

Debunking Myths about Gender and Mathematics Performance

Jonathan M. Kane and Janet E. Mertz

Gender differences in mathematics participation rate, mean and high-end performance, and variance in distribution of performance have been reported on numerous occasions. The reasons for these findings have been the subject of much debate. For example, the *greater male variability hypothesis*, originally proposed by Ellis in 1894 [42] and reiterated in 2005 by Lawrence Summers when he was president of Harvard University [48], states that variability in intellectual abilities is intrinsically greater among males. If true, it could account for the fact that all Fields medalists have been male. If gender differences in means and variances are primarily a consequence of innate, biologically determined differences between the sexes, one would expect these differences to be similar among countries regardless of their culture and to remain fairly constant across time. Such a finding would suggest that little can be done to diminish these differences. In support of this hypothesis, Machin and Pekkarinen [26] claimed that greater male variance in mathematics performance was a “robust phenomenon”, that is, observed among fifteen-year-olds in thirty-five out of the forty countries that participated in the 2003 Programme for International Student Assessment (PISA). In addition, women’s nature might include a tendency to prefer the more nurturing fields, such as nursing and teaching young children, to the more quantitative ones, such as mathematics, physics, and engineering. If so, it might not make

sense to encourage and direct any but the unusual female toward studying and seeking employment in these latter fields. This viewpoint has led some folks to propose that it may be a waste of time and money to expend resources directed toward trying to increase participation of women in these mathematics-intensive fields (e.g., [5], [6], [46], [49], [50]).

Alternatively, boys and girls may be born similar in their innate intellectual potential but end up displaying differences due to a variety of sociocultural factors present in their environment, for example, gender-stratification ([2]). If true, one might see differences among countries and changes over time in mathematics variances and mean performances. This *gender-stratified hypothesis* is consistent with several recent findings. For example, Hyde and collaborators ([20], [25]) reported that girls have now reached parity with boys in mean mathematics performance in the United States, even in high school, where a significant gap in mean performance existed in the 1970s. Likewise, both Brody and Mills ([3]) and Wai et al. ([51]) noted a drop in nonrandom samples of students under thirteen years of age, from 13:1 in the 1970s down to approximately 3:1 by the 1990s in the ratio of U.S. boys to girls scoring above 700 on the quantitative section of the college-entrance SAT examination. The percentage of Ph.D.’s in the mathematical sciences awarded to U.S. citizens who are women has increased from 6 percent in the 1960s to 30 percent in the past decade ([4], [9]). Sociocultural, legal, and educational changes that took place during this time span may account for these dramatic improvements in mathematics performance and participation by U.S. females.

Gender differences in opportunities and outcomes within countries have been quantified by a variety of measures. The Gender Gap Index (GGI) is a composite, weighted measure of the gap

Jonathan M. Kane is professor of mathematics and computer science at the University of Wisconsin-Whitewater. His email address is kanej@uww.edu.

*Janet E. Mertz is professor of oncology at the McArdle Laboratory for Cancer Research, University of Wisconsin School of Medicine and Public Health, Madison. Her email address is mertz@oncology.wisc.edu.
DOI: <http://dx.doi.org/10.1090/noti790>*

between men and women with respect to economic participation, educational attainment, political empowerment, and health. Looking across countries, Guiso et al. ([16]) reported that the GGI negatively correlates with both the gap in mean mathematics performance between boys and girls and the ratio of boys to girls scoring above the 95th percentile on the 2003 PISA; the gap and 95th percentile ratio were essentially zero and unity, respectively, in some countries with high GGI indexes. Even considering the extreme right side of the distribution, one notes that the percentage of worldwide participants who are girls in the International Mathematical Olympiad (IMO), an extremely difficult, proof-based, essay-style examination in mathematical problem solving, has increased from approximately 2½ percent in the 1970s to 10 percent in recent years ([22]). In addition, we ([1], [21]) reported that a positive correlation also exists between GGI and identification of girls with profound mathematical talent as measured by percentage of girls who participated in the IMO as members of high-ranked teams. These findings are consistent with the idea that the gap between boys' and girls' mathematics performance is due to differences in opportunities available to males versus females, which we will call here the *gap due to inequity hypothesis*. If either the gap due to inequity hypothesis or the gender-stratified hypothesis is true, it would suggest that a variety of actions could be taken to increase both average and high-end mathematics performance of females and participation rates by women in science, technology, engineering, and mathematics (STEM) fields.

Of course, outcome could be the end result of complex interplay between nature and nurture (for examples, see [43]). In this case, it would still make sense to devote resources toward increasing participation and performance of females in the mathematical sciences given that they make up half of the population and that U.S. STEM workers will likely be in short supply in the future.

However, Fryer and Levitt ([14]) and Ellison and Swanson ([10]) recently suggested a *Muslim culture hypothesis* given their finding that these above-mentioned correlations disappear when the samples include predominately Muslim countries with very low GGIs where most children attend single-gender schools. The Fryer-Levitt findings were based on data from eighth graders who participated in the 2003 Trends in International Mathematics and Science Study (TIMSS). They found that "in countries like Bahrain, which are among the worst in terms of gender equality, girls are actually outperforming boys on math". Alternatively, they propose a *single-gender classroom hypothesis* that "mixed-gender classrooms are a necessary component for gender inequality to translate into poor female math performance, although it is

difficult to distinguish single-gender classrooms from Islamic religion."

In this article, we tested each of these above-stated hypotheses by analyzing data from the 2007 TIMSS, 2009 PISA, and 2001–2010 IMOs, which included more countries from a variety of cultures, educational systems, and degrees of wealth than did these earlier studies. In support of the gender-stratified hypothesis, we show here that greater male variability and gender gap in mathematics performance, when present, are both largely artifacts of a complex variety of sociocultural factors rather than intrinsic differences, co-educational schooling, or specific religious following per se. Importantly, we document that mathematics performance for both boys and girls exhibits a strong positive correlation with some measures of gender equity, especially participation rates and salaries of women in the paid labor force relative to men.

Methods

Measures of Mathematics Performance

Most measures of mathematics performance presented here are based on the TIMSS, a quadrennial study that includes a mathematics assessment administered to samplings of countries' students. Approximately 138,000 fourth graders from twenty-six countries and 256,000 eighth graders from forty-eight countries participated in the TIMSS in 2003, with twenty-four of the latter countries also participating in the 2003 PISA. Approximately 183,000 fourth graders from thirty-eight countries and 242,000 eighth graders from fifty-two countries participated in the TIMSS in 2007. Summaries of the data are presented in the TIMSS 2003 and 2007 International Mathematics Reports ([32], [33]); other details are available in [12], [28], [29], and [32]–[34]. The benchmark participants from Basque Country, Dubai, and Ontario were used as representative measures for Spain, United Arab Emirates, and Canada, respectively.

The TIMSS sets the overall mean score among the benchmark participants at 500 with a standard deviation of 100 so scores can be compared across studies; scores of all other participants are presented relative to these benchmarks. We examined both the overall mean on the mathematics test for each country and the means and variances for each gender within each country. Gender gap was calculated as the difference in means, with positive values indicating that boys outscored girls. Normalized gender gap, called the effect size (d), was calculated as this difference divided by the standard deviation of the combined boys' and girls' scores within the country. The variance ratio (VR) for each country was calculated as the variance in their boys' scores divided by the variance in their girls' scores. To obtain the distribution of scores within a country and to see how individual scores relate to gender and other demographic

attributes, we examined the raw data available through the TIMSS 2003 and 2007 websites, using the first of the five plausible values for individual students' scores. Other available data included each student's school, sex, age, and answers to survey questions, such as his or her attitude toward learning mathematics.

