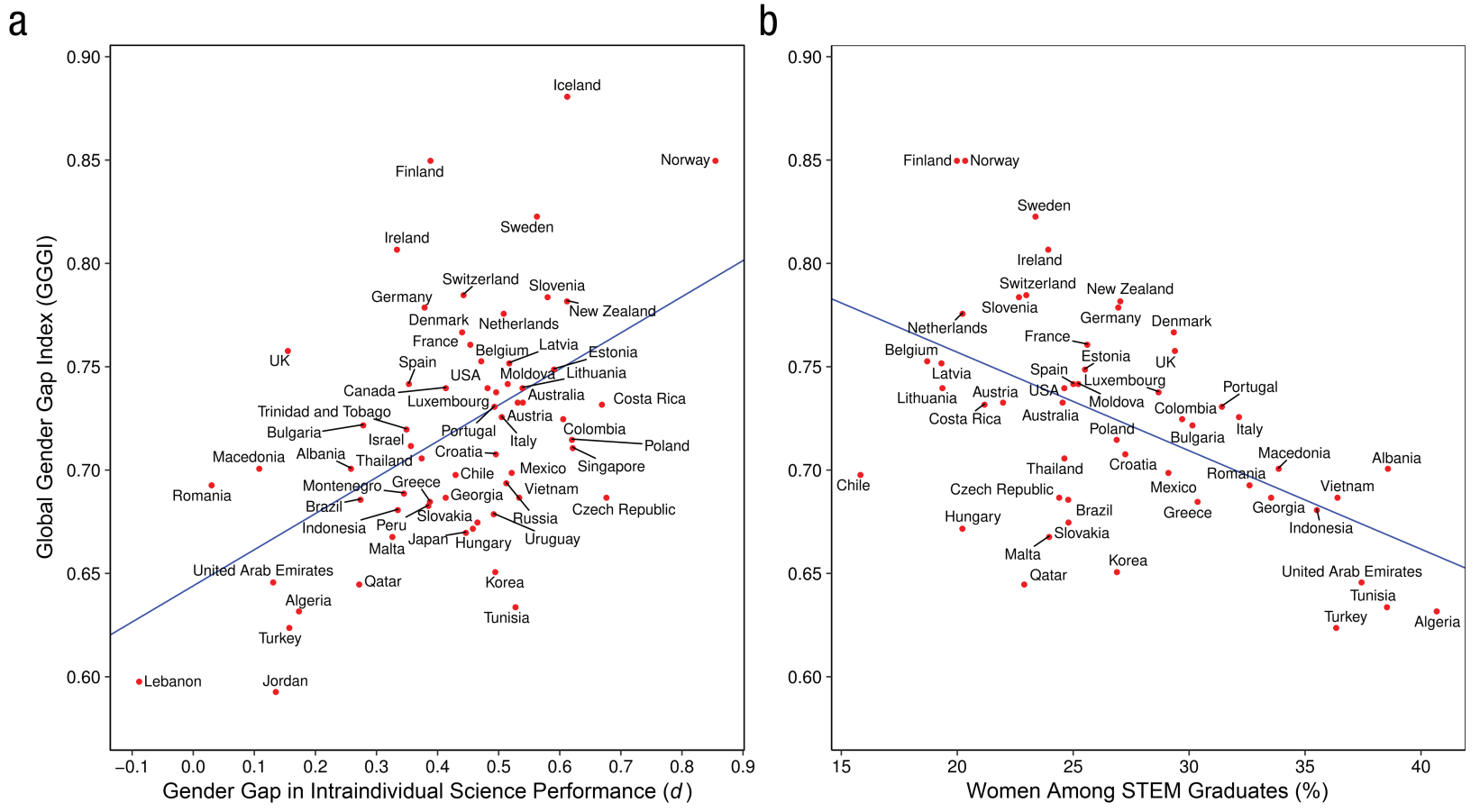

Fig. 3. Scatterplots (with best-fitting regression lines) showing the relation between gender equality and sex differences in (a) intraindividual science performance and (b) the percentage of women among science, technology, engineering, and math (STEM) graduates. Gender equality was measured with the Global Gender Gap Index (GGGI), which assesses the extent to which economic, educational, health, and political opportunities are equal for women and men. The gender gap in intraindividual science scores (a) was larger in more gender-equal countries (rs = .42). The percentage of women with degrees in STEM fields (b) was lower in more gender-equal countries (rs = −.47).

