

# Investigating Gender Differences in Mathematics and Science: Results from the 2011 Trends in Mathematics and Science Survey

David Reilly<sup>1</sup>, David L. Neumann<sup>1,2</sup>, Glenda Andrews<sup>1,2</sup>

This is a pre-print reproduction of the manuscript, and may differ from the final published edition. Please refer to the final published version of the article available directly from the publisher's website (doi: 10.1007/s11165-017-9630-6). The final publication is available at <http://link.springer.com>

This may be cited as

Reilly, D., Neumann, D. L., Andrews, G. (2017). Investigating gender differences in mathematics and science: Results from the 2011 Trends in Mathematics and Science Survey. *Research in Science Education*. doi: 10.1007/s11165-017-9630-6)

## Abstract

The underrepresentation of women in science, technology, engineering and mathematics (STEM)-related fields remains a concern for educators and the scientific community. Gender differences in mathematics and science achievement play a role, in conjunction with attitudes and self-efficacy beliefs. We report results from the 2011 Trends in Mathematics and Science Study (TIMSS), a large international assessment of eighth grade students' achievement, attitudes and beliefs among 45 participating nations ( $N = 261,738$ ). Small to medium sized gender differences were found for most individual nations (from  $d = -.60$  to  $+.31$  in mathematics achievement, and  $d = -.60$  to  $+.26$  for science achievement), although the direction varied and there were no global gender differences overall. Such a pattern cross-culturally is incompatible with the notion of immutable gender differences. Additionally, there were different patterns between OECD and non-OECD nations, with girls scoring higher than boys in mathematics and science achievement across non-OECD nations. An association was found between gender differences in science achievement and national levels of gender equality, providing support for the gender segregation hypothesis. Furthermore, the performance of boys was more variable than that of girls in most nations, consistent with the greater male variability hypothesis. Boys reported more favorable attitudes towards mathematics and science and girls reported lower self-efficacy beliefs. While the gender gap in STEM achievement may be closing, there are still large sections of

**Keywords:** gender differences, mathematics, science, education, meta-analysis



David Reilly  
davidreilly.com@gmail.com

<sup>1</sup> School of Applied Psychology, Griffith University, Gold Coast, Australia

<sup>2</sup> Behavioural Basis of Health Program, Menzies Health Institute, Gold Coast, Australia

While progress has been made in recent decades, the underrepresentation of women in science, technology, engineering and mathematics (STEM)-related fields remains a serious issue for the scientific and technical community across many developed nations (OECD, 2011). Although exceptions exist for the fields of medical science and psychology, more men than women complete bachelors and masters degrees in STEM fields (National Science Foundation, 2011). Disparities in the proportion of women completing doctoral degrees in STEM are even greater (National Science Foundation, 2011). Global gender disparities in science remain a serious concern (Sugimoto, Larivière, Ni, Gingras, & Cronin, 2013), and women in research remain a minority across most nations despite the significant progress made in gender equality across other areas of society (UNESCO, 2011).

We argue the importance of further research into gender differences in mathematics and science for two main reasons. Firstly, as noted across many developed nations there is a shortage of science graduates to meet the needs of industry, which Handelsman et al. (2005) have argued can best be addressed by broadening the pool of applicants entering the science pipeline to include a greater representation of women. Many national science boards have identified increasing the number of girls and women studying STEM-related degrees as an important intervention target. For example, in the United States, the National Science Foundation produces a bi-annual report into the number of girls and women studying STEM-related subjects and degrees (National Science Foundation, 2017), mandated by the US Science and Engineering Equal Opportunities Act. Secondly, gender equality of educational outcomes remains an issue of importance. Regardless of whether or not a student chooses to pursue a STEM-related field, educators stress the importance of basic mathematical skills and science literacy for full participation in today's society (OECD, 2016), and that equality of education outcomes is a desirable social good.

The causal factors responsible for the underrepresentation of women in STEM fields are complex and contentious (Ceci & Williams, 2011). But two major factors that have been suggested to make a contribution are a) gender differences in mathematics and science achievement during compulsory schooling (Else-Quest, Hyde, & Linn, 2010), and b) self-efficacy beliefs and attitudes towards mathematics and science (Nosek, Banaji, & Greenwald, 2002). These factors are important, as a large body of research has shown that they contribute to the decision-making processes that lead adolescent students to undertake or avoid further STEM-based studies and careers (Eccles, 2013; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Simpkins, Davis-Kean, & Eccles, 2006). While some researchers have investigated gender differences in these factors at the endpoint of compulsory schooling (Guiso, Monte,

Sapienza, & Zingales, 2008; Reilly, Neumann, & Andrews, 2015), it is also important to consider differences at *earlier* time-points in child development. Attitudes towards mathematics and science are acquired by children early in their education during primary and secondary school (Marginson, Tytler, Freeman, & Roberts, 2013; Simpkins et al., 2006), and are also supported by early socialization experiences and enrichment activities provided by parents and caregivers (Simpkins, Davis-Kean, & Eccles, 2005). However developmental literature has identified that the eighth grade represents a critical time for the development of self-efficacy beliefs in STEM subjects, as well as formation of attitudes towards mathematics and science as a career (Eccles, 2013; Riegle-Crumb, Moore, & Ramos-Wada, 2011). Additionally, a sizeable body of research has documented widening gender differences in mathematics achievement and beliefs in adolescence after puberty, making eighth grade ideally suited for documenting developmental gender differences (Goldman & Penner, 2014). We review evidence of gender differences in eighth grade students, before turning to cross-cultural investigations.

#### **Gender differences in mathematics and science achievement**

Gender differences in mathematics and science achievement have been reported in the literature across many decades (Hedges & Nowell, 1995; Maccoby & Jacklin, 1974). However, researchers are divided over how significant these gender gaps may be, with some arguing that the gaps are small but still meaningful (Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000; Gallagher & Kaufman, 2005; Reilly et al., 2015), and others that these gaps are so small, in fact, as to be characterized as 'trivial' (Hyde, Lindberg, Linn, Ellis, & Williams, 2008; Spelke, 2005). Gender differences are less frequently investigated in science than mathematics achievement and are somewhat larger in magnitude, although some researchers dispute their existence as well, especially at younger ages (e.g. Haworth, Dale, & Plomin, 2010; Hyde & Linn, 2006).

How are we to reconcile these conflicting views? The lack of consistency in findings may partially reflect differences in the size and type of samples that are recruited. As it is impractical to measure *all* members of a target population, researchers typically draw inferences based on a small sample of participants, and make a conclusion about the population from which they came. Such an approach can be problematic though if the sample studied is different in some way from the general community. Indeed, a large body of gender difference research has shown interactions between cognitive abilities and demographic characteristics, including socioeconomic background, age, ethnicity, and level of parental

education. These interactions may lead to an overestimation or underestimation of the magnitude of gender differences in some studies.

Hedges and Nowell (1995) argued that the strongest and most reliable evidence for gender differences in cognitive abilities comes from samples that are representative of the general population. One such source is large-scale educational assessments of students on standardized tests. In the first such study of its kind, Hedges and Nowell (1995) drew on several decades of nationally representative assessments of student ability in the USA, and found small to medium effect sizes favoring males in mathematics and science. These findings have been replicated in USA samples in subsequent decades (Reilly et al., 2015). Hedges and Nowell also reported support for the *greater male variability* hypothesis (Feingold, 1992). Because the performance of males is more variable than that of females (Hyde, 2005; Machin & Pekkarinen, 2008), there is a greater proportion of males at the far right tail of the ability distribution (for example, the 95<sup>th</sup> percentile). This may have important implications for STEM decision-making, as students gifted in mathematics and science are more likely to consider pursuing a career in these fields (Eccles, 2013; Jacobs et al., 2002). While contributing to the debate on gender differences in STEM, such studies are U.S.-centric and findings may not necessarily generalize to children from other educational systems and cultures.

Observation of gender differences in educational achievement raises the question of etiology. Unger (1979) cautioned that while gender can be a subject variable (reflecting an internal trait), it can also be a stimulus variable that elicits a change in how one is perceived and treated. While there are still some researchers who argue for biological explanations of gender differences from evolutionary psychological perspective (e.g. Geary, 2010), most researchers endorse a biopsychosocial model of gender differences that stresses the importance of sociocultural factors in shaping differences between males and females (Halpern, Benbow, et al., 2007). Boys and girls typically differ in early socialization experiences (Lytton & Romney, 1991; C. L. Martin & Ruble, 2004), and quickly acquire cultural stereotypes and norms about the sex-typing of intellectual pursuits such as mathematics and science. Even when explicit gender stereotypes about ability are not endorsed, there is strong evidence of implicit beliefs associating maleness with mathematics (Nosek et al., 2002), and national differences in the strength of gender-science stereotypes are correlated with national gender gaps in science achievement scores (Nosek et al., 2009). A large body of literature has demonstrated that parental and teacher beliefs about gender stereotypes in STEM affect children's subsequent STEM attitudes and achievement scores (Gunderson, Ramirez, Levine, &

Beilock, 2012). The observed differences between males and females might be due to gender, or some other factor that is correlated with gender. Therefore caution must be taken before attributing a causal explanation.

#### **Attitudes towards mathematics and science, and self-efficacy**

Researchers have also begun to focus on the important role that mathematics and science attitudes play in the decision to undertake further mathematics and science study in high school, at college and university, and eventually in pursuing a STEM-based career. Boys report more positive attitudes towards learning mathematics and science (Else-Quest et al., 2010; Weinburgh, 1995), and girls report lower self-confidence and greater mathematics anxiety (Beilock, Gunderson, Ramirez, & Levine, 2010; Jacobs et al., 2002). This is despite many studies showing that girls actually obtain better grades than boys in mathematics courses (Dwyer & Johnson, 1997; Kenney-Benson, Pomerantz, Ryan, & Patrick, 2006). From an early age, children show rigid gender-typing of mathematics and science as inherently masculine, while reading and language are regarded as feminine (Fennema, Peterson, Carpenter, & Lubinski, 1990; Halim & Ruble, 2010). Even when such stereotypes are not explicitly endorsed, implicit gender stereotypes persist into adulthood and are found cross-culturally (Nosek et al., 2002). It is a particularly serious issue, as negative stereotypes can: a) impair girls' performance on standardized tests through the mechanism of stereotype threat (Spencer, Steele, & Quinn, 1999; Steele, 1997); and b) undermine self-efficacy beliefs about STEM and reduce achievement motivation. Some researchers argue that gender differences in attitudes may be even more important than gender differences in aptitude for explaining the underrepresentation of women in STEM-fields (Else-Quest, Mineo, & Higgins, 2013; Smeding, 2012).