The Program for International Student Assessment (PISA) is a triennial study of fifteen-year-old schoolchildren's scholastic performance that measures reading, mathematics, and science literacy. In 2003, approximately 275,000 students from forty-one countries, thirty of which were members of the Organisation for Economic Co-operation and Development (OECD), participated in a PISA that primarily focused on math literacy, testing problem solving and real-life situations that use math ([35]). In 2009, over 475,000 students from sixty-five countries (thirty-four from the OECD) participated in a PISA that focused on reading but included a mathematics section as well ([37]). We examined PISA data to compare it with prior findings of other researchers and our own findings with the TIMSS data.

Several differences exist between the PISA and TIMSS data sets:

(a) Countries. Prior to the 2009 PISA, countries participating in the TIMSS for eighth graders were more diverse with respect to wealth, religion, and gender-related schooling practices. Thus our findings presented here primarily come from analyses performed with the 2007 TIMSS eighth-grade data set. Afterward, we used the recently released 2009 PISA data to determine whether these findings were reproducible with a different examination in which only thirty-one out of eighty-six countries participated in both studies.

(b) Students' ages. The PISA samples students between the ages of fifteen years, three months and sixteen years, two months regardless of their grade in school, with most being in the equivalent of U.S. tenth grade. The TIMSS samples eighth graders, regardless of their age. Although most are 13.3 to 15.3 years of age, a few are younger, and the number of students over the age of 15.3 exceeded 10 percent in one-third of the countries. We also examined the 2003 and 2007 TIMSS fourth-grade data sets because in these, presumably, fewer students would have already dropped out of school, reducing potential sample bias.

(c) Religion. In eighteen of the countries that participated in the 2007 TIMSS eighth-grade examination, a majority of citizens come from Muslim backgrounds. To test the Muslim culture hypothesis, we sometimes separately analyzed data from countries with greater than 75 percent Muslims, indicating them by open circles or white bars in the figures.

(d) Schooling. To test the single-gender classroom hypothesis, we examined data from

countries in which 17 percent or more of students attended single-gender schools, separately comparing students attending gender-segregated and coeducational ones.

During the past decade, over eighty countries per year from throughout the world have sent six-member teams of precollegiate students to participate in the IMO ([22]). To examine mathematics performance at an extremely high level, we analyzed IMO gender data from countries with mean team member performance among the top sixty that had at least thirty students participate during 2001 through 2010.

Measures of Gender Equity

This study used two measures of the degree to which women within a country have yet to reach full equality with men living in the same country. The first, the World Economic Forum's GGI, is a composite, weighted measure of the gap between men and women with respect to: (i) economic participation and opportunity (EPO); (ii) educational attainment (ED); (iii) political empowerment (POL); and (iv) health and survival (H&S). It is measured on a 0-1 scale, with 1.00 being complete gender equity. The GGI data used here for year 2007 ([17]) ranged from a low of 0.4510 for Yemen to a high of 0.8146 for Sweden, with the United States scoring 0.7002 for a rank of 31st highest out of the 128 countries for which data were available. For correlations with the 2009 PISA data, we used 2009 GGI rankings ([18]).

Because the H&S subcomponent of the GGI is to a considerable degree a reflection of the wealth of a country and its citizens, we also analyzed data available through the Social Watch Group. Their Gender Equity Index (GEI) is a composite, weighted measure with respect to only: (i) economic participation rate and income earned (EPI); (ii) educational literacy rate and school enrollment (ED); and (iii) empowerment as reflected by percentage of women in technical, management, and government positions (TMG). It typically yielded slightly higher correlations with mathematics performance. It is measured on a 0-100 scale, with 100 being complete gender equity. The GEI data used here for year 2007 ([44]) ranged from a low of 31 for Yemen to a high of 89 for Sweden, with the United States scoring 74 for a rank of 24th highest out of the 154 countries for which data were available. For correlations with the 2009 PISA data, we used 2009 GEI rankings ([45]).

Other Measures

(i) Economic wealth. The gross domestic product (GDP) per capita used here was the 2007 GDP per capita in real terms deflated with Laspeyres Price Index. It was taken from the Penn World Table ([19]).

(ii) Religious affiliation. Each country's predominant religion was obtained from the CIA's *The World Factbook* ([7]); the percentage of citizens

affiliated as Muslim was taken from the Pew Research Center's 2009 report ([31]).

Statistical Analyses

Comparison of measures of mathematics performance and gender equity were performed using Pearson correlations (r) and regressions. Comparisons of attitudes toward mathematics education among groupings of countries were done by constructing contingency tables and performing chi-square tests. Statistics were considered significant if associated with a p -value of at most 0.10; all such correlations are displayed with † for a p -value < 0.10, * for < 0.05, ** for < 0.01, and *** for < 0.001.

Results and Discussion

Gender Gap in Math Performance versus Equity Indexes

Except for the 2007 eighth graders, among whom girls outperformed boys by five points ($p < 0.05$), no statistically significant gender gap existed overall in the mean scores of fourth and eighth graders on the 2003 and 2007 TIMSS ([32], [33]). To test the gap due to inequity hypothesis, we compared various measures of countries' gender equity (i.e., their equity indexes and subcomponents) with their gender gaps in mean mathematics performance on these examinations by calculating effect sizes, d (see Table 1). As already noted by others ([11], [14]), an insignificant Pearson correlation of -0.027 was observed between countries' GGIs and effect sizes calculated using the 2003 TIMSS eighth-grade data set. This correlation was 0.295 ($p < 0.05$) using the larger 2007 TIMSS eighth-grade data set; that is, the gap tended to increase in countries with greater gender equity as measured by their GGIs, a finding opposite of the one previously reported using the 2003 PISA data set ([16]). The positive correlation was even greater and

highly significant ($r = 0.577$; $p < 0.001$) using the 2007 TIMSS fourth-grade data set. Replacement of the 2007 GGI ratings with the 2007 GEI ones led to fairly similar correlations (Table 1). Again, they were more positive in fourth grade than in eighth grade and in 2007 than in 2003. No correlation was found between countries' effect sizes in mean mathematics performance on the 2009 PISA and their 2009 GGIs ($r = 0.083$) or GEIs ($r = 0.136$). This irreproducibility in the relationship between gender gap and equity indexes negates the gap due to inequity hypothesis.

Gender Gap versus Variance Ratio in Math Performance

Next, we tested the greater male variance hypothesis. If true, the variance ratios (VRs) for all countries should be greater than unity and similar in value. This is not what we observed. The VR measured for any given nation was quite reproducible, that is, it rarely differed by more than 20 percent from one test administration year to the next, among students in different grades, or between the PISA and TIMSS; typically, it differed by at most 10 percent (see Table 2). For example, the VRs for Australia, England, Hungary, and the United States ranged from 1.10 to 1.21, 1.05 to 1.12, 1.03 to 1.10, and 1.08 to 1.19, respectively, among the five tests analyzed here. These findings agree well with the VR of 1.08 reported from a large meta-analysis involving data from 242 studies involving over 1 million Americans ([25]). However, for Indonesia, Morocco, and Tunisia, the VRs ranged from 0.95 to 1.02, 0.96 to 1.04, and 0.93 to 1.09, respectively; that is, they were essentially unity. For Singapore and Taiwan, they ranged from 1.21 to 1.25 and 1.25 to 1.31, respectively. In fact, the VRs calculated using the 2007 TIMSS eighth-grade data set studied in detail here varied widely among countries, ranging all the way from

Table 1. Correlations between measures of gender equity and mean mathematics scores.