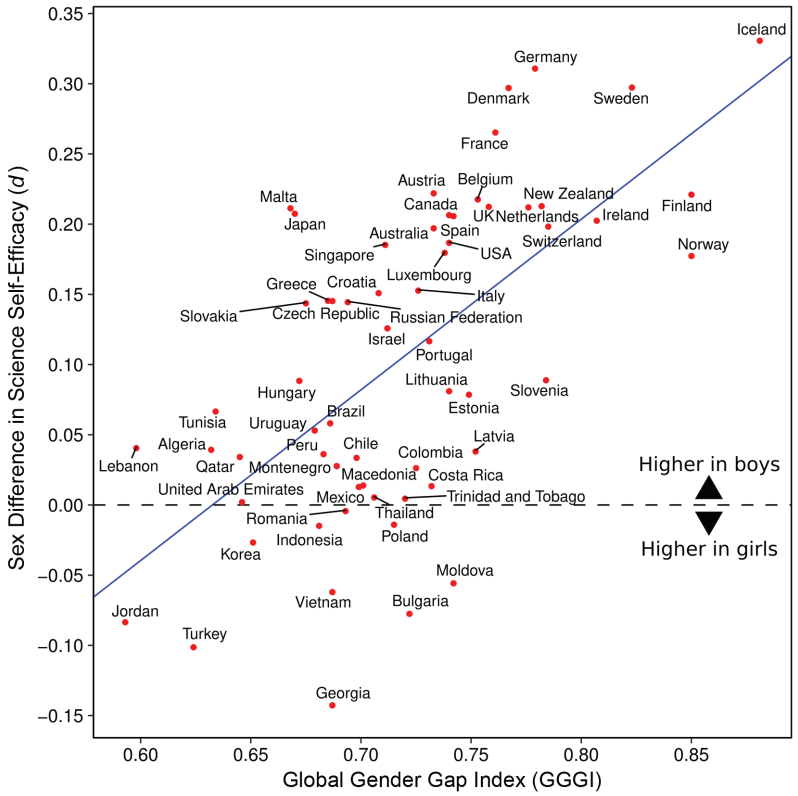

Fig. 4. Scatterplot (with best-fitting regression line) showing the relation between sex difference in science self-efficacy and the Global Gender Gap Index.

science, and enjoyment of science. Boys' science self-efficacy was higher than that of girls in 39 of 67 (58%) countries, and especially so in more gender-equal countries, r_s = .60, 95% CI = [.41, .74], p < .001, n = 61 (Fig. 4). Similarly, boys expressed a stronger broad interest in science than girls in 51 (76%) countries, and again this was particularly true in more gender-equal countries, r_s = .41, 95% CI = [.15, .62], p = .003, n = 50. And finally, the same was found for students' enjoyment of science; boys reported more joy in science than girls in 29 (43%) countries, and more so in gender-equal countries, r_s = .46, 95% CI = [.23, .64], p < .001, n = 61. Further, these attitude gaps were correlated with the intraindividual science gap (self-efficacy: r_s = .24, 95% CI = [−.00, .46], p = .052, n = 66; enjoyment of science: r_s = .31, 95% CI = [.07, .52], p = .010, n = 66; broad interest: r_s = .27, 95% CI = [.01, .51], p = .043, n = 54).

Science self-efficacy was relatively weakly correlated with science performance (across participating nations, r = .17, 95% CI = [.16, .18], n = 472,242, p < .001). This means that the deviation between science self-efficacy and science performance is of interest (e.g., students might under- or overestimate their own performance, and this could influence later choices). We calculated for each student the difference between standardized science self-efficacy scores and standardized science performance scores (this is a measure of the component of self-efficacy that is independent from actual performance; see Method). Using this metric, we found that in 34 (49%) countries, boys overestimated their science self-efficacy and deviated significantly from girls, in comparison with 5 (7%) countries where girls overestimated their science self-efficacy and deviated significantly from boys. Paradoxically, boys' overestimation of their competence in science was larger in countries with higher GGGI scores (M = 0.739, SD = 0.06) relative to countries in which there was no sex difference in the estimation of science competence (M = 0.697, SD = 0.04), t(54) = 2.66, p = .010.