#### **Cross-cultural tests of gender differences and similarities hypotheses**

A key limitation of much research into the underrepresentation of women in STEM-fields is that it draws on evidence from a single geographical region or nation, such as the USA. Cross-cultural investigations of gender differences and similarities in STEM achievement provide stronger evidence for debates about the role of gender in educational outcomes more generally. While most researchers endorse a biopsychosocial model of gender differences (see Halpern, Benbow, et al., 2007), there remain some researchers who argue for biological determinism (e.g. Kimura, 2000, 2002). If gender differences in mathematics and science are the result of innate and immutable biological differences between males and

females, then it follows that significant differences in quantitative reasoning would be found universally as claimed by the evolutionary and biological psychology perspectives (Geary, 2010; Kimura, 2000). Large fluctuations in the magnitude and direction of gender differences between nations would bolster evidence for the role that culture and environmental factors play in their development, supporting sociocultural perspectives (Eagly, Wood, & Diekmann, 2000). Cross cultural analysis of gender differences can therefore shed some light on the competing biological and sociocultural perspectives on the development of gender-typed cognitive abilities (Wood & Eagly, 2002).

Two international assessments of students' achievements in STEM-related areas provide opportunities to examine gender differences cross-culturally. These assessments are the Trends in International Mathematics and Science Study (TIMSS) which examines students in eighth grade, and the Programme for International Student Assessment (PISA) which examines students in their final year of compulsory schooling. Each of these assessments recruits large nationally representative samples of students using stratified sampling, and administers standardized tests within a vigorous and comprehensive assessment framework (Linn, 2002). These large datasets are frequently employed by educational researchers to study the impact of educational environments, student attitudes and beliefs (Anderson, Lin, Treagust, Ross, & Yore, 2007), as well as the contributions of demographic factors such as gender on educational outcomes in STEM.

International assessments of mathematics and science also provide the opportunity to examine the contributions of macro-level cultural factors such as gender equality (Else-Quest & Grabe, 2012). There is considerable diversity in the rights and status of women cross-culturally, as well as their degree of participation in economic, educational and political domains as a reflection of cultural beliefs. Baker and Jones (1993) proposed the *gender segregation hypothesis*, which suggests that pronounced gender inequality may negatively influence the cognitive and educational development of girls. For example, Guiso et al. (2008) examined gender differences in mathematics achievement for students participating in the PISA 2003 wave. They found an association between gender equity and the gender gap. Smaller gender differences were observed in more gender-equal countries. This finding supports the gender segregation hypothesis, and has been replicated with later waves of PISA testing data (Reilly, 2012).

Other researchers (e.g. Else-Quest et al., 2010) have investigated gender differences in younger children who participated in TIMSS. Administered every four years, TIMSS seeks to provide rigorous

testing of students' abilities in mathematics and science. It samples students from a variety of countries in their eighth year of compulsory schooling. Mullis et al. (2000) found a modest but significant gender difference in mathematics favoring males which was consistent with previous waves of TIMSS testing. But more recently, Else-Quest et al. (2010) examined the 2003 TIMSS cycle. They found small-to-medium gender differences in mathematics achievement within countries, but no differences globally. They also tested the gender segregation hypothesis, but contrary to other researchers (e.g. Guiso et al., 2008) found no significant association between measures of gender equity and mathematics achievement in this age group. Despite no global gender differences in achievement, the researchers did report meaningful gender differences in mathematics attitude and self-efficacy beliefs (Else-Quest et al., 2010). Boys reported more positive attitudes and higher self-efficacy than girls. However, this study did not examine whether there might also be gender differences in science achievement, attitude and self-efficacy beliefs that were collected during the TIMSS 2003 wave of assessment. These might also be relevant in understanding the STEM course-taking decisions of students.

Over a decade has passed since the 2003 TIMSS cycle, and the present study seeks to determine whether these earlier findings are replicated with the 2011 wave of TIMSS. Furthermore, we extend our analyses to cover science achievement, attitudes, and self-efficacy beliefs. We set out with four key research questions for the domains of mathematics and science. First, are there still gender differences in overall mathematics and science achievement across the globe, or has the gender gap at the beginning of high school been eliminated? Second, do males show greater variance in STEM achievement than females, as predicted by the greater male variability hypothesis? Third, is there a relationship between measures of national gender equality and gender differences in students' achievement in STEM-related areas? Fourth, are there gender differences in attitudes towards STEM and in self-efficacy beliefs? In addition to testing for global gender differences, we also report comparisons for individual nations which may be of benefit to these regions in the planning of educational interventions.

## Method

### Trends in International Mathematics and Science Study (TIMSS) 2011

The 2011 TIMSS assessment is a large multinational study conducted by the International Association for the Evaluation of Educational Achievement (IEA). Assessment data was collected for eighth grade students with a mean age of 14.60 years ( $SD = 0.90$ ) from 45 countries. These countries and the

sample sizes for each are shown in Table 1 ( $N = 261,738$ ). We also considered whether there were differences in outcomes between OECD ( $n = 15$ ) and non-OECD countries ( $n = 30$ ).

While a longitudinal study that tracked mathematics achievement, attitudes and self-efficacy over time would offer the strongest evidence of gender differences or similarities, at present there are no nationally-representative international assessments. However, the large sample sizes and generalizability of the 2011 TIMSS does provide insight into the current cohort of eighth grade students, and can be supplemented by evidence from similar assessments such as PISA which measure students in their final year of high school (Else-Quest et al., 2010; Guiso et al., 2008; Reilly, 2012).

TIMSS assesses students in a range of mathematical areas (number, algebra, geometry, probability) and content from a broad range of scientific domains (biology, chemistry, physics, and earth science), providing good coverage of the foundation skills expected of students as they enter more advanced mathematics and science streams. For more information on the specific framework under which mathematics and science are assessed in TIMSS, see the IEA Technical Report (M. O. Martin & Mullis, 2012).

### Assessment Measures

The primary variables of interest were the mathematics and science achievement scores from the 2011 TIMSS. In addition to completing the mathematics and science achievement test, students completed paper-and-pencil survey instruments measuring home environment, teacher and school environment, and student characteristics. Of particular interest in the present study was the latter category, which includes students' attitudes and self-efficacy beliefs about mathematics and science. Students endorsed a series of statements about learning mathematics and science on a 4 point Likert-type scale ranging from 'Agree a lot' to 'Disagree a lot', with no midpoint. These were combined to form composite scales on the constructs of interest. Survey instruments were translated into the student's native language for equivalency, and administered by trained professionals (M. O. Martin & Mullis, 2012).

Mathematics attitudes reflects students' attitudes and enjoyment of learning mathematics, with higher values indicating more positive attitudes. Representative sample items include "I enjoy learning mathematics" and "Mathematics is boring". Some items are reverse coded. This measure contained 6 items drawn from Q14A-F of the student survey which shows high internal consistency (Cronbach's  $\alpha = .85$ ).

Table 1

*National Sample Size Information for TIMSS 2011, and Gender-Gap Index (GGI) Scores*

Country	Gender Gap Index (GGI)	N of males	N of Females	Total Sample
<i>(OECD Nations)</i>				
Australia	0.729	3890	3747	7556
Chile	0.703	2702	3133	5835
England	0.746	1981	1861	3842
Finland	0.838	2181	2085	4266
Hungary	0.664	2608	2570	5178
Israel	0.693	2327	2372	4699
Italy	0.680	2054	1925	3979
Japan	0.651	2231	2183	4414
New Zealand	0.781	2788	2548	5336
Norway	0.840	1985	1877	3862
South Korea	0.628	2503	2663	5166
Sweden	0.804	2889	2679	5568
United States	0.741	5180	5297	10477
<i>(Non-OECD)</i>				
Armenia	0.665	2952	2894	5846
Bahrain	0.623	2352	2288	4640
Botswana	0.683	2628	2772	5400

Georgia	0.662	2358	2205	4563
Ghana	0.681	3822	3501	7323
Honduras	0.695	1971	2447	4418
Hong Kong	<i>n/a</i>	2018	1997	4015
Indonesia	0.659	2823	2972	5795
Iran	0.589	3213	2816	6029
Jordan	0.612	3574	4120	7694
Kazakhstan	0.701	2210	2180	4390
Lebanon	0.608	1846	2128	3974
Lithuania	0.713	2398	2349	4747
Macedonia	0.697	2068	1994	4062
Malaysia	0.653	2815	2918	5733
Morocco	0.580	4675	4310	8985
Oman	0.587	4851	4691	9542
Palestinian National Authority	<i>n/a</i>	3317	4495	7812
Qatar	0.623	2259	2163	4422
Romania	0.681	2821	2702	5523
Russia	0.704	2479	2414	4893
Saudi Arabia	0.575	2142	2202	4344
Singapore	0.691	2993	2934	5927
Slovenia	0.704	2257	2158	4415
South Africa	0.748	6078	5888	11966
Syria	0.590	2176	2237	4413
Taiwan	<i>n/a</i>	2594	2448	5042
Thailand	0.689	2708	3416	6124
Tunisia	0.626	2508	2620	5128
Turkey	0.595	3514	3414	6928
Ukraine	0.686	1655	1723	3378
United Arab Emirates	0.645	7054	7035	14089

Note: Countries which were missing gender equality data in the Global Gender Gap Report are indicated by *n/a* subscript.

Mathematics self-efficacy quantified student’s perceptions of competency in mathematics and self-efficacy beliefs. Higher values indicate greater self-efficacy. This comprised a 9 item scale drawn from Q16A-I, with representative sample items including “I learn things quickly in mathematics” and “Mathematics makes me confused and nervous”. The scale showed high internal consistency (Cronbach’s  $\alpha = .86$ ).

Science attitudes reflects students’ interest and enjoyment of learning the science curriculum. Higher values indicate more positive science attitudes. This measure contained 7 items, comprising Q17A-G of the student survey with high internal consistency (Cronbach’s  $\alpha = .86$ ). Sample items include “I wish I did not have to study science” and “I learn many interesting things in science”.

Self-efficacy in science assessed students’ perceptions of their competency in science, using a 6

item scale comprising Q19A-F (Cronbach’s  $\alpha = .81$ ). Representative items include “I learn things quickly in science”, and “Science makes me confused and nervous”. Higher values reflect greater self-efficacy beliefs for science.

In order to test the gender segregation hypothesis, objective measures of national-level gender equity were obtained from the World Economic Forum’s “Global Gender Gap Report for 2011”. The Gender Gap Index (GGI) instrument provides an objective measure of gender equality for most nations in the developed and developing world, drawing on independently verified sources of information from four domains: economic participation and opportunity, educational attainment, political representation, and health and survival statistics (Hausmann, Tyson, & Zahidi, 2011). Aggregate scores across these domains are normed, with a minimum possible score of 0 (complete inequality) and maximum

possible score of 1 (complete gender equality). GGI values were available for 42 of the 45 counties participating in TIMSS (see Table 1). Within the countries assessed by the 2011 TIMSS, scores ranged from 0.58 (Saudi Arabia) to 0.84 (Norway and Finland). The decision to use the GGI rather than other available indicators was due to the broad coverage of equality metrics it provides (Else-Quest & Grabe, 2012). In addition it was the measure also used by Guiso et al. (2008) in samples of older students. By using the same measure, we were able to test whether Guiso et al.'s findings would be replicated with a younger age group.

