Attitudes to learning – questioning the PISA data

Tom Benton

COLOUR PRINTING RECOMMENDED

Cambridge Assessment Research Report
2nd September 2015

University of Cambridge
Local Examinations Syndicate

Author contact details:
Tom Benton
ARD Research Division
Cambridge Assessment
1 Regent Street
Cambridge
CB2 1GG
Benton.t@cambridgeassessment.org.uk

http://www.cambridgeassessment.org.uk/

Cambridge Assessment is the brand name of the University of Cambridge Local Examinations Syndicate, a department of the University of Cambridge. Cambridge Assessment is a not-for-profit organisation.

How to cite this publication:

Benton, T. (2015). *Attitudes to learning – questioning the PISA data*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Introduction

This aim of this report is to highlight some of the issues around gender differences that are revealed by data from the Programme for International Student Assessment (PISA). The OECD themselves have undertaken a very thorough review of the available data on this issue (see OECD, 2015) and as such this report draws heavily upon the report that they have already published. Unless stated otherwise, all references to “the OECD report” within this report will refer to OECD (2015).

The OECD report covers a number of issues including some beyond PISA such as participation in higher education and further careers. This brief report cannot possibly review all of the issues raised in the OECD report and interested readers should look at this more detailed analysis for themselves. Rather, the aim of this report is to present some of the major findings from the OECD’s work. Where appropriate, this report also discusses the limitations of their analyses and highlights alternative interpretations of the data beyond those originally presented.

The report is broken down into four sections:

1) A general review of gender differences in mathematics, reading and science internationally including consideration of whether high performing jurisdictions tend to exhibit smaller levels of inequality between genders.
2) An analysis of some of the apparent reasons for underperformance amongst boys – particularly in reading.
3) An analysis of some of the apparent reasons for widespread underperformance of girls in mathematics.
4) An analysis of the types of PISA science items appearing to favour each gender.

1. Is gender equality a by-product of high performance?

The countries and economies with strong performances in PISA have a high profile. In the past, discussions of high performance often centred on Finland whereas more recently the emphasis has been on high performing east Asian economies such as Korea, Singapore and in particular Shanghai-China. It is of obvious interest to understand whether these high performing economies also achieve gender equality in their performances.

The OECD report avoids making any straightforward statements indicating that high performance will imply gender equality. However, in a number of places it hints that the two go hand in hand. For example, the foreword to the report states that

> "*in the top performing countries in PISA, such as Shanghai-China, Singapore, Hong-Kong China and Chinese Taipei, girls perform on a par with their male classmates in mathematics and attain higher scores in mathematics than boys in most other countries*" (OECD, 2015, page 3)

Most bluntly, the report later states that

> "*the average girl in Shanghai-China scores…well above boys' average performance in every other country…and, crucially, just as well as the average boy in Shanghai-China. Similarly girls in Finland, Macao-China, Singapore and Chinese Taipei perform as well as boys in mathematics – despite the fact that (**or maybe because**) standards of performance in these countries are among the highest in the world*" (OECD, 2015, page 64, emphasis added)

Similar statements are made on pages 15, 153 and 155. Such statements seem to hint that gender differences would disappear if only countries were any good at the subjects being tested in PISA. As such, they are worthy of further scrutiny. Figure 1.1 shows, for all countries, how overall performance in PISA mathematics relates to the size of the gender difference. All results are shown on the PISA scale of ability which was defined to have a mean of 500 and a standard deviation of 100 on a scale defined for OECD countries participating in PISA 2003. This chart includes dotted lines for each country showing a 95 per cent confidence interval (in other words a range of reasonable uncertainty) for the gender differences. As can be seen from the chart, in the majority of countries - though by no means all - boys tend to outperform girls on average.

Figure 1.1 shows that high performing economies do not have any consistent tendency to display lower gender differences than low performing ones. Out of the top 5 performers in PISA, two (Hong-Kong China and Korea) display gender differences that are larger than the OECD average[1], and three (Chinese-Taipei, Singapore and Shanghai-China) display differences that are smaller than the OECD average. More importantly, in four out of five of these cases the gender difference is not significantly different from the OECD average. For example, this implies that, once uncertainties in the estimates are allowed for, there is no strong evidence that the gender gap in Shanghai-China is any smaller than in the average OECD country. Furthermore, as can clearly be seen from the chart, the level of uncertainty over the gender gap in Chinese Taipei is so large that it's impossible to make any precise statements about gender inequality within this country on the basis of PISA data. Only Singapore displays a significantly different gender gap from the OECD average, with girls performing slightly better than boys on average – but not significantly so.

Having said this, even the apparently low level of inequality in Singapore has a negative side. Table 1.3a of the OECD report shows that the underperformance of boys in mathematics becomes statistically significant for low performers. Specifically, the 10th percentile of achievement for boys is 21 points below the 10th percentile for girls. This means that, in

[1] Indicating that the statements relating to Hong-Kong China on pages 3, 15 and 155 of the OECD report are simply incorrect. Furthermore, Table 1.3a of the OECD report shows that, amongst high performing students (the 90th percentile), Hong-Kong actually has the 7th largest gender gap in the world.

Singapore, the gender gap among low performers is the 4th biggest in the world. Only Qatar and Jordan and display larger gender gaps in favour of girls, and only Liechtenstein displays a gender gap of at least this magnitude in favour of boys (albeit with an extremely high standard error).

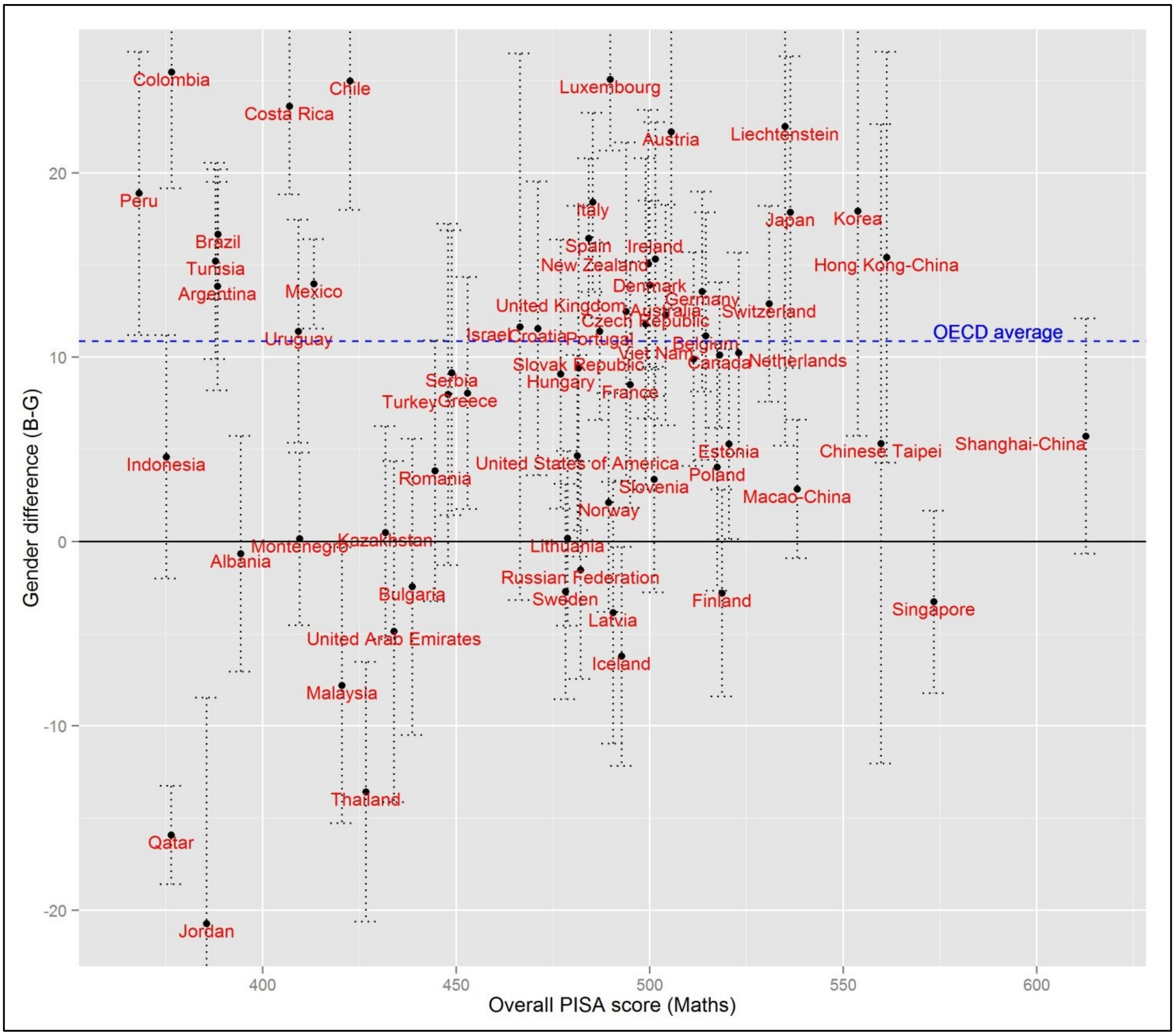

Figure 1.1: Overall performance and gender differences (with 95% confidence intervals) in Mathematics in PISA 2012.