Next, we used the science performance data and attitude data (broad interest in science and enjoyment of science) to determine the percentage of female students who, in principle, could be successful in tertiary education in STEM fields. For this, we defined suitability as follows: A student would need to have a proficiency level of at least 4 in all three PISA domains (science, mathematics, and reading; see Method). Using these ability criteria, we would expect far more women among STEM graduates (international mean = 49%, SD = 4) than are actually found in any country (international mean = 28%, SD = 6; Fig. 5a). In regard to attitudes, we assumed that they should at least have the

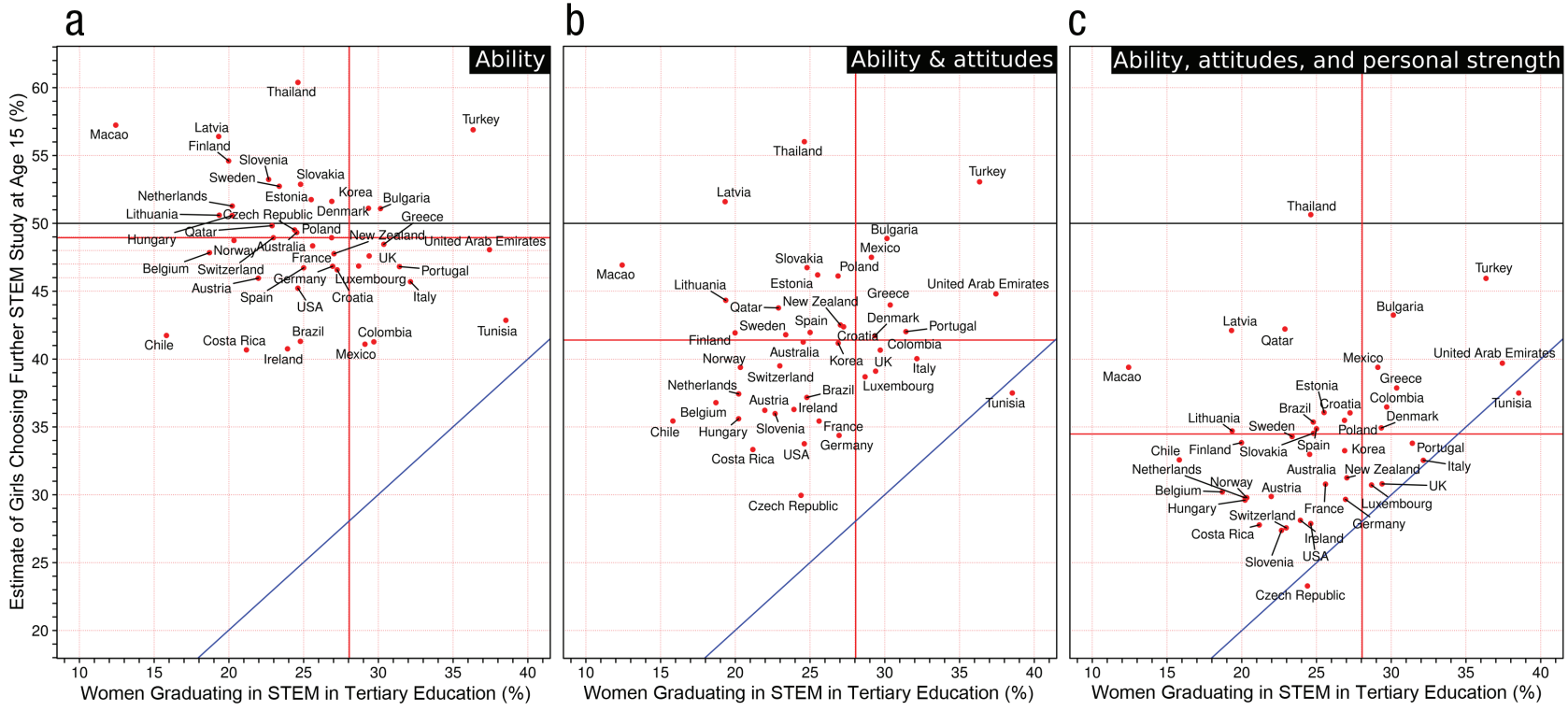

Fig. 5. Scatterplots showing the relation between the percentage of female students estimated to choose further science, technology, engineering, and math (STEM) study after secondary education and the estimated percentage of female STEM graduates in tertiary education. Red lines indicate the estimated (horizontal) and actual (vertical) average graduation percentage of women in STEM fields. For instance, in (c), we estimated that 34% of women would graduate college with a STEM degree (internationally), but only 28% did so. Identity lines (i.e., 45° lines) are colored blue; points above the identity lines indicate fewer women STEM graduates than expected. Panel (a) displays the percentage of female students estimated to choose STEM study on the basis of ability alone (see the text for criteria). Although there was considerable cross-cultural variation, on average around 50% of students graduating in STEM fields could be women, which deviates considerably from the actual percentage of women among STEM graduates. The estimate of women STEM students shown in (b) was based on both ability, as in (a), and being above the international median score in science attitudes. The estimate shown in (c) is based on ability, attitudes, and having either mathematics or science as a personal strength.