### Meta-Analytic Procedure

Descriptive statistics for males and females were obtained from the International Association for the Evaluation of Educational Achievement (IEA). Software provided by the IEA and the SPSS package were used to undertake the jackknife statistical procedure for the imputation of student achievement scores as per the TIMSS technical manual. From this, the mean and standard deviation of males and females were calculated for each nation, allowing for the calculation of national and cross-national gender differences. Effect sizes are reported as the mean difference between males and females in standardized units (Cohen, 1988; Hedges, 2008), which is commonly referred to as Cohen's *d*. Positive values for *d* indicate higher male performance while negative values indicate higher female performance (Hyde, 2005).

Although statistically significant differences are often found with large sample sizes, greater emphasis was placed on the effect sizes as these give an indication of the actual magnitude of group differences (Hedges, 2008; Wilkinson, 1999). As a guideline for interpretation, Hyde (2005, p. 582) advocated that even statistically significant effect sizes of  $d < .10$  be considered 'trivial' in magnitude. We note instances where effect sizes exceed Hyde's criterion for non-trivial gender differences.

Comprehensive Meta Analysis (CMA) V2 software was used to perform the meta-analyses (Borenstein & Rothstein, 1999). Meta-analysis typically employs either a fixed-effects or random-effects model for combining samples (Rosenthal & DiMatteo, 2001). Because the 2011 TIMSS assessment draws independent samples from a range of host countries and educational environments, a random-effects model was chosen for combining samples (Borenstein, Hedges, Higgins, & Rothstein, 2009). The random effects model produces wider confidence intervals than a fixed-effects model, but yields a more appropriate estimate of the degree of variability across samples (Hunter & Schmidt, 2000; Rosenthal & DiMatteo, 2001).

Based on the findings of previous studies (e.g. Guiso et al., 2008), we also investigated the contribution of national-level indicators of gender equality to each

country's gender differences in mathematics and science achievement. We employed a random effects model (method of moments) for conducting meta-regression to test whether the gender equality acted as a moderator (Borenstein et al., 2009; Thompson & Higgins, 2002). Meta-regression was only performed on achievement scores (and not the other measures of attitude and self-efficacy) in order to minimize the cumulative Type 1 error rate that would be present from performing six separate meta-regressions.

Additionally, we examined whether there were gender differences in variance as predicted by the greater male variability hypothesis. Following the method of Feingold (1992), variance ratios (VR) were defined as the ratio of male to female variance in achievement, and calculated for each country. A VR of 1.00 would indicate that males and females are equal in variance. VR values less than 1.00 indicate that females show more variability than males, while VR values greater than 1.00 reflect greater male variability (Priess & Hyde, 2010). Feingold (1994) has argued that values between .90 and 1.10 ought to be regarded as negligible (i.e., homogeneity of variance), and this convention is adopted herein.

## Results

### Mathematics Achievement

Across all nations, there was no significant gender difference in mathematical achievement globally, with a weighted mean effect size of  $d = -.04$  [95%CI  $-.09$  to  $+.02$ ],  $Z = -1.27$ ,  $p = .205$ . There was however considerable variation from nation to nation, with substantial heterogeneity in effect sizes,  $Q(44) = 2155.68$ ,  $p < .001$ ,  $I^2 = 97.96$  (see Figure 1). As can be seen in Table 2, some nations showed sizeable differences favoring males, whereas others showed sizeable differences favoring females (range  $d = -.60$  to  $+.31$ ). Of the nations that showed greater female performance, the largest differences were found in Oman, Bahrain and Jordan. The largest gender differences where males scored higher were Honduras, Ghana and New Zealand. Thus, there is no clear advantage to either gender when viewed globally but important differences are present at the national level.

Previous research showed that gender equality was associated with smaller gender differences in mathematics in older students (Guiso et al., 2008; Reilly, 2012), but this finding was not replicated with the eighth grade sample examined here,  $Z = 1.16$ ,  $p = .246$ . Surprisingly, some of the largest gender differences favoring females (i.e., Oman, Bahrain and Jordan) occurred in nations that score comparatively low on measures of gender equality relative to other nations.

To further explore why some nations showed gender differences and others did not, we tested whether there were differences between developing and developed nations in mathematics achievement. Prior research had identified markedly different patterns of gender

differences between OECD and non-OECD countries under the PISA assessments (Reilly, 2012). Accordingly, subgroup analysis was performed. While there was a slight trend towards higher mathematics scores for boys in OECD nations, it fell short of statistical significance  $d = +.04$  [95%CI  $-.05$  to  $+.13$ ],  $Z = .92$ ,  $p = .357$ . However, there was a statistically significant difference for non-OECD nations favoring females,  $d = -.07$  [95%CI  $-.14$  to  $-.02$ ],  $Z = -2.29$ ,  $p = .022$ . Gender variance ratios for individual nations are also reported in Table 2. In support of the greater male variability hypothesis, almost all nations showed greater variability in the performance of males than females. Chile, Honduras, Macedonia, Morocco and Romania were the only exceptions. The weighted mean average variance ratio for mathematics achievement was 1.16, which exceeds Feingold's (1994) threshold for non-trivial gender ratios.

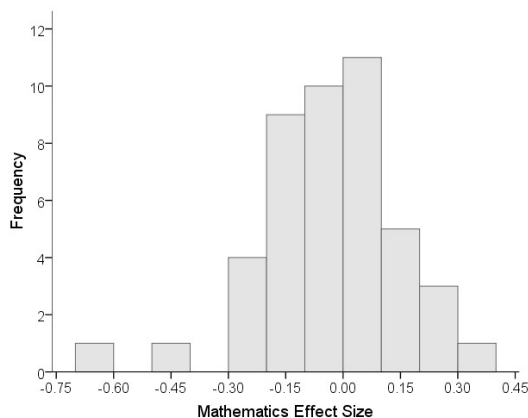


Figure 1. Histogram showing distribution of national gender difference effect sizes in mathematics achievement for the 2011 TIMSS Assessment

**Mathematics Attitudes and Self-Efficacy Beliefs**

Overall males reported slightly more positive attitudes towards mathematics than females, with a weighted effect size of  $d = .05$  [95%CI  $.01$  to  $.09$ ],  $Z = 2.35$ ,  $p = .019$  (see Table 3 for individual nations). However, larger gender differences were present for many individual nations (ranging from  $d = -.32$  to  $+.36$ ), with Japan, Hong Kong and Oman showing the largest gender differences in attitudes towards mathematics. Subgroup analysis showed larger effects within OECD nations,  $d = +.11$ ,  $Z = 3.30$ ,  $p = .001$ , but non-significant effects across non-OECD nations. These findings are consistent with previous studies finding more positive attitudes towards mathematics in males generally (Else-Quest et al., 2013; Su, Rounds, & Armstrong, 2009), but it is interesting to note that these were not universal in our sample. Under the right educational and cultural circumstances, gender differences in attitudes towards mathematics can emerge for either gender.

Gender variance ratios for individual nations are also reported in Table 2. In support of the greater male variability hypothesis, almost all nations showed greater variability in the performance of males than females. Chile, Honduras, Macedonia, Morocco and Romania were the only exceptions. The weighted mean average variance ratio for mathematics achievement was 1.16, which exceeds Feingold's (1994) threshold for non-trivial gender ratios.

**Mathematics Attitudes and Self-Efficacy Beliefs**

Overall males reported slightly more positive attitudes towards mathematics than females, with a weighted effect size of  $d = .05$  [95%CI  $.01$  to  $.09$ ],  $Z = 2.35$ ,  $p = .019$  (see Table 3 for individual nations). However, larger gender differences were present for many individual nations (ranging from  $d = -.32$  to  $+.36$ ), with Japan, Hong Kong and Oman showing the largest gender differences in attitudes towards mathematics. Subgroup analysis showed larger effects within OECD nations,  $d = +.11$ ,  $Z = 3.30$ ,  $p = .001$ , but non-significant effects across non-OECD nations. These findings are consistent with previous studies finding more positive attitudes towards mathematics in males generally (Else-Quest et al., 2013; Su et al., 2009), but it is interesting to note that these were not universal in our sample. Under the right educational and cultural circumstances, gender differences in attitudes towards mathematics can emerge for either gender.

Next we examined whether there were gender differences in self-efficacy beliefs about mathematical ability (Else-Quest et al., 2010; Fennema et al., 1990). There were statistically significant gender differences in mathematics self-efficacy in most nations, with a weighted mean effect size favoring males of  $d = +.17$  [95%CI  $.12$  to  $.21$ ],  $Z = 7.93$ ,  $p < .001$ , which exceeds Hyde's (2005) criterion for trivial effect sizes (see Table 3). Subgroup analysis showed larger effects within OECD nations,  $d = +.26$ ,  $Z = 8.22$ ,  $p < .001$  than non-OECD nations,  $d = +.12$ ,  $Z = 5.15$ ,  $p < .001$ . While boys in most individual nations reported greater self-efficacy in mathematics than girls, the largest effect sizes were found in England, Japan, New Zealand and Hong Kong. In only two nations (Oman and Bahrain) did females report meaningfully higher self-efficacy beliefs for mathematics, and this may be related to the superior female achievement in mathematics noted above for these countries (see Table 2).

**Science Achievement**

Although most individual nations showed statistically significant gender differences favoring either males or females, when averaged across all nations, there was no statistically significant gender difference in science achievement, with a weighted mean effect size of  $d = -.05$  [95%CI  $-.11$  to  $+.02$ ],  $Z = -1.39$ ,  $p = .164$ . There was considerable cross-cultural variability in gender



differences (ranging from  $d = -.74$  to  $d = +.27$ ) with some nations showing a male advantage and others a female advantage (see Table 4). This was reflected in heterogeneity statistics for science achievement,  $Q(44) = 3234.33, p < .001, I^2 = 98.64$  (see Figure 2). Some of the largest gender differences favoring males were found for Ghana, Honduras, Tunisia and New Zealand, while medium to large gender differences in favor of females were found in Oman, Bahrain, Jordan and Saudi Arabia. As with mathematics, the lack of uniformity in direction is inconsistent with biologically-based explanations for innate gender differences.