Similar analysis for reading is shown in Figure 1.2. To begin with it should be noted that every country in the world displays a gender gap in favour of girls in reading and that this gap (38 points on average for OECD countries) is much bigger than even the largest gender gaps displayed for boys in mathematics. It is also worth noting that, despite widespread concern over underachievement amongst boys in the UK, in fact, the UK shows one of the smallest gender gaps of any country.

Figure 1.2 provides a more compelling case for high performance being associated with greater gender equality. All five of the top performers in reading display gender differences that are significantly below the OECD average. There is some logic to this finding. Since so much low performance in reading is found amongst boys, it is surely very difficult to achieve the highest performances without tackling this and this would itself lead to a reduction in the gender gap.

Having said this, it should also be noted that very large gender gaps exist in even the highest performing jurisdictions. Indeed all of these economies display gaps very similar to that which is found in the UK (and that remains a concern among policy makers). Furthermore, it is also clear that a number of very *low* performing countries (especially Albania) display very low levels of inequality between the genders. In addition, high-performing Finland is notable for the fact that its high level of performance in reading overall goes alongside one of the largest gender gaps in the world. Finally, it is worth noting that, when Reading was assessed more thoroughly in PISA 2009[2], only Singapore of the top 5 performers from 2012[3] displayed a gender gap significantly different from the OECD average[4].

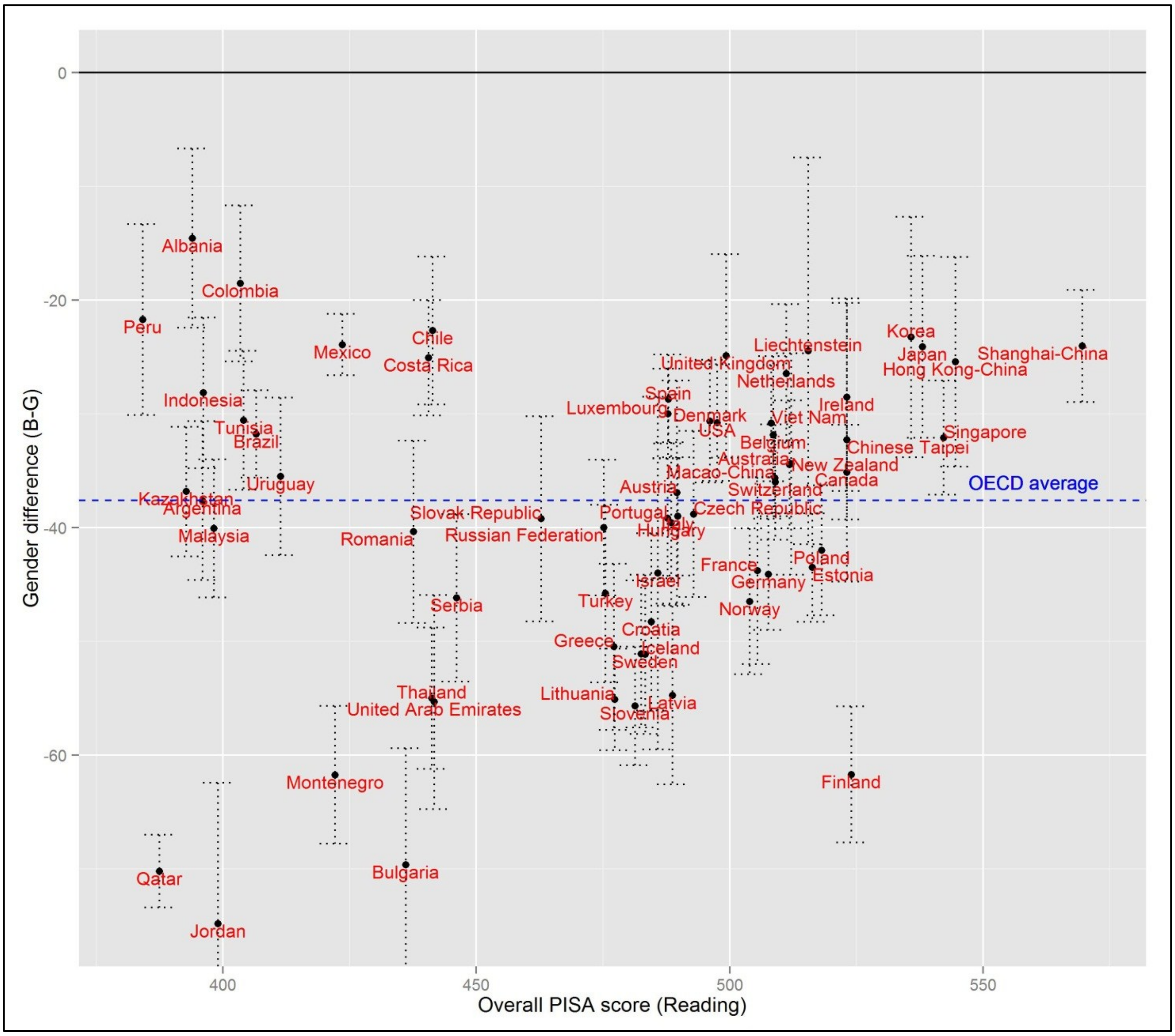

Figure 1.2: Overall performance and gender differences (with 95% confidence intervals) in Reading in PISA 2012.

For the sake of completeness, Figure 1.3 shows the same relationship for science. It is worth noting that most countries do not show any significant differences in performance between genders[5] and that across OECD countries boys perform just 1.5 points better than girls on average. As such, issues of gender inequality are less pressing in science performance and the OECD report makes no reference to higher performing economies displaying lower levels of

[2] Reading was the main focus for PISA 2009. Mathematics was the main focus in 2012.
[3] That is, Shanghai-China, Hong-Kong China, Singapore, Japan, and Korea.
[4] See OECD (2010) Table I.2.3.
[5] The UK is one of the few countries where boys significantly outperform girls in science. In fact, across OECD countries only Luxembourg displays a larger disparity in favour of boys. Only Finland and Greece display larger inequalities in favour of girls.

gender inequality for science. Nonetheless, it is perhaps worth noting in passing that high-performing Finland also displays one of the largest gender gaps in the world.

Figure 1.3: Overall performance and gender differences (with 95% confidence intervals) in Science in PISA 2012.

As has been demonstrated above therefore, across the three subjects, there is no compelling reason to believe that gender differences are absent amongst high-performing economies. Rather, they affect high and low performing countries almost equally. Furthermore, in the case of reading, the gender gap is universal to all PISA participants.

Having said the above, one of the reasons the OECD presents their findings is to demonstrate that “gender gaps in school performance are not determined by innate differences in ability” (OECD, 2015, page 15). This statement is indisputably true. For example, if boys perform well in reading in Korea then it cannot be true that boys are incapable of being good at reading. However, it is not necessary to appeal to high performing economies to demonstrate this. In fact, every country in the world contains examples of high performing girls in mathematics and high performing boys in reading. This itself makes it obvious that gender is not itself an inherent barrier to high performance in any subjects. The differences in the average performance of boys and girls are dwarfed by the variation in performance within genders. For example, the expected difference in reading performance between two randomly chosen girls will be larger than the difference between the average score for boys and the average score for girls. In this sense, we can clearly see that “aptitude knows no gender” (OECD, 2015, page 13) without any need to refer to performance in high performing economies.