	2003 TIMSS						2007 TIMSS						2009 PISA		
	4th Graders			8th Graders			4th Graders			8th Graders			15-year-olds		
	Gap ¹	Girls	Boys	Gap	Girls	Boys	Gap	Girls	Boys	Gap	Girls	Boys	Gap	Girls	Boys
GGI ³	.144	.302	.304	-.027	.281† ²	.283†	.577***	.624***	.660***	.295*	.408**	.462***	.083	.400***	.405***
EPO	.174	.547**	.546**	-.044	.327*	.330*	.511***	.757***	.780***	.258†	.442**	.488***	-.025	.398***	.387**
POL	.233	.019	.039	.073	.007	.020	.557***	.315†	.362*	.262†	.262†	.314*	.123	.340**	.352**
ED	.126	.420*	.417*	-.292†	.246	.207	.317†	.550***	.562***	-.168	.248†	.200	.043	.003	.010
H & S	.385†	.053	.089	.247	.098	.136	.433**	.136	.181	.279†	.177	.231	.146	-.129	-.102
No.Countries ⁴ :	24			43			36			46			61		
GEI ⁵	.210	.525**	.529**	-.075	.425**	.424**	.637***	.698***	.738***	.176	.479***	.510***	.136	.333**	.345**
EPI	.259	.572**	.581**	-.029	.420**	.427**	.564***	.691***	.723***	.261†	.453**	.506***	.015	.412***	.405***
TMG	.121	.423*	.420*	-.039	.341*	.342*	.681***	.603***	.651***	.160	.422**	.450**	.196	.320*	.342**
ED	.187	.381†	.386†	-.235	.268†	.237	.386*	.543***	.566***	-.160	.263†	.219	.058	-.366**	-.350**
No.Countries:	24			44			35			46			61		

¹Gender gap in mean score, calculated as effect size, d ; positive value means boys scored higher than girls.

² p -values: †, <0.1; *, <0.05; **, <0.01; ***, <0.001.

³GGI, World Economic Forum's Gender Gap Index; EPO, economic participation and opportunity subcomponent; POL, political empowerment subcomponent; ED, educational attainment subcomponent; and H & S, health and survival subcomponent.

⁴Number of participating countries for which equity data exist.

⁵GEI, Social Watch's Gender Equity Index; EPI, economic participation rate and income earned subcomponent; TMG, empowerment as reflected by percentage of women in technical, management, and government positions; and ED, educational literacy rate and school enrollment.

Table 2. Variance ratios in mathematics performance among countries in different examinations, grades, and years¹.

Country	2003			2007	
	PISA ² 15-year-olds	TIMSS 4th Graders	TIMSS 8th Graders	TIMSS 4th Graders	TIMSS 8th Graders
Armenia	-	1.12	1.13	0.98	1.00
Australia	1.18	1.10	1.16	1.21	1.20
Belgium	1.19	1.11	1.06	-	-
Canada	1.24	1.09	1.06	1.12	1.19
Cyprus	-	1.15	1.16	-	1.25
Czech Rep	1.07	-	-	1.18	1.03
England	1.06	1.12	1.05	1.12	1.07
Hong Kong	1.36	1.21	1.12	1.20	1.29
Hungary	1.10	1.03	1.08	1.07	1.07
Indonesia	0.95	-	1.02	-	0.98
Iran	-	1.10	1.15	1.21	1.15
Italy	1.27	1.08	1.14	1.05	1.08
Japan	1.29	1.28	1.19	1.17	1.02
Korea	1.10	-	1.10	-	1.14
Latvia	1.19	1.18	1.18	1.15	-
Lithuania	-	1.08	1.13	1.17	1.08
Morocco	-	0.96	1.00	1.04	1.00
Netherlands	1.00	0.96	1.03	1.00	-
New Zealand	1.16	1.10	1.20	1.20	-
Norway	1.20	1.11	1.09	1.00	1.17
Russian Fed	1.20	1.00	1.14	1.10	1.16
Scotland	-	1.17	1.14	1.22	1.08
Serbia	1.30	-	1.14	-	1.12
Singapore	-	1.21	1.25	1.21	1.22
Slovak Rep	1.10	-	1.16	1.00	-
Slovenia	-	1.17	1.18	1.18	1.15
Spain	1.20	-	1.28	-	1.23
Sweden	1.10	-	1.06	1.13	1.12
Taiwan	-	1.25	1.25	1.26	1.31
Tunisia	1.03	1.06	0.93	1.09	0.91
United States	1.19	1.11	1.11	1.08	1.08

¹Calculated as boys' variance divided by girls' variance within country. Data shown only for countries that participated in 3 or more of these 5 examinations.

²Taken from Machin and Pekkarinen (26).

0.91 to 1.52 (Figure 1A). This distribution of VRs is similar to the one previously reported using the 2003 PISA data set ([26]) except for being even broader with the inclusion of more predominantly Muslim countries (white bars in Figure 1A), many of which exhibit usually low or high VRs.

Variances for girls and boys also varied widely throughout a threefold range (Figure 1B). Countries with small variances typically had VRs within 0.2 of unity. Most of the countries with large VRs were ones that also had unusually large boys' variances. Thus, for the 2007 TIMSS eighth-grade data set, the correlation between variance and VR was 0.297 ($p < 0.05$); the correlation between boys' variance and VR was 0.414 ($p < 0.01$). Therefore, we conclude that both variance and VR in mathematics performance vary greatly among countries. Confirming our earlier finding ([21]), we also conclude that VR is reproducibly essentially unity for some countries. These findings are inconsistent with the greater male variability hypothesis.

Interestingly, a strong negative correlation ($r = -0.640$, $p < 0.001$) was observed between VR and

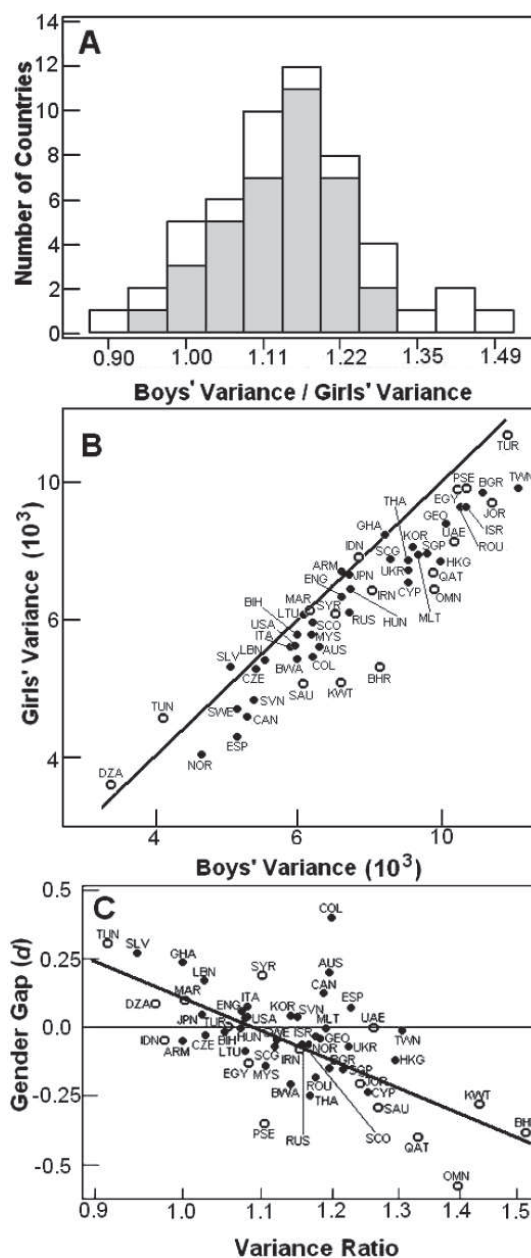


Figure 1. Variance ratio, variances, and gender gap on 2007 TIMSS for eighth graders vary widely among countries and strongly correlate.

(A) Histogram showing distribution of measured variance ratios among predominantly Muslim (i.e., >75%; white bars) and other participating countries (gray bars). X axis is standard variance ratios presented in evenly spaced intervals in log₁₀. (B) Scatter plot showing girls' versus boys' variance in score distributions plotted using log scales. The diagonal line indicates equity. (C) Scatter plot showing relationship between normalized gender gap, i.e., effect size d , and variance ratio plotted on a log scale. Pearson correlation, $r = -0.640$ ($p < 0.001$); $r = -0.790$ ($p < 0.001$) for countries >75% Muslim countries, indicated by open circles. Regression line is indicated. The 3-letter ISO codes adjacent to each circle indicate the countries they represent.

