

RESEARCH

Open Access



# Sex differences in variability across nations in reading, mathematics and science: a meta-analytic extension of Baye and Monseur (2016)

Helen Gray<sup>1</sup>, Andrew Lyth<sup>1</sup>, Catherine McKenna<sup>1</sup>, Susan Stothard<sup>3</sup>, Peter Tymms<sup>1</sup> and Lee Copping<sup>2\*</sup>

\*Correspondence:

l.copping@tees.ac.uk

<sup>2</sup> School of Social Sciences,  
Humanities and Law,  
Teesside University,  
Middlesbrough TS1 3BX, UK  
Full list of author information  
is available at the end of the  
article

## Abstract

A recent study by Baye and Monseur (Large Scale Assess Educ 4:1–16, 2016) using large, international educational data sets suggest that the “greater male variation hypothesis” is well supported. Males are often over-represented at the tails of the ability distribution despite similarity in measures of central tendency and the gradual closing of the attainment gap relative to females. In this study, we replicate and expand Baye and Monseur’s work, and explore greater male variability by country using meta-analysis and meta-regression. While we broadly confirm that variability is greater for males internationally, we find that there is significant heterogeneity between countries, and that much of this can be quantified using variables applicable across these assessments (such as test, year, male–female effect size, mean country score and Global Gender Gap Indicators). While it is still not possible to make any causal conclusions regarding why males are more varied than females in academic assessments, it is possible to show that some national level variables effect the magnitude of this variation. Results and suggestions for further work are discussed.

## Introduction

Sex differences in cognitive abilities is a contentious issue, yet one that continues to draw the attention of the public and the research community alike. 21st Century society is motivated to ensure issues of equity between the sexes are adequately addressed, particularly within the sphere of educational opportunity (Marks 2008; UNESCO 2011). Despite best efforts however, inequities still exist internationally, for example, with females underrepresented in our most prestigious educational institutions and males overrepresented in school underperformance, particularly in core areas such as reading and mathematics (Baye and Monseur 2016; Dubet 2010; Jacobs 1996; Morgan and Kett 2003; Quinn and Wagner 2013).

While the evidence shows that the gap between men and women is closing on average across many educational outcomes (Hyde et al. 1990; OECD 2015) and, in some cases, it now favours women (Lietz 2006; OECD 2015), this shift does not appear to have translated directly into ensuring parity across higher professions and positions, a

phenomenon which appears somewhat paradoxical. Baye and Monseur (2016) suggested that this may be due to the way in which sex differences have been historically examined, focussing on mean results which assume homogeneity of variance across the achievement distribution. In a study using international assessment data, they demonstrated that the magnitude of the sex differences in achievement across literacy, mathematics and science varied across the range of results, and that the largest differences are seen at the extreme tails of the distribution. Girls tended to outperform boys at both tails of the distribution on reading measures, and in the lower percentiles of mathematics and science, while boys outperformed girls in the higher percentiles of mathematics and science. While the differences at the top of the distribution were of note, they called attention to the fact that inequities in the lower percentiles of the distribution were much more striking.

Baye and Monseur (2016) also examined the variance ratios of boys and girls on these assessments and found that in 93% of cases, variances for boys were higher. The finding of greater male variances in assessments here is not in and of itself original and has been noted in studies for many decades (although rarely as a core focus). The “greater male variability” hypothesis in fact has its roots in the 19th century (Ellis 1894). However, if we are to understand differences between the sexes at different points of the distribution, we must attempt to determine how their respective distributions differ. It is to the issue of differences in variability, not average performance, that the rest of this paper attempts to address building on earlier work.

### **Male and female variability**

Differences in the spread of scores between males and females have been noted in educational assessments for a long time, although often with contrasting findings. Maccoby and Jacklin (1974) showed that males were more variable than females in mathematical and spatial abilities, whereas variances showed parity in verbal measures. Feingold (1992) found larger male variances in the domains of general reasoning, mechanical reasoning, abstract reasoning, quantitative and spatial abilities, perceptual speed, memory and on verbal test batteries. Strand et al. (2006) found similar patterns in the domains of verbal, quantitative and non-verbal reasoning on a representative sample of 11-year olds in the UK, with greater male variances ranging between 7 and 17%. Similar results on U.S. students were found by Lohman and Lakin (2009) and later, Lakin (2013). IQ scores have also shown to reflect the same pattern (Johnson et al. 2008). Finally, assessments of non-cognitive and behavioural domains such as creativity (He et al. 2013; Karowski et al. 2016), sensation seeking (Cross et al. 2011), personality (Borkenau et al. 2013) and aggression (Archer and Mehdikhani 2003) appear subject to the effect. Combine these findings with the work reported earlier from Baye and Monseur and the fact that the above represents only a fraction of reported findings, one can see why many consider greater male variability to be ubiquitous.

Yet despite the volume of work related to differences in variances between the sexes, there has been little systematic attempt to explain this phenomenon (either partially or in its entirety). This is likely in part due to the contention that studies on sex differences in abilities tends to bring with it. Feingold (1992) noted that the explanation for greater male variability has become a polarised nature versus nurture debate. As a result,

many empirical papers avoid proposing an explanation. Johnson et al. (2008) point out that although results have often seemed clear, studies are often attacked on methodological grounds pertaining to sample size, representativeness, sample selectivity and age amongst other things. While it is not our intent to repeat the full history of the greater male variability hypothesis (see Johnson et al. for an in-depth review) we will briefly consider some of the proposed explanations for this effect.

### **Explanations for greater male variability**

As Feingold claimed, arguments regarding biological innateness are often invoked for theories of sex differences in cognitive and behavioural domains. Early theories (Ounsted and Taylor 1972) focused on the Y chromosome, claiming that differences in gene expression resulted in slower development and expressed more harmful as well as more beneficial traits, which would presumably lead to more variability in males. Gualtieri and Hicks (1985) suggested such differences could emerge from differences in the uterine environment, making males more differentially susceptible to physical and psychological disorders over the lifespan.

Evolutionary theories suggest that ancient adaptive mechanisms produced greater male variability to enhance survival in ancestral environments and that they are still in operation today. Evolutionary theories are based on sexual selection theory and parental investment theories (see Archer and Mehdikhani 2003 for a comprehensive review) and they would ultimately result in males showing greater variation across a range of traits in order to ensure reproductive fitness. Hill (2017) proposed two mathematical models simulating how one sex could have become more variable over evolutionary time if one sex in our ancestral past (presumably females in the case of homo sapiens, although Hill makes no explicit assumption) is more selective of the other for the purposes of mating, and that this greater variability will be independent of other measures of central tendency. Hill also suggested that in such circumstances where the selective sex is no longer being as selective, greater variability in the selected sex may in fact decline over successive generations. No direct test of this latter hypothesis has been made however.

While many support the biological and evolutionary basis for greater male variability, there are some shortfalls in this interpretation, as well as additional potential explanations as to why males are perhaps more variable. Miller (2001) claimed that susceptibility to defects resulting from prenatal conditions would only explain why males are overrepresented in the lower, not the higher tail of a distribution. As early as (1922), Hollingworth argued for an explanation based on gender roles, claiming that male employment, compared to the more restricted home role of women, allowed them the opportunity for greater diversification in education and environmental experiences. Noddings (1992) highlighted the issue of conformity, claiming that while most girls worked hard enough to avoid being in the bottom of the distribution in class, brighter girls are often pressured into not demonstrating the full extent of their abilities. Ceci et al. (2009) argued that biological accounts of differences in quantitative fields between the sexes are largely inconsistent and suggested that female preferences were a better explanation of underrepresentation in some professions. Critics of the evolutionary perspective also argue that if this phenomenon resulted from innate, evolved mechanisms, invariance of this effect across cultures would be expected. Several previous studies indicate that some

nations show greater male variation, others greater female variation and many show homogeneity of variance (Feingold 1994). Feingold went on to attribute heterogeneity in his data to social and cultural factors rather than any innate biological mechanism. Feingold (1992) also argued that national test norms alone may not be sufficiently generalizable to afford definitive proof of a biological origin of greater male variability. However, more recent studies using international assessments such as PISA, PIRLS and TIMSS do seem to suggest that variability is greater for males in the domains of reading and mathematics across cultures (Baye and Monseur 2016; Machin and Pekkarinen 2008).

There has been some suggestion that elements of test design may also play a role in magnifying sex differences in terms of measures of central tendency and variances. Spelke (2005) claimed that supposed differences in ability, particularly in mathematics and science, resulted largely from item and test biases favouring males, and that research generally fails to support the greater male variability hypothesis in these domains. Lakin (2013) supports this to an extent, suggesting that changes to Cognitive Ability Tests (specifically, the introduction of new quantitative reasoning items with a lesser verbal load) may have been responsible for shifting more males into the upper echelons of the distribution compared with earlier versions of the assessment. Strand et al. however found few substantive sex differences related to item difficulties in non-verbal and verbal batteries and suggested that test construction was unlikely the root cause of differences in variability. They made a tentative suggestion that a speed-accuracy trade off favouring boys may account for some of the variability differences in quantitative domains, but cautiously note that that previous research has mirrored these effects in untimed assessments (such as Feingold 1992). Lakin also noted that the consistent trend of increasing variance ratios between cognitive ability tests at grades 4 and 7 is likely to be something more systematic than simple test design and potentially reflects changes to society in terms of educational opportunity and personal educational preferences. Arguments focussing purely on test construction and procedure are thus hard to substantiate in the current literature.

Machin and Pekkarinen (2008) highlighted a compositional effect of sex differences in central tendency and distribution of scores. In their analysis of TIMSS and PIRLS data in 15-year olds, they noted that greater male variance in maths was attributable to overrepresentation of males in the higher part of the test distribution, with males outperforming females on average. In reading, male overrepresentation was largely at the bottom of the distribution, with females outperforming males on average. Indeed, Nowell and Hedges (1998) found a correlation of 0.74 between variance ratios and male–female effect sizes. Baye and Monseur found a smaller overall correlation of 0.42. However, they noted that the strength of the relationship varied by the point in the distribution. At the 5th percentile, the relationship was 0.50. At the 95th percentile, this had declined to 0.31. These results seem to suggest that variability for males increases in line with superior female performance, particularly at the lower end of the distribution.

### **The current study**

While Feingold's work (1994) failed to show a consistent greater male variance in international test scores, this could be attributable to the methodology. He conducted a meta-analysis by searching the literature for reading, mathematics and spatial measures, which

carries many issues with it including many different tests, test administrations, issues of representation etc. Baye and Monseur (2016), using more recently available international assessments (PISA, PIRLS and TIMSS) found different results, suggesting that greater male variability was effectively universal. They found that variances (on average) were 15% greater for males in reading, 12% greater in maths and 14% greater in science. Even using Feingold's (1994) conservative estimate of any ratio falling between 0.90 and 1.10 as not representing evidence of greater variance, Baye and Monseur's work is suggestive of greater male variability. The advantage of using these international assessments is that they are designed to be internationally comparable, with representative samples of children selected in each country and administered in a standardised fashion. This helps remove potentially confounding factors that may impact on assessment results.