2. What can we learn about underachievement amongst boys?

The OECD report notes that a greater number of boys than girls display very low levels of attainment. Specifically, “across OECD countries in 2012, 14 per cent of boys and 9 per cent of girls did not attain the baseline level of proficiency in any of the three core subjects” (OECD, 2015, page 20). However it should be noted that this difference is driven by reading. If we calculate the percentage of students failing to attain the baseline level in either mathematics or science (but ignore reading), then across the OECD there is hardly any difference between the genders. However, it is once reading is included in this calculation that the underperformance of boys emerges. This is unsurprising since across OECD countries 24 per cent of boys fail to achieve the baseline level in reading compared to only 12 per cent of girls (OECD, 2013, Table I.4.2a). For this reason, this section will focus exclusively on differences in reading attainment.

The OECD report examines numerous differences in the attitudes and behaviours of boys and girls that might possibly explain the differences in reading attainment. These include differences in the way they use computers, the amount of time spent reading, the amount of time spent on homework and attitudes towards school amongst other things. By far the biggest differences emerge in the amount of time spent playing video games. On average across OECD countries 61 per cent of boys say they play one-player games every day compared to 41 per cent of girls. Similarly, 51 per cent of boys say they play collaborative online games every day compared to 27 per cent of girls (OECD, 2015, Figure 2.4). Furthermore, some association is found between playing video games and performance; with one-player games associated with slightly increased performance and collaborative online games associated with decreased performance[6]. However, even if we accept these associations as causal links, then the analysis suggests that these differences would only account for around 5 points of the difference of roughly 40 between genders in reading performance[7].

Another striking difference between the genders is in the amount of time spent doing homework. The OECD report describes boys as being “overwhelmingly less likely than girls to spend time doing homework” (OECD, 2015, page 49). Specifically, analysis PISA 2012 reveals that on average across OECD countries girls spend around an hour a week more on doing homework than boys (average of 5.5 hours versus 4.5 hours) and in a number of countries this difference exceeds 2 hours. However, further analysis suggests that even if boys spent just as much time on homework as girls, all of this additional effort would only reduce the gender gap by around 8 points (OECD, 2015, Table 2.10b).

A further difference comes in in terms of attitudes to school with boys’ attitudes being considerably worse. For example, on average across OECD countries, 16 per cent of boys agree that “school has been a waste of time” compared to 8 per cent of girls (OECD, 2015, Table 2.15). However, a quick analysis of the data (not described in detail here) suggested this only accounted for around 4 points of the gender gap in reading[8].

Although the above gender differences are interesting, the factor with the greatest potential to explain gender differences in performance is the extent to which students read for enjoyment. Although, this was not measured as part of PISA 2012 (which focused on mathematics) these variables were captured within the surveys from PISA 2009.

For the purposes of illustration, in this report we shall focus upon a single question from PISA 2009 examining reading for pleasure. Students were asked how often they read a number of

[6] Although in both cases the relationship is non-linear and complex.

[7] Estimated by comparing raw gender differences in Table 1.2a of the OECD report to those calculated after accounting for the effect of video games in Table 2.8a of the same report.

[8] Based on analysis of data from PISA 2009.

genres of writing[9] “**because they want to**” (emphasis in original questionnaire). The percentage of students giving each of the available responses when asked about how often they read fiction are shown in Figure 2.1. Results are shown for the UK and for Finland with these countries being chosen because of the UK context in which this report will be presented and because of the large gender gap for Finland found in Figure 1.2. As can be seen, in both countries, boys are far more likely than girls to say they “never or almost never” read fiction for pleasure. At the other end of the scale, girls are overwhelmingly more likely read fiction several times a week. Although differences are clear in both countries, the gender gaps are more pronounced in Finland. This aligns with the earlier finding (Figure 1.2) of a very large gender gap in performance in this country.

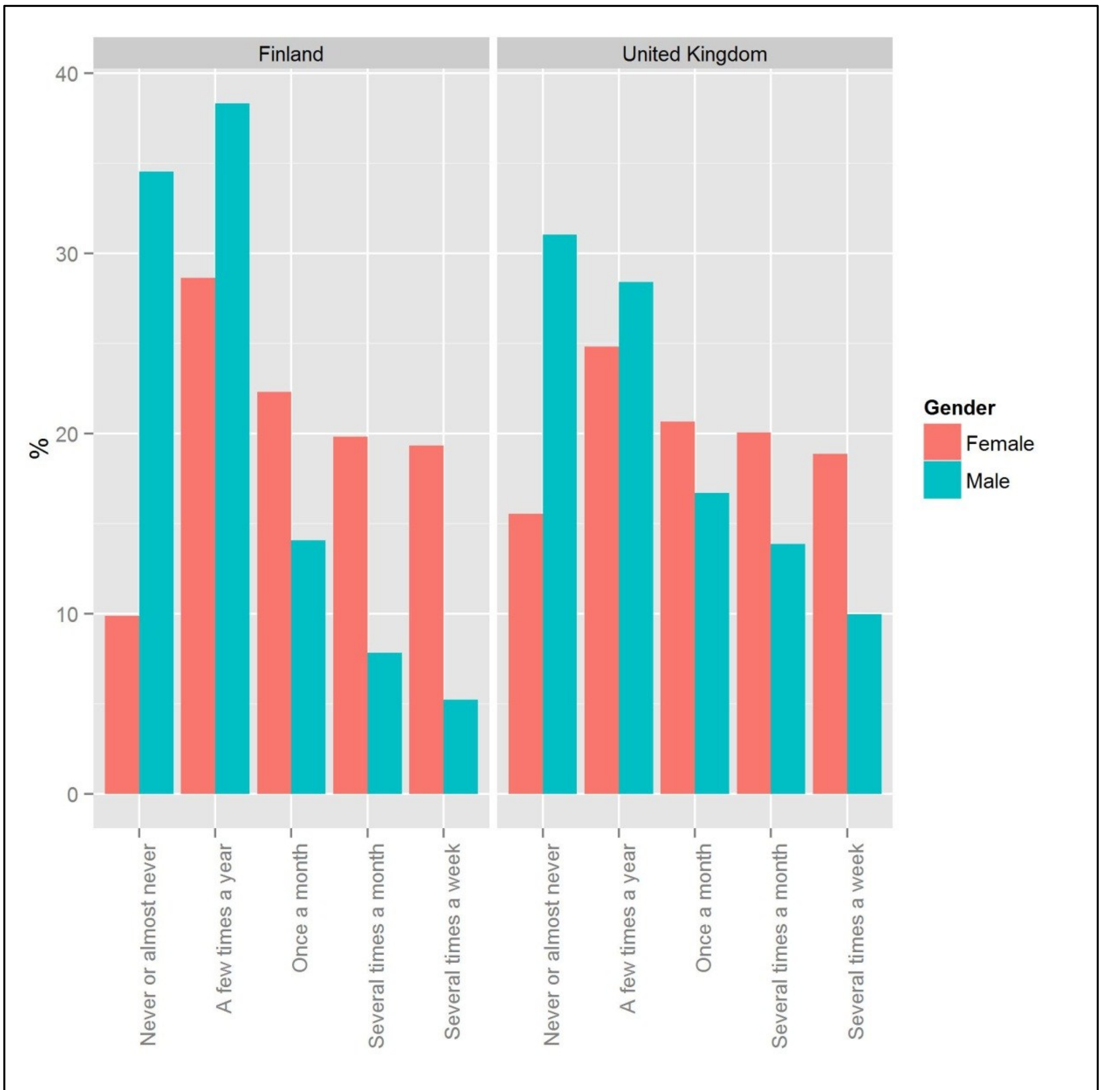

Figure 2.1: How often do you read fiction (novels, narratives, stories) because you want to (PISA 2009)?