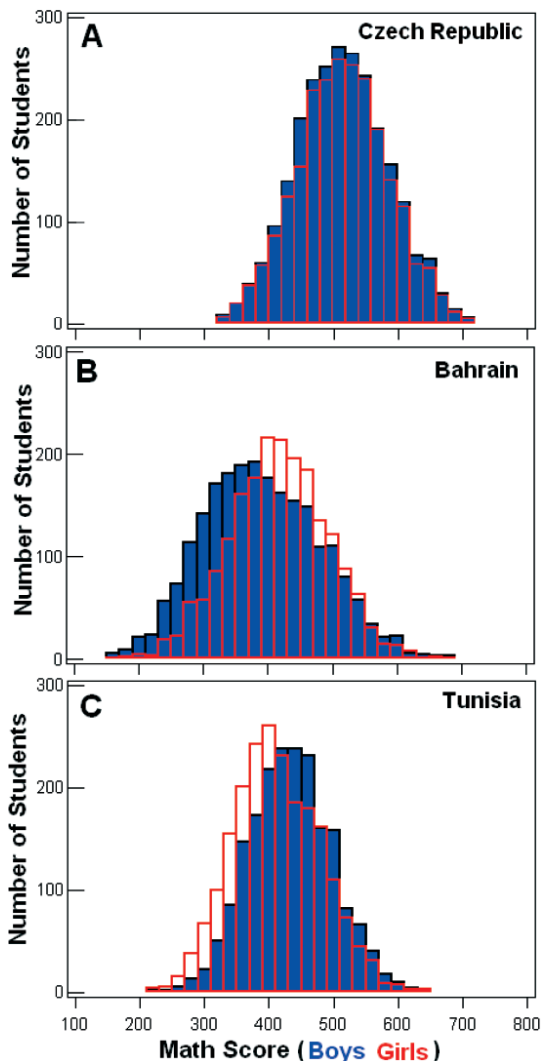


Figure 2. Distributions of eighth-grade girls' and boys' mathematics scores on 2007 TIMSS in (A) the Czech Republic, (B) Bahrain, and (C) Tunisia. The Czech Republic had 2,335 girls and 2,510 boys participate, with $VR = 1.02$ and $d = -0.027$. Bahrain had 1,974 girls and 2,256 boys participate, with $VR = 1.52$ and $d = -0.38$. Tunisia had 2,121 girls and 1,915 boys participate, with $VR = 0.91$ and $d = 0.31$.

effect size (Figure 1C). This negative correlation was even stronger ($r = -0.790$; $p < 0.001$) among the predominantly (that is, > 75 percent) Muslim countries (open circles in Figure 1C), with effect size ranging from quite negative, i.e., favoring girls, to somewhat positive. Furthermore, the percentage of Muslims in a country's population did not correlate with effect size [$r = -0.029$ for 2003 TIMSS eighth graders; $r = -0.285$ ($p < 0.05$) for 2007 TIMSS eighth graders, but $r = -0.107$ when the four countries with outlier VRs > 1.30 are omitted]. These findings negate the Muslim culture hypothesis. Rather, we conclude that gender gap

in mathematics performance exhibits a strong negative correlation with variance ratio.

Why might effect size correlate with VR? Comparison of actual boys' and girls' score distributions for individual countries was informative. For example, they were essentially coincident in some countries, such as the Czech Republic, where VR and gender gap were near unity and zero, respectively (Figure 2A). On the other hand, in Bahrain, the country with the largest VR, the right sides of the distributions were similar, whereas the left sides contained many more boys than girls (Figure 2B). This country also had 282 fewer eighth-grade girls than boys participate in this test. One hypothesis consistent with these data is that Bahraini girls who potentially might have scored poorly were more likely to have left school prior to eighth grade than were potentially poorly scoring boys due to sociocultural factors; thus they would not have been among this country's eighth-grade test-takers. If true, it would explain why Bahraini girls' variance was much smaller than the boys' variance (Figure 1B). It would also account for the fact that Bahraini girls significantly outscored boys on average when, in reality, they might not have done so if the population sampled had included school dropouts. Another potential contributing factor is that some Bahraini boys receive their education at religious schools [24], which may devote less time to mathematics education than do secular schools. Among Bahraini eighth graders attending mixed-gender schools, the VR and gender gap were near unity and zero, respectively (Table 3).

Gender differences in schooling practices may also explain variance ratios that are less than unity and gender gaps that favor boys. For example, boys were preferentially missing from the left half of the score distributions for eighth graders in Tunisia (Figure 2C), the country with the smallest VR (Figure 1C) and somewhat fewer male than female participants. Thus we conclude that math variance ratios and gender gaps significantly different from unity and zero, respectively, are both largely consequences of the same sociocultural factors that differ among countries, some of which lead to different educational experiences and patterns of school attendance. This latter finding is reminiscent of the Hyde et al. [20] conclusion that the gender gap on the U.S. college entrance ACT examination disappears when sample bias is eliminated by testing all eleventh-grade students rather than just college-bound ones.

Single- versus Mixed-Gender Schooling

We next tested the single-gender classroom hypothesis by examining data at the level of individual eighth-grade students from the seventeen countries in which 17 percent or more of students attended single-gender schools (Table 3). In agreement with numerous previous studies including findings from 2003 TIMSS data [52], 2006 PISA

Table 3. Math performance on 2007 TIMSS among eighth graders attending single- versus mixed-gender schools.

Country	% Predom.		Single-gender schools				Mixed-gender schools				Difference school type	
	SG ¹	Religion	Girls Mean	Boys Mean	Gap ²	VR ³	Girls Mean	Boys Mean	Gap	VR	Girls ⁴	Boys ⁵
Australia	18	Christian	520 (5.2) ⁶	544 (3.6)	0.30	0.90	483 (1.8)	488 (1.8)	0.07	1.14	-0.47	-0.71
Bahrain	91	Muslim	412 (1.7)	380 (1.9)	-0.39	1.40	473 (6.0)	480 (5.1)	0.08	1.04	0.73	1.19
Dubai, UAE	54	Muslim	425 (2.8)	401 (3.6)	-0.24	1.44	487 (3.3)	484 (3.6)	-0.03	1.27	0.61	0.82
Egypt	66	Muslim	423 (2.0)	393 (2.2)	-0.29	1.17	406 (3.1)	423 (3.1)	0.17	1.11	-0.17	0.29
Hong Kong	17	None	619 (3.3)	612 (6.0)	-0.08	2.44	572 (2.3)	563 (2.5)	-0.10	1.23	-0.52	-0.54
Iran	97	Muslim	415 (2.0)	417 (2.0)	0.02	1.26	-	-	-	-	-	-
Israel	23	Jewish	492 (4.2)	486 (5.1)	-0.06	1.25	456 (2.7)	454 (3.0)	-0.02	1.14	-0.36	-0.33
Jordan	85	Muslim	435 (1.8)	408 (2.4)	-0.27	1.25	430 (6.8)	410 (4.3)	-0.21	1.10	-0.05	0.02
Korea	43	None	589 (3.0)	596 (3.0)	0.08	1.10	598 (2.6)	600 (2.7)	0.01	1.16	0.10	0.04
Kuwait	82	Muslim	367 (1.6)	338 (2.3)	-0.37	1.39	361 (5.0)	343 (3.8)	-0.23	1.27	-0.07	0.07
Malta	58	Catholic	485 (2.1)	451 (3.5)	-0.38	1.53	503 (3.2)	516 (1.9)	0.14	0.77	0.20	0.72
Oman	83	Muslim	399 (1.8)	346 (2.3)	-0.56	1.48	397 (8.0)	346 (3.8)	-0.54	0.82	-0.02	0.00
Palestine	81	Muslim	391 (2.1)	345 (2.7)	-0.45	1.25	372 (5.2)	359 (4.7)	-0.12	0.95	-0.19	0.13
Qatar	17	Muslim	335 (2.4)	269 (8.2)	-0.71	1.29	322 (1.7)	292 (1.7)	-0.33	1.29	-0.14	0.24
Saudi Arabia	98	Muslim	346 (1.5)	322 (1.8)	-0.31	1.32	-	-	-	-	-	-
Singapore	17	Buddhist	650 (3.7)	639 (5.0)	-0.12	1.40	585 (2.0)	569 (2.0)	-0.18	1.16	-0.71	-0.77
Syria	64	Muslim	383 (1.9)	405 (2.1)	0.27	1.18	397 (3.0)	401 (3.0)	0.05	1.00	0.18	-0.04

¹Percentage of participating students who were attending a single-gender school.