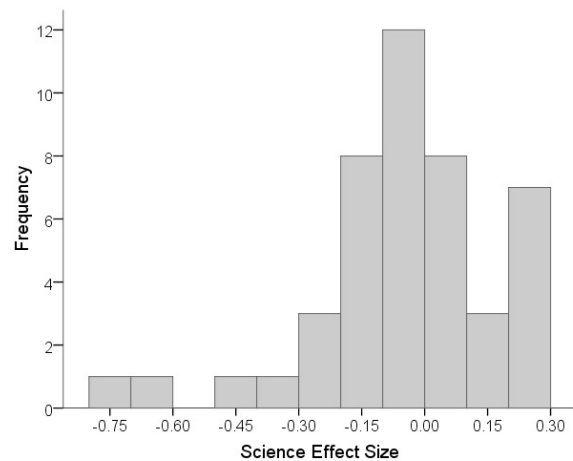


Figure 2. Histogram showing distribution of national gender difference effect sizes in science achievement for the 2011 TIMSS Assessment

Table 2  
National Gender Differences in TIMSS Mathematics Achievement and Variance Ratios (VR)

Country	Cohen's $d$	95% Confidence Interval		Variance Ratio	Test of null (2-tail)	
		Lower limit	Upper limit	(VR)	Z-value	P-value
<i>(OECD Nations)</i>						
Australia	.10 <sup>+</sup>	.06	.15	1.13 <sup>+</sup>	4.45	< .001***
Chile	.18 <sup>+</sup>	.13	.23	0.97	6.94	< .001***
England	-.03	-.10	.03	1.15 <sup>+</sup>	-1.01	.313
Finland	-.07	-.13	-.01	1.07	-2.24	.025*
Hungary	.06	.01	.12	1.02	2.26	.024*
Israel	-.09	-.14	-.03	1.31 <sup>+</sup>	-2.92	.004**
Italy	.15 <sup>+</sup>	.08	.21	1.03	4.60	< .001***
Japan	.09	.03	.15	1.22 <sup>+</sup>	3.01	.003**
New Zealand	.22 <sup>+</sup>	.16	.27	1.13 <sup>+</sup>	7.89	< .001***
Norway	-.04	-.10	.02	1.05	-.13	.205
South Korea	.07	.02	.13	1.08	2.56	.010*
Sweden	-.06	-.11	-.00	1.02	-2.12	.034*
United States	.05	.01	.08	1.04	2.35	.019*
<i>(Non-OECD)</i>						
Armenia	-.11 <sup>+</sup>	-.16	-.06	1.16 <sup>+</sup>	-4.08	< .001***
Bahrain	-.44 <sup>+</sup>	-.50	-.38	1.48 <sup>+</sup>	-14.83	< .001***
Botswana	-.18 <sup>+</sup>	-.23	-.12	1.11 <sup>+</sup>	-6.50	< .001***
Georgia	.03	-.03	.08	1.18 <sup>+</sup>	.89	.371

Ghana	.28 <sup>+</sup>	.23	.32	1.01	11.78	< .001***
Honduras	.31 <sup>+</sup>	.25	.37	0.98	10.03	< .001***
Hong Kong	-.07	-.13	-.01	1.18 <sup>+</sup>	-2.13	.033*
Indonesia	-.15 <sup>+</sup>	-.21	-.10	1.05	-5.88	< .001***
Iran	.07	.02	.12	1.10 <sup>+</sup>	2.78	.005**
Jordan	-.29 <sup>+</sup>	-.33	-.24	1.46 <sup>+</sup>	-12.53	< .001***
Kazakhstan	.02	-.04	.08	1.10 <sup>+</sup>	.70	.483
Lebanon	.16 <sup>+</sup>	.10	.23	1.11 <sup>+</sup>	5.15	< .001***
Lithuania	-.12 <sup>+</sup>	-.17	-.06	1.14 <sup>+</sup>	-4.06	< .001***
Macedonia	-.06	-.13	-.00	0.99	-2.02	.043*
Malaysia	.20 <sup>+</sup>	-.25	-.15	1.28 <sup>+</sup>	-7.66	< .001***
Morocco	.00	-.04	.04	0.95	.05	.960
Oman	-.60 <sup>+</sup>	-.64	-.56	1.44 <sup>+</sup>	-28.79	< .001***
Palestinian National Authority	-.23 <sup>+</sup>	-.28	-.19	1.26 <sup>+</sup>	-10.10	< .001***
Qatar	-.10 <sup>+</sup>	-.16	-.04	1.10 <sup>+</sup>	-3.37	.001**
Romania	-.11 <sup>+</sup>	-.16	-.05	0.96	-3.96	< .001***
Russia	.01	-.05	.06	1.11 <sup>+</sup>	.30	.762
Saudi Arabia	-.16 <sup>+</sup>	-.22	-.10	1.52 <sup>+</sup>	-5.26	< .001***
Singapore	-.10 <sup>+</sup>	-.15	-.05	1.34 <sup>+</sup>	-3.90	< .001***
Slovenia	.07	.01	.12	1.03	2.17	.030*
South Africa	-.04	-.08	-.00	1.15 <sup>+</sup>	-2.18	.029*
Syria	.11 <sup>+</sup>	.05	.17	1.04	3.62	< .001***
Taiwan	-.06	-.11	-.00	1.23 <sup>+</sup>	-2.12	.034*
Thailand	-.21 <sup>+</sup>	-.26	-.16	1.22 <sup>+</sup>	-7.99	< .001***
Tunisia	.22 <sup>+</sup>	.17	.28	1.00	7.88	< .001***
Turkey	-.08	-.12	-.03	1.05	-3.16	.002**
Ukraine	.04	-.03	.10	1.25 <sup>+</sup>	1.04	.301
United Arab Emirates	-.19 <sup>+</sup>	-.23	-.16	1.41 <sup>+</sup>	-11.46	< .001***

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

<sup>+</sup> indicates effect size exceeds Hyde's criterion for non-trivial differences ( $d \geq .10$  or  $d \leq -.10$ ) or in the case of variance ratios, a ratio that exceeds Feingold's suggested cutoffs for non-trivial gender ratios. Positive values for  $d$  indicate males score higher than females, while negative values indicate females score higher than males.

Table 3

*Gender Differences in Attitudinal Ratings and Self-Efficacy Beliefs for Mathematics*

Country	Mathematics Attitude	Test of null (2-tail)		Mathematics Self-Efficacy	Test of null (2-tail)	
	Cohen's <i>d</i>	Z-value	P-value	Cohen's <i>d</i>	Z-value	P-value
<i>(OECD Nations)</i>						
Australia	.14 <sup>+</sup>	6.25	< .001***	.32 <sup>+</sup>	13.83	< .001***
Chile	.21 <sup>+</sup>	8.09	< .001***	.33 <sup>+</sup>	12.31	< .001***
England	.28 <sup>+</sup>	8.61	< .001***	.45 <sup>+</sup>	13.64	< .001***
Finland	-.02	-0.65	.518	.21 <sup>+</sup>	6.99	< .001***
Hungary	.00	0.03	.973	.23 <sup>+</sup>	8.11	< .001***
Israel	-.01	-0.18	.855	.10 <sup>+</sup>	3.34	.001**
Italy	.19 <sup>+</sup>	6.12	< .001***	.34 <sup>+</sup>	9.99	< .001***
Japan	.36 <sup>+</sup>	11.72	< .001***	.46 <sup>+</sup>	15.18	< .001***
New Zealand	.29 <sup>+</sup>	10.54	< .001***	.43 <sup>+</sup>	15.50	< .001***
Norway	.00	-0.13	.900	.18 <sup>+</sup>	5.46	< .001***
South Korea	.14 <sup>+</sup>	5.04	< .001***	.23 <sup>+</sup>	8.17	< .001***
Sweden	.07	2.57	.010**	.26 <sup>+</sup>	9.47	< .001***
United States	.08	4.13	< .001***	.21 <sup>+</sup>	10.50	< .001***
<i>(Non-OECD)</i>						
Armenia	-.08	-2.96	.003	.10 <sup>+</sup>	3.80	< .001***
Bahrain	-.02	-.81	.415	-.12 <sup>+</sup>	-4.07	< .001***
Botswana	-.06	-2.32	.020*	.15 <sup>+</sup>	5.50	< .001***
Georgia	.04	1.37	.171	.22 <sup>+</sup>	7.30	< .001***
Ghana	.16 <sup>+</sup>	6.72	< .001***	.28 <sup>+</sup>	11.88	< .001***
Honduras	.14 <sup>+</sup>	4.49	< .001***	.22 <sup>+</sup>	7.30	< .001***
Hong Kong	.32 <sup>+</sup>	9.93	< .001***	.43 <sup>+</sup>	13.44	< .001***
Indonesia	-.13 <sup>+</sup>	-4.84	< .001***	.07	2.73	.006**
Iran	.15 <sup>+</sup>	5.86	< .001***	.13 <sup>+</sup>	4.94	< .001***
Jordan	.14 <sup>+</sup>	5.92	< .001***	.01	.32	.752
Kazakhstan	-.15 <sup>+</sup>	-4.96	< .001***	-.04	-1.21	.228
Lebanon	.10 <sup>+</sup>	3.15	< .001***	.13 <sup>+</sup>	4.05	< .001***
Lithuania	-.02	-0.56	.577	.11 <sup>+</sup>	3.95	< .001***

Macedonia	-.00	-0.09	.929	.05	1.67	.094
Malaysia	-.19 <sup>+</sup>	-7.34	< .001***	.13 <sup>+</sup>	4.77	< .001***
Morocco	.06	2.71	.007**	.23 <sup>+</sup>	10.67	< .001***
Oman	-.32 <sup>+</sup>	-15.49	< .001***	-.17 <sup>+</sup>	-8.31	< .001***
Palestinian National Authority	-.05	-2.18	.029*	-.03	-1.39	.166
Qatar	.16 <sup>+</sup>	5.26	< .001***	.04	1.44	.150
Romania	-.08	-2.93	.003*	.09	3.49	< .001***
Russia	.03	.96	.338	.07	2.29	.022*
Saudi Arabia	.09	2.88	.004**	.11 <sup>+</sup>	3.63	< .001***
Singapore	-.01	-0.32	.748	.19 <sup>+</sup>	7.39	< .001***
Slovenia	.00	0.02	.985	.19 <sup>+</sup>	6.23	< .001***
South Africa	.02	1.01	.314	.12 <sup>+</sup>	6.82	< .001***
Syria	.05	1.55	.120	.09	2.83	< .001***
Taiwan	.19 <sup>+</sup>	6.82	< .001***	.37 <sup>+</sup>	13.00	< .001***
Thailand	-.14 <sup>+</sup>	-5.57	< .001***	.18 <sup>+</sup>	6.98	< .001***
Tunisia	.10 <sup>+</sup>	3.45	< .001***	.21 <sup>+</sup>	7.67	< .001***
Turkey	-.07	-2.75	.006**	.05	2.11	.003**
Ukraine	.03	1.88	.061	.10 <sup>+</sup>	6.05	< .001***
United Arab Emirates	.00	0.07	.943	.02	.50	.619

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

<sup>+</sup> indicates effect size exceeds Hyde's criterion for non-trivial differences ( $d \geq .10$  or  $d \leq -.10$ ). Positive values for  $d$  indicate males score higher than females, while negative values indicate females score higher than males.