Having established the above difference in reading fiction for pleasure, it is now of interest to determine the extent to which this difference accounts for the performance gap. In order to do this we estimate what the performance of boys would be if the extent to which they read fiction for pleasure was the same as was indicated by girls. In order to estimate this, we reweight the PISA data, giving more weight to boys who say they read fiction often and less weight to those who never or rarely read it. The performance of boys in this reweighted sample can then be compared to performance in the original sample (as well as to the performance of girls) to see how much difference this would make. This analysis is completed in Figure 2.2. The top of the

[9] Magazines, comic books, fiction, non-fiction books, and newspapers.

chart shows the difference in the original performance distribution in both countries. As can be seen in both cases, but particularly in Finland, girls are far more likely to have high levels of performance and boys are more likely to display poor performance. The extent to which differences reduce, once differences in the extent of reading fiction for pleasure are removed, is shown in the bottom part of the chart. The performance distribution for girls is unchanged as only the boys' data has been reweighted. As can be seen, amongst samples of pupils who do equal amounts of reading for pleasure, the gender difference in reading performance is much smaller. In fact, in the UK, the performance distribution becomes very similar between the two genders. Overall, accounting for the differences in reading fiction for pleasure reduces the gender gap in average performance from 24 to 8 points in the UK and from 55 to 26 points in Finland.

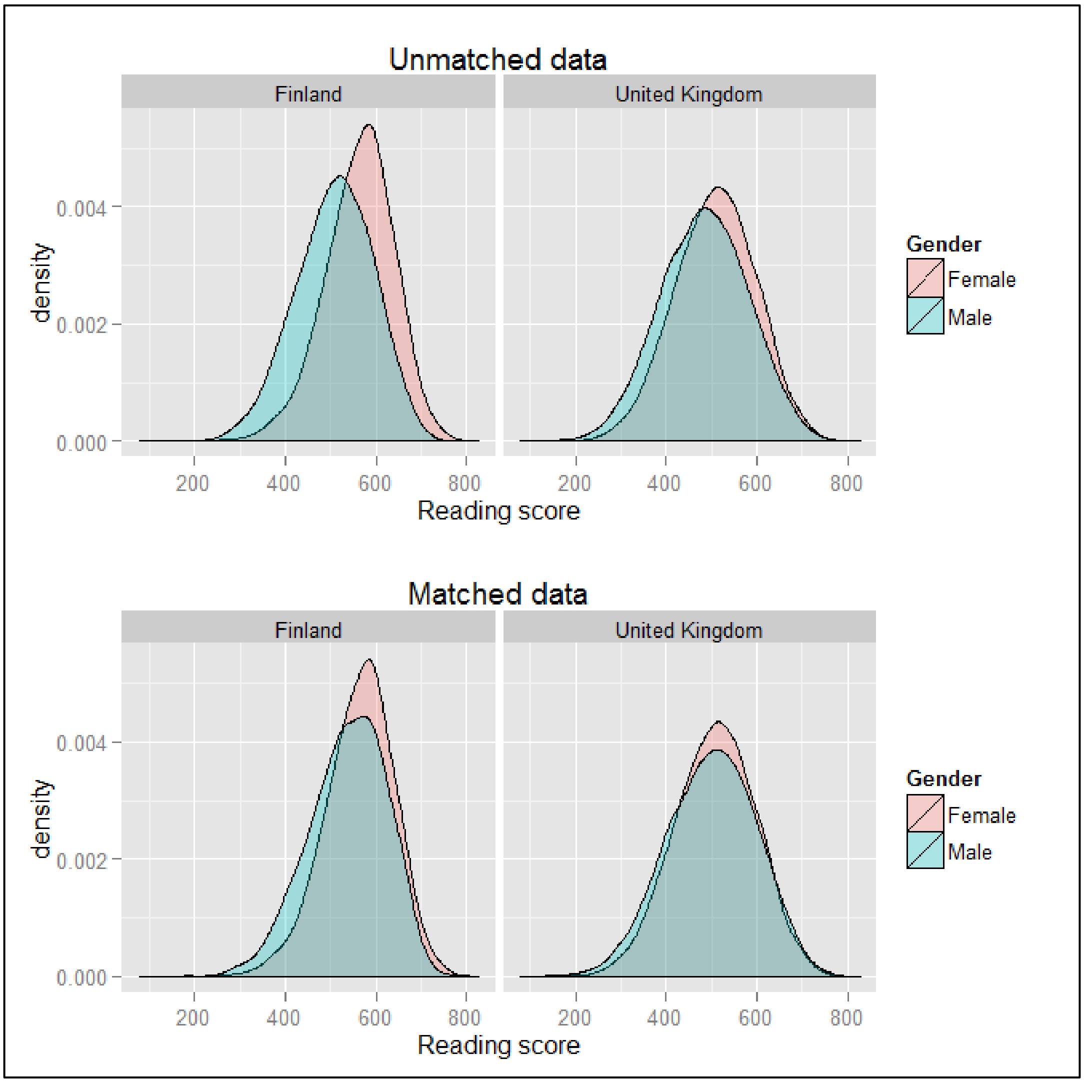

Figure 2.2: Distribution of reading performance in the United Kingdom and Finland for original sample (unmatched data) and for sample where boys are as likely to read fiction for pleasure as girls (matched data).

Using the same methodology, the effect of accounting for differences in reading fiction for pleasure on the performance gap is shown across all OECD countries and several East Asian countries of interest in Table 2.1. On average across OECD countries, the difference in the extent of reading fiction for pleasure explains roughly half of the gender gap in reading performance – reducing it from 38 points to 22 points. More thorough analysis in the OECD report (OECD, 2015, Figure 2.11), using a fuller measure the extent of reading enjoyment, suggests that if boys enjoyed reading to the same extent as girls the gender gap would reduce a little further to 16 points. In fact, in some countries such as the UK, the USA and the Netherlands, the gender gap can be almost entirely explained by differences in the extent to which boys and girls enjoy reading. The results from this analysis are reproduced in the final column of Table 2.1.

Although the above findings are interesting, we must be careful in attributing the cause of the gender gap so quickly. In particular although the above analyses establish a link between reading enjoyment (or extent of reading fiction) and reading performance we cannot be certain from this that the former causes the latter. Whilst it is very plausible that a greater enjoyment of reading (amongst girls) will lead to greater motivation and improved performance, it is also possible that the causality works the other way around – greater ability in reading allows a greater level of enjoyment and causes the increased levels of reading for pleasure. Equally it is possible that both reading performance and reading enjoyment are influenced by some other external factors leading to the associations between the two. Thus, whilst encouraging greater levels of reading for enjoyment amongst boys is doubtless a sensible policy (OECD, 2015, page 156), the PISA data cannot absolutely prove that such an approach will necessarily reduce the gender gap.

This section has noted a number of factors that may explain the gender gap in reading. In each case we have provided an estimate of the number of points of difference that might be explained by each factor. However, it is important to note that these effects are not independent of one another and cannot simply be added up. For example, the greater amount of time spent playing computer games by boys may relate to the fact they are less likely to read for enjoyment and spend less time doing homework. Similarly, the more negative attitudes to school displayed by boys may also explain why they make less effort with homework. Furthermore, since data on many of these factors were collected in different sweeps of PISA it is not possible to combine them all into a single analysis to get an overall picture of the extent to which the gender gap can be reduced by simultaneously accounting for many of these different factors. This remains a possible area for further research.

To finish this section, it is interesting to note that, although girls outperform boys in reading on average across all countries in PISA, differences in literacy ability are far less evident amongst adults. In fact, across a number of countries (including the UK) men aged 16-29 outperform women of the same age (OECD, 2015, Figure 4.16, page 123). This again suggests that gender differences in reading ability are not innate. It is possible that the workplace provides greater motivation for these skills in boys and that, once in this environment rather than in school, the universal issue of boys' underachievement in reading disappears.