²Gender gap in mean score among students attending same school type, calculated as effect size, *d*, with normalization to total within-country standard deviation; positive means boys scored higher than girls.

³Calculated as boys' variance divided by girls' variance within country among students attending same school type.

⁴Mean score of girls attending mixed-gender schools minus mean score of girls attending single-gender schools, with normalization to total within-country standard deviation.

⁵Mean score of boys attending mixed-gender schools minus mean score of boys attending single-gender schools, with normalization to total within-country standard deviation.

⁶Standard errors of means indicated within parentheses.

data [8], and a 2005 U.S. Department of Education report that systematically reviewed dozens of articles with data on single-sex versus coeducational schools [27], we observed no consistent trends among either the predominantly Muslim or non-Muslim countries. Importantly, in South Korea, where middle school students are randomly assigned to single- or mixed-gender public schools [23] and the demographics of the students attending them are quite similar (data not shown), both girls and boys performed approximately one standard deviation above the benchmark mean of 500 regardless of school type. In predominantly Muslim Dubai, boys and girls performed similarly in the coeducational schools, much better than either of them did in the single-gender ones. Girls did significantly outperform boys in gender-segregated schools in some of the Muslim countries. However, it was not because their girls performed well; rather, their boys performed quite poorly. In fact, a strong positive correlation ($r = 0.687$; $p < 0.01$) was found between boys' mean score and gender gap among countries > 75 percent Muslim, that is, the gap favoring girls became smaller as boys' mean score improved; above 400, the gap was either not significantly different from zero or favored boys for all predominantly Muslim countries except Jordan.

Why did boys perform so poorly in some countries? Attendance at single-gender schools per se

is not the answer, given that boys attending such schools in Hong Kong, South Korea, and Singapore averaged a standard deviation or more above the benchmark mean, doing as well if not significantly better than their peers in coeducational schools. Conversely, boys attending mixed-gender schools in Ghana, Kuwait, and Oman had mean scores below 350; in Qatar, the mean was only 292.

Thus we also reject the single-gender classroom hypothesis, concluding, instead, that other factors are the main determinants of mean mathematics performance for both girls and boys; gender gap, when present in either direction, is a consequence of these other factors that differentially affect performance.

Math Performance versus Equity Indexes

What are these factors? One of them is poverty, a fact observable in the relationship between log real gross domestic product (GDP) per capita and eighth-grade 2007 TIMSS mean mathematics performance (Figure 3E). A strong positive correlation ($r = 0.622$; $p < 0.01$) exists among the twenty-one participating countries with real GDPs per capita below US\$11,500. However, this correlation was not significant ($r = -0.183$; $r = 0.076$ with outlier Qatar omitted) for the twenty-nine wealthier participating countries. Analysis of the data from the 2009 PISA confirmed these findings: (i) Among the sixteen poorer countries with known GDPs, a strong positive correlation ($r = 0.605$; $p < 0.05$) was

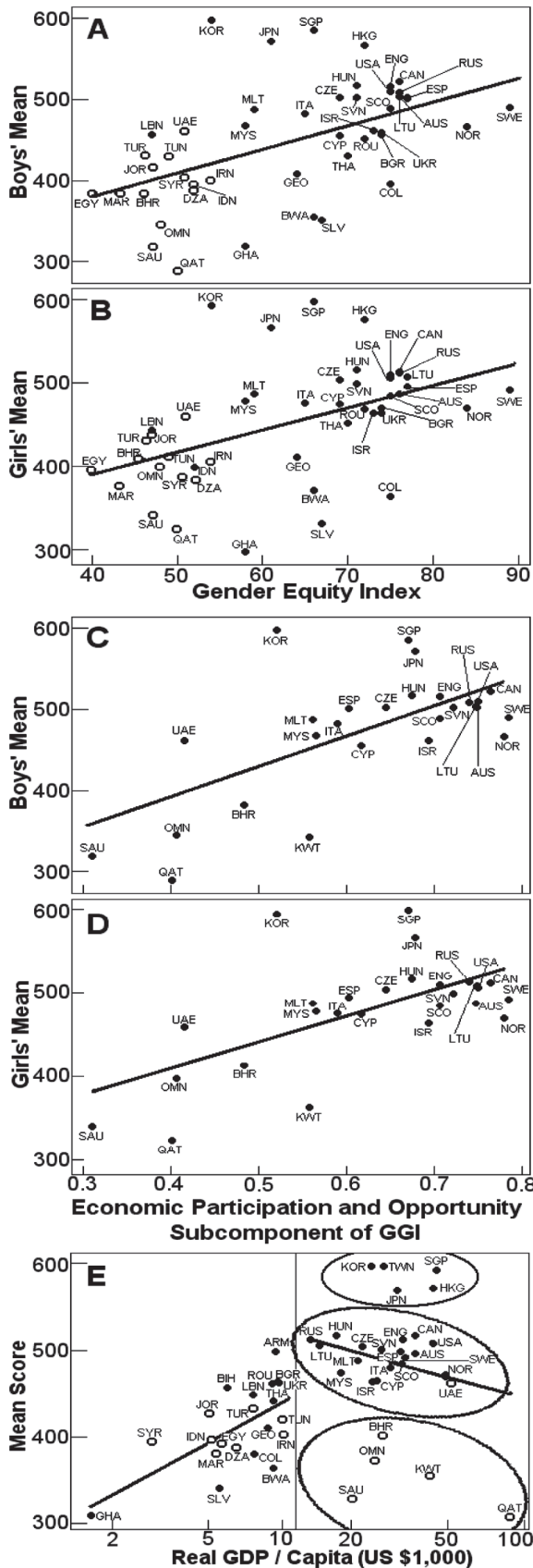


Figure 3 (left). Scatter plots showing relationships among countries in mathematics performance of eighth graders on 2007 TIMSS. (A, B) GEI versus boys' ($r = 0.510$, $p < 0.001$) and girls' ($r = 0.479$, $p < 0.001$) mean, respectively. (C, D) EPO subcomponent of GGI among countries with real GDP per capita above US\$11,500 versus boys' ($r = 0.658$, $p < 0.001$) and girls' ($r = 0.618$, $p < 0.001$) mean, respectively. (E) Nations' mean versus their real GDP per capita plotted on a log scale. For the 29 countries with GDP per capita above US\$11,500, $r = 0.183$; $r = 0.076$ with outlier Qatar omitted. For the 21 countries with GDP per capita below US\$11,500, $r = 0.622$ ($p < 0.01$); $r = 0.420$ ($p < 0.10$) with outlier Ghana omitted. Ovals indicate the three subgroups of wealthier countries described in the text and Table 6 as Middle Eastern, East Asian, and other plus Dubai.