international median level of enjoyment of science, interest in science, and science self-efficacy. Using these additional criteria, the percentage of girls likely to enjoy, feel capable of participating in, and be successful in tertiary STEM programs is still considerably higher in every country (international mean = 41%, *SD* = 6), except Tunisia, than was actually found (Fig. 5b).

As argued above, we believe that factors other than attitude and motivation play a role—namely personal academic strengths. When we added this factor to our estimate (Fig. 5c), we saw that the difference between expected and actual STEM graduates became smaller (international mean = 34%, *SD* = 6), although it is still the case that in most countries women's STEM graduation rates are lower than we would anticipate (see Discussion).

Mediation model

Thus far, we have shown that the sex differences in STEM graduation rates and in science literacy as an academic strength become larger with gains in gender equality and that schools prepare more girls for further STEM study than actually obtain a STEM college degree. We will now consider one of the factors that might explain why the graduation gap may be larger in the more gender-equal countries. Countries with the highest gender equality tend to be welfare states (to varying degrees) with a high level of social security for all its citizens; in contrast, the less gender-equal countries have less secure and more difficult living conditions, likely leading to lower levels of life satisfaction (Pittau et al., 2010). This may in turn influence one's utility beliefs about the value of science and pursuit of STEM occupations, given that these occupations are relatively high paying and thus provide the economic security that is less certain in countries that are low in gender equality. We used OLS as a measure of overall life circumstances; this is normally distributed and is a good proxy for economic opportunity and hardship and social and personal well-being (Pittau et al., 2010).

In more equal countries, overall life satisfaction was higher (r_s = .55, 95% CI = [.35, .70], $p < .001$, n = 62). Accordingly, we tested whether low prospects for a satisfied life may be an incentive for girls to focus more on science in school and, as adults, choose a career in a relatively higher paid STEM field. If our hypothesis is correct, then OLS should at least partially mediate the relation between gender equality and the sex differences in STEM graduation. A formal mediation analysis using a bootstrap method with 5,000 iterations confirmed the mediational model path of life satisfaction for STEM graduation (mean indirect effect = −0.19, *SE* = 0.08, Sobel's z = −2.24, $p < .025$, 95% CI of bootstrapped samples = [−0.39, −0.04]). The effect of the direct path in the mediation model was statistically significant (mean direct effect = −0.34, *SE* = 0.135, 95% CI of bootstrapped samples = [−0.65, −0.02], p = .038), and the mediation was considered partial (proportion mediated = 0.35, 95% CI = [0.06, 0.95], p = .013; Table S3 in the Supplemental Material). A sensitivity analysis of this mediation (Imai, Keele, & Tingley, 2010; Tingley, Yamamoto, Hirose, Keele, & Imai, 2014) showed the point at which the average causal mediation effect (ACME) was approximately zero (ρ = −0.4, 95% CI = [−0.11, 0.15], $R_M^{2*}R_Y^{2*} = 0.16$, $\tilde{R}_M^2\tilde{R}_Y^2 = 0.07$; Fig. S1 in the Supplemental Material). The latter finding suggests that an unknown third variable may have confounded the mediation model (see Discussion).

Discussion

Using the most recent and largest international database on adolescent achievement, we confirmed that girls performed similarly or better than boys on generic science literacy tests in most nations. At the same time, women obtained fewer college degrees in STEM disciplines than men in all assessed nations, although the magnitude of this gap varied considerably. Further, our analysis suggests that the percentage of girls who would likely be successful and enjoy further STEM study was considerably higher than the percentage of women graduating in STEM fields, implying that there is a loss of female STEM capacity between secondary and tertiary education.

One of the main findings of this study is that, paradoxically, countries with lower levels of gender equality had relatively more women among STEM graduates than did more gender-equal countries. This is a paradox, because gender-equal countries are those that give girls and women more educational and empowerment opportunities and that generally promote girls' and women's engagement in STEM fields (e.g., Williams & Ceci, 2015).