Country	Boys (before matching)	Boys (matched)	Girls	Difference before matching	Difference after matching	Difference after accounting for reading enjoyment (OECD 2015, Table 2.9k)
Australia	498.9	519.7	534.1	-35.2	-14.4	-8.7
Austria	454.7	481.0	495.9	-41.2	-15.0	-12.2
Belgium	501.4	520.9	524.4	-23.1	-3.5	-5.0
Canada	509.2	526.6	542.4	-33.2	-15.7	-4.7
Chile	439.9	444.4	461.8	-21.9	-17.4	-9.3
Czech Republic	460.2	493.7	507.6	-47.4	-13.8	-19.5
Denmark	481.9	494.4	510.1	-28.2	-15.7	-7.2
Estonia	480.9	490.6	524.6	-43.7	-34.0	-15.6
Finland	509.0	538.1	563.7	-54.7	-25.6	-20.4
France	477.3	497.7	516.3	-39.0	-18.6	-18.0
Germany	486.6	513.4	523.5	-36.9	-10.1	-7.9
Greece	459.3	476.1	506.4	-47.1	-30.3	-21.5
Hungary	476.0	485.0	514.0	-38.1	-29.1	-13.3
Iceland	479.3	494.5	523.7	-44.3	-29.2	-18.6
Ireland	480.1	491.0	518.2	-38.1	-27.2	-20.5
Israel	460.1	466.5	497.4	-37.3	-31.0	-20.4
Italy	465.0	480.6	510.0	-45.0	-29.4	-19.6
Japan	502.0	512.4	540.7	-38.6	-28.2	-25.2
Korea	523.2	531.3	558.5	-35.3	-27.2	-24.4
Luxembourg	455.0	483.4	492.7	-37.7	-9.3	-11.6
Mexico	414.3	413.3	438.7	-24.3	-25.3	-17.9
Netherlands	500.0	526.9	523.1	-23.1	3.7	0.5
New Zealand	501.3	519.6	545.1	-43.8	-25.4	-16.5
Norway	481.2	501.8	527.8	-46.6	-26.0	-22.4
Poland	477.5	493.7	525.6	-48.0	-31.9	-24.8
Portugal	471.3	486.8	508.2	-36.9	-21.4	-14.2
Slovak Republic	453.3	469.1	503.8	-50.5	-34.7	-32.7
Slovenia	459.6	473.6	512.5	-52.9	-38.9	-31.5
Spain	468.1	482.5	496.3	-28.2	-13.8	-7.7
Sweden	477.2	503.4	521.7	-44.5	-18.3	-14.7
Switzerland	482.7	508.9	520.7	-38.0	-11.7	-9.0
Turkey	445.1	443.6	487.0	-41.9	-43.4	-31.6
United Kingdom	484.0	500.0	508.0	-23.9	-7.9	-2.5
United States	489.0	501.9	513.4	-24.4	-11.6	-1.3
OECD average	**476.6**	**493.1**	**514.6**	**-38.0**	**-21.5**	**-15.6**
Hong Kong-China	518.1	525.5	550.5	-32.4	-25.0	-17.3
Macao-China	470.2	477.2	504.1	-33.9	-26.9	-21.3
Shanghai-China	536.1	539.2	575.7	-39.6	-36.5	-27.0
Singapore	511.2	523.4	542.0	-30.8	-18.6	-5.6
Chinese Taipei	478.3	485.8	514.1	-35.8	-28.3	-16.7

Table 2.1: Gender differences in reading performance before and after accounting for either extent of reading fiction for pleasure or more general reading enjoyment (PISA 2009)[10].

[10] Pupils answering the question on reading fiction for pleasure only.

3. Why do we think girls tend to perform worse than boys in Mathematics?

The OECD's chapter on this subject is entitled "Girls' lack of self-confidence" and suggests that this is at the heart of their underperformance (on average and in some countries) relative to boys. A selection of questions indicating the tendency towards lower confidence and higher anxiety amongst girls when it comes to mathematics is shown in Table 3.1[11]. Average results across OECD countries are shown alongside results for Finland, the UK as well as high-performing Shanghai-China and Singapore.

Percentage of students…	**OECD average**		**Finland**		**UK**		**Shanghai-China**		**Singapore**	
…confident that they could….	**Boys**	**Girls**	**Boys**	**Girls**	**Boys**	**Girls**	**Boys**	**Girls**	**Boys**	**Girls**
Calculate how much cheaper a TV would be after a 30% discount	84	75	81	63	88	80	95	95	95	93
Calculate how many square metres of tiles are needed to cover a floor	75	61	68	48	77	61	93	91	83	77
Solve an equation like 3x+5=17	85	86	82	86	88	85	96	98	93	94
Calculate the petrol-consumption rate of a car	67	44	64	28	62	40	84	76	77	69
…agreeing that…										
I learn mathematics quickly	59	45	65	48	66	50	60	38	67	58
I am just not good at mathematics	37	48	36	47	24	41	36	57	35	40
I feel helpless when doing a mathematics problem	25	35	20	35	15	24	22	33	25	29

Table 3.1: Confidence and anxiety in mathematics

Girls are less likely to be confident that they would be able to complete mathematics tasks. The biggest difference is seen in the percentages of pupils saying they would be confident that they could calculate the petrol consumption rate of a car with more than two-thirds of boys across OECD countries saying they would be confident compared to less than half of girls. The difference is particularly striking in Finland where, despite the fact that girls actually outperform boys in mathematics, only just over a quarter of girls say they would be confident in completing this task compared to nearly two-thirds of boys. One exception to this rule comes in relation to confidence relating to solving a specific equation of the type that students may have seen in class. For this question there is hardly any difference in confidence between boys and girls, and in Finland girls are significantly more likely to say that they could complete this task.

Across more general questions asking about pupils' feeling to mathematics, for every country in Table 3.1, girls were less positive than boys. In every case they were more likely to agree that they are "just not good at mathematics" and "feel helpless when doing a mathematics problem". They were also less likely to agree that they "learn mathematics quickly". As such it is certainly clear that girls display less confidence in their mathematics ability than boys.

Having established that gender differences in confidence exist, it is next of interest to understand the extent to which these contribute the gender gap in mathematics. For the purposes of this

[11] Examples of further questions covering these areas are shown in OECD (2015) Figures 3.8, 3.9 and 3.10.

section, we will focus upon analysis of just one of the questions listed above – the extent to which pupils agree that they are "just not good at mathematics".

A more detailed breakdown of responses to this question in the UK, Singapore and Shanghai-China is shown in Figure 3.1[12]. In addition to the results from Table 3.1, it can be seen from this graph that, in each country, girls are far more likely to strongly agree that they are not good at mathematics and boys are far more likely to strongly disagree.