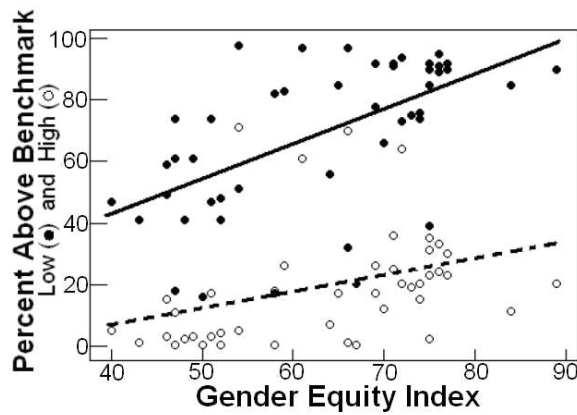


Figure 4. Scatter plots showing relationships between GEI and percentage of eighth graders scoring at or above the low (400, filled circles; $r = 0.570$, $p < 0.001$) and high (550, open circles; $r = 0.366$, $p < 0.05$; $r = 0.635$, $p < 0.001$ with the four East Asian outlier countries omitted) benchmarks on the 2007 TIMSS. Regression lines indicated.

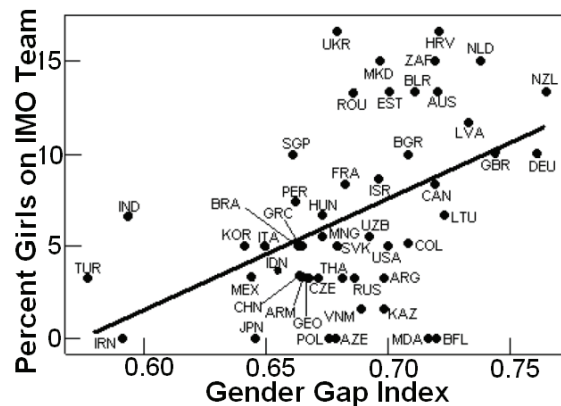


Figure 5. Scatter plot showing relationship between GGI and percentage of girls on IMO teams whose mean participant scores ranked among the top 50 countries during 2001-2010 ($r = 0.474$, $p < 0.001$). Regression line indicated.

Table 4. Correlations between measures of gender equity and mean mathematics scores for poorer and richer countries.¹

	8 th graders 2007 TIMSS						15-year-olds 2009 PISA		
	19 poorer countries			27 richer countries			47 richer countries		
	Gap	Girls	Boys	Gap	Girls	Boys	Gap	Girls	Boys
GGI	.017	.092	.116	.591***	.418*	.478*	.095	.206	.217
EPO	-.139	.087	.062	.643***	.618***	.658***	-.043	.344**	.329*
POL	.314	-.337	-.294	.423*	.229	.283	.137	.150	.168
GEI	-.105	.162	.152	.598***	.454*	.516**	.048	.190	.193
EPI	-.062	.023	.010	.652***	.574**	.630***	-.017	.406***	.393**
TMG	-.056	.104	.101	.501**	.392*	.442*	.074	.118	.126

¹Demarcation between poorer and richer countries was drawn at a Real GDP per capita of US\$11,500.

found between a country's log real GDP per capita and its mean mathematics score; (ii) Among the forty-seven richer countries, this correlation was, again, not significant ($r = 0.186$), having changed from negative to positive largely due to the absence of several Middle Eastern countries that had participated in the 2007 TIMSS. The wealthy countries included ones such as Qatar whose students performed among the very lowest in the world on both tests. Since the GEI is less influenced by wealth than the GGI, we used it here as our primary measure of gender equity.

Strikingly, a strong positive correlation ($r = 0.510$, $p < 0.001$) was found between GEI and eighth-grade boys' mean performance on the 2007 TIMSS (Figure 3A; Table 1). This correlation is slightly stronger than the one observed for eighth-grade girls ($r = 0.479$, $p < 0.001$). Similar correlations were obtained using data only from the wealthier countries (Table 3). Even stronger correlations were observed between GEI and fourth-grade boys' ($r = 0.738$, $p < 0.001$) and girls' ($r = 0.698$, $p < 0.001$) mean scores. Good correlations between these factors also exist using the 2003 TIMSS data sets (Table 1). Positive correlations with GEI were also observed with respect to percentage of boys and girls scoring at or above low and high benchmark scores (Table 5; Figure 4). Analysis of the 2009 PISA data confirmed these findings (Tables 1, 2, and 4). Noteworthy is the fact that 26 and 27 percent of the girls and boys, respectively, from Shanghai, China, scored above 669 on the 2009 PISA; the corresponding numbers for the U.S. girls and boys were 1.2 and 2.5 percent,

Table 5. Correlations between economic participation and income subcomponent of GEI and percentage of students scoring at or above benchmarks.

Exam:	8 th graders 2007 TIMSS ¹			15-year-olds 2009 PISA ²		
Score:	400	475	550	420	482	607
Girls	.519***	.536***	.386**	.446***	.419***	.266*
Boys	.571***	.544***	.408**	.439***	.410***	.274*

¹The international mean score was 475; standard deviation was set at 100.

²The average score was 496; it was determined by averaging the mean scores of the OECD countries, with each country weighted equally. The scores of 420, 482, and 607 correspond to levels 2, 3, and 5, respectively, out of six levels of proficiency.

respectively, below the 2.8 percent overall for OECD countries. Thus mathematics performance at the low, median, and high levels for both boys and girls strongly correlates with equity indexes; gender inequity may be one of the reasons boys do poorly in some wealthy countries.

This finding explains why equity indexes and gender gap in mean mathematics performance do not reproducibly correlate; that is, while girls' scores increase as equity indexes increase, boys' scores do likewise. The eighth-grade 2007 TIMSS data set included some wealthy countries with low equity indexes, such as Bahrain, Kuwait, Oman, Qatar, and Saudi Arabia, where boys significantly underperformed relative to girls (Figure 1C). Thus boys' mean mathematics scores on this examination increased with equity indexes even more than girls' scores, leading to the unintuitive finding of a positive correlation between gender gap favoring boys and GGI (Table 1). This correlation is highly positive ($r = 0.591$, $p < 0.001$) when the data from the poorer countries are omitted (Table 4). However, the correlation between equity indexes and gender gap in mathematics performance is insignificantly different from zero on the 2009 PISA, in which only one of these five Middle Eastern countries participated (Tables 1 and 4).

Noteworthy is the specific subcomponent of the gender equity indexes that largely accounts for the strong positive correlation with boys' and girls' mathematics performance at the low, intermediate, and high levels. While educational quality and equity may still be an important factor in poor countries (e.g., [15], [40]), it no longer is among most wealthy countries, where females have now largely reached if not exceeded parity with males in this area ([17], [18]). Instead, the primary contributor is the economic subcomponent of the GGI and GEI (EPO and EPI, respectively) that includes women's income and rates of participation in the work force relative to men's (Tables 1 and 4). The strong positive correlation between this economic subcomponent and mean mathematics performance is slightly stronger for boys than girls on the 2007 TIMSS and similar between the genders on the 2003 TIMSS and 2009 PISA. Thus, even though family income is, presumably, not limiting the ability of most children in the wealthier countries to obtain an education through at least eighth grade, quality employment opportunities for women nevertheless strongly parallel the mathematics performance of their students. This finding is consistent with the gender similarities hypothesis [20], with maternal education and employment opportunities likely having indirect effects on learning by their offspring regardless of gender (e.g., [13], [30], [38]).

Also noteworthy is the fact that culturally related wealthier countries cluster into largely nonoverlapping groups with respect to mean

mathematics scores: Middle Eastern predominantly Muslim (Bahrain, Kuwait, Oman, Qatar, Saudi Arabia), East Asian (Hong Kong, Japan, South Korea, Singapore, Taiwan), and remaining countries (ovals in Figure 3E). A multiple regression indicates that the mean scores among these twenty-nine wealthier countries can be determined to a very high degree ($R^2 = 94$ percent) by simply knowing the country's EPO, its real GDP per capita, and into which of these three clusters the country falls (Table 6). In particular, the mean scores show a significant ($p < 0.05$) strong positive dependence on EPO and negative dependence on GDP, with East Asian countries scoring higher and Middle Eastern ones, except for Dubai, scoring lower than the other countries. Among the many cultural factors on which these groups of countries differ are parents' and students' attitudes regarding the importance of mastering school mathematics and whether the ability to do so requires innate ability or effort (e.g., [36], [47], [50], [53]). Interestingly, a country's mean mathematics score tends to be higher when its students attribute a lower importance to mastering mathematics (i.e., chi-square with six degrees of freedom on students' answers to each of four questions varies from 6,100 to 9,700; $p < 0.0001$). Thus mathematics performance of students largely reflects the academic standards and expectations of the community in which they are raised. These findings are consistent with and extend prior findings ([2], [11], [16], [39], [41]) of correlations between educational and economic gender equity and mathematics performance.