However, Baye and Monseur's work leaves many questions unanswered. How similar are countries to each other in terms of variance ratios, and are there some that are much more male biased than others? If countries vary in terms of male and female variances, are there any recorded factors that may account for this? Baye and Monseur did make some attempt to look at differences between primary and secondary school measures, as well as by IEA and OECD membership, but beyond this, no systematic heterogeneity analysis was conducted. Yet analysing heterogeneity is important and can be revealing. Furthermore, this international data could be linked to cross-country metrics that may elucidate meaningful patterns of variation. For example, Borkenau et al. (2013) showed that differences across countries in variances in personality were significantly linked to national measures of gender inequality and human development. Given earlier suggestions by Hollingworth (1922) that variances favouring males are largely due to gender roles, and later works (Ceci et al. 2009; Lakin 2013) suggesting that societal practices and female choice are likely to have a major impact on variance ratios, international indices of societal development, particularly forms of gender inequality, are potential sources that could be used to explain any cross-national heterogeneity. To our knowledge, this has not been examined in the context of large-scale international assessments.

In this study, we attempt to answer these questions and extend our knowledge surrounding the nature of greater male variability. We examined the same data sets used by Baye and Monseur, with the addition of more recent test administrations from years 2015 and 2016, to (1) replicate their findings using meta-analysis, (2) determine if greater male variability is homogenous both within and between countries and (3) quantify any meaningful sources of heterogeneity. For the purposes of the third aim, we link these data to international metrics on human progress (Human Development Index) and male–female participation in education, labour forces and politics (Global Gender Gap Index) as well as examining test specific factors such as grade, test, OECD membership, the size of the male–female difference at the mean and national means.

## **Method**

### **Data sources**

Data from three major international assessments were selected to allow an examination of variance ratios across countries: OECD PISA (Programme for International Student Assessment; 2000, 2003, 2006, 2009, 2012, 2015), IEA PIRLS (Progress in International Reading Literacy Study; 2001, 2006, 2011) and IEA TIMSS (Trends in International

Mathematics and Science Study; 1995, 1998, 1999, 2003, 2007, 2008, 2011, 2015). These were selected due to having multiple testing points over time and having a wide coverage of countries across the globe. All data is freely available from the OECD website (<http://www.pisa.oecd.org>) and IEA Study Data Repository (<http://rms.iea-dpc.org>). Methodological information is available in the technical reports on each survey (Adams and Wu 2002; Martin et al. 2000, 2003, 2004, 2007, 2016; Martin and Kelly 1996, 1997; Martin and Mullis 1996, 2012; OECD 2005, 2009a, 2014, 2016; Olson et al. 2008).

International data on Human Development was also collected where available for each country. The Human Development Index (HDI) is made up of four sub-factors: expected years of schooling for children of school entry age, mean years of schooling for adults aged 25 and above, life expectancy and gross national income per capita (GNI). This data is freely available from the United Nations Development program website (<http://hdr.undp.org/en/data>).

International data on gender inequality was also gathered from the Global Gender Gap project. The Global Gender Gap Index (GGGI) is made of four sub-factors: economic participation, educational attainment, health and survival and political empowerment. Each factor represents an outcome and is measured on a scale of 0 to 1, where a score of 1 would represent parity between males and females. Data is freely available from the World Economic Forum's website (<http://reports.weforum.org>).

### Sample

Data from each country surveyed within each of the assessments was included in this analysis. For the purposes of this study, we used measures from three content areas: literacy, maths literacy and science literacy. In total, we included 564 cases for literacy, 1054 cases for mathematics literacy and 991 cases for science literacy gathered from over 100 nations worldwide (where each case represents a national test occurrence within a given year and within a specific content area). In terms of population size across all cases, in mathematics literacy it consists of 2,507,046 males and 2,512,273 females, for reading 1,471,698 males and 1,486,578 females and for science literacy 2,512,559 males and 2,515,645 females. It should be noted that for science literacy, we did not use data from TIMSS Advanced as these measures focussed on concepts from Physics only.

### Data calculations

Statistics were calculated by generating means and standard deviations for males and females within each country for each measure within each assessment. These were calculated using each of the five plausible values within each database and aggregated according to the methodologies supplied by the OECD and IEA in their analyses manuals (OECD 2009b; Martin et al. 2016). Standard errors for these statistics were calculated using replicate weights within each database (80 Fay weights in PISA and 75 JK2 replicates in PIRLS and TIMSS). SPSS (V22; IBM Corp 2013) was used to calculate these statistics (see OECD 2009b; Martin et al. 2016 for technical details regarding the SPSS macros used to compute these statistics).

Variances were calculated from the standard deviations. The ratio of male to female variances was taken by dividing the male variance by the female variance. A variance ratio greater than one would indicate that the male variance is higher than the female

variance. Variance ratios are a common method of examining variability between the sexes (see Hedges and Friedman 1993; Baye and Monseur 2016). In keeping with previous authors (Hedges and Friedman 1993; Katzman and Alliger 1992), but not Baye and Monseur (2016), ratios were logarithmically transformed to increase precision of the estimates and to avoid overestimation, as it ensures a normal distribution. Assuming that the log of the variances follows a normal distribution, the variances of these ratios were then calculated as:

$$v = 2/(n_f - 1) + 2/(n_m - 1)$$

As we are examining variance ratios by country, some of the data points were combined for the purposes of the analysis. Countries such as Italy, Spain, Canada and the United States often report data for sub-regions but not consistently over assessments. These were collapsed for the purposes of this study. Where a nation has national and regional data within a given test administration, the subnational data points were used. China and the United Kingdom also report at the level of autonomous states (England, Scotland, Northern Ireland, Taipei, Machao, Shangai and Hong Kong). Countries falling into these states are denoted in the table but are not considered separately for aggregation. Assessments were considered together regardless of whether they were done in the primary or secondary years.

### **Meta-analysis**

To examine the overall size of the variance ratio and to meaningfully quantify heterogeneity, meta-analyses were conducted using Comprehensive Meta-Analysis Version 3 (Borenstein et al. 2013). Many traditional analyses assume that effect size parameters are fixed and relatively homogenous. In this study, we are not assuming homogeneity of these parameters and are thus implementing a random effects model, assuming that effect size parameters are randomly sampled. The use of a random effects model is appropriate where heterogeneity is expected. In this study, we examined heterogeneity by country, whether the countries were OECD member states, test and grade.

Heterogeneity is examined by calculating Q statistics, which can be used to test for equality of effect sizes within and between analysis categories and follow the formulae below:

$$Q = \sum_{i=1}^k w(d_i - \bar{d})^2,$$

where  $w = 1/v$ ,  $v = (N_{\text{male}} + N_{\text{female}})/N_{\text{total}} + d^2/2(N_{\text{total}})$ , and  $k$  is the number of effect sizes.

Q statistics follow a Chi square distribution of  $k - 1$  degrees of freedom (Hedges and Olkin 1985). While significant Q statistics can detect the presence of homogeneity, they are not indicative of its magnitude. They are also sensitive to sample size (Hardy and Thompson 1998; Higgins and Thompson 2002) and its presence is generally expected when analysing large numbers of studies (Higgins 2008).

The mean of the log variance ratios, standard errors and confidence intervals for each country were then calculated (and presented in their un-transformed format for ease of understanding). For each country, we also tabulated the proportion of studies where; (1)

the variances were significantly larger for males, (2) the variances were larger for males but not significantly so, (3) the variances were greater for females but not significantly so and finally (4) the variances were significantly greater for females.

### **Meta-regression**

Meta-regression was used to explore and quantify potential sources of heterogeneity. We recorded the mean test score for each country in each year and calculated a weighted effect size of the gender difference between male and female means, as previous work has suggested that this effect size is related to the variance ratio (Baye and Monseur 2016). This was taken as the female mean subtracted from the male mean (a negative score therefore suggests higher scores for females). Using SPSS, this was converted into a standardised effect size (Hedges  $g$ ) calculated from the effect size  $d$  multiplied by the correction factor  $J$  (correcting for small sample sizes):

$$d = \frac{\mu_1 - \mu_2}{SD_{pooled}}$$

$$J = 1 - \frac{3}{4df - 1}$$

Other additional moderators were derived from test administrations. Previous researchers (discussed earlier) have suggested that some differences may result from test design. As such, the test type, year, test grade and OECD membership were included as moderators to determine if these had a substantial impact on heterogeneity. Baye and Monseur (2016) found small differences in variance ratios between these variables and thus they may be contributing to some of the heterogeneity. Alongside these, the subfactors of the HDI and the GGGI were included to see if other country level contributing factors could account for variation across countries. As consistent data for both these indices is only available from 2006, meta-regression was performed only on cases from test administrations from 2006 onwards.

## **Results**

Analysis of each content domain is presented separately. Countries with only one or two data points are included in the analysis although conclusions about the stability of their variance ratios must be treated cautiously. Variance ratios and their confidence intervals are presented in their un-transformed form for ease of interpretation. The percentage of cases that have a variance ratio below (significantly and non-significantly) and above (significantly and non-significantly) 1, with ratios above 1 representing greater male variance, are also presented. Q statistics and their significance are also reported for each nation.

### **Mathematics literacy**

Table 1 shows the results for this analysis on international mathematics literacy data sources. Each of the 102 individual participating nations is listed in alphabetical order.