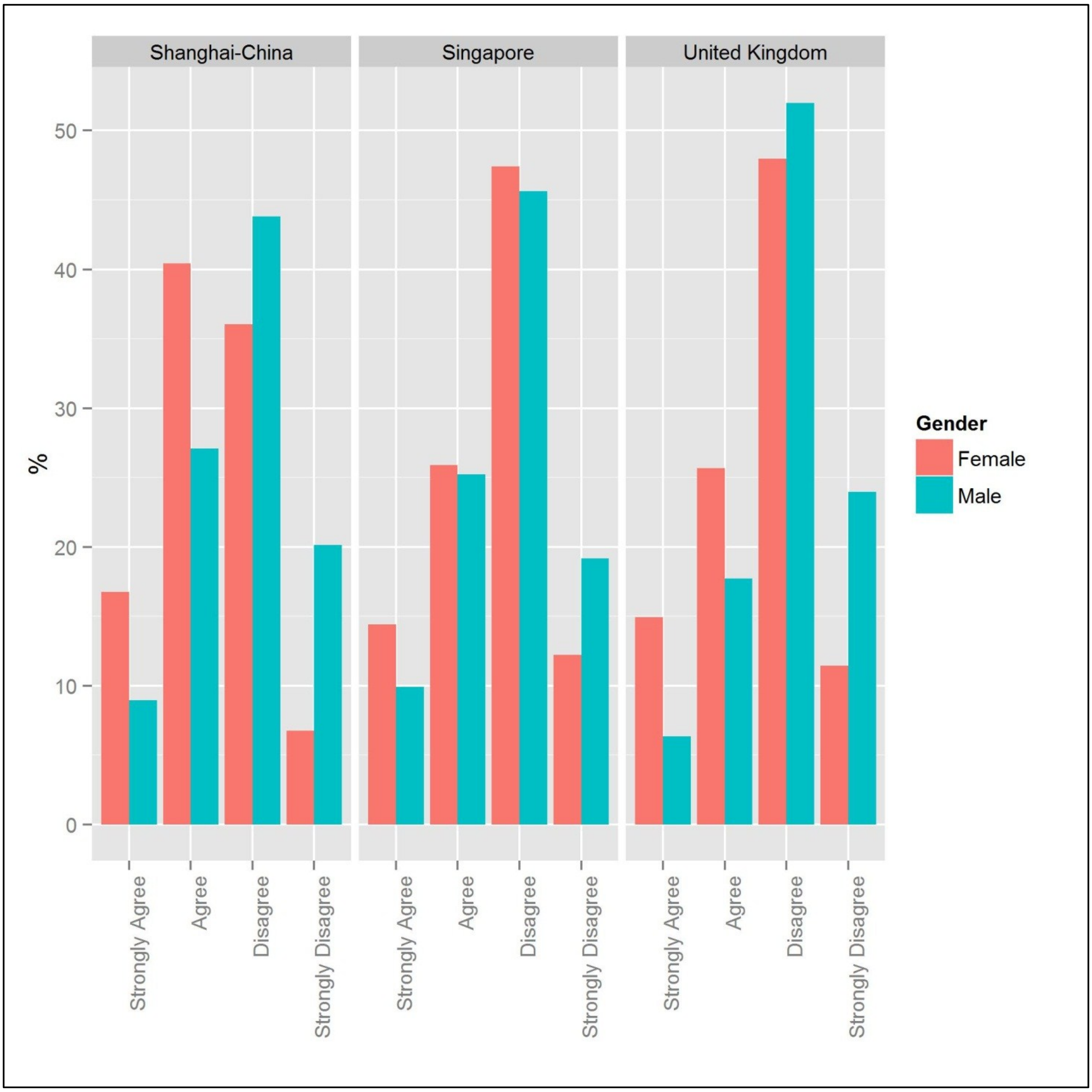

Figure 3.1: Responses to the statement "I am just not good at mathematics" in three jurisdictions.

[12] For reasons of space, Finland is not included in this and subsequent figures.

As with our earlier analysis of reading performance, we now reweight the data to estimate what the performance of girls would be if they were just as likely to disagree with this statement as boys[13]. This analysis is shown in Figure 3.2. The top part of this chart compares the distribution of mathematics performance in the original samples. As can be seen, in each country (but particularly in the UK) there are marginally more boys at the top end of the performance distribution. However, in Singapore there are also more boys at the bottom end of the distribution so that the average score for girls marginally exceeds the average for boys. The bottom half of the Figure shows the estimated score distributions if girls responded to the statement “I am just not good at mathematics” in the same way as boys. In all three countries, the gender gap is changed so that, on average, it is estimated that girls would outperform boys.

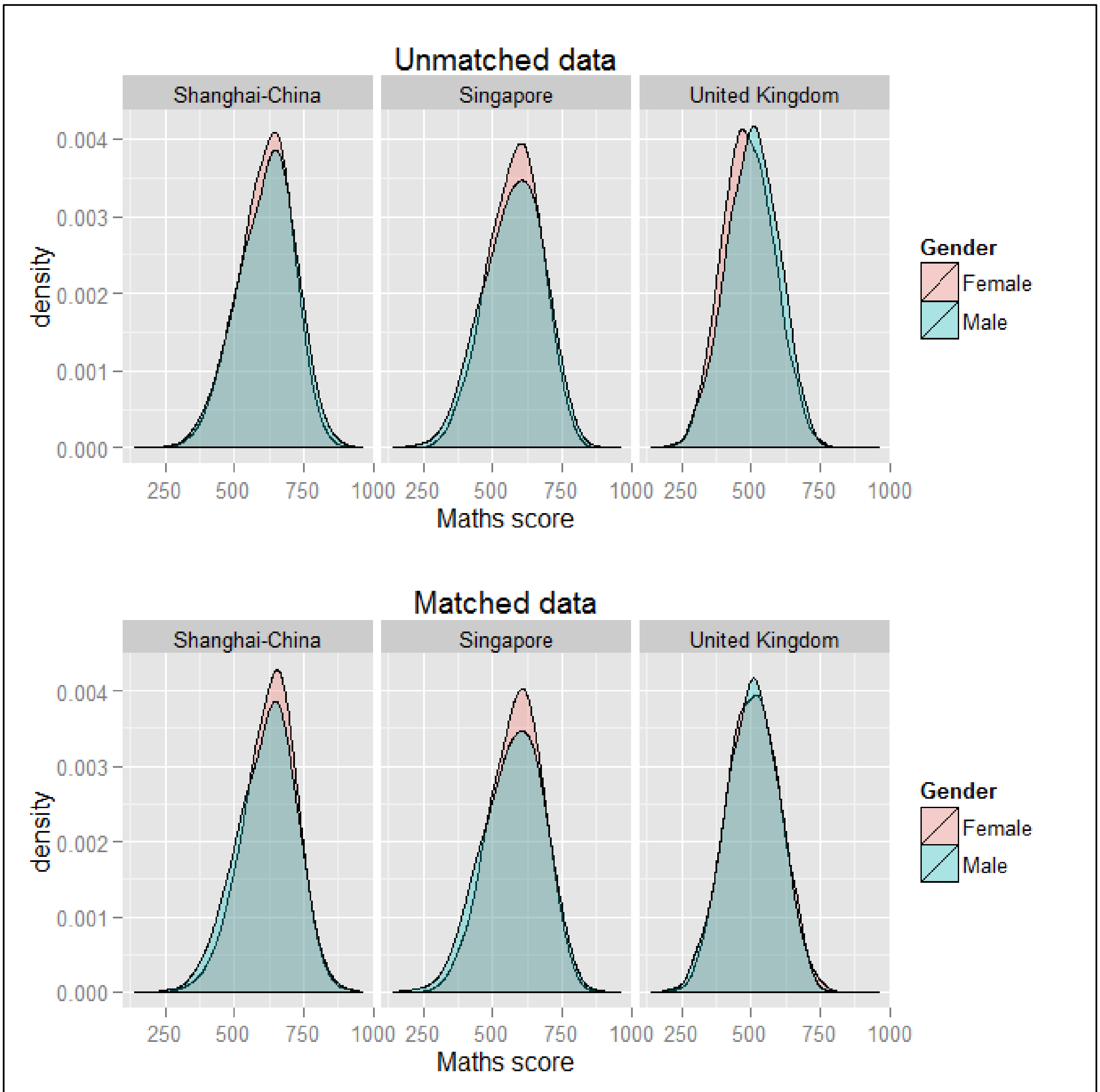

Figure 3.2: Distribution of mathematics performance in three jurisdictions for original sample (unmatched data) and for sample where girls respond in the same way to the statement “I am just not good at mathematics” as boys (matched data).

[13] In fact we estimate what their performance would be if they were equally likely as boys to give each of the four possible responses.

Using the same methodology, the effect of accounting for differences in mathematics confidence is shown across all OECD countries and several East Asian countries of interest in Table 3.2. On average across OECD countries, accounting for differences mathematics confidence reduces the gender gap from 13 points in favour of boys to 3 points. More thorough analysis in the OECD report (OECD, 2015, Table 3.6b), using fuller measures of confidence, anxiety and self-efficacy in mathematics, suggests that accounting for these factors changes the gender gap to being 4 points in favour of girls. The results from this analysis are reproduced in the final two columns of Table 3.2.