GGI and IMO

Our previously reported correlation between GGI and percentage of IMO team members who were girls [21] was based upon analysis of data from teams with median ranks among the top thirty during 1989–2008. Although female IMO participants were quite rare in the past, they have composed approximately 10 percent of participants in recent years [22]. Also, many more countries now participate: 90 to 104 per year since 2005, up from 50 in 1989. Ellison and Swanson analyzed IMO data from 100 percent of the countries that participated at any time during 2007–2009 [10]. They did not observe a significant correlation. However, their analysis included some countries with low gender equity indexes whose teams were not of IMO caliber, rarely participated, and contained multiple female students, for example, the 2009 all-girl U.A.E. team that scored only three points. We redid this calculation using the 2001–2010 IMO data, omitting countries with either (i) < 50 percent participation rate during this time period or (ii) mean student scores below 8.6 out of 42 possible points. Again, we obtained a strong positive correlation between GGI and percentage of girls on teams, with $r = 0.407$ ($p < 0.01$), $r = 0.474$ ($p < 0.001$), and $r = 0.415$ ($p < 0.001$) for

Table 6. Multiple linear regression analysis of mean mathematics performance on 2007 TIMSS of eighth graders from the 27 wealthier countries.¹

Predictor	β^2	s^3	T^4	p
Constant	651.27	95.49	6.82	0.000
Log real GDP/capita	-22.47	9.18	-2.45	0.023
GGI-EPO	104.44	40.61	2.57	0.017
Indicator Middle Eastern ^{*1}	-109.51	13.97	-7.84	0.000
Indicator East Asian ^{*1}	103.39	11.98	8.63	0.000
$R^2 = 94.0\%$				

¹Countries correspond to the ones indicated in the ovals in Fig. 3E.

² β , least squares estimate for the coefficient of the predictor term in the regression model.

³ s , standard deviation of the estimate β .

⁴ T , associated Student-t statistic β/s .

teams with mean student scores of at least 12.1, 10.5, and 8.6 corresponding to a rank among the top forty, fifty, and sixty participating countries, respectively (Figure 5). This finding confirms our conclusion that GGI correlates with identification of females who excel in mathematics performance at an extremely high level.

Conclusions

In summary, we conclude that gender equity and other sociocultural factors, not national income, school type, or religion per se, are the primary determinants of mathematics performance at all levels for both boys and girls. Our findings are consistent with the gender stratified hypothesis, but not with the greater male variability, gap due to inequity, single-gender classroom, or Muslim culture hypotheses. At the individual level, this conclusion suggests that well-educated women who earn a good income are much better positioned than are poorly educated women who earn little or no money to ensure that the educational needs of their children of either gender with regard to learning mathematics are well met. It is fully consistent with socioeconomic status of the home environment being a primary determinant for success of children in school. At the national level, the United States ranked only thirty-first in mean mathematics performance out of the sixty-five countries that participated in the 2009 PISA. Eliminating gender discrimination in pay and employment opportunities could be part of a win-win formula for producing an adequate supply of future workers with high-level competence in mathematics. Wealthy countries that fail to provide gender equity in employment are at risk of producing too few citizens of either gender with the skills necessary to compete successfully in a knowledge-based economy driven by science and technology.

Acknowledgments

We thank Douglas Bates and his statistics class for suggesting some hypotheses to test; William

Mickelson for formatting the TIMSS data sets; and Glenn Ellison, Janet Hyde, Jeremy Kane, and Cathy Kessel for suggestions regarding the manuscript.

References

- [1] T. ANDREESCU, J. GALLIAN, J. M. KANE, and J. E. MERTZ, Cross-cultural analysis of students with exceptional talent in mathematical problem solving, *Notices of the AMS* **55** (2008), 1248–1260.
- [2] D. P. BAKER and D. P. JONES, Creating gender equality: Cross-national gender stratification and mathematical performance, *Sociol. Ed.* **66** (1993), 91–103.
- [3] L. BRODY and C. MILLS, Talent search research: What have we learned? *High Ability Studies* **16** (2005), 101.
- [4] J. BURRELLI, *Thirty-Three Years of Women in S&E Faculty Positions*, InfoBrief, Science Resources Statistics, NSF 08-308, National Science Foundation Directorate for Social, Behavioral, and Economic Sciences, 2008. Retrieved from <http://www.nsf.gov/statistics/infbrief/nsf08308/nsf08308.pdf>.
- [5] S. J. CECI and W. M. WILLIAMS, *The Mathematics of Sex: How Biology and Society Conspire to Limit Talented Women and Girls*, Oxford University Press, 2010.
- [6] _____, Understanding current causes of women's underrepresentation in science, *Proc. Nat. Acad. Sci. USA* **108** (2011), 3157–3162.
- [7] Central Intelligence Agency, *The World Factbook*. Retrieved from <https://www.cia.gov/library/publications/the-world-factbook/fields/2122.html>.
- [8] J. CRESSWELL, M. IKEDA, M. JAKUBOWSKI, A. SCHLEICHER, S. VAYSETTES, and P. ZOIDO, *Equally Prepared for Life? How 15-Year-Old Boys and Girls Perform in School*, OECD, Brussels, 2009. Retrieved from <http://www.oecd.org/dataoecd/59/50/42843625.pdf>.
- [9] R. J. DAVERMAN (AMS Secretary), Statistics on women mathematicians compiled by the AMS, *Notices of the AMS* **58** (2011), 1310.
- [10] G. ELLISON and A. SWANSON, The gender gap in secondary school mathematics at high achievement levels: Evidence from the American Mathematics Competitions, *J. Econ. Per.* **24** (2010), 109–128.
- [11] N. M. ELSE-QUEST, J. S. HYDE, and M. C. LINN, Cross-national patterns of gender differences in mathematics: A meta-analysis, *Psychol. Bull.* **136** (2010), 103–127.
- [12] P. FOY and J. F. OLSON (eds.), *TIMSS 2007 International Database and User Guide*, TIMSS and PIRLS International Study Center, Boston College, Chestnut Hill, MA, 2009. Retrieved from http://timss.bc.edu/timss2007/idb_ug.html.
- [13] P. M. FROME and J. S. ECCLES, Parents' influence on children's achievement-related perceptions, *J. Personality Social Psychol.* **74** (1998), 435–452.
- [14] R. G. FRYER and S. D. LEVITT, An empirical analysis of the gender gap in mathematics, *Am. Econ. J.: Applied Econ.* **2** (2010), 210–240.
- [15] T. FUCHS and L. WÖSSMANN, What accounts for international differences in student performance? A re-examination using PISA data, *Empirical Econ.* **32** (2007), 433–464.
- [16] L. GUISO, F. MONTE, P. SAPIENZA, and L. ZINGALES, Culture, gender, and math, *Science* **320** (2008), 1164–1165.
- [17] R. HAUSMANN, L. D. TYSON, and S. ZAHIDI, *The Global Gender Gap Report 2007*, World Economic Forum, Geneva, Switzerland, 2008. Retrieved from http://www3.weforum.org/docs/WEF_GenderGap_Report2007.pdf.
- [18] _____, *The Global Gender Gap Report 2009*, World Economic Forum, Geneva, Switzerland, 2010. Retrieved from <https://members.weforum.org/pdf/gendergap/report2009.pdf>.
- [19] A. HESTON, R. SUMMERS, and B. ATEN, Penn World Table Version 6.3, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, August 2009. Retrieved from http://pwt.econ.upenn.edu/php_site/pwt63/pwt63_form.php.
- [20] J. S. HYDE, S. M. LINDBERG, M. C. LINN, A. ELLIS, and C. WILLIAMS, Gender similarities characterize math performance, *Science* **321** (2008), 494–495.
- [21] J. S. HYDE and J. E. MERTZ, Gender, culture, and mathematics performance, *Proc. Nat. Acad. Sci. USA* **106** (2009), 8801–8807.
- [22] International Mathematical Olympiad home page, <http://official.imo2011.nl/>.
- [23] T. KIM, J.-H. LEE, and Y. LEE, Mixing versus sorting in schooling: Evidence from the equalization policy in South Korea, *Econ. Ed. Rev.* **27** (2008), 697–711.
- [24] Kingdom of Bahrain Ministry of Education, *Education in Bahrain*. Retrieved from <http://www.moe.gov.bh/en/education/religious.aspx>.
- [25] S. M. LINDBERG, J. S. HYDE, J. L. PETERSEN, and M. C. LINN, New trends in gender and mathematics performance: A meta-analysis, *Psychol. Bull.* **136** (2010), 1123–1135.
- [26] S. MACHIN and T. PEKKARINEN, Global sex differences in test score variability, *Science* **322** (2008), 1331–1332.
- [27] F. MAEL, A. ALONSO, D. GIBSON, K. ROGERS, and M. SMITH, *Single-sex versus Coeducational Schooling: A Systematic Review*, U.S. Department of Education Office of Planning, Evaluation and Policy Development, Washington, D.C., 2005. Retrieved from <http://www2.ed.gov/rschstat/eval/other/single-sex/single-sex.pdf>.
- [28] M. O. Martin (ed.), *TIMSS 2003 User Guide for the International Database*, TIMSS and PIRLS International Study Center, Boston College, Chestnut Hill, MA, 2005. Retrieved from <http://timss.bc.edu/timss2003i/userguide.html>.
- [29] M. O. MARTIN, I. V. S. MULLIS, and S. J. CHROSTOWSKI (eds.), *TIMSS 2003 Technical Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*, IEA TIMSS and PIRLS International Study Center, Boston College, Chestnut Hill, MA, 2004. Retrieved from http://timss.bc.edu/PDF/t03_download/T03TECHRPT.pdf.
- [30] E. C. MELHUIS, K. SYLVA, P. SAMMONS, I. SIRAJ-BLATCHFORD, B. TAGGART, et al., Preschool influences on mathematics achievement, *Science* **321** (2008), 1161–1162.
- [31] T. Miller (Ed.), *Mapping the Global Muslim Population: A Report on the Size and Distribution of the World's Muslim Population*, Pew Research Center, October 2009. Retrieved from <http://pewforum.org/newassets/images/reports/Muslimpopulation/Muslimpopulation.pdf>.