In our explanation of this paradox, we focused on decisions that individual students may make and decisions and attitudes that are likely influenced by broader socioeconomic considerations. On the basis of expectancy-value theory (Eccles, 1983; Wang & Degol, 2013), we reasoned that students should at least, in part, base educational decisions on their academic strengths. Independently of absolute levels of performance, boys on average had personal academic strengths in science and mathematics, and girls had strengths in reading comprehension. Thus, even when girls' absolute science scores were higher than those of boys, as in Finland, boys were often better in science relative to their overall academic average. Similarly, girls might have scored higher than boys in science, but they were often

even better in reading. Critically, the magnitude of these sex differences in personal academic strengths and weaknesses was strongly related to national gender equality, with larger differences in more gender-equal nations. These intraindividual differences in turn may contribute, for instance, to parental beliefs that boys are better at science and mathematics than girls (Eccles & Jacobs, 1986; Gunderson, Ramirez, Levine, & Beilock, 2012).

We also found that boys often expressed higher self-efficacy, more joy in science, and a broader interest in science than did girls. These differences were also larger in more gender-equal countries and were related to the students' personal academic strength. We discuss some implications below (Interventions).

Explanations

We propose that when boys are relatively better in science and mathematics while girls are relatively better at reading than other academic areas, there is the potential for substantive sex differences to emerge in STEM-related educational pathways. The differences are expected on the basis of expectancy-value theory and are consistent with prior research (Eccles, 1983; Wang & Degol, 2013). The differences emerge from a seemingly rational choice to pursue academic paths that are a personal strength, which also seems to be common academic advice given to students, at least in the United Kingdom (e.g., Gardner, 2016; Universities and Colleges Admissions Service, 2015).

The greater realization of these potential sex differences in gender-equal nations is the opposite of what some scholars might expect intuitively, but it is consistent with findings for some other cognitive and social sex differences (e.g., Lippa, Collaer, & Peters, 2010; Pinker, 2008; Schmitt, 2015). One possibility is that the liberal mores in these cultures, combined with smaller financial costs of foregoing a STEM path (see below), amplify the influence of intraindividual academic strengths. The result would be the differentiation of the academic foci of girls and boys during secondary education and later in college, and across time, increasing sex differences in science as an academic strength and in graduation with STEM degrees.

Whatever the processes that exaggerate these sex differences, they are abated or overridden in less gender-equal countries. One potential reason is that a well-paying STEM career may appear to be an investment in a more secure future. In line with this, our mediation analysis suggests that OLS partially explains the relation between gender equality and the STEM graduation gap. Some caution when interpreting this result is needed, though. Mediation analysis depends on a number of assumptions, some of which can be tested using a sensitivity analysis, which we conducted (Imai, Keele, & Yamamoto, 2010). The sensitivity analysis gives an indication of the correlation between the statistical error component in the equations used for predicting the mediator (OLS) and the outcome (STEM graduation gap); this includes the effect of unobserved confounders. Given the range of ρ values in the sensitivity analysis (Fig. S1), it is possible that a third variable could be associated with OLS and the STEM graduation gap. A related limitation is that the sensitivity analysis does not explore confounders that may be related to the predictor variable (i.e., GGGI). Future research that includes more potential confounders is needed, but such data are currently unavailable for many of the countries included in our analysis.

Relation to previous studies of gender equality and educational outcomes

Our current findings agree with those of previous studies in that sex differences in mathematics and science performance vary strongly between countries, although we also believe that the link between measures of gender equality and these educational gaps (e.g., as demonstrated by Else-Quest, Hyde, & Linn, 2010; Guiso, Monte, Sapienza, & Zingales, 2008; Hyde & Mertz, 2009; Reilly, 2012) can be difficult to determine and is not always found (Ellison & Swanson, 2010; for an in-depth discussion, see Stoet & Geary, 2015).

We believe that one factor contributing to these mixed results is the focus on sex differences in absolute performance, as contrasted with sex differences in academic strengths and associated attitudes. As we have shown, if absolute performance, interest, joy, and self-efficacy alone were the basis for choosing a STEM career, we would expect to see more women entering STEM career paths than do so (Fig. 5).

It should be noted that there are careers that are not STEM by definition, although they often require STEM skills. For example, university programs related to health and health care (e.g., nursing and medicine) have a majority of women. This may partially explain why even fewer women than we estimated pursue a college degree in STEM fields despite obvious STEM ability and interest.