For mathematics literacy, variance ratios across nations range between 0.96 (Algeria) and 1.43 (Saudi Arabia), the average being 1.12. Data from 102 nations clearly shows

**Table 1 Variance ratios, confidence intervals and heterogeneity statistics for countries in the domain of mathematics literacy**

Country	k	VR	LCI	UCI	p	Q	Q (p)	Below 1 (sig) (%)	Below 1 (nsig) (%)	Above 1 (nsig) (%)	Above 1 (sig) (%)
Albania	4	1.13	1.06	1.20	0.00	6.28	0.10	0.00	0.00	25.00	75.00
Algeria	3	0.96	0.92	1.01	0.10	1.25	0.53	0.00	100.00	0.00	0.00
Argentina	9	1.12	1.02	1.23	0.01	69.53	0.00	0.00	33.30	11.10	55.60
Armenia	7	1.09	1.03	1.16	0.00	20.16	0.00	0.00	14.30	14.30	71.43
Australia	21	1.13	1.09	1.16	0.00	87.27	0.00	0.00	4.80	28.56	66.64
Austria	14	1.09	1.05	1.14	0.00	42.49	0.00	7.14	0.00	42.85	50.01
Azerbaijan	3	1.14	0.99	1.31	0.07	20.26	0.00	0.00	0.00	33.30	66.70
Bahrain	6	1.40	1.33	1.48	0.00	11.55	0.04	0.00	0.00	0.00	100.00
Belgium	19	1.10	1.07	1.13	0.00	56.95	0.00	5.26	0.00	47.37	47.37
Bosnia/Herzegovina	1	1.06	0.97	1.15	0.20	0.00	1.00	0.00	0.00	100.00	0.00
Botswana	5	1.13	1.06	1.20	0.00	11.91	0.02	0.00	0.00	20.00	80.00
Brazil	6	1.11	1.06	1.16	0.00	12.29	0.03	0.00	0.00	16.70	83.30
Bulgaria	9	1.12	1.06	1.17	0.00	20.53	0.01	0.00	0.00	44.40	55.60
Canada	36	1.10	1.07	1.13	0.00	161.38	0.00	0.00	0.00	44.45	55.55
Chile	11	1.10	1.07	1.14	0.00	23.15	0.01	0.00	9.10	18.20	72.70
China	38	1.17	1.14	1.20	0.00	126.48	0.00	0.00	0.00	21.06	78.94
<sup>a</sup> Taipei	13	1.17	1.13	1.22	0.00	40.06	0.00	0.00	0.00	7.70	92.30
<sup>a</sup> Macao	5	1.15	1.07	1.23	0.00	14.09	0.01	0.00	0.00	20.00	80.00
<sup>a</sup> Hong Kong	19	1.18	1.13	1.22	0.00	64.45	0.00	0.00	0.00	31.57	68.43
Colombia	8	1.11	1.07	1.15	0.00	10.44	0.17	0.00	12.50	12.50	75.00
Costa Rica	3	1.14	1.10	1.19	0.00	1.95	0.38	0.00	0.00	0.00	100.00
Croatia	6	1.16	1.13	1.20	0.00	4.94	0.42	0.00	0.00	0.00	100.00
Cyprus	12	1.15	1.10	1.20	0.00	23.11	0.02	0.00	0.00	50.00	50.00
Czech Republic	17	1.09	1.05	1.13	0.00	43.87	0.00	0.00	0.00	58.84	41.16
Denmark	13	1.07	1.04	1.11	0.00	15.96	0.19	0.00	7.69	53.85	38.46
Dominican Republic	1	1.11	1.02	1.20	0.01	0.00	1.00	0.00	0.00	0.00	100.00
Egypt	3	1.08	1.02	1.14	0.01	4.43	0.11	0.00	0.00	33.30	66.70

**Table 1 (continued)**

Country	k	VR	LCI	UCI	P	Q	Q (p)	Below 1 (sig) (%)	Below 1 (nsig) (%)	Above 1 (nsig) (%)	Above 1 (sig) (%)
El Salvador	2	1.01	0.88	1.16	0.88	5.22	0.02	0.00	50.00	50.00	0.00
Estonia	5	1.09	0.99	1.19	0.08	28.26	0.00	0.00	40.00	0.00	60.00
Finland	10	1.13	1.07	1.20	0.00	38.20	0.00	0.00	0.00	40.00	60.00
France	12	1.12	1.07	1.18	0.00	29.78	0.00	0.00	8.33	33.33	58.34
Georgia	8	1.18	1.14	1.21	0.00	7.23	0.41	0.00	0.00	12.50	87.50
Germany	13	1.09	1.04	1.14	0.00	33.52	0.00	0.00	7.69	53.85	38.46
Ghana	3	1.02	0.97	1.06	0.48	1.16	0.56	0.00	33.30	66.70	0.00
Greece	11	1.15	1.07	1.24	0.00	57.69	0.00	0.00	9.09	54.56	36.35
Honduras	2	1.01	0.94	1.09	0.73	1.35	0.25	0.00	50.00	50.00	0.00
Hungary	20	1.09	1.07	1.12	0.00	32.46	0.03	0.00	5.00	50.00	45.00
Iceland	11	1.11	1.04	1.18	0.00	32.37	0.00	0.00	9.11	63.62	27.27
India	2	1.19	1.07	1.31	0.00	0.77	0.38	0.00	0.00	50.00	50.00
Indonesia	11	1.05	1.01	1.10	0.03	36.51	0.00	0.00	27.30	54.50	18.20
Iran	14	1.14	1.08	1.20	0.00	78.23	0.00	0.00	0.00	35.71	64.29
Ireland	13	1.12	1.08	1.16	0.00	21.52	0.04	0.00	0.00	46.15	53.85
Israel	12	1.18	1.11	1.26	0.00	58.02	0.00	0.00	16.67	0.00	83.33
Italy	38	1.20	1.14	1.25	0.00	187.23	0.00	2.63	0.00	34.20	63.17
Japan	19	1.15	1.11	1.18	0.00	40.72	0.00	0.00	0.00	31.58	68.42
Jordan	9	1.26	1.17	1.36	0.00	65.59	0.00	0.00	0.00	11.10	88.90
Kazakhstan	8	1.13	1.06	1.20	0.00	30.63	0.00	0.00	0.00	50.00	50.00
Korea	17	1.12	1.08	1.17	0.00	51.06	0.00	0.00	0.00	29.41	70.59
Kosovo	1	1.13	1.04	1.22	0.00	0.00	1.00	0.00	0.00	0.00	100.00
Kuwait	6	1.25	1.16	1.36	0.00	26.56	0.00	0.00	0.00	0.00	100.00
Kyrgyzstan	2	1.21	1.14	1.28	0.00	0.44	0.51	0.00	0.00	0.00	100.00
Latvia	14	1.13	1.09	1.18	0.00	26.54	0.01	0.00	0.00	42.86	57.14
Lebanon	7	1.09	1.05	1.14	0.00	6.22	0.40	0.00	0.00	57.10	42.90

**Table 1 (continued)**

Country	k	VR	LCI	UCI	P	Q	Q (p)	Below 1 (sig) (%)	Below 1 (nsig) (%)	Above 1 (nsig) (%)	Above 1 (sig) (%)
Liechtenstein	5	1.19	1.07	1.32	0.00	12.12	0.02	0.00	20.00	20.00	60.00
Lithuania	17	1.10	1.07	1.13	0.00	24.87	0.07	0.00	11.76	35.29	52.95
Luxembourg	6	1.18	1.13	1.23	0.00	7.71	0.17	0.00	0.00	16.70	83.30
Macedonia	5	1.07	1.02	1.11	0.00	4.56	0.34	0.00	20.00	40.00	40.00
Malaysia	8	1.13	1.08	1.18	0.00	22.95	0.00	0.00	0.00	12.50	87.50
Malta	5	1.16	1.11	1.20	0.00	0.69	0.95	0.00	0.00	0.00	100.00
Mauritius	1	1.10	1.02	1.19	0.01	0.00	1.00	0.00	0.00	0.00	100.00
Mexico	6	1.19	1.14	1.24	0.00	16.33	0.01	0.00	16.70	0.00	83.30
Miranda-Venezuela	1	1.04	0.96	1.12	0.30	0.00	1.00	0.00	0.00	100.00	0.00
Moldova	5	1.09	1.04	1.16	0.00	7.74	0.10	0.00	0.00	40.00	60.00
Mongolia	2	1.02	0.96	1.08	0.52	0.18	0.67	0.00	0.00	100.00	0.00
Montenegro	4	1.19	1.05	1.35	0.01	31.79	0.00	0.00	0.00	25.00	75.00
Morocco	9	1.01	0.98	1.04	0.46	11.91	0.16	0.00	44.44	44.44	11.11
Netherlands	18	1.05	1.00	1.11	0.05	105.60	0.00	0.00	16.66	61.10	22.24
New Zealand	19	1.14	1.12	1.17	0.00	30.89	0.03	0.00	0.00	31.58	68.42
Norway	21	1.12	1.08	1.16	0.00	52.24	0.00	0.00	9.52	38.08	52.40
Oman	5	1.29	1.17	1.41	0.00	46.98	0.00	0.00	0.00	0.00	100.00
Palestine	3	1.13	0.99	1.28	0.06	17.48	0.00	0.00	0.00	33.30	66.70
Panama	1	1.11	1.03	1.20	0.01	0.00	1.00	0.00	0.00	0.00	100.00
Peru	4	1.11	1.04	1.18	0.00	6.90	0.08	0.00	0.00	50.00	50.00
Philippines	4	1.07	0.97	1.19	0.20	22.75	0.00	0.00	25.00	25.00	50.00
Poland	8	1.18	1.13	1.23	0.00	13.50	0.06	0.00	0.00	12.50	87.50
Portugal	13	1.15	1.07	1.22	0.00	77.96	0.00	0.00	0.00	38.46	61.54
Qatar	10	1.24	1.18	1.30	0.00	56.54	0.00	0.00	0.00	10.00	90.00
Romania	12	1.13	1.04	1.22	0.00	115.15	0.00	0.00	16.67	25.00	58.33
Russia	23	1.09	1.06	1.13	0.00	85.47	0.00	0.00	8.70	34.78	56.52

**Table 1 (continued)**

Country	k	VR	LCI	UCI	P	Q	Q (p)	Below 1 (sig) (%)	Below 1 (nsig) (%)	Above 1 (nsig) (%)	Above 1 (sig) (%)
Saudi Arabia	6	1.43	1.25	1.63	0.00	72.75	0.00	0.00	0.00	0.00	100.00
Serbia	7	1.16	1.10	1.23	0.00	17.39	0.01	0.00	0.00	14.30	85.70
Singapore	18	1.16	1.12	1.20	0.00	72.40	0.00	0.00	5.52	16.67	77.80
Slovak Republic	12	1.09	1.06	1.13	0.00	29.34	0.00	0.00	0.00	50.00	50.00
Slovenia	21	1.11	1.07	1.15	0.00	61.58	0.00	0.00	4.80	42.82	52.38
South Africa	7	1.06	1.01	1.12	0.03	27.25	0.00	0.00	14.29	42.86	42.86
Spain	71	1.15	1.12	1.18	0.00	299.92	0.00	1.41	2.82	33.79	61.98
Sweden	19	1.09	1.05	1.12	0.00	40.51	0.00	0.00	5.26	52.63	42.11
Switzerland	10	1.10	1.04	1.16	0.00	44.32	0.00	0.00	0.00	60.02	39.98
Syria	3	1.07	1.02	1.12	0.01	0.92	0.63	0.00	0.00	66.70	33.30
Thailand	15	1.12	1.07	1.18	0.00	79.52	0.00	0.00	13.34	26.67	59.99
Trinidad and Tobago	2	1.08	1.02	1.14	0.01	0.67	0.41	0.00	0.00	50.00	50.00
Tunisia	12	1.03	0.98	1.08	0.20	47.16	0.00	16.70	16.70	33.30	33.30
Turkey	11	1.11	1.06	1.16	0.00	45.62	0.00	0.00	0.00	45.50	54.50
Ukraine	3	1.20	1.14	1.26	0.00	1.76	0.41	0.00	0.00	0.00	100.00
UAE	22	1.27	1.22	1.32	0.00	194.96	0.00	0.00	0.00	9.10	90.90
United Kingdom	34	1.11	1.08	1.14	0.00	106.36	0.00	2.91	0.00	38.26	58.84
England <sup>a</sup>	13	1.09	1.04	1.14	0.00	36.44	0.00	7.68	0.00	46.15	46.17
Northern Ireland <sup>a</sup>	2	1.07	0.97	1.19	0.19	2.45	0.12	0.00	0.00	50.00	50.00
Scotland <sup>a</sup>	13	1.13	1.08	1.19	0.00	41.60	0.00	0.00	0.00	46.15	53.85
United States	29	1.09	1.07	1.11	0.00	56.82	0.00	3.45	3.45	27.59	65.52
Uruguay	5	1.18	1.14	1.22	0.00	2.84	0.59	0.00	0.00	0.00	100.00
Viet Nam	2	1.10	1.04	1.16	0.00	0.35	0.55	0.00	0.00	0.00	100.00
Yemen	4	1.06	1.02	1.10	0.00	0.19	0.98	0.00	0.00	100.00	0.00
Yugoslavia	1	1.33	0.58	3.06	0.50	0.00	1.00	0.00	0.00	0.00	100.00
Overall	1054	1.118	1.114	1.123	0.000	5307.991	0.000	0.80	5.01	33.69	60.50