Country	Boys	Girls	Girls (matched)	Difference before matching	Difference after matching	Difference after accounting for efficacy, confidence and anxiety (OECD)	SE of adjusted difference (OECD)
Australia	510.9	498.2	513.6	12.7	-2.7	-12.9	2.9
Austria	520.6	497.0	507.1	23.6	13.5	2.7	5.1
Belgium	530.7	512.8	521.6	17.9	9.1	-9.3	4.5
Canada	526.1	514.5	526.9	11.6	-0.8	-5.7	2.3
Chile	436.8	410.6	419.4	26.3	17.4	9.9	4.4
Czech Republic	509.6	498.9	507.5	10.7	2.0	-6.2	4.6
Denmark	513.9	496.7	517.1	17.3	-3.2	-9.9	4.1
Estonia	524.0	517.3	524.5	6.6	-0.5	-7.8	4.0
Finland	523.6	522.3	536.6	1.3	-13.1	-20.5	3.3
France	503.3	492.8	505.4	10.5	-2.1	-17.2	4.5
Germany	531.0	515.5	526.9	15.4	4.1	-14.0	4.6
Greece	461.5	450.6	456.8	10.9	4.7	-3.4	4.2
Hungary	484.2	472.2	480.5	12.0	3.7	-5.4	5.4
Iceland	498.1	496.1	508.0	2.0	-9.9	-17.8	5.0
Ireland	510.1	492.1	504.1	18.0	6.0	7.1	4.5
Israel	482.2	465.9	468.7	16.3	13.5	7.9	6.1
Italy	496.5	477.1	481.8	19.4	14.7	3.9	2.7
Japan	546.9	526.6	533.6	20.3	13.2	5.0	3.5
Korea	562.6	544.9	553.6	17.7	9.0	-0.2	4.7
Luxembourg	502.1	478.2	489.2	23.9	12.9	3.1	4.0
Mexico	422.8	406.9	411.8	15.9	11.0	5.1	1.8
Netherlands	536.3	522.9	529.4	13.4	6.9	2.1	4.5
New Zealand	507.6	491.7	510.9	15.9	-3.2	-11.4	5.1
Norway	494.6	486.5	501.8	8.1	-7.3	-17.4	3.7
Poland	522.6	515.2	520.5	7.5	2.1	4.0	3.6
Portugal	497.4	482.5	488.9	14.9	8.5	3.6	3.0
Slovak Republic	489.8	480.0	490.7	9.8	-0.9	4.5	5.2
Slovenia	506.4	503.3	509.3	3.2	-2.9	-4.2	4.9
Spain	494.2	477.4	484.5	16.7	9.6	6.4	3.0
Sweden	482.9	481.6	495.4	1.3	-12.5	-20.1	3.9
Switzerland	539.0	525.1	538.8	13.9	0.1	-12.6	4.2
Turkey	453.8	444.6	446.3	9.2	7.5	4.0	5.1
United Kingdom	505.2	491.0	509.5	14.2	-4.4	-13.4	5.0
United States	486.6	481.8	488.0	4.8	-1.4	-3.2	4.4
OECD average	**503.4**	**490.3**	**500.3**	**13.0**	**3.1**	**-4.2**	**0.7**
Hong Kong-China	570.3	556.0	570.0	14.3	0.3	-6.1	5.9
Macao-China	541.4	538.0	553.0	3.5	-11.5	-10.4	3.4
Shanghai-China	614.7	610.5	626.1	4.2	-11.4	-12.2	4.9
Singapore	574.4	577.2	582.8	-2.9	-8.5	-10.1	4.6
Chinese Taipei	562.6	557.8	574.7	4.8	-12.1	-14.8	7.1

Table 3.2: Gender differences in mathematics performance before and after accounting for either responses to “I am just no good at mathematics” or more general measures of confidence, efficacy and anxiety.

Although the above results are interesting, as with the analysis of reading, we must be careful not to assume that this proves that differences in confidence are the cause of the gender difference. In particular it seems highly likely that high performance in maths causes people to exhibit confidence and become less likely to agree with statements such as “I am just not good at maths”. Indeed, the fact that, as shown above, if there were more girls who thought they were good at maths then girls would be better at maths is bordering on tautology. As such, it cannot be concluded from the PISA data that building girls’ self-confidence will necessarily directly translate into them outperforming boys in mathematics. To reinforce this point, it can be seen that in countries such as Finland and Singapore girls already perform at least as well as boys despite that the fact that they tend to exhibit lower levels of confidence.

To illustrate the possible reversed causality, another way to look at the link between achievement and confidence is shown in Figure 3.3. This chart shows how the probability of pupils agreeing that they are “just not good at maths” varies in Shanghai-China, Singapore and the UK depending upon the proportion of the PISA mathematics items they attempted that they answered correctly. Only pupils who attempted at least 20 mathematics items are included in this analysis. In all three countries pupils who answer more of the test correctly are less likely to believe they are not good at maths. However, for any given percentage of items correctly answered, girls are more likely than boys to say they are not good at maths. This is particularly true in Shanghai-China, where even if they have answered more than 90 per cent of items correctly, around 40 per cent of girls say they are just not good at maths. Given the same level of achievement, less than one in five boys in Shanghai-China agree with this statement.

Having said all of the above, some research referenced by the OECD (e.g. Marsh and Martin, 2011) indicates that academic self-concept has causal effects on subsequent achievement. This suggests that efforts to build girl’s confidence in mathematics may well be worthwhile. However it would be unwise to suggest that this will necessarily remove the gender gap in its entirety.

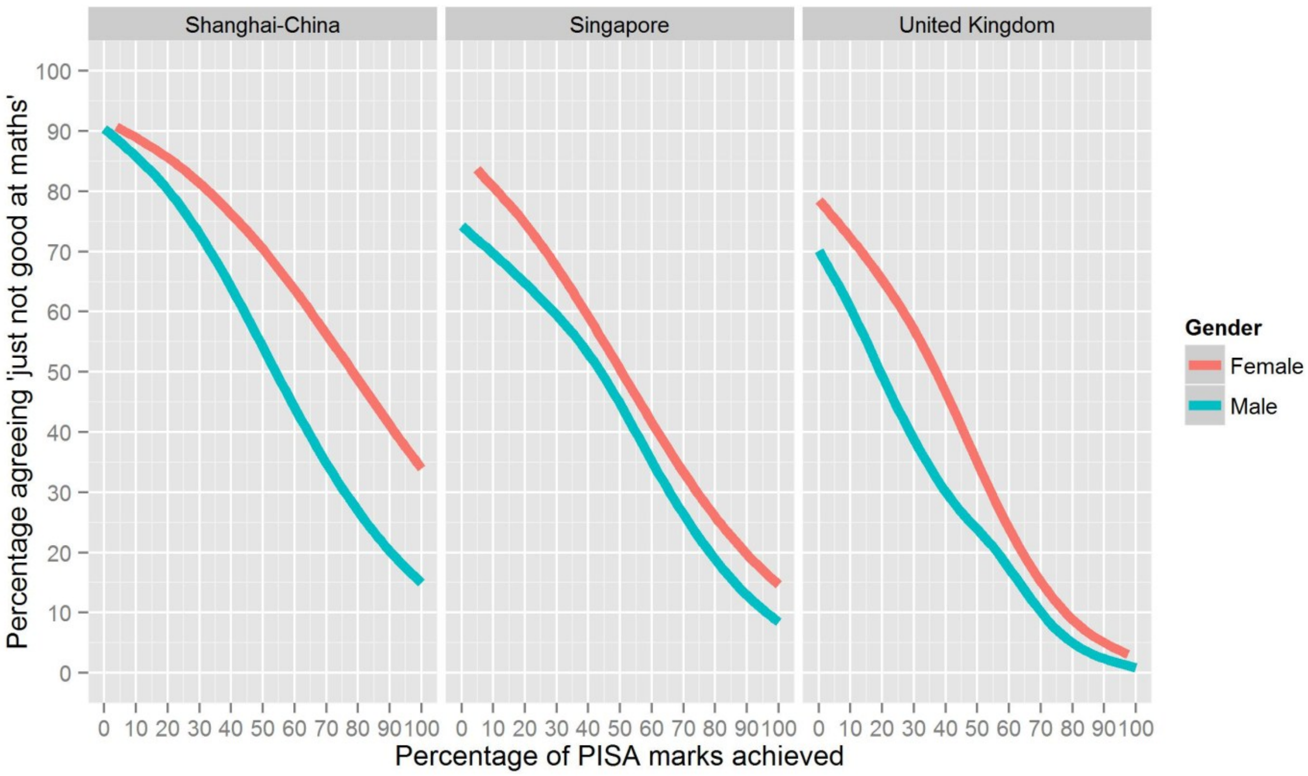

Figure 3.3: Probability of agreeing that “just not good at mathematics” given difference percentage of PISA mathematics item answered correctly in three jurisdictions.