Table 4

*National Gender Differences in TIMSS Science Achievement and Variance Ratios (VR)*

Country	Cohen's <i>d</i>	95% Confidence Interval		Variance Ratio (VR)	Test of null (2-tail)	
		Lower limit	Upper limit		Z-value	P-value
<i>(OECD Nations)</i>						
Australia	.19 <sup>+</sup>	.15	.24	1.20 <sup>+</sup>	8.39	< .001***
Chile	.21 <sup>+</sup>	.16	.27	1.02	8.14	< .001***
England	-.02	-.09	.04	1.20 <sup>+</sup>	-.74	.456
Finland	-.07	-.13	-.01	1.17 <sup>+</sup>	-2.44	.015*
Hungary	.22 <sup>+</sup>	.16	.27	1.02	7.80	< .001***
Israel	-.07	-.13	-.02	1.37 <sup>+</sup>	-2.54	.011*
Italy	.20 <sup>+</sup>	.14	.26	1.05	6.32	< .001***
Japan	.11 <sup>+</sup>	.05	.16	1.20 <sup>+</sup>	3.52	< .001***
New Zealand	.24 <sup>+</sup>	.18	.29	1.14 <sup>+</sup>	8.68	< .001***
Norway	-.02	-.08	.05	1.15 <sup>+</sup>	-.47	.638
South Korea	.06	.01	.11	1.19 <sup>+</sup>	2.15	.031*
Sweden	-.04	-.09	.02	1.11 <sup>+</sup>	-1.37	.171
United States	.14 <sup>+</sup>	.10	.18	1.11 <sup>+</sup>	7.08	< .001***
<i>(Non- OECD)</i>						
Armenia	-.19 <sup>+</sup>	-.25	-.14	1.12 <sup>+</sup>	-7.43	< .001***
Bahrain	-.60 <sup>+</sup>	-.66	-.54	1.59 <sup>+</sup>	-20.07	< .001***
Botswana	-.10 <sup>+</sup>	-.16	-.05	1.15 <sup>+</sup>	-3.79	< .001***
Georgia	-.12 <sup>+</sup>	-.17	-.06	1.28 <sup>+</sup>	-3.88	< .001***
Ghana	.27 <sup>+</sup>	.22	.31	1.03	11.38	< .001***
Honduras	.26 <sup>+</sup>	.20	.32	1.03	8.61	< .001***
Hong Kong	-.02	-.08	.04	1.18 <sup>+</sup>	-.70	.483
Hungary	.22 <sup>+</sup>	.16	.27	1.02	7.80	< .001***
Indonesia	-.09	-.14	-.04	1.07	-3.43	.001**
Iran	-.06	-.11	-.01	1.12 <sup>+</sup>	-2.29	.022*
Jordan	-.43 <sup>+</sup>	-.48	-.39	1.65 <sup>+</sup>	-18.76	< .001***
Kazakhstan	-.05	-.11	.01	1.16 <sup>+</sup>	-1.60	.111
Lebanon	.04	-.02	.10	1.15 <sup>+</sup>	1.28	.201
Lithuania	-.11 <sup>+</sup>	-.17	-.05	1.16 <sup>+</sup>	-3.77	< .001***

Macedonia	-.16 <sup>+</sup>	-.22	-.10	1.05	-5.01	< .001***
Malaysia	-.15 <sup>+</sup>	-.20	-.10	1.28 <sup>+</sup>	-5.62	< .001***
Morocco	-.04	-.08	-.00	1.03	-1.96	.050
Oman	-.74 <sup>+</sup>	-.78	-.70	1.60 <sup>+</sup>	-35.09	< .001***
Palestinian National Authority	-.26 <sup>+</sup>	-.31	-.22	1.30 <sup>+</sup>	-11.46	< .001***
Qatar	-.21 <sup>+</sup>	-.27	-.15	1.02	-7.07	< .001***
Romania	-.02	-.08	.03	1.05	-.86	.389
Russia	.09	.03	.14	1.13 <sup>+</sup>	3.07	.002**
Saudi Arabia	-.32 <sup>+</sup>	-.38	-.26	1.53 <sup>+</sup>	-10.48	< .001***
Singapore	.02	-.04	.07	1.29 <sup>+</sup>	.058	.561
Slovenia	.05	-.00	.11	1.13 <sup>+</sup>	1.82	.069
South Africa	-.06	-.10	-.02	1.10 <sup>+</sup>	-3.26	.001**
Syria	.07	.01	.13	1.05	2.29	.022*
Taiwan	.00	-.05	.06	1.26 <sup>+</sup>	0.09	.972
Thailand	-.19 <sup>+</sup>	-.24	-.14	1.32 <sup>+</sup>	-7.35	< .001***
Tunisia	.25 <sup>+</sup>	.20	.31	1.01	8.93	< .001***
Turkey	-.15 <sup>+</sup>	-.20	-.11	1.16 <sup>+</sup>	-6.35	< .001***
Ukraine	.05	-.02	.12	1.25 <sup>+</sup>	1.46	.143
United Arab Emirates	-.26 <sup>+</sup>	-.29	-.22	1.46 <sup>+</sup>	-15.20	< .001***

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

+ indicates effect size exceeds Hyde's criterion for non-trivial differences ( $d \geq .10$ ), or in the case of variance ratios, a ratio that exceeds Feingold's suggested cutoffs for non-trivial gender ratios. Positive values for  $d$  indicate males score higher than females, while negative values indicate females score higher than males.

Previous cross-cultural research had noted a significant difference in science achievement in OECD nations favoring males and the reverse pattern in non-OECD countries (Reilly, 2012). When grouped according to OECD status, this effect was partially replicated with eighth grade students. Amongst non-OECD nations, girls scored significantly higher in science achievement than boys,  $d = -.11$  [95%CI = -.18 to -.03],  $Z = -2.78$ ,  $p = .005$ . This effect size also exceeds Hyde (2005)'s criterion for non-trivial gender differences. For OECD nations, there was a slight trend in the opposite direction but this fell short of statistical significance,  $d = +.07$  [95%CI = -.04 to +.18],  $Z = 1.30$ ,

$p = .192$ . Thus it would appear that gender differences in science are highly culturally bound.

To further explore the cultural contributions to gender differences in science skills, a meta-regression was conducted with gender equality as the predictor. There was a significant positive association between gender equality and the size of gender differences in science achievement,  $Z = 2.11$ ,  $p = .035$ , indicating that in more gender equal countries we see a pattern of higher male achievement while for nations with lower gender equality we see a pattern of higher female achievement (see Figure 3). Though it may seem a counter-intuitive finding, we address this later in the discussion.

In support of the greater male variability hypothesis, all nations in the 2011 TIMSS sample showed greater male variability (i.e. VRs > 1) in science achievement. The weighted mean average variance ratio for all countries was 1.21, which is considerably greater than Feingold’s suggested cutoffs for non-trivial gender ratios.

**Science Attitudes and Self-Efficacy**

While all nations completed the mathematics and science assessments, only a subset of countries ( $n = 26$ ) administered the complete student survey booklet assessing attitudes towards science and science self-

efficacy (see Table 5). Boys reported more positive attitudes towards learning science than girls, with a small difference of  $d = +.07$  [95%CI .01 to .14],  $Z = 2.25$ ,  $p = .024$ . There was also wide variation between countries in attitudes towards science (ranging from  $d = -.27$  to  $+.48$ ). Japan, Taiwan, England and Oman showed the largest gender differences in attitudes towards learning science. Subgroup analysis showed that greater male interest in science was confined mainly to OECD nations,  $d = +.17$ ,  $Z = 3.70$ ,  $p < .001$ , with no significant difference in science interest for non-OECD nations collectively.

Table 5  
*Gender Differences in Attitudinal Ratings and Self-Efficacy Beliefs for Science*

Country	Science Attitudes	Test of null (2-tail)		Science Self-Efficacy	Test of null (2-tail)	
	Cohen’s $d$	Z-value	P-value	Cohen’s $d$	Z-value	P-value
<i>(OECD Nations)</i>						
Australia	.19 <sup>+</sup>	8.44	< .001***	.23 <sup>+</sup>	10.06	< .001***
Chile	.03	.96	.339	.07	2.54	.011*
England	.30 <sup>+</sup>	9.11	< .001***	.38 <sup>+</sup>	11.66	< .001***
Israel	.10 <sup>+</sup>	3.28	< .001***	.05	1.82	.068
Italy	.13 <sup>+</sup>	4.04	< .001***	.13 <sup>+</sup>	4.07	< .001***
Japan	.48 <sup>+</sup>	15.81	< .001***	.49 <sup>+</sup>	15.94	< .001***
New Zealand	.17 <sup>+</sup>	6.24	< .001***	.23 <sup>+</sup>	8.27	< .001***
Norway	.25 <sup>+</sup>	7.78	< .001***	.27 <sup>+</sup>	8.38	< .001***
South Korea	.23 <sup>+</sup>	8.38	< .001***	.21 <sup>+</sup>	7.68	< .001***
United States	.20 <sup>+</sup>	10.31	< .001***	.20 <sup>+</sup>	10.38	< .001***
<i>(Non-OECD)</i>						
Bahrain	.01	.43	.670	-.08	-2.70	.007**
Botswana	-.07	-2.50	.012*	.16 <sup>+</sup>	5.72	< .001***
Ghana	.10 <sup>+</sup>	4.37	< .001***	.18 <sup>+</sup>	7.65	< .001***
Honduras	-.06	-2.02	.044*	.02	.62	.536
Hong Kong	.26 <sup>+</sup>	8.25	< .001***	.30 <sup>+</sup>	9.48	< .001***
Iran	.01	.51	.607	-.06	-2.48	.013*
Jordan	-.18 <sup>+</sup>	-7.96	< .001***	-.28 <sup>+</sup>	-12.26	< .001***
Malaysia	-.04	-1.64	.102	.14 <sup>+</sup>	5.29	< .001***
Oman	-.27 <sup>+</sup>	-13.18	< .001***	-.29 <sup>+</sup>	-13.95	< .001***
Palestinian Authority	-.08	-3.63	< .001***	-.08	-3.34	.001**
Qatar	.06	2.03	.043*	-.06	-1.86	.063
Saudi Arabia	.00	.15	.881	-.10 <sup>+</sup>	-3.14	.002**
Singapore	.24 <sup>+</sup>	9.20	< .001***	.39 <sup>+</sup>	14.79	< .001***
South Africa	.03	1.64	.101	.04	2.32	.021*
Taiwan	.36 <sup>+</sup>	12.66	< .001***	.43 <sup>+</sup>	15.08	< .001***
Thailand	-.13 <sup>+</sup>	-5.23	< .001***	.10 <sup>+</sup>	3.91	< .001***
Tunisia	-.10 <sup>+</sup>	-3.62	< .001***	-.10 <sup>+</sup>	-3.72	< .001***

Turkey	-.16 <sup>+</sup>	-6.63	< .001***	-.07	-2.87	.004**
United Arab Emirates	.04	2.40	.016*	-.04	-2.58	.010**

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

<sup>+</sup> indicates effect size exceeds Hyde’s criterion for non-trivial differences ( $d \geq .10$  or  $d \leq -.10$ ). Positive values for  $d$  indicate males score higher than females, while negative values indicate females score higher than males.

Next, we tested for gender differences in science self-efficacy. Boys reported higher self-efficacy in science than girls,  $d = +.10$  [95%CI +.03 to +.17],  $Z = 2.69$ ,  $p = .007$ . Again there was considerable variability in national gender differences in science self-efficacy, ranging from  $-.29$  to  $+.49$ . In general, girls have lower science self-efficacy beliefs across nations despite there being no significant difference in global science achievement scores. Subgroup analysis found the effect was stronger across OECD nations,  $d = +.20$ ,  $Z = 3.62$ ,  $p < .001$ , but fell short of statistical significance across non-OECD nations.