<sup>a</sup> Nations of the UK administered separately/<sup>ns</sup> Chinese districts administered separately

that in mathematics, less than 6% of recorded cases show larger variances for females. Almost 61% show significantly larger variances for males than females and less than 1% of ratios are significantly female biased. In 91 countries, the variances are significantly larger for males than females. In only one country (Algeria) was the opposite pattern found to be true (and with no evidence of heterogeneity) although this result is not significant. In 36 nations (35%), there is no significant evidence of heterogeneity. Heterogeneity is present in the remaining 65% of cases however and is present overall. Figure 1 demonstrates the ratios and 95% CIs graphically (in order from smallest to largest). As is evident, while many significantly differ from 1.00 and countries vary considerably, few countries significantly vary from each other in the domain of mathematics literacy.

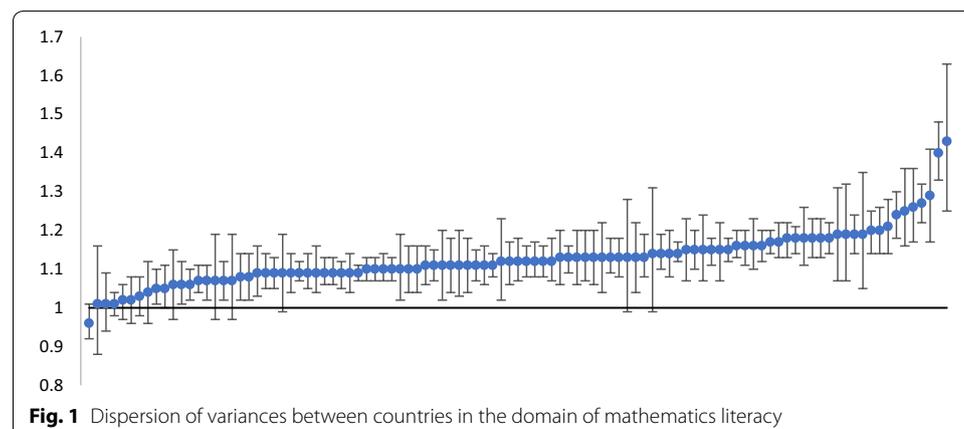
### Reading

Table 2 shows the results of the same analysis on international measures of reading. Note that while many countries are common to both assessments, this is not true of all of them.

In reading, around 95% of all assessments taken had wider variances for males than for females (almost 79% significantly so). Ratios range from 0.96 (Algeria) to 1.75 (Saudi Arabia). Only 4% of all cases were female biased (<1% significantly so). The average across countries was 1.16. Only two countries have a wider variance for females and these are each based on only one assessment point (Algeria and Belize). Across 87 countries, the variances are significantly male biased at the 5% level. Only 34 nations (37%) however don't show significant heterogeneity in their Q scores, and of these, 14 are single case nations where the figure can't be calculated. As the remaining 63% show significant heterogeneity, the data cannot be considered homogenous. Figure 2 demonstrates the ratios and 95% CIs graphically (in order from smallest to largest). As is evident, while many significantly differ from 1.00 and countries vary considerably, few countries have ratios that significantly vary from each other in the domain of reading. Those that do differ significantly from each other tend to be positioned at the tails of the distribution.

### Science literacy

For science literacy, we conducted the same analysis as in the previous two cognitive domains (Table 3).



**Table 2 Variance ratios, confidence intervals and heterogeneity statistics for countries in the domain of reading**

Country	k	VR	LCI	UCI	p	Q	p(Q)	Below (sig) (%)	Below (nsig) (%)	Above (nsig) (%)	Above (sig) (%)
Albania	4	1.17	1.04	1.32	0.01	24.45	0.00	0.00	0.00	25.00	75.00
Algeria	1	0.96	0.89	1.04	0.32	0.00	1.00	0.00	100.00	0.00	0.00
Argentina	7	1.22	1.08	1.39	0.00	68.47	0.00	0.00	14.29	14.29	71.43
Australia	7	1.20	1.15	1.26	0.00	27.21	0.00	0.00	0.00	0.00	100.00
Austria	8	1.15	1.09	1.20	0.00	24.00	0.00	0.00	0.00	12.50	87.50
Azerbaijan	3	1.20	1.00	1.43	0.05	32.38	0.00	0.00	0.00	33.33	66.67
Belgium	12	1.16	1.12	1.20	0.00	30.91	0.00	0.00	0.00	25.00	75.00
Belize	1	0.97	0.88	1.07	0.55	0.00	1.00	0.00	100.00	0.00	0.00
Botswana	1	1.19	1.09	1.28	0.00	0.00	1.00	0.00	0.00	0.00	100.00
Brazil	6	1.17	1.13	1.21	0.00	7.89	0.16	0.00	0.00	0.00	100.00
Bulgaria	8	1.14	1.07	1.22	0.00	27.37	0.00	0.00	0.00	25.00	75.00
Canada	18	1.11	1.05	1.15	0.00	173.05	0.00	0.00	11.11	38.89	50.00
Chile	5	1.13	1.07	1.19	0.00	11.13	0.03	0.00	0.00	20.00	80.00
China	23	1.20	1.16	1.25	0.00	123.64	0.00	0.00	0.00	8.70	91.30
<sup>a</sup> Taipei	6	1.15	1.09	1.22	0.00	23.65	0.00	0.00	0.00	16.67	83.33
<sup>a</sup> Macao	5	1.23	1.12	1.36	0.00	28.07	0.00	0.00	0.00	0.00	100.00
<sup>a</sup> Hong Kong	9	1.25	1.19	1.31	0.00	19.90	0.01	0.00	0.00	0.00	100.00
<sup>a</sup> Shanghai	2	1.21	1.12	1.31	0.00	8.00	0.00	0.00	0.00	0.00	100.00
Colombia	6	1.07	1.01	1.14	0.03	24.20	0.00	0.00	16.67	33.33	50.00
Costa Rica	3	1.09	1.03	1.17	0.01	8.07	0.02	0.00	0.00	33.33	66.67
Croatia	5	1.16	1.12	1.20	0.00	3.59	0.46	0.00	0.00	0.00	100.00
Cyprus	1	1.07	0.97	1.19	0.16	0.00	1.00	0.00	0.00	100.00	0.00
Czech Republic	8	1.15	1.11	1.20	0.00	13.88	0.05	0.00	0.00	0.00	100.00
Denmark	8	1.08	1.05	1.12	0.00	8.99	0.25	0.00	0.00	25.00	75.00
Dominican Republic	1	1.11	1.02	1.20	0.01	0.00	1.00	0.00	0.00	0.00	100.00
Estonia	4	1.13	1.07	1.17	0.00	4.17	0.24	0.00	0.00	25.00	75.00
Finland	7	1.16	1.07	1.25	0.00	31.03	0.00	0.00	0.00	28.57	71.43

**Table 2 (continued)**

Country	k	VR	LCI	UCI	p	Q	p(Q)	Below (sig) (%)	Below (nsig) (%)	Above (nsig) (%)	Above (sig) (%)
France	9	1.16	1.09	1.23	0.00	31.55	0.00	0.00	0.00	33.33	66.67
Georgia	4	1.14	1.08	1.20	0.00	4.12	0.25	0.00	0.00	0.00	100.00
Germany	9	1.14	1.07	1.21	0.00	39.35	0.00	0.00	0.00	33.33	66.67
Greece	7	1.30	1.17	1.43	0.00	42.82	0.00	0.00	0.00	0.00	100.00
Honduras	1	1.04	0.94	1.15	0.42	0.00	1.00	0.00	0.00	100.00	0.00
Hungary	9	1.14	1.08	1.20	0.00	25.18	0.00	0.00	0.00	22.22	77.78
Iceland	9	1.20	1.12	1.28	0.00	23.18	0.00	0.00	0.00	22.22	77.78
India	2	1.34	1.20	1.48	0.00	0.88	0.35	0.00	0.00	0.00	100.00
Indonesia	8	1.09	1.03	1.16	0.00	26.54	0.00	0.00	0.00	37.50	62.50
Iran	3	1.03	0.98	1.07	0.22	0.56	0.76	0.00	0.00	100.00	0.00
Ireland	7	1.20	1.13	1.28	0.00	19.38	0.00	0.00	0.00	14.29	85.71
Israel	8	1.14	1.03	1.27	0.01	71.46	0.00	0.00	0.00	37.50	62.50
Italy	28	1.26	1.19	1.34	0.00	166.89	0.00	0.00	3.57	10.71	85.71
Japan	6	1.27	1.19	1.36	0.00	19.04	0.00	0.00	0.00	0.00	100.00
Jordan	4	1.38	1.17	1.60	0.00	45.54	0.00	0.00	0.00	0.00	100.00
Kazakhstan	2	1.32	1.19	1.48	0.00	1.07	0.30	0.00	0.00	0.00	100.00
Korea	6	1.25	1.17	1.32	0.00	12.56	0.03	0.00	0.00	0.00	100.00
Kosovo	1	1.17	1.08	1.27	0.00	0.00	1.00	0.00	0.00	0.00	100.00
Kuwait	3	1.13	1.01	1.27	0.03	10.62	0.00	0.00	0.00	33.33	66.67
Kyrgyzstan	2	1.21	1.14	1.27	0.00	0.03	0.86	0.00	0.00	0.00	100.00
Latvia	8	1.15	1.09	1.21	0.00	21.44	0.00	0.00	12.50	0.00	87.50
Lebanon	1	1.13	1.04	1.22	0.00	0.00	1.00	0.00	0.00	0.00	100.00
Liechtenstein	5	1.23	1.16	1.31	0.00	4.31	0.37	0.00	20.00	20.00	60.00
Lithuania	7	1.12	1.06	1.17	0.00	16.22	0.01	0.00	0.00	28.57	71.43
Luxembourg	7	1.19	1.11	1.26	0.00	26.27	0.00	0.00	0.00	14.29	85.71
Macedonia	3	1.02	0.96	1.08	0.55	2.48	0.29	0.00	33.33	33.33	33.33
Malaysia	2	1.22	1.13	1.31	0.00	1.56	0.21	0.00	0.00	0.00	100.00