Interventions

Our results indicate that achieving the goal of parity in STEM fields will take more than improving girls' science education and raising overall gender equality. The generally overlooked issue of intraindividual differences in academic competencies and the accompanying influence on one's expectancies of the value of pursuing one type of career versus another need to be incorporated into approaches for encouraging more women to enter

the STEM pipeline. In particular, high-achieving girls whose personal academic strength is science or mathematics might be especially responsive to STEM-related interventions.

In closing, we are not arguing that sex differences in academic strengths or wider economic and life-risk issues are the only factors that influence the sex difference in the STEM pipeline. We are confirming the importance of the former (Wang et al., 2013) and showing that the extent to which these sex differences manifest varies consistently with wider social factors, including gender equality and life satisfaction. In addition to placing the STEM-related sex differences in broader perspective, the results provide novel insights into how girls' and women's participation in STEM might be increased in gender-equal countries.

Action Editor

Timothy J. Pleskac served as action editor for this article.

Author Contributions

G. Stoet and D. C. Geary collaboratively designed the study and contributed equally to the writing of the article. G. Stoet analyzed all data, which were processed and interpreted by both authors. Both authors approved the final version of the manuscript for submission.

ORCID iD

Gijsbert Stoet https://orcid.org/0000-0002-7557-483X

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797617741719

Open Practices

All materials are publicly available (for links, see the Open Practices Disclosure form). The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797617741719. This article has received the badge for Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

References

Burke, R. J., & Mattis, M. C. (2007). *Women and minorities in science, technology, engineering, and mathematics: Upping the numbers*. Cheltenham, England: Edward Elgar.

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*, 75–141.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences, USA*, *108*, 3157–3162.

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, *135*, 218–261.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, *143*, 1–35.

Eccles, J. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: W. H. Freeman.

Eccles, J., & Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs*, *11*, 367–380.

Ellison, G., & Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels. *Journal of Economic Perspectives*, *28*, 109–128.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127.

Gardner, A. (2016). *How important are GCSE choices when it comes to university?* Retrieved from http://university.which.co.uk/advice/gcse-choices-university/how-important-are-gcse-choices-when-it-comes-to-university

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, *320*, 1164–1165.

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, *66*, 153–166.

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences, USA*, *106*, 8801–8807.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*, 309–334.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*(1), 51–71.

Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Archives of Sexual Behavior*, *39*, 990–997.

National Science Foundation. (2017). *Women, minorities, and persons with disabilities in science and engineering: 2017* (Special Report NSF 17-310). Arlington, VA: National Center for Science and Engineering Statistics.

OECD. (2016a). *PISA 2015 assessment and analytical framework: Science, reading, mathematic and financial literacy*. Paris, France: Author.

OECD. (2016b). *PISA 2015 results: Excellence and equity in education* (Vol. 1). Paris, France: Author.

Pinker, S. (2008). *The sexual paradox: Men, women and the real gender gap*. New York, NY: Simon & Schuster.

Pittau, M. G., Zelli, R., & Gelman, A. (2010). Economic disparities and life satisfaction in European regions. *Social Indicators Research*, *96*, 339–361.

Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLOS ONE*, *7*(7), Article e39904. doi:10.1371/journal.pone.0039904

Schmitt, D. P. (2015). The evolution of culturally-variable sex differences: Men and women are not always different, but when they are . . . It appears *not* to result from patriarchy or sex role socialization. In T. K. Shackelford & R. D. Hansen (Eds.), *The evolution of sexuality* (pp. 221–256). London, England: Springer.

Stoet, G., Bailey, D. H., Moore, A. M., & 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), Article e0153857. doi:10.1371/journal.pone.0153857

Stoet, G., & Geary, D. C. (2015). Sex differences in academic achievement are not related to political, economic, or social equality. *Intelligence*, *48*, 137–151.

Su, R., & Rounds, J. (2016). All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in Psychology*, *6*, Article 189. doi:10.3389/fpsyg.2015.00189

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5). doi:10.18637/jss.v059.i05

United Nations Development Programme. (2016). *Human development report 2016*. New York, NY: Palgrave Macmillan.

Universities and Colleges Admissions Service. (2015). *Tips on choosing A level subjects*. Retrieved from https://www.ucas.com/sites/default/files/tips_on_choosing_a_levels_march_2015_0.pdf

Wang, M., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy-value perspective to understand individual and gender differences in STEM fields. *Developmental Review*, *33*, 304–340.

Wang, M., 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*, 770–775.

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, USA*, *112*, 5360–5365.

World Economic Forum. (2015). *The global gender gap report 2015*. Geneva, Switzerland: Author.