**Discussion**

National and international gender difference comparisons for the 2011 wave of TIMSS show a complex and nuanced picture, with small to medium effect sizes in mathematics and science achievement at the national level but no clear advantage to either gender when viewed from a global perspective. Thus, there is only partial support for the gender differences hypothesis in this age group, with meaningful gender differences present for some - but not all - countries. For example there was a small gender difference favoring males in science for the United States, Australia and New Zealand, but moderately large differences favoring females in the Middle-Eastern countries of Oman, Bahrain and Jordan. The lack of uniformity in the direction of mathematics (Figure 1) and science achievement (Figure 2) would be inconsistent with the notion of large innate differences in quantitative reasoning, and instead reflect sociocultural factors.

Earlier rounds of TIMSS had found that boys performed significantly higher than girls in mathematics achievement (Mullis et al., 2000), but a more recent study by Else-Quest (2010) did not. Thus we were able to replicate Else-Quest et al.’s finding in the subsequent round of TIMSS. A previous study by Neuschmidt et al. (2008) also reported that in earlier waves of TIMSS, most countries showed significant gender differences in science achievement favoring males. A markedly different pattern of gender differences emerged here for the TIMSS 2011 dataset, with a large number of countries reporting a significant female advantage in science achievement. This was

sufficient to lead to an elimination of global gender differences in science achievement, and may reflect the greater representation of Middle-Eastern nations noted above that were participating in TIMSS for the first time.

The absence of a gender difference in mathematics and science achievement under TIMSS stands in contrast to findings with older students in PISA (Guiso et al., 2008; Reilly, 2012). While it is possible that fundamental differences in test content between TIMSS and PISA may be a contributor, both assessment frameworks have shown good construct validity in their comprehensive assessment of mathematics and science. Eighth grade may simply be too early to detect emerging gender differences in mathematics and science achievement cross-culturally, but it appears that what minimal gender gaps exist in eighth grade widen considerably by the final year of compulsory schooling when PISA assessments occur (Guiso et al., 2008; Reilly, 2012) The issue could be best settled by longitudinal evidence that tracks students as they move throughout high school, but at present there are no large scale cross-cultural assessments available that would meet these criteria.

An alternate explanation for the presence of national but not cross-national gender differences in achievement might also be the composition of countries that participate in PISA and TIMSS assessments. PISA includes all OECD member countries and only a limited number of partners, and therefore recruits a higher proportion of Western countries. In contrast TIMSS represent a larger proportion of developing economies, with greater coverage of Asian, African and Middle-Eastern countries. It is possible that the findings of meaningfully sized gender differences from more affluent and developed economies do not generalize well to non-Western cultures. For example, the TIMSS data showed that Oman, Bahrain and Jordan had considerably sized gender differences in mathematics and science achievement favoring girls. However, these countries do not participate in PISA testing, and hence would not be reflected in the findings of studies that use this data source.

In terms of gender differences at the level of individual countries, 38 of the 45 nations assessed showed significant gender differences in mathematics,



although the direction and magnitude varied. For many of these nations the magnitude of effect sizes exceeded Hyde's criterion for meaningfully-sized gender gaps, warranting further attention. For science, significant differences in achievement were found in 35 of the 45 nations surveyed with a tendency towards somewhat larger effect sizes than in mathematics. There was also considerable overlap between nations with significant gender differences in mathematics and science. This information may be useful for the planning of educational interventions as a matter of gender equity, particularly in those countries with larger gender differences. Furthermore, within OECD countries there was a slight trend towards higher mathematics and science achievement for males, but across non-OECD nations the opposite was found with significantly higher female scores. Teasing apart the many social, educational and cultural factors that contribute to the development of gender differences remains an ongoing research objective.

### Gender segregation hypothesis

In the TIMSS dataset, we were not able to replicate the earlier findings of Guiso et al. (2008) of an association between national-level indicators of gender equality and the magnitude of gender differences in mathematics. We offer two possible reasons for this anomaly. Firstly the age of the two samples could be a factor. Gender equality may exert a minimal effect as children enter high school, but may grow in increasing importance as students draw closer to the end of compulsory schooling. Secondly, as stated earlier, the composition of participating nations in PISA and TIMSS are different. The association between gender equality and the size of gender differences in mathematics might not generalize to all cultures and economies, as Else-Quest et al., (2008) found under the 2003 TIMSS assessment.

There was, however, a significant positive association between gender equality and the development of science achievement. Though it may seem counterintuitive, girls' science achievement was actually *higher* than their male peers in countries with comparatively lower levels of gender equality, while the gender gap grew smaller and then favored males in more gender equal nations. The observed pattern replicates the finding of Reilly (2012) who found a positive association between national GGI ratings and science in 15 year old students under PISA. STEM-related careers can be a pathway to a higher economic standard of living and increased job security. It may be the case that growing up in an environment where there are stark inequalities in the status of women, it is adaptive for girls to divert their efforts towards educational achievement in STEM when they are

afforded access to education and employment. In countries where women are not highly valued, girls might be more motivated to "prove" their value by doing well in school. If their countries lack infrastructure (roads, water supply, medical services) they might also be attracted to STEM related areas of study because they see them as relevant to improving conditions for society as well as providing a decent income for themselves. Conversely, growing up in a more gender-equal culture may grant girls the freedom to pursue other socially valued occupational skills such as language and the arts (Charles & Bradley, 2009). Gender equality is also strongly associated with economic prosperity (measured as gross domestic product or GDP), and in more affluent economies opportunities in the language and arts are more available.

Boys also reported greater self-efficacy beliefs for mathematics cross-culturally ( $d = +.17$ ), with somewhat larger effect sizes in the small to medium range across most nations. Once again there were some notable exceptions to the general trend, with a number of Middle-Eastern countries showing either negligible differences or a small female advantage in self-efficacy beliefs. It is also worth noting in those countries with greater female self-efficacy, girls also performed considerably higher than boys in mathematics achievement. However, the general trend across these nations was for boys to report being more confident than girls in mathematics which is consistent with prior research findings across recent decades (Huang, 2013; Hyde, Fennema, Ryan, Frost, & Hopp, 1990).

Eighth grade is a particularly sensitive time period for students, as students in most nations will soon have to make important decisions about course selection and whether to pursue more academically challenging subjects such as intermediate or advanced mathematics. Students may also begin to tentatively consider the prospect of future occupational choices (Jacobs et al., 2002), and whether mathematics will be part of their intellectual identity (Nosek et al., 2002; Steele, 1997). Personal attitudes towards mathematics and self-efficacy beliefs play a large part in such decisions, and research has shown that these are a significant factor in course selection (Eccles, 1987, 1994; Feniger, 2011; Riegle-Crumb, Farkas, & Muller, 2006). However, there is also a body of evidence suggesting that girls' lower self-efficacy beliefs also undermines achievement motivation and performance, through the mechanism of stereotype threat (Nguyen & Ryan, 2008; Spencer et al., 1999; Steele, 1997). This makes attitudes and self-efficacy beliefs ideal targets for educational intervention before the gender gap in mathematics widens in later years (Guiso et al., 2008; Reilly, 2012). Liben and Coyle (2014) review a

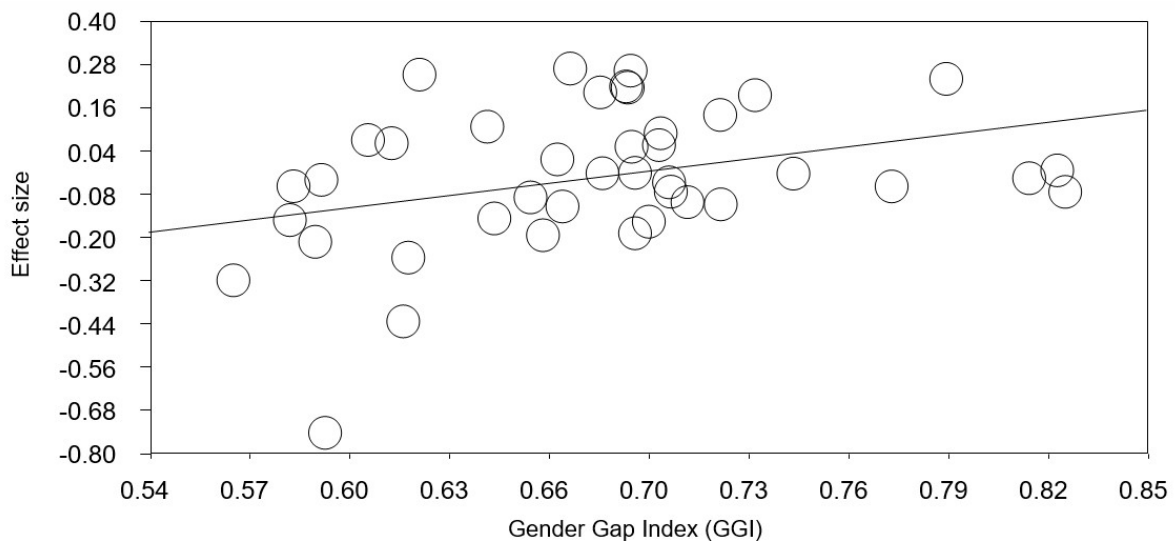


Figure 3. Meta-regression between gender equality (predictor) and gender gap in science achievement scores.

number of developmental interventions for girls in STEM that focus on affect and self-efficacy, as well as aiming to build resilience against stereotype threat (Miyake et al., 2010). These include providing additional spatial experiences and training, enrichment activities for students and STEM mentoring. Miyake et al. (2010) also argue that a brief values affirmation intervention exercise can also be protective, which involves students writing a brief essay on their personal character strengths. This activity appears to buffer against gender stereotypes associating STEM with males, and has been used across a range of educational settings (Kost-Smith et al., 2012).

#### Attitudes towards Science and Self-Efficacy Beliefs

Consistent with an earlier meta-analysis into student attitudes towards science (Weinburgh, 1995), boys reported more positive attitudes towards learning science across cultures although the mean effect size was small ( $d = +.07$ ). As with attitudes towards mathematics, there was wide variation across nations, with some nations showing a small-to-medium effect size, and other nations showing negligible differences in attitude or even a slight female advantage. Attitudes towards science can predict later academic performance and course taking decisions (Barnes, McInerney, & Marsh, 2005; Leibham, Alexander, & Johnson, 2013), making them important targets for interventions. Childrens' views of learning science are shaped by various sociocultural factors, as well as their own inherent interests and talents. However, they are subject to influence, with parents and teachers shaping

a child's view of science through attitudes and the provision of science-positive experiences (Bhanot & Jovanovic, 2009; Crowley, Callanan, Tenenbaum, & Allen, 2001; Frome & Eccles, 1998; Jodl, Michael, Malanchuk, Eccles, & Sameroff, 2001; Lynch, 2002). While the present study investigated attitudes and self-efficacy in Grade 8 students, these are also likely shaped by experiences earlier in their education (Halpern, Aronson, et al., 2007; Simpkins et al., 2005).