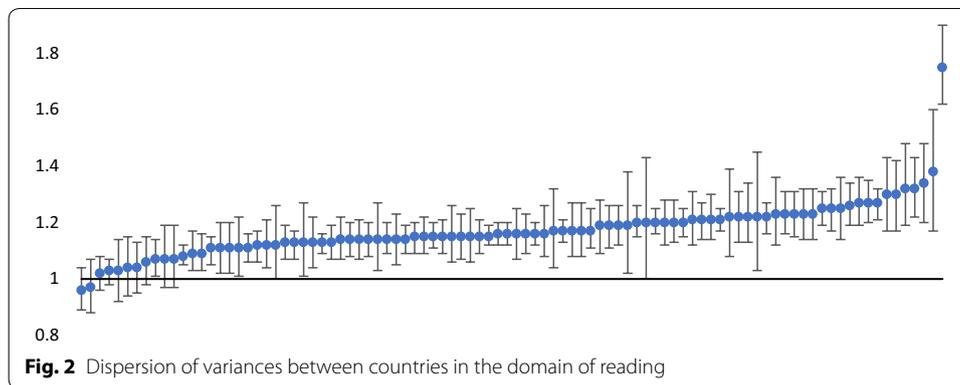
**Table 2 (continued)**

Country	k	VR	LCI	UCI	p	Q	p(Q)	Below (sig) (%)	Below (nsig) (%)	Above (nsig) (%)	Above (sig) (%)
Malta	4	1.14	1.09	1.20	0.00	1.22	0.75	0.00	0.00	25.00	75.00
Mauritius	1	1.06	0.98	1.15	0.13	0.00	1.00	0.00	0.00	100.00	0.00
Mexico	6	1.22	1.13	1.34	0.00	52.55	0.00	0.00	0.00	16.67	83.33
Miranda-Venezuela	1	1.11	1.02	1.20	0.01	0.00	1.00	0.00	0.00	0.00	100.00
Moldova	4	1.15	1.06	1.26	0.00	10.64	0.01	0.00	0.00	25.00	75.00
Montenegro	4	1.25	1.14	1.36	0.00	14.38	0.00	0.00	0.00	0.00	100.00
Morocco	4	1.11	1.01	1.22	0.03	18.77	0.00	0.00	0.00	25.00	75.00
Netherlands	9	1.12	1.04	1.21	0.00	71.51	0.00	0.00	0.00	44.44	55.56
New Zealand	9	1.20	1.15	1.25	0.00	17.58	0.02	0.00	0.00	11.11	88.89
Norway	10	1.19	1.12	1.26	0.00	40.06	0.00	0.00	10.00	10.00	80.00
Oman	1	1.27	1.20	1.35	0.00	0.00	1.00	0.00	0.00	0.00	100.00
Panama	1	1.14	1.05	1.23	0.00	0.00	1.00	0.00	0.00	0.00	100.00
Peru	4	1.12	1.00	1.26	0.05	24.39	0.00	0.00	25.00	0.00	75.00
Poland	8	1.21	1.14	1.30	0.00	32.30	0.00	0.00	0.00	12.50	87.50
Portugal	7	1.30	1.17	1.42	0.00	51.33	0.00	0.00	0.00	14.29	85.71
Qatar	6	1.23	1.15	1.31	0.00	32.84	0.00	0.00	0.00	0.00	100.00
Romania	8	1.19	1.02	1.38	0.03	198.23	0.00	0.00	0.00	37.50	62.50
Russia	10	1.15	1.07	1.23	0.00	71.15	0.00	0.00	10.00	20.00	70.00
Saudi Arabia	1	1.75	1.62	1.90	0.00	0.00	1.00	0.00	0.00	0.00	100.00
Serbia	3	1.22	1.03	1.45	0.02	28.67	0.00	0.00	0.00	0.00	100.00
Singapore	6	1.16	1.12	1.20	0.00	4.03	0.54	0.00	16.67	0.00	83.33
Slovak Republic	8	1.14	1.09	1.19	0.00	24.71	0.00	0.00	0.00	37.50	62.50
Slovenia	7	1.23	1.14	1.32	0.00	37.38	0.00	0.00	0.00	0.00	100.00
South Africa	2	1.03	0.92	1.14	0.64	4.17	0.04	0.00	50.00	50.00	0.00
Spain	51	1.21	1.17	1.25	0.00	378.02	0.00	3.92	5.88	5.88	84.31
Sweden	10	1.11	1.06	1.16	0.00	22.19	0.01	0.00	10.00	20.00	70.00
Switzerland	6	1.13	1.09	1.16	0.00	3.26	0.66	0.00	0.00	16.67	83.33

**Table 2 (continued)**

Country	k	VR	LCI	UCI	p	Q	p(Q)	Below (sig) (%)	Below (nsig) (%)	Above (nsig) (%)	Above (sig) (%)
Thailand	6	1.23	1.14	1.32	0.00	27.71	0.00	0.00	0.00	0.00	100.00
Trinidad and Tobago	4	1.13	1.07	1.19	0.00	3.28	0.35	0.00	0.00	25.00	75.00
Tunisia	5	1.15	1.06	1.25	0.00	20.43	0.00	0.00	0.00	20.00	80.00
Turkey	6	1.16	1.08	1.26	0.00	27.33	0.00	0.00	0.00	16.67	83.33
UAE	9	1.27	1.21	1.32	0.00	38.36	0.00	0.00	0.00	0.00	100.00
United Kingdom	16	1.15	1.09	1.21	0.00	97.27	0.00	0.00	6.25	25.00	68.75
England <sup>a</sup>	3	1.04	0.95	1.13	0.00	4.85	0.09	0.00	25.00	25.00	50.00
Scotland <sup>a</sup>	6	1.17	1.08	1.27	0.00	23.75	0.00	0.00	0.00	33.33	66.67
Northern Ireland <sup>a</sup>	1	1.07	0.97	1.19	0.00	0.00	0.00	0.00	0.00	100.00	0.00
United States	11	1.15	1.12	1.19	0.00	13.11	0.22	0.00	0.00	27.27	72.73
Uruguay	5	1.22	1.16	1.27	0.00	6.43	0.17	0.00	0.00	0.00	100.00
Viet Nam	2	1.17	1.11	1.25	0.00	0.13	0.72	0.00	0.00	0.00	100.00
Yugoslavia	1	1.32	1.22	1.43	0.00	0.00	1.00	0.00	0.00	0.00	100.00
Overall	564	1.16	1.15	1.16	0.00	3650.87	0.00	0.35	3.55	17.02	79.08

<sup>a</sup> Nations of the UK administered separately/<sup>a</sup> Chinese districts administered separately



In science literacy, around 95% of all assessments taken had wider variances for males than for females (almost 69% significantly so). Ratios range from 0.96 (Algeria) to 1.48 (Saudi Arabia). Less than 4% of all cases were female biased (< 1% significantly so). The average across all countries was 1.13. Only two countries have a wider variance for females. Across 86 countries, the variances are significantly male biased at the 5% level. 44 nations (48%) don't show significant heterogeneity in their Q scores, and of these, 2 are single case nations where the figure can't be calculated. While less heterogeneous than the other two content domains, as 52% are showing significant heterogeneity, the data cannot be considered homogenous. Figure 3 demonstrates the ratios and 95% CIs graphically (in order from smallest to largest). As is evident, while many significantly differ from 1.00 and countries vary considerably, few countries significantly vary from each other in the domain of science literacy.

### Meta-regression

While meta-analysis gives us an approximation of overall ratios and points to the presence of heterogeneity, on its own it does not advance our understanding of where the heterogeneity is coming from. A novel approach to attempt to apportion the variance attributable to known sources of heterogeneity is to use a form of linear regression often termed meta-regression. This procedure produces outputs recognisable as regression coefficients for covariates and an amount of variance explained synonymous with the traditional  $R^2$  value. Table 4 illustrates a predictive model of variance ratios across countries which is built from the following covariates: Year, test, being an OECD country, mean score for each country, the average male–female effect size (calculated as Hedges  $g$ ) GGGI economic participation, GGGI educational attainment, GGGI health, GGGI survival and political empowerment, expected years of schooling, mean years of schooling, life expectancy and GNI. Academic grade could not be considered in the model for mathematics literacy and reading as it was collinear with test. The referent categories were PISA and non OECD. Due to availability of matched HDI and GGGI variables only cases from tests administered from 2006 onwards are included in the analysis. Table 4 shows the results of this analysis.

These covariates predicted 31% of heterogeneity in Mathematics Literacy, 46% of the heterogeneity in Science Literacy and 54% of the heterogeneity in Reading. Many of the factors included in the model explain significant amounts of variance in effect

**Table 3 Variance ratios, confidence intervals and heterogeneity statistics for countries in the domain of science literacy**

Country	k	VR	LCI	UCI	P	Q	Q (p)	Below 1 (sig) (%)	Below 1 (nsig) (%)	Above 1 (nsig) (%)	Above 1 (sig) (%)
Albania	4	1.10	1.00	1.20	0.04	12.90	0.01	0.00	25.00	25.00	50.00
Algeria	3	0.96	0.92	1.01	0.09	0.41	0.81	0.00	100.00	0.00	0.00
Argentina	6	1.07	1.02	1.12	0.01	6.32	0.28	0.00	0.00	50.00	50.00
Armenia	7	1.08	1.03	1.13	0.00	13.97	0.03	0.00	14.29	28.57	57.14
Australia	20	1.17	1.14	1.19	0.00	40.80	0.00	0.00	0.00	5.00	95.00
Austria	13	1.12	1.07	1.16	0.00	32.55	0.00	0.00	7.69	30.77	61.54
Azerbaijan	2	0.99	0.94	1.05	0.72	0.12	0.72	0.00	50.00	50.00	0.00
Bahrain	7	1.45	1.38	1.52	0.00	13.03	0.04	0.00	0.00	0.00	100.00
Belgium	21	1.12	1.10	1.15	0.00	29.35	0.08	0.00	0.00	28.57	71.43
Bosnia and Herzegovina	1	1.05	0.97	1.15	0.24	0.00	1.00	0.00	0.00	100.00	0.00
Botswana	6	1.17	1.11	1.23	0.00	11.53	0.04	0.00	0.00	0.00	100.00
Brazil	6	1.14	1.11	1.16	0.00	5.68	0.34	0.00	0.00	0.00	100.00
Bulgaria	8	1.10	1.04	1.17	0.00	26.66	0.00	0.00	12.50	25.00	62.50
Canada	33	1.12	1.10	1.14	0.00	95.93	0.00	0.00	0.00	30.30	69.70
Chile	12	1.09	1.05	1.13	0.00	25.69	0.01	0.00	8.33	41.67	50.00
China	40	1.17	1.15	1.20	0.00	97.08	0.00	0.00	0.00	17.50	82.50
<sup>a</sup> Taipei	14	1.17	1.12	1.21	0.00	45.22	0.00	0.00	0.00	21.43	78.57
<sup>a</sup> Macao	5	1.18	1.11	1.26	0.00	9.70	0.05	0.00	0.00	20.00	80.00
<sup>a</sup> Hong Kong	20	1.18	1.15	1.21	0.00	32.44	0.03	0.00	0.00	15.00	85.00
Colombia	8	1.14	1.10	1.19	0.00	17.30	0.02	0.00	0.00	12.50	87.50
Costa Rica	2	1.18	1.12	1.24	0.00	1.03	0.31	0.00	0.00	0.00	100.00
Croatia	7	1.16	1.12	1.20	0.00	7.77	0.26	0.00	0.00	14.29	85.71
Cyprus	10	1.18	1.14	1.23	0.00	18.01	0.04	0.00	0.00	10.00	90.00
Czech Republic	16	1.08	1.03	1.12	0.00	52.46	0.00	6.25	0.00	50.00	43.75
Denmark	12	1.11	1.07	1.14	0.00	14.24	0.22	0.00	0.00	33.33	66.67
Dominican Republic	1	1.09	1.00	1.18	0.04	0.00	1.00	0.00	0.00	0.00	100.00
Egypt	3	1.09	1.04	1.15	0.00	3.27	0.20	0.00	0.00	33.33	66.67