- [32] I. V. S. MULLIS, M. O. MARTIN, and P. FOY (with J. F. OLSON, C. E. PREUSCHOFF, E. ERBERBER, A. ARORA, and J. GALIA), *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*, IEA TIMSS and PIRLS International Study Center, Boston College, Chestnut Hill, MA, 2008. Retrieved from <http://timss.bc.edu/TIMSS2007/mathreport.html>.
- [33] I. V. S. MULLIS, M. O. MARTIN, E. J. GONZALEZ, and S. J. CHROSTOWSKI (eds.), *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*, IEA TIMSS and PIRLS International Study Center, Boston College, Chestnut Hill, MA, 2004. Retrieved from http://timss.bc.edu/PDF/t03_download/T03INTLMATRPT.pdf.
- [34] J. F. OLSON, M. O. MARTIN, and I. V. S. MULLIS (eds.), *TIMSS 2007 Technical Report*, IEA TIMSS and PIRLS International Study Center, Boston College, Chestnut Hill, MA, 2008. Retrieved from <http://timss.bc.edu/timss2007/techreport.html>.
- [35] Organisation for Economic Co-operation and Development, *Learning for Tomorrow's World: First Results from PISA 2003*, OECD, Paris, France, 2004. Retrieved from http://www.oecd.org/pages/0,3417,en_32252351_32236173_1_1_1_1_1,00.html.
- [36] Y.-S. PARK and U. KIM, Family, parent-child relationship, and academic achievement in Korea: Indigenous, cultural, and psychological analysis, *Indigenous and Cultural Psychology: Understanding People in Context* (U. Kim, K.-S. Yang, and K.-K. Hwang, eds.), Springer, 2006, 421-444.
- [37] *PISA 2009 Results*. Retrieved from www.oecd.org/edu/pisa/2009.
- [38] D. G. POPE and J. R. SYDNOR, Geographic variation in the gender differences in test scores, *J. Econ. Perspectives* **24** (2010), 95-108.
- [39] C. RIEGLE-CRUMB, The cross-national context of the gender gap in math and science, in *The Social Organization of Schooling* (L. Hedges and B. Schneider, eds.), Russell Sage Fdn., New York, NY, 2005, 227-243.
- [40] H. RINDERMANN and S. J. CECI, Educational policy and country outcomes in international cognitive competence studies, *Perspectives Psychol. Sci.* **4** (2009), 551-577.
- [41] C. SHEN and H. P. TAM, The paradoxical relationship between student achievement and self-perception: A cross-national analysis based on three waves of TIMSS data, *Ed. Res. Eval.* **14** (2008), 87-100.
- [42] S. A. SHIELDS, The variability hypothesis: The history of a biological model of sex differences in intelligence, *Signs: J. Women Cult. Soc.* **7** (1982), 769-797.
- [43] J. P. SHONKOFF and D. A. PHILLIPS (eds.), *From Neurons to Neighborhoods: The Science of Early Childhood Development*, Committee on Integrating the Science of Early Childhood Development, National Research Council and Institute of Medicine, 2000. Retrieved from http://www.nap.edu/openbook.php?record_id=9824.
- [44] *Social Watch Gender Equity Index 2007*. Retrieved from <http://www.socialwatch.org/node/9355>.
- [45] *Social Watch Gender Equity Index 2009*. Retrieved from <http://www.socialwatch.org/node/11561>.
- [46] C. H. Sommers (ed.), *The Science on Women and Science*, American Enterprise Institute Press, 2009.
- [47] H. W. STEVENSON, S.-Y. LEE, C. CHEN, J. W. STIGLER, C.-C. HSU, S. KITAMURA, and G. HATANO, *Contexts of Achievement: A Study of American, Chinese, and Japanese Children*, Mono. Soc. Res. Child Devel., **55**, No. 1/2, Blackwell, 1990, i-iii-vi+1-119.
- [48] L. SUMMERS, *Remarks at NBER Conference on Diversifying the Science and Engineering Workforce*, 2005, January 14. Retrieved from http://www.harvard.edu/president/speeches/summers_2005/nber.php.
- [49] J. TIERNEY, Daring to discuss women in science, *New York Times* (2010, June 8), D1; Legislation won't close gender gap in science, *New York Times*, (2010, June 15), D4.
- [50] D. H. UTTAL, Beliefs about genetic influences on mathematics achievement: A cross-cultural comparison, *Genetica* **99** (1997), 165-172.
- [51] J. WAI, M. CACCHIO, M. PUTALLAZ, and M. C. MAKEL, Sex differences in the right tail of cognitive abilities: A 30-year examination, *Intelligence* **38** (2010), 412-423.
- [52] A. W. WISEMAN, A culture of (in)equality?: A cross-national study of gender parity and gender segregation in national school systems, *Res. Comp. Int. Ed.* **3** (2008), 179-201.
- [53] Z. Y. ZADEH, E. GEVA, and M. A. ROGERS, The impact of acculturation on the perception of academic achievement by immigrant mothers and their children, *School Psych. Int.* **29** (2008) 39-70.