Given that some cultures showed negligible gender differences in science attitudes or even a small female preference, educational interventions and a concerted effort to provide early science learning opportunities may be warranted for those countries where girls report less interest than boys in learning science (Halpern, Aronson, et al., 2007; Shapiro & Williams, 2012). This is particularly important for countries where the general level of science interest is already low. Eccles (2007) noted that interest in science is often related to perceived relevance to future careers, and the compatibility between science and feminine life goals and values. Parents can play a role in encouraging girls to cultivate an interest in science through early science-enrichment activities for their daughters (Newcombe & Frick, 2010), and stressing the relevance of learning science to future occupational aspirations that have an impact on people and communities (Rozek, Hyde, Svoboda, Hulleman, & Harackiewicz, 2014).

Gender differences were also found for science self-efficacy beliefs, with a small but non-trivial effect size across countries favoring males ( $d = +.10$ ), and

small-to-medium effect sizes for individual nations. There was a trend towards larger gender differences in OECD nations. This finding is consistent with reports from other large scale international assessments (OECD, 2014). Low self-efficacy beliefs have the potential to undermine girls' confidence in undertaking science tasks and using lab equipment (Burkam, Lee, & Smerdon, 1997), and may also impair performance on standardized tests via the mechanism of stereotype threat (Spencer et al., 1999; Steele, 1997). However, lower female self-efficacy in science is not inevitable, as demonstrated by the higher female self-efficacy beliefs than males in several countries. Furthermore, educational interventions designed to improve girls' self-efficacy beliefs about STEM also show that it is not immutable (Luzzo, Hasper, Albert, Bibby, & Martinelli Jr, 1999). Higher self-efficacy translates into improved interest and achievement. For example, Miyake and colleagues have proposed a brief and simple values affirmation exercise for science classes (Miyake et al., 2010), which buffers against the deleterious effects of stereotype threat and raises women's achievement considerably. Their study has been subsequently replicated (Kost-Smith et al., 2012; Lauer et al., 2013) with large-sized effects between treatment and control groups. Encouraging a sense of self-confidence and mastery in science is crucial for students to reach their potential *regardless* of gender, but girls may benefit more from such interventions.

### Strengths and Limitations

Analysis of nationally-representative samples of student performance such as TIMSS offers greater confidence in the validity of any tests of gender differences and similarities than smaller selective samples (Anderson et al., 2007; Hedges & Nowell, 1995). It also provides the opportunity to examine sociocultural contributions to gender gaps. TIMSS uses an assessment framework that gives a good foundation for assessing the types of skills and tasks required for proficiency in later STEM subjects. It also allows us to examine attitudes towards mathematics and science, and self-efficacy beliefs, which some researchers argue may be more important than actual differences in ability (Else-Quest et al., 2013).

However, this study is not without its limitations. Firstly, it can only speak for the experiences of students in their eighth grade of schooling, and there are no large scale international assessments of mathematics and science between eighth grade under TIMSS, and the final year of compulsory schooling under PISA. In addition, a chief limitation of any study investigating international patterns of gender differences is how comprehensive the coverage of world nations is. The TIMSS 2011 assessment of mathematics and science represents only 45 countries

across the world, with many of the countries that participate in PISA not represented in TIMSS. However, it does include a higher proportion of Asian, African and Middle-Eastern nations, giving us greater cultural diversity and the capability to explore the influence of gender in developing economies. Since the overlap between countries participating in both TIMSS and PISA is small, we cannot discount the possibility that this may have contributed to the lack of global gender differences in mathematics and science achievement with this age group.

Additionally, while such studies can identify which countries and educational environments show larger gender differences in mathematics and science achievement, attitudes and self-efficacy, it cannot definitively determine *causality*. Many educational interventions and initiatives have been proposed or instituted (Buckley, 2016; Sander, Endepohls-Ulpe, & Quaiser-Pohl, 2016; Walters, 2010), but may lack efficacy evidence showing they translate into actual STEM outcomes. Further research is required into brief educational interventions such as values affirmation (Miyake et al., 2010), and spatial training (Newcombe et al., 2009; Uttal, Miller, & Newcombe, 2013) which are low-cost, and may provide benefits to older students as well as those beginning their education.

### Conclusions

While global gender gaps in mathematics and science achievement were negligible for eighth grade students, small to medium sized gender differences were found for many individual nations though the direction of gender differences varied. Furthermore, the performance of males showed greater variability than for females, which leads to a higher proportion of high-achieving male students in STEM. Boys also reported more favorable attitudes towards learning mathematics and science than girls, while girls reported lower self-efficacy beliefs. However, gender differences were not universal for all nations. For a number of Middle-Eastern nations, girls reported more positive perceptions towards STEM than boys. Considerable cross-cultural variability is supportive of the role that environment and culture play in the development – or suppression – of gender differences in STEM. Finally, we found mixed support for the gender segregation hypothesis for this age group. Significant differences in achievement were also found between OECD and non-OECD nations. Taken together these findings show that gender differences in attitudes, self-efficacy and achievement for mathematics and science are not universal, but that important differences also remain for specific countries.

## References

- Anderson, J., Lin, H.-S., Treagust, D., Ross, S., & Yore, L. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education*, 5(4), 591-614. doi: 10.1007/s10763-007-9090-y
- Baker, D. P., & Jones, D. P. (1993). Creating gender equality: Cross-national gender stratification and mathematical performance. *Sociology of Education*, 66(2), 91-103. doi: 10.2307/2112795
- Barnes, G., McInerney, D. M., & Marsh, H. W. (2005). Exploring sex differences in science enrolment intentions: An application of the general model of academic choice. *The Australian Educational Researcher*, 32(2), 1-23. doi: 10.1007/BF03216817
- Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, 107(5), 1860-1863. doi: 10.1073/pnas.0910967107
- Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological Science*, 11(6), 474-480. doi: 10.1111/1467-9280.00291
- Bhanot, R. T., & Jovanovic, J. (2009). The links between parent behaviors and boys' and girls' science achievement beliefs. *Applied Developmental Science*, 13(1), 42-59. doi: 10.1080/10888690802606784
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.
- Borenstein, M., & Rothstein, H. R. (1999). *Comprehensive meta-analysis: A computer program for research synthesis*. Englewood, NJ: BioStat.
- Buckley, S. (2016). *Gender and sex differences in student participation, achievement and engagement in mathematics*. Melbourne, Australia: Australian Council for Educational Research.
- Burkam, D. T., Lee, V. E., & Smerdon, B. A. (1997). Gender and science learning early in high school: Subject matter and laboratory experiences. *American Educational Research Journal*, 34(2), 297-331. doi: 10.3102/00028312034002297
- Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, 108(8), 3157-3162. doi: 10.1073/pnas.1014871108
- Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*, 114(4), 924-976. doi: 10.1086/595942
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Crowley, K., Callanan, M. A., Tenenbaum, H. R., & Allen, E. (2001). Parents explain more often to boys than to girls during shared scientific thinking. *Psychological Science*, 12(3), 258-261. doi: 10.1111/1467-9280.00347
- Dwyer, C. A., & Johnson, L. (1997). Grades, accomplishments, and correlates. In W. Willingham & N. S. Cole (Eds.), *Gender and fair assessment* (pp. 127-156). Mahwah, NJ: Erlbaum.
- Eagly, A. H., Wood, W., & Diekmann, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123-174). Mahwah, NJ: Lawrence Erlbaum Associates.
- Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly*, 11(2), 135-172. doi: 10.1111/j.1471-6402.1987.tb00781.x
- Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of Women Quarterly*, 18(4), 585-609. doi: 10.1111/j.1471-6402.1994.tb01049.x
- Eccles, J. S. (2007). Where are all the women? Gender differences in participation in physical science and engineering. In S. J. Ceci (Ed.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 199-210). Washington, D.C.: American Psychological Association.
- Eccles, J. S. (2013). Gender and achievement choices. In E. T. Gershoff, R. S. Mistry & D. Crosby (Eds.), *Societal Contexts of Child Development: Pathways of Influence and Implications for Practice and Policy* (pp. 19-34). New York: Oxford University Press.
- Else-Quest, N. M., & Grabe, S. (2012). The political is personal: Measurement and application of nation-level indicators of gender equity in psychological research *Psychology of Women*

- Quarterly*, 36(2), 131-144. doi: 10.1177/0361684312441592
- 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(1), 103-127. doi: 10.1037/a0018053
- Else-Quest, N. M., Mineo, C. C., & Higgins, A. (2013). Math and science attitudes and achievement at the intersection of gender and ethnicity. *Psychology of Women Quarterly*, 37(3), 293-309. doi: 10.1177/0361684313480694
- Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research*, 62(1), 61-84. doi: 10.3102/00346543062001061
- Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross-cultural perspective. *Sex Roles*, 30(1), 81-92. doi: 10.1007/BF01420741
- Feniger, Y. (2011). The gender gap in advanced math and science course taking: Does same-sex education make a difference? *Sex Roles*, 65(9), 670-679. doi: 10.1007/s11199-010-9851-x
- Fennema, E., Peterson, P. L., Carpenter, T. P., & Lubinski, C. A. (1990). Teachers' attributions and beliefs about girls, boys, and mathematics. *Educational Studies in Mathematics*, 21(1), 55-69. doi: 10.1007/BF00311015
- Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology*, 74(2), 435-452. doi: 10.1037/0022-3514.74.2.435
- Gallagher, A. M., & Kaufman, J. C. (Eds.). (2005). *Gender differences in mathematics*. New York: Cambridge University Press.
- Geary, D. C. (2010). *Male, female : the evolution of human sex differences* (2nd ed.). Washington, DC: American Psychological Association.
- Goldman, A. D., & Penner, A. M. (2014). Exploring international gender differences in mathematics self-concept. *International Journal of Adolescence and Youth*, 1-16. doi: 10.1080/02673843.2013.847850
- Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, 320(5880), 1164-1165. doi: 10.1126/science.1154094
- Gunderson, E., Ramirez, G., Levine, S. C., & Beilock, S. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, 66(3), 153-166. doi: 10.1007/s11199-011-9996-2
- Halim, M. L., & Ruble, D. N. (2010). Gender identity and stereotyping in early and middle childhood. In J. C. Chrisler & D. R. McCreary (Eds.), *Handbook of Gender Research in Psychology* (pp. 495-525). New York: Springer.
- Halpern, D. F., Aronson, J., Reimer, N., Simpkins, S., Star, J. R., & Wentzel, K. (2007). *Encouraging girls in math and science*. Washington, DC: National Center for Education Research, U.S. Department of Education.
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, 8(1), 1-51. doi: 10.1111/j.1529-1006.2007.00032.x
- Handelsman, J., Cantor, N., Carnes, M., Denton, D., Fine, E., Grosz, B., . . . Sheridan, J. (2005). More women in science. *Science*, 309(5738), 1190-1191. doi: 10.1126/science.1113252
- Hausmann, R., Tyson, L. D., & Zahidi, S. (2011). *The Global Gender Gap Report 2011*. Geneva, Switzerland: World Economic Forum.
- Haworth, C. M. A., Dale, P. S., & Plomin, R. (2010). Sex differences in school science performance from middle childhood to early adolescence. *International Journal of Educational Research*, 49(2-3), 92-101. doi: 10.1016/j.ijer.2010.09.003
- Hedges, L. V. (2008). What are effect sizes and why do we need them? *Child Development Perspectives*, 2(3), 167-171. doi: 10.1111/j.1750-8606.2008.00060.x
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269(5220), 41-45. doi: 10.1126/science.7604277
- Huang, C. (2013). Gender differences in academic self-efficacy: a meta-analysis. *European Journal of Psychology of Education*, 28(1), 1-35. doi: 10.1007/s10212-011-0097-y
- Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8(4), 275-292. doi: 10.1111/1468-2389.00156
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581-592. doi: 10.1037/0003-066X.60.6.581
- Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of