**Table 3 (continued)**

Country	k	VR	LCI	UCI	P	Q	Q (p)	Below 1 (sig) (%)	Below 1 (nsig) (%)	Above 1 (nsig) (%)	Above 1 (sig) (%)
El Salvador	2	1.01	0.89	1.13	0.93	3.81	0.05	0.00	50.00	50.00	0.00
Estonia	5	1.09	1.01	1.19	0.03	20.90	0.00	0.00	20.00	20.00	60.00
Finland	11	1.17	1.14	1.21	0.00	14.33	0.16	0.00	0.00	18.18	81.82
France	9	1.15	1.10	1.21	0.00	23.92	0.00	0.00	11.11	0.00	88.89
Georgia	8	1.19	1.15	1.22	0.00	5.94	0.56	0.00	0.00	0.00	100.00
Germany	12	1.08	1.05	1.11	0.00	10.86	0.46	0.00	0.00	58.33	41.67
Ghana	3	1.00	0.96	1.05	0.82	0.13	0.94	0.00	0.00	100.00	0.00
Greece	10	1.16	1.10	1.23	0.00	34.58	0.00	0.00	0.00	20.00	80.00
Honduras	3	1.05	0.99	1.10	0.09	2.29	0.32	0.00	0.00	66.67	33.33
Hungary	20	1.08	1.06	1.11	0.00	29.51	0.06	0.00	10.00	50.00	40.00
Iceland	10	1.18	1.14	1.22	0.00	3.82	0.92	0.00	0.00	10.00	90.00
Indonesia	11	1.02	0.98	1.06	0.38	39.15	0.00	0.01	45.45	36.36	18.18
Iran	14	1.14	1.09	1.18	0.00	38.53	0.00	0.00	0.00	21.43	78.57
Ireland	14	1.15	1.11	1.18	0.00	20.91	0.08	0.00	0.00	21.43	78.57
Israel	12	1.21	1.13	1.29	0.00	71.99	0.00	0.00	8.33	16.67	75.00
Italy	16	1.12	1.07	1.16	0.00	103.27	0.00	0.00	6.25	43.75	50.00
Japan	20	1.18	1.15	1.21	0.00	40.16	0.00	0.00	0.00	10.00	90.00
Jordan	9	1.28	1.19	1.37	0.00	68.80	0.00	0.00	0.00	11.11	88.89
Kazakhstan	9	1.12	1.09	1.14	0.00	6.54	0.59	0.00	0.00	11.11	88.89
Korea	18	1.15	1.11	1.19	0.00	60.53	0.00	0.00	0.00	22.22	77.78
Kosovo	1	1.12	1.03	1.21	0.01	0.00	1.00	0.00	0.00	0.00	100.00
Kuwait	7	1.32	1.19	1.47	0.00	62.63	0.00	0.00	0.00	0.00	100.00
Latvia	14	1.15	1.11	1.19	0.00	22.47	0.05	0.00	0.00	28.57	71.43
Lebanon	5	1.09	1.04	1.15	0.00	6.92	0.14	0.00	0.00	40.00	60.00
Liechtenstein	2	1.28	1.00	1.63	0.05	0.01	0.93	0.00	0.00	100.00	0.00
Lithuania	16	1.10	1.07	1.12	0.00	20.17	0.17	0.00	0.00	43.75	56.25
Luxembourg	6	1.17	1.13	1.21	0.00	4.47	0.48	0.00	0.00	0.00	100.00

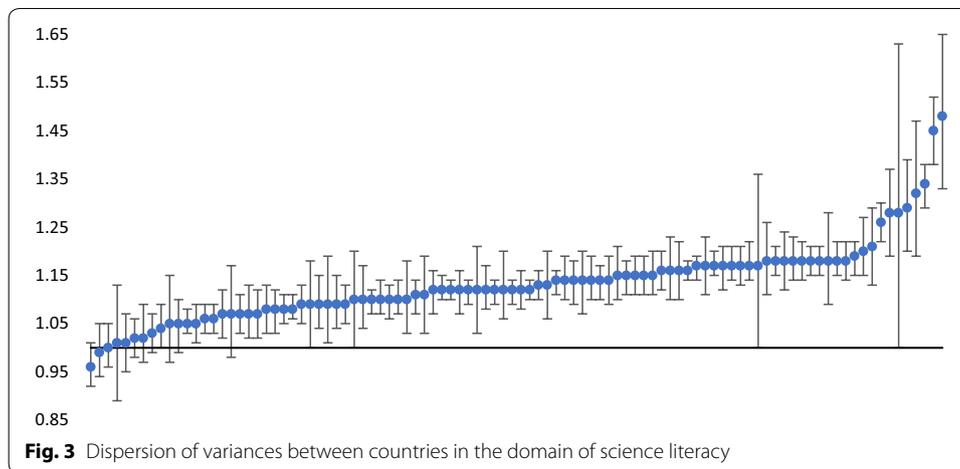
**Table 3 (continued)**

Country	k	VR	LCI	UCI	P	Q	Q (p)	Below 1 (sig) (%)	Below 1 (nsig) (%)	Above 1 (nsig) (%)	Above 1 (sig) (%)
Macedonia	5	1.11	1.03	1.19	0.01	13.58	0.01	0.00	20.00	0.00	80.00
Malaysia	7	1.14	1.07	1.20	0.00	29.36	0.00	0.00	0.00	28.57	71.43
Malta	5	1.14	1.10	1.19	0.00	2.07	0.72	0.00	0.00	20.00	80.00
Mexico	6	1.13	1.10	1.16	0.00	13.11	0.02	0.00	0.00	0.00	100.00
Moldova	4	1.04	1.00	1.09	0.04	1.68	0.64	0.00	0.00	100.00	0.00
Mongolia	2	1.02	0.97	1.09	0.42	0.00	0.99	0.00	0.00	100.00	0.00
Montenegro	4	1.13	1.06	1.20	0.00	6.80	0.08	0.00	0.00	25.00	75.00
Morocco	10	1.03	0.99	1.07	0.19	27.43	0.00	10.00	20.00	50.00	20.00
Netherlands	16	1.05	1.03	1.08	0.00	18.79	0.22	0.00	6.25	75.00	18.75
New Zealand	19	1.18	1.15	1.21	0.00	33.91	0.01	0.00	0.00	0.00	100.00
Norway	19	1.14	1.10	1.17	0.00	35.81	0.01	0.00	0.00	26.32	73.68
Oman	6	1.29	1.20	1.39	0.00	50.14	0.00	0.00	0.00	0.00	100.00
Palestine	3	1.18	1.09	1.28	0.00	7.52	0.02	0.00	0.00	0.00	100.00
Peru	4	1.01	0.95	1.07	0.78	7.01	0.07	0.00	25.00	50.00	25.00
Philippines	3	1.07	0.98	1.17	0.15	9.46	0.01	0.01	33.33	33.33	33.33
Poland	9	1.18	1.15	1.22	0.00	5.65	0.69	0.00	0.00	0.00	100.00
Portugal	13	1.12	1.08	1.17	0.00	30.26	0.00	0.00	0.00	30.77	69.23
Qatar	11	1.26	1.22	1.30	0.00	26.20	0.00	0.00	0.00	0.00	100.00
Romania	13	1.07	1.03	1.11	0.00	37.19	0.00	0.00	23.08	30.77	46.15
Russia	18	1.10	1.07	1.14	0.00	40.35	0.00	0.00	0.00	44.44	55.56
Saudi Arabia	7	1.48	1.33	1.65	0.00	69.49	0.00	0.00	0.00	0.00	100.00
Serbia	6	1.16	1.10	1.22	0.00	11.59	0.04	0.00	0.00	0.00	100.00
Singapore	17	1.17	1.14	1.22	0.00	58.90	0.00	0.00	0.00	17.65	82.35
Slovak Republic	13	1.06	1.03	1.09	0.00	18.45	0.10	0.00	0.00	53.85	46.15
Slovenia	18	1.12	1.09	1.14	0.00	19.47	0.30	0.00	0.00	27.78	72.22
South Africa	6	1.07	1.02	1.13	0.01	21.61	0.00	0.00	0.00	50.00	50.00