4. Are girls better at some elements of science and boys at others?
As seen earlier, on average across OECD countries there is very little difference in the average performances of boys and girls in science. However, when the different elements of science measured within PISA are analysed separately some differences emerge. Specifically, in PISA 2006[14], on average across OECD countries, boys perform 17 points worse than girls in the scale i*dentifying scientific issues*, boys perform 15 points better on the scale *explaining phenomena scientifically*, and the two genders perform roughly equally on the scale *using scientific evidence* with a gap of just 3 points in favour of girls (OECD, 2015, Figure 1.6). The OECD report summarises these findings by saying that 'girls appear to underperform considerably when they are required to "think like scientists"' (OECD, 2015, page 64)[15]. This section examines this finding in a little more detail.

In order to examine this finding we take a closer look at the precise items being used to measure each of these scales. Only 23 of the science items from PISA 2006 are publicly available. The exact details of these questions can be found in OECD (2009b). These 23 released items are grouped within 8 broader questions. For example, a question about the greenhouse effect provides some brief text with background information on this effect, two graphs detailing global carbon dioxide emissions and global temperatures over time, and has three items asking pupils some scientific questions relating to this information and this topic more generally.

Table 4.1 provides highly simplified versions of what is asked by each question, which science scale the each item is associated with and the performance of boys on girls on average across OECD countries in PISA 2006. As would be expected given the findings in the OECD report:

- For 6 out of the 11 items in the e*xplaining phenomena scientifically* domain boys performed better than girls[16].
- For all 8 of the items in the i*dentifying scientific issues* domain girls performed better than boys.
- In the *using scientific evidence* domain, boys performed better on two of the items and girls performed better on the other two.

[14] PISA 2006 was the last occasion on which science was the major domain of study. As such, performance on the separate scales of science was not estimated in either PISA 2009 or PISA 2012.
[15] This OECD summary relates not only the findings in science but also to the fact that girls perform worse than boys on the mathematics scale *formulating situations mathematically* (OECD, 2015, Figure 1.5). However, the gender difference in this scale is only 5 points wider than the average gap for mathematics as a whole and is not considered in greater detail within this report.
[16] For one of the items the performance of boys and girls is within 1 percentage point and on 4 others girls' performance clearly exceeds boys.

Question subject	Question (simplified and shortened versions)	Associated science scale	% of students answering correctly (OECD)	
			Boys	Girls
	What is it about the two graphs [of global temperatures and CO2 emissions over time] that supports the conclusion [that the increase in the earth's temperature is due to CO2 emissions]?	Using scientific evidence	52.8	55.1
	Give an example of a part of the graph that does not support this conclusion.	Using scientific evidence	34.6	34.3
Greenhouse	Name one of the [other] factors [that could also influence global temperatures].	Explaining phenomena scientifically	18.7	19.1
	Can the following claims [about "intelligent" clothes] be tested scientifically (Yes/No for each)? The material can be: washed without damage, wrapped around objects without damage, scrunched up without damage, produced cheaply?	Identifying scientific issues	45.1	50.7
Clothes	Which piece of equipment would you need to check that the fabric conducts electricity? (Voltmeter/Light box/Micrometer/Sound meter)	Explaining phenomena scientifically	81.7	77.0
	How do temperature changes and water help to speed up the breakdown of rocks?	Explaining phenomena scientifically	68.3	66.9
	What happened millions of years ago that explains why there are many fossils in [a Limestone layer of] the Grand canyon? (Multiple choice)	Explaining phenomena scientifically	76.9	74.6
The Grand Canyon	Can the following questions be answered by scientific investigation (Yes/No for each): Amount of erosion caused by use of walking tracks, is the park as beautiful as 100 years ago?	Identifying scientific issues	59.9	62.8
	[Description of experiment testing effectiveness of sunscreens]. Which of the following statements is a scientific description of the role of mineral oil and zinc oxide in comparing the effectiveness of sunscreens? (Multiple choice)	Identifying scientific issues	38.1	43.0
	Which one of these questions was [the experiment] attempting to answer? (Multiple choice)	Identifying scientific issues	56.3	60.3
	Why was the second sheet of plastic [in the experiment] pressed down? (Multiple choice)	Identifying scientific issues	41.7	44.3
Sunscreens	Which diagram (of four shown) displays the likely result of the experiment?	Using scientific evidence	25.8	28.4
	What kinds of diseases can be vaccinated against? (Multiple choice)	Explaining phenomena scientifically	72.6	77.2
	Why does a particular type of bacteria only tend to make a person sick just once?	Explaining phenomena scientifically	75.3	74.9
Mary Montagu	Why should young children and old people in particular be vaccinated against the flu?	Explaining phenomena scientifically	58.5	65.0
	Where do sulphur oxides and nitrogen oxides in the air come from?	Explaining phenomena scientifically	59.8	55.5
	What will the mass of a marble chip be after it has been left in vinegar overnight [compared to its original mass]? (Multiple choice)	Using scientific evidence	68.3	65.1
Acid Rain	Why should such an experiment also include leaving marble chips in water overnight?	Identifying scientific issues	33.5	37.6
	Which of the following are advantages of regular exercise (Yes/No for each one): Prevent heart and circulation illnesses, leads to healthy diet, avoids becoming overweight?	Explaining phenomena scientifically	51.9	53.5
	Which of the following happens when muscles are exercised (Yes/No for each one): Muscles get an increased flow of blood, Fats are formed in the muscles?	Explaining phenomena scientifically	81.7	83.1
Physical Exercise	Why do you have to breathe more heavily when you are doing physical exercise?	Explaining phenomena scientifically	49.3	40.8
	[Description of experiment testing effect of GM crops on insect populations]. Which of the following factors were varied in the experiment (Yes/No for each one): The number of insects in the environment, the types of herbicides used?	Identifying scientific issues	58.5	63.5
Genetically Modified Crops	Why did the scientists use more than one site in the study? (Multiple choice)	Identifying scientific issues	72.7	74.4

Table 4.1: Brief descriptions of released items from PISA 2006 with performance in OECD countries.

When examined in this way therefore, the biggest gender gap appears to be in the scale *identifying scientific issues*. Close inspection of the items in this scale reveals that most of them require careful reading of the description of an experiment and then correctly identifying what factors have been manipulated, why this has been done and what further experimental work could be undertaken. To a small extent, girls' high performance may be attributable to their higher performance in reading generally, since careful reading of the provided text will help in answering some of these questions. However, the high performance also indicates that girls are showing higher ability than boys on average to carefully plan experiments and understand the scientific method. In this sense there is no doubt that girls are showing plenty of ability to "think like scientists".

The (marginal) underperformance of girls is seen in the domain *explaining phenomena scientifically*. Inspection of the released items suggests that many of these items simply test pupils' recall of scientific facts. For example, the largest gender difference comes in a question on physical exercise asking why exercise causes people to breathe more heavily[17]. This item simply requires pupils to demonstrate their knowledge of the purpose of breathing and boys are 9 percentage points more likely to answer this question correctly. Similarly, the second biggest difference in favour of boys occurs in an item that simply requires pupils to identify which piece of equipment (out of a possible four) can be used to detect an electric current.

It is hard to know the extent to which the 11 released items in this domain are representative of all items in this domain in PISA 2006. However, from what has been released it appears that the majority of girls' underperformance in science is within their ability to recall relevant scientific facts under (low-stakes) examination conditions rather than from any inability to "think like scientists".

References

Marsh H.W. and Martin, A.J. (2011). Academic self-concept and academic achievement: Relations and causal ordering, *British Journal of Educational Psychology*, 81, 59-77.

OECD (2009a). *PISA 2009 Assessment Framework: Key competencies in reading, mathematics and science*, PISA, OECD Publishing. http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/pisa2009assessmentframework-keycompetenciesinreadingmathematicsandscience.htm.

OECD (2009b). *Take the test: Sample Questions from OECD's PISA Assessments*, PISA, OECD Publishing. http://www.oecd-ilibrary.org/education/take-the-test_9789264050815-en.

OECD (2010). *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I)*, PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264091450-en.

OECD (2013). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I)*, PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264201118-en.

OECD (2015). *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*, PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264229945-en.

[17] Saying either the need to provide additional oxygen to the body or the need to remove carbon dioxide from the body are both deemed correct responses.