- mathematics attitudes and affect. *Psychology of Women Quarterly*, 14(3), 299-324. doi: 10.1111/j.1471-6402.1990.tb00022.x
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, 321(5888), 494-495. doi: 10.1126/science.1160364
- Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science*, 314(5799), 599-600. doi: 10.1126/science.1132154
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73(2), 509-527. doi: 10.1111/1467-8624.00421
- Jodl, K. M., Michael, A., Malanchuk, O., Eccles, J. S., & Sameroff, A. (2001). Parents' roles in shaping early adolescents' occupational aspirations. *Child Development*, 72(4), 1247-1266. doi: 10.1111/1467-8624.00345
- Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology*, 42(1), 11-26. doi: 10.1037/0012-1649.42.1.11
- Kimura, D. (2000). *Sex and cognition*. Cambridge, MA: MIT Press.
- Kimura, D. (2002). Sex hormones influence human cognitive pattern. *Neuroendocrinology letters*, 23, 67-77.
- Kost-Smith, L. E., Pollock, S. J., Finkelstein, N. D., Cohen, G. L., Ito, T. A., Miyake, A., . . . Singh, C. (2012). *Replicating a self-affirmation intervention to address gender differences: Successes and challenges*. Paper presented at the AIP Conference Proceedings-American Institute of Physics.
- Lauer, S., Momsen, J., Offerdahl, E., Kryjevskaja, M., Christensen, W., & Montplaisir, L. (2013). Stereotyped: investigating gender in introductory science courses. *CBE-Life Sciences Education*, 12(1), 30-38. doi: 10.1187/cbe.12-08-0133
- Leibham, M. B., Alexander, J. M., & Johnson, K. E. (2013). Science interests in preschool boys and girls: Relations to later self-concept and science achievement. *Science Education*, 97(4), 574-593. doi: 10.1002/sc.21066
- Liben, L. S., & Coyle, E. F. (2014). Developmental interventions to address the STEM gender gap: Exploring intended and unintended consequences. In L. S. Liben & R. S. Bigler (Eds.), *The Role of Gender in Educational Contexts and Outcomes* (Vol. 47, pp. 77-115). San Diego, CA: Academic Press.
- Linn, R. L. (2002). The measurement of student achievement in international studies. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 27-57). Washington, DC: National Academy Press.
- Luzzo, D. A., Hasper, P., Albert, K. A., Bibby, M. A., & Martinelli Jr, E. A. (1999). Effects of self-efficacy-enhancing interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students. *Journal of counseling psychology*, 46(2), 233-243. doi: 10.1037//0022-0167.46.2.233
- Lynch, J. (2002). Parents' self-efficacy beliefs, parents' gender, children's reader self-perceptions, reading achievement and gender. *Journal of Research in Reading*, 25(1), 54-67. doi: 10.1111/1467-9817.00158
- Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin*, 109(2), 267-296. doi: 10.1037/0033-2909.109.2.267
- 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(5906), 1331-1332. doi: 10.1126/science.1162573
- Marginson, S., Tytler, R., Freeman, B., & Roberts, K. (2013). *STEM Country comparisons: International comparisons of science, technology, engineering and mathematics (STEM) education*. Melbourne, Australia: Australian Council of Learned Academies.
- Martin, C. L., & Ruble, D. N. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, 13(2), 67-70. doi: 10.1111/j.0963-7214.2004.00276.x
- Martin, M. O., & Mullis, I. V. S. (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330(6008), 1234-1237. doi: 10.1126/science.1195996
- Mullis, I. V., Martin, M. O., & Gonzalez, E. J. (2000). *TIMSS 1999: International Mathematics Report: Findings from IEA's repeat of the*

- Third International Mathematics and Science Study at the eighth grade*: International Study Center.
- National Science Foundation. (2011). *Women, minorities, and persons with disabilities in science and engineering: 2011*. Arlington, VA: National Science Foundation, Retrieved from <http://www.nsf.gov/statistics/wmpd/pdf/wmpd2011.pdf>.
- National Science Foundation. (2017). *Women, minorities, and persons with disabilities in science and engineering: 2017*. Arlington, VA: National Science Foundation Retrieved from <https://www.nsf.gov/statistics/2017/nsf17310/>.
- Neuschmidt, O., Barth, J., & Hastedt, D. (2008). Trends in gender differences in mathematics and science (TIMSS 1995–2003). *Studies in Educational Evaluation*, 34(2), 56-72. doi: 10.1016/j.stueduc.2008.04.002
- Newcombe, N. S., Ambady, N., Eccles, J. S., Gomez, L., Klahr, D., Linn, M., . . . Mix, K. (2009). Psychology's role in mathematics and science education. *American Psychologist*, 64(6), 538-550. doi: 10.1037/a0014813
- Newcombe, N. S., & Frick, A. (2010). Early education for spatial intelligence: Why, what, and how. *Mind, Brain, and Education*, 4(3), 102-111. doi: 10.1111/j.1751-228X.2010.01089.x
- Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314-1334. doi: 10.1037/a0012702
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math= male, me= female, therefore math≠ me. *Journal of Personality and Social Psychology*, 83(1), 44-59. doi: 10.1037//0022-3514.83.1.44
- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., . . . Gonsalkorale, K. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106(26), 10593–10597. doi: 10.1073/pnas.0809921106
- OECD. (2011). *Education at a Glance 2011: OECD Indicators*. Paris: Organisation for Economic Co-Operation and Development. Retrieved from [http://www.oecd-ilibrary.org/education/education-at-a-glance-2011\\_eag-2011-en](http://www.oecd-ilibrary.org/education/education-at-a-glance-2011_eag-2011-en).
- OECD. (2014). *Are boys and girls equally prepared for life?* Paris: Organisation for Economic Co-Operation and Development. Retrieved from <http://www.oecd.org/pisa/pisaproducts/PIF-2014-gender-international-version.pdf>.
- OECD. (2016). *PISA 2015 Results - Excellence and Equity in Education*. OECD Publishing.
- Priess, H. A., & Hyde, J. S. (2010). Gender and academic abilities and preferences. In J. C. Chrisler & D. R. McCreary (Eds.), *Handbook of Gender Research in Psychology* (pp. 297-316). New York: Springer.
- Reilly, D. (2012). Gender, culture and sex-typed cognitive abilities. *PLoS ONE*, 7(7), e39904. doi: 10.1371/journal.pone.0039904
- Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology*, 107(3), 645-662. doi: 10.1037/edu0000012
- Riegle-Crumb, C., Farkas, G., & Muller, C. (2006). The role of gender and friendship in advanced course taking. *Sociology of Education*, 79(3), 206-228. doi: 10.1177/003804070607900302
- Riegle-Crumb, C., Moore, C., & Ramos-Wada, A. (2011). Who wants to have a career in science or math? Exploring adolescents' future aspirations by gender and race/ethnicity. *Science Education*, 95(3), 458-476. doi: 10.1002/sce.20431
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52(1), 59-82. doi: 10.1146/annurev.psych.52.1.59
- Rozek, C. S., Hyde, J. S., Svoboda, R. C., Hulleman, C. S., & Harackiewicz, J. M. (2014). Gender differences in the effects of a utility-value intervention to help parents motivate adolescents in mathematics and science. *Journal of Educational Psychology*, 107(1), 195-206. doi: 10.1037/a0036981
- Sander, E., Endepohls-Ulpe, M., & Quaiser-Pohl, C. (2016). Adult education in science, technology, engineering and mathematics under the gender aspect - A critical overview of programs and strategies in Germany. In Maksimović, M., J. Ostrouch-Kamińska, K. Popović & A. Bulajić (Eds.), *Contemporary Issues and Perspectives on Gender Research in Adult Education* (pp. 211-223). Belgrade: Institute for Pedagogy and Andragogy.
- Shapiro, J., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM

- fields. *Sex Roles*, 66(3-4), 175-183. doi: 10.1007/s11199-011-0051-0
- Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2005). Parents' socializing behavior and children's participation in math, science, and computer out-of-school activities. *Applied Developmental Science*, 9(1), 14-30. doi: 10.1207/s1532480xads0901\_3
- Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, 42(1), 70-83. doi: 10.1037/0012-1649.42.1.70
- Smeding, A. (2012). Women in Science, Technology, Engineering, and Mathematics (STEM): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex Roles*, 67(11-12), 617-629. doi: 10.1007/s11199-012-0209-4
- Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: A critical review. *American Psychologist*, 60(9), 950-958. doi: 10.1037/0003-066X.60.9.950
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4-28. doi: 10.1006/jesp.1998.1373
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613-629. doi: 10.1037/0003-066X.52.6.613
- Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological Bulletin*, 135(6), 859-884. doi: 10.1037/a0017364
- Sugimoto, C., Larivière, V., Ni, C., Gingras, Y., & Cronin, B. (2013). Global gender disparities in science. *Nature*, 504(7479), 211-213. doi: 10.1038/504211a
- Thompson, S. G., & Higgins, J. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in medicine*, 21(11), 1559-1573. doi: 10.1002/sim.1187
- UNESCO. (2011). *Fact Sheet: Women in Science (2011)*. UNESCO Institute for Statistics. Retrieved from <http://www.uis.unesco.org/ScienceTechnology/Documents/fs14-women-science-2011-en.pdf>.
- Unger, R. K. (1979). Towards a redefinition of sex and gender. *American Psychologist*, 34(11), 1085-1094. doi: 10.1037/0003-066X.34.11.1085
- Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking links to achievement in Science, Technology, Engineering, and Mathematics? *Current Directions in Psychological Science*, 22(5), 367-373. doi: 10.1177/0963721413484756
- Walters, J. (2010). Recasting Title IX: Addressing gender equity in the science, technology, engineering, and mathematics professoriate. *Review of Policy Research*, 27(3), 317-332. doi: 10.1111/j.1541-1338.2010.00444.x
- Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching*, 32(4), 387-398. doi: 10.1002/tea.3660320407
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604. doi: 10.1037/0003-066X.54.8.594
- Wood, W., & Eagly, A. H. (2002). A cross-cultural analysis of the behavior of women and men: Implications for the origins of sex differences. *Psychological Bulletin*, 128(5), 699-727. doi: 10.1037//0033-2909.128.5.699