**Table 3 (continued)**

Country	k	VR	LCI	UCI	P	Q	Q (p)	Below 1 (sig) (%)	Below 1 (nsig) (%)	Above 1 (nsig) (%)	Above 1 (sig) (%)
Spain	64	1.16	1.14	1.18	0.00	76.95	0.11	0.00	1.56	34.38	64.06
Sweden	16	1.10	1.06	1.13	0.00	32.81	0.04	0.00	0.00	50.00	50.00
Switzerland	8	1.09	1.05	1.13	0.00	15.22	0.03	0.00	0.00	50.00	50.00
Syria	3	1.07	1.02	1.12	0.01	1.05	0.59	0.00	0.00	66.67	33.33
Thailand	16	1.15	1.11	1.20	0.00	62.09	0.00	0.00	0.00	18.75	81.25
Trinidad and Tobago	2	1.12	1.06	1.20	0.00	1.14	0.29	0.00	0.00	0.00	100.00
Tunisia	13	1.05	1.01	1.09	0.01	30.72	0.00	7.69	7.69	53.85	30.77
Turkey	12	1.10	1.07	1.14	0.00	22.77	0.02	0.00	0.00	41.67	58.33
Ukraine	3	1.20	1.15	1.27	0.00	1.68	0.43	0.00	0.00	0.00	100.00
UAE	20	1.34	1.29	1.38	0.00	90.15	0.00	0.00	0.00	0.00	100.00
United Kingdom	34	1.12	1.09	1.14	0.00	75.24	0.00	0.01	5.88	32.35	61.76
England <sup>a</sup>	14	1.14	1.09	1.19	0.00	34.10	0.00	0.00	7.14	28.57	64.29
Scotland <sup>a</sup>	12	1.12	1.08	1.16	0.00	18.19	0.08	0.00	8.33	41.67	50.00
Northern Ireland <sup>a</sup>	3	1.10	1.03	1.18	0.00	2.81	0.25	0.00	0.00	33.33	66.67
United States	28	1.12	1.10	1.14	0.00	36.35	0.11	0.00	0.00	32.14	67.86
Uruguay	5	1.18	1.14	1.22	0.00	0.63	0.96	0.00	0.00	0.00	100.00
Viet Nam	2	1.17	1.00	1.36	0.05	7.92	0.01	0.00	0.00	0.00	100.00
Yemen	6	1.06	1.03	1.09	0.00	1.24	0.94	0.00	0.00	83.33	16.67
Overall	991	1.13	1.12	1.13	0.00	4701.03	0.00	0.30	3.43	27.75	68.52

<sup>a</sup> Nations of the UK administered separately/<sup>a</sup> Chinese districts administered separately



**Table 4** Meta regression coefficients for mathematics literacy, science literacy and reading

Covariate	B		
	Mathematics	Science	Reading
Intercept	4.842* (0.694/8.990)	-1.264 (-5.135/2.606)	1.278 (-4.146/6.701)
Test: TIMSS	-0.047*** (-0.061/-0.033)	-0.004 (-0.025/0.017)	-
Test: TIMSS ADVANCED	0.011 (-0.066/0.043)	-	-
Test: PIRLS	-	-	-0.045*** (-0.068/-0.022)
Year	0.002* (-0.0040/-0.0003)	0.001 (-0.001/0.002)	-0.001 (-0.003/0.002)
Grade	-	0.004 (-0.0002/0.0080)	-
GGGI economic participation	-0.166** (-0.244/-0.088)	-0.197*** (-0.270/-0.125)	-0.142** (-0.247/-0.036)
GGGI educational attainment	0.313** (0.070/0.556)	0.344*** (0.159/0.528)	-0.131 (-0.557/0.294)
GGGI health	-0.396 (-1.03/0.236)	-0.118 (-0.690/0.454)	-0.050 (-0.813/0.712)
GGGI political empowerment	-0.063 (-0.126/0.001)	-0.024 (-0.081/0.033)	-0.077* (-0.148/-0.005)
HDI expected years of schooling	0.004 (-0.002/0.010)	0.004 (-0.001/0.010)	0.004 (-0.003/0.011)
HDI mean years of schooling	0.003 (-0.002/0.008)	-0.002 (-0.006/0.002)	-0.001 (-0.007/0.005)
HDI life expectancy	0.001 (-0.001/0.004)	0.001 (-0.001/0.003)	0.002 (-0.0004/0.0048)
HDI GNI	6.32E-07** (2.62E-07/1E-06)	8.78E-06*** (5.4E-06/1.22E-5)	1.32E-06 (3.6E-06/6.2E-06)
Hedges g	-0.235** (-0.289/-0.181)	-0.279*** (-0.328/-0.231)	-0.345*** (-0.403/-0.287)
Country mean score	0.0001 (-0.0001/0.0002)	0.0003*** (0.0001/0.0004)	0.0004*** (0.0002/0.0005)
OECD membership	-0.022 (-0.0438/0.0002)	-0.015 (-0.035/0.005)	-0.013 (-0.041/0.014)
Model R <sup>2</sup>	0.31	0.46	0.54
k	636	636	404

\*p < .05  
 \*\*p < .01  
 \*\*\*p < .001

sizes however, this varies by domain. By far the most significant predictor is the size of the gender difference in scores (across all three domains). As the gap becomes larger in favour of females, the variance for males increases. The mean score of the country is statistically significant for reading and science literacy but has a very small, positive impact. The same can be said for the test year in mathematics literacy. There are small and significant effects for the tests (with TIMSS and PIRLS showing slightly less male variance) but this is harder to interpret, as it is confounded by age. HDI indicators seem to have little impact on variance ratios, although GNI has a very small positive but statistically significant effect on mathematics literacy and science literacy. GGGI indicators have a stronger, negative impact on national variance ratios however. Countries with higher Economic Participation for women have ratios favouring females across all domains. Better Educational Attainment for women significantly increases the ratios in favour of males however in mathematics literacy and science literacy. Increased political empowerment for women also seems to increase variances for females in literacy.

## Discussion

Results broadly confirm the previous works of Baye and Monseur (2016) and suggest that male variances are greater than female variances internationally. This was largely expected as, although the methodology differed, most of the data used in this study was the same. Baye and Monseur showed variances for males were greater by 15% in reading, 12% in maths and 14% in science. Our results indicated that these ratios are 16%, 12% and 13% respectively, and suggest that the inclusion of more recent international surveys has not altered them substantively. Similarly, the correlation between male–female effect sizes and variance ratios was in line with those found by previous authors, with superior female performance increasing the gap in variance between the sexes. As such, we can broadly support the findings of past research and conclude that over the studied period, male variances in the domains of reading, mathematics literacy and science literacy are almost universally greater.

However, these results suggest that we can take this conclusion a step further. Feingold (1994) suggested that a difference of about 10% in variance ratios should be considered a substantive difference. Tables 1, 2 and 3 clearly show that for most countries engaging with PISA, PIRLS and TIMSS assessments, male variances are greater by often more than this threshold in all three domains. There are no geographical areas in this study that show significantly greater female variances. It would seem therefore that the question currently should no longer be, do male and female variances differ, but by how much more varied are males compared with females?

While in over 95% of cases, males show greater amounts of variance, there is a significant heterogeneity in these results, both within and between countries. While we can say with confidence that males are certainly more varied and generate a fairly precise estimate of a global average, we cannot come to an absolute value for each country individually and must contend with a large amount of dispersion. This dispersion is telling however and shows that not only do countries differ (significantly in some cases, as is evident in Figs. 1, 2 and 3) but that they vary internally as well. There is a significant amount of heterogeneity across these data in most countries examined in this study which requires explaining.

Our meta-regression within each domain has gone some way in explaining close to half of the heterogeneity observed in the dataset for reading and science literacy and about a third for mathematics literacy. Some of the findings are harder to interpret than others. The variable with the largest impact is the male–female effect size. This is the most substantive factor across all three domains and suggests that as girls outperform boys, the variability of boys increases. This seems to support earlier works that demonstrated a correlation between effect sizes and variance ratios (Baye and Monseur 2016; Nowell and Hedges 1998). The mean score for the country also has a significant albeit smaller impact in the same direction for science literacy and reading. Countries that perform better on average are therefore more likely to have greater variability for boys.

PISA tests appear to result in slightly more variance for males than TIMSS and PIRLS. Baye and Monseur (2016) found slightly smaller ratios in the primary years across all three domains. As TIMSS and PIRLS assess younger children, it may be that this simply reflects an age or maturity effect. However, we cannot rule out that the actual tests themselves are not causing some of the heterogeneity or, that there may be a compositional effect between the two.

Interestingly, most of the HDI indicators were not significantly predictive of variance ratios across domains. The exception to this appears to be the GNI indicator (an adjusted form of GDP per capita) for mathematics literacy and science literacy but not reading. Reading is a specific skill that requires mastery and is often contingent on home environments for reinforcement. While this is to an extent true of basic mathematical concepts, later mathematics and science are likely tied more strongly to whatever specific curriculum is delivered, and this is largely coordinated at a national level. This may explain why national wealth may impact more upon maths and science as opposed to reading. However, it should be noted that, despite its statistical significance, it has only a minute impact on increasing male variance.

Measures from the Global Gender Gap Index however seem to have a larger impact on variance ratios. Increasing female economic participation appears to increase levels of female variance across all three domains. This suggests that countries actively incorporating more women into the labour force has an impact on educational outputs. Increased political empowerment for women also increases female variances in reading. Increased educational attainment for women has mixed impacts however. It has a significant effect of increasing male variances in mathematics and science but a non-significant effect of increasing variance for women in reading. Taken together, it suggests that cultural practices tied to increasing female participation generally appear to increase variances for females and suggests that greater male variance in educational outcomes may be practically reduced on national levels. While this study cannot isolate what specific national level practices are responsible for this, it does lead to interesting further questions regarding the processes underlying male/female variability.

The year of the test also had a very small but statistically significant effect on variance ratios in mathematics literacy. As with the test variable itself, why precisely this should be the case is difficult to rationalise. As mentioned earlier, there could be specific test administrations which have differences that create a small, positive effect. Alternatively, it could be that national educational systems have been adapting educational practices in order to improve their position in international rankings, and that these new practices

are impacting upon the spread of scores. From this data alone, we can only speculate on the specifics as to why this may be the case.

#### **Limitations and future work**

There are several limitations to the data and the procedure we have used to explore it. First, a meta-analysis of international assessments such as PISA, PIRLS and TIMSS, while it controls for many extraneous variables not possible to account for in a meta-analysis via a literature search, does limit generalizability to alternative educational assessments. There could be something specific to these assessments that creates this effect. A limitation perhaps related to this applies to the assessments themselves. In PISA, the content being assessed is heavily based in literacy abilities. Even mathematics and science components are rooted in the ability to read and poor readers are unlikely to achieve if they cannot interpret the questions posed. As is evident from Table 2, the domain with the greatest amount of male variability is reading. As such, it is possible that mathematics and science show comparable overall ratios simply because they are rooted heavily in the ability to read. It is interesting to note that previous works using different assessments have shown greater variabilities in quantitative domains compared to verbal ones (Lakin 2013; Lohman and Lakin 2009). Thus, what this data may perhaps be showing is the greater variability in reading generally. This is still important and would pose the question ‘why are males more variable at reading’ but we must therefore be cautious regarding the conclusions we draw from the mathematics and science domains.

This study tentatively suggests (as does Baye and Monseur 2016) that age may be a factor, and that variability for males increases as candidates get older. To our knowledge, no study specifically examines this, either longitudinally or cross-sectionally (with perhaps the exception of Lakin 2013). Alternatively, attempting to quantify nation specific factors that could be included in additional regression analyses may be a future avenue worth exploring (particularly considering the impact of GGGI variables on ratios), potentially allowing us to quantify greater levels of heterogeneity in these results.

A final avenue of exploration would be to examine this effect over additional academic assessments. Research historically focuses on core domains of reasoning (Baye and Monseur 2016; Lohman and Lakin 2009; Strand et al. 2006). While this is important, do we get similar patterns across curricular subject examinations (anything from art to zoology), or different modes of assessment (pencil and paper tests compared to practical performance assessments)? These are often studied less, in part due to reasons of sample representation, or the fact that specific subjects are often self-selecting. As it stands from the data and the literature reviewed here, we would expect to see similar patterns across assessments generally. It would be telling if this was not the case. If there are exceptions, what are they and why do they differ?

#### **Implications for theory and policy**

From a theoretical perspective, we cannot contribute causal explanations for why males are more variable. Data suggests the effect is almost universal, which, while supportive of biological and evolutionary theories, doesn’t rule out specific cultural, educational, political, social or religious practices. Indeed, the fact that we can quantify substantial

variation as dependent on increased female participation in society suggests that, at least in educational outcomes, it is not necessarily the case that males should vary more.

However, without a clear understanding of *why* males vary more and *how* this difference is maintained, we acknowledge that a meaningful discussion regarding what can be done to ensure parity is difficult. Increased female participation in the economy, education and political empowerment significantly reduce the size of the discrepancy in variances between males and females across the three educational domains studied here. If these increase, we might expect the variance gap to decrease. Which specific practices within countries are enabling this however are not discernible from the existing data, and more comparative, in-depth work within nations (with closer attention to specific educational practices) would be required before specific policy recommendations could be formulated to ensure parity between males and females across the ability distribution.

Differences in the spread of abilities are important for society. If, for example, we want to increase the representation of women in top positions and educational institutions, so that parity between the sexes exists at this level, it is important that males and females are equally represented in the higher percentiles of whatever qualifications or ability metrics that constitute the selection processes. Similarly, the large gap in reading ability between boys and girls in the lower percentiles (Baye and Monseur 2016) suggests that some boys are likely to be at a serious disadvantage in later education (and potentially later life outcomes). Whilst implementing measures that strive for parity in the right tail of the distribution are important, we must also be mindful to not neglect the left.

## Conclusions

Our analysis seems to suggest that greater male variability is currently universal in internationally comparable assessments implemented over the past decade. However, this effect is far from homogenous, and there are quantifiable differences that exist over nations. Furthermore, some of this heterogeneity can be attributed to some yet unspecified practices or policies targeted at increasing male–female equality, general male–female performance as well as potentially the age of candidates and the type of test. Further work however is required to examine these factors in more detail, and analyses within nations may be informative to examine more specific practices that can explain national patterns. Comparative work examining high and low scoring GGGI countries may be informative in this endeavour. In doing so, it may be possible to determine if the root cause of these differences in distributions are attributable to some species universal mechanism or some other social or cultural phenomenon.

### Authors' contributions

HG, AL, CM, SS, PT and LC contributed to developing the research design and writing the manuscript. Statistical analysis was conducted by LC and AL. All authors read and approved the final manuscript.

### Author details

<sup>1</sup> Centre for Evaluation and Monitoring, Ushaw College, University of Durham, Durham DH7 7DW, UK. <sup>2</sup> School of Social Sciences, Humanities and Law, Teesside University, Middlesbrough TS1 3BX, UK. <sup>3</sup> Newcastle, UK.

### Acknowledgements

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Availability of data and materials

All data sets used in this study are accessible online via the provided links within the manuscript.

**Funding**

Not applicable.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 25 January 2018 Accepted: 28 January 2019

Published online: 12 February 2019

**References**

- Adams, R., & Wu, M. (Eds.). (2002). *PISA 2000 technical report*. Paris: OECD.
- Archer, J., & Mehdkhani, M. (2003). Variability among males in sexually selected attributes. *Review of General Psychology*, 7, 219–236. <https://doi.org/10.1037/1089-2680.7.3.219>.
- Baye, A., & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large Scale Assessments in Education*, 4, 1–16.
- Borenstein, M., Hedges, L., Higgins, J. & Rothstein, H. (2013). Comprehensive meta analysis version 3.0. Biostat, Englewood, NJ.
- Borkenau, P., McCrae, R. R., & Terracciano, A. (2013). Do men vary more than women in personality? A study in 51 cultures. *Journal of Research in Personality*, 47, 135–144.
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, 135, 218–261.
- Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex differences in impulsivity: A meta-analysis. *Psychological Bulletin*, 137, 97–130. <https://doi.org/10.1037/a0021591>.
- Dubet, F. (2010). L'école "embarrassée" par la mixité. *Revue Française de Pédagogie*, 171, 77–86.
- Ellis, H. (1894). *Man and woman: A study of human sexual characters*. London: Walter Scott.
- Feingold, A. (1992). Sex differences in variability in intellectual abilities: a new look at an old controversy. *Review of Educational Research*, 62, 61–84.
- Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross cultural perspective. *Sex Roles: A Journal of Research*, 30, 81–92.
- Gualtieri, T., & Hicks, R. E. (1985). The immunoreactive theory of selective male affliction. *Behavioral and Brain Sciences*, 8, 427–477.
- Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine*, 17, 841–856.
- He, W.-J., Wong, W.-C., Li, Y., & Xu, H. (2013). A study of the greater male variability hypothesis in creative thinking in Mainland China: Male superiority exists. *Personality and Individual Differences*, 55, 882–886.
- Hedges, L. V., & Friedman, L. (1993). Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research*, 63, 94–105.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. London: Academic Press.
- Higgins, J. P. T. (2008). Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology*, 37, 1158–1160. <https://doi.org/10.1093/ije/dyn204>.
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539–1558. <https://doi.org/10.1002/sim.1186>.
- Hill, T.P. (2017). A theory for gender differences in variability. *Cornell University Library*. <https://arxiv.org/abs/1703.04184>. Accessed Oct 2017.
- Hollingsworth, L. S. (1922). Differential action upon the sexes of forces which tend to segregate the feebleminded. *Journal of Abnormal and Social Psychology*, 17, 35–57.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance—a meta-analysis. *Psychological Bulletin*, 107, 139–155.
- IBM Corp. (2013). *IBM SPSS statistics for windows, version 22.0*. Armonk: IBM Corp.
- Jacobs, J. (1996). Gender inequality and higher education. *Annual Review of Sociology*, 22, 153–185.
- Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives on Psychological Science*, 3, 518–531. <https://doi.org/10.1111/j.1745-6924.2008.00096.x>.
- Karowski, M., Jankowska, D. M., Gajda, A., Marczak, M., Groyecka, A., & Sorokowski, P. (2016). Greater male variability in creativity outside the WEIRD world. *Creativity Research Journal*, 28, 467–470.
- Katzman, S., & Alliger, G. M. (1992). Averaging untransformed variance ratios can be misleading. A comment on Feingold. *Review of Educational Research*, 62, 427–428.
- Lakin, J. M. (2013). Sex differences in reasoning abilities: Surprising evidence that male-female ratios in the tails of the quantitative reasoning distribution have increased. *Intelligence*, 41, 263–274.
- Lietz, P. (2006). Issues in the change in gender differences in reading achievement in cross-national research studies since 1992: a meta-analytic view. *International Education Journal*, 7, 127–149.
- Lohman, D. F., & Lakin, J. M. (2009). Consistencies in sex differences on the Cognitive Abilities Test across countries, grades, test forms, and cohorts. *British Journal of Educational Psychology*, 79, 289–407.
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford: Stanford University Press.
- Machin, S., & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science*, 322, 1331–1332.
- Marks, G. N. (2008). Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. *Oxford Review of Education*, 34, 89–109.
- Martin, M. O., Gregory, K. D., & Stempler, S. E. (Eds.). (2000). TIMSS 1999 technical report: IEA's repeat of the third international.

- Martin, M. O., & Kelly, D. L. (Eds.). (1997). *TIMSS technical report. Volume II. Implementation and analysis, primary and middle school years*. Chestnut Hill: Boston College.
- Martin, M. O., Kelly, D. L., & Chrostowski, S. J. (Eds.). (1996). *TIMSS technical report: Volume I design and development*. Chestnut Hill: Boston College.
- Martin, M. O., & Mullis, I. V. S. (Eds.). (1996). *TIMSS: quality assurance in data collection*. Chestnut Hill: Boston College.
- Martin, M. O., & Mullis, I. V. S. (Eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., & Chrostowski, S. J. (Eds.). (2004). *TIMSS 2003 technical report*. Chestnut Hill: Boston College.
- Martin, M. O., Mullis, I. V. S., & Hooper, M. (Eds.). (2016). *Methods and procedures in TIMSS 2015*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (Eds.). (2003). *PIRLS 2001 technical report*. Chestnut Hill: Boston College.
- Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (Eds.). (2007). *PIRLS 2006 technical report*. Chestnut Hill: Boston College.
- Miller, E. M. (2001, June). Ultimate and proximate causes of greater male variability: An attempt to explain female underrepresentation in certain fields. Paper presented at the conference of the Human Behavior and Evolution Society, Salt Lake City, UT.
- Morgan, M., & Kett, M. (2003). *The prison adult literacy survey: Results and implications*. Dublin: Irish Prison Service.
- Noddings, N. (1992). Variability: A pernicious hypothesis. *Review of Educational Research*, 62, 85–88.
- Nowell, A., & Hedges, L. V. (1998). Trends in gender differences in academic achievement from 1960 to 1994: an analysis of differences in mean, variance, and extreme scores. *Sex Roles*, 39, 21–43.
- OECD. (2005). *PISA 2003 technical report*. Paris: OECD.
- OECD. (2009a). *PISA 2006 technical report*. Paris: OECD.
- OECD. (2009b). *PISA Data Analysis Manual* (2nd ed.). Paris: OECD.
- OECD. (2014). *PISA 2012 technical report*. Paris: OECD.
- OECD. (2015). *Closing the Gender Gap: Act Now*. Paris: OECD.
- OECD. (2016). *PISA 2015 Results*. Paris: OECD.
- Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). (2008). *TIMSS 2007 technical report*. Chestnut Hill: Boston College.
- Ounsted, C., & Taylor, D. C. (1972). The Y chromosome message: A point of view. In C. Ounsted & D. C. Taylor (Eds.), *Gender differences: Their ontogeny and significance* (pp. 241–262). London: Churchill.
- Quinn, J. M., & Wagner, R. K. (2013). Gender differences in reading impairment and in the identification of impaired readers: Results from a large-scale study of at-risk readers. *Journal of Learning Disabilities*, 48, 433–445. <https://doi.org/10.1177/0022219413508323>.
- Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? *American Psychologist*, 60, 950–958.
- Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in Cognitive Abilities Test scores: A UK national picture. *British Journal of Educational Psychology*, 76, 463–480.
- UNESCO. (2011). *The hidden crisis: armed conflict and education*. Paris: United Nations Educational, Scientific and Cultural Organization.
- United Nations Development Program: Human Development Reports (Retrieved August 2018). <http://hdr.undp.org/en/data>.
- World Economic Forum: The Global Gender Gap Index (Retrieved August 2018). <http://reports.weforum.org>.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---