Early Gender Test Score Gaps Across OECD Countries

Kelly Bedard
Department of Economics
University of California, Santa Barbra

Insook Cho
School of Technology Management
Ulsan National Institute of Science and Technology

September 2009

Abstract

The results reported in this paper contribute to the debate about gender skill gaps in at least three ways. First, we document the large differences in early gender gaps across developed countries using a large scale, modern, representative data source. Second, we show that countries with pro-female sorting, countries that place girls in classes with higher than average scores have smaller gender test score gaps, at least in math. Third, we show that the degree of academic tracking is correlated with observed gender gaps across developed countries.

We thank Heather Antecol, Michael Conlin, Bill Fischel, Brian Krauth, Peter Kuhn, Krishna Pendakur, Cathy Weinberger, Simon Woodcock and seminar participants at the UCSB labor lunch, University of Kentucky, Michigan State University, Simon Fraser University, and the 2007 SOLE meetings for helpful comments.

1. Introduction

Social scientists have long tried to understand the underlying forces that led to a substantial under representation of women in science and engineering fields (see Weinberger 2005 for an overview) and its impact on the gender pay gap among college graduates (Eide 1997; Brown and Corcoran 1997; Weinberger 1998 and 2001; Weinberger and Kim 1999). Since math and science skills are highly valued in the economy, it seems important to understand the origin of any skill gap that might lead to differential education and labor market choices across genders. Unfortunately, the root causes of math/science gender skill-gaps observed during adolescence and adulthood are likely a complex combination of biology and the environment from birth onward. In other words, innate math/science ability, family environment, teacher interactions and evaluations, peer interactions, student expectations about the importance of math and science skills, and class/program/stream assignments[1] are all embedded in test scores. This complex set of interactions makes it very difficult to isolate the role played by any particular component.

While in theory instrumental variables and/or natural experiments offer potential ways forward, convincing examples are difficult to come by. A more descriptive approach using internationally comparable test scores to examine differences in math and science gender gaps across countries is therefore a useful first step that may give us some insight into the possible roles played by culture and institutions. The primary objective of this paper is to document early gender test score gaps across OECD countries to see what we can learn from the observed differences. For example, we examine the relationship between observed gender gaps and proactive policies that place girls in higher achieving math and science classes. In a similar vein, we also explore the relationship between other educational structures, such as same-sex education and educational streaming, and early gender test score gaps across OECD countries.

2. Existing international evidence

Depending on the test, time-period, and country, researchers have estimated female-male math test score gaps ranging from -39.5 to 4.6 (see Appendix Table 2). The lack of agreement regarding the size of the gender test score gap is evident even if one restricts attention to the United States. Using Early Childhood Longitudinal Study (ECLS) data, Freeman (2004) finds that boys and girls have similar math scores at the end of grade one but that by grade three boys

[1] Streaming refers to students being assigned to different educational tracks (e.g. college, trade school, etc.).

out score girls by three percentage points. In contrast, using 1999 National Assessment of Educational Progress (NAEP) data, Dee (2007) finds no evidence of a gender gap in math or science among nine year olds. He does however find a statistically significant male premium in science among thirteen year olds. Using 1996 NAEP data, Coley (2001) estimates a male score advantage for fourth graders and a male science advantage among eighth and twelfth graders. Finally, using six data sets collected between 1960 and 1992,[2] Hedges and Nowell (1995) find that boys perform slightly better in math and science than girls.

The gender gaps in math and science are not, of course, limited to the United States. However, just as in the United States, gender gaps in other countries appear to have complicated patterns. For example, Kaur (1990) reports that 16-year-old Singaporean boys outperform girls in O-level math. In contrast, Lavy (2008) finds that Israeli girls in their final year of high school outperform boys in math and science. Lummis and Stevenson (1990) conducted math tests in Taiwan, Japan, and the United States. Their general finding is that there are few gender differences in grade one and only a few small male advantages in cognitive mathematics tasks by grade five. Further, the gender gaps that exist are consistent across countries. They therefore argue that culture has little effect on the gender math gap. In a similar vein, Engelhard (1990) finds a similar gender math score gap in the U.S. and Thailand.

While the studies discussed above focus on at most a small number of countries, there have also been several large-scale international testing exercises in math and science in recent years. These include the First International Mathematics Study (FIMS) and the First International Science Study (FISS) conducted in 1964 and 1971, the Second International Mathematics Study (SIMS) and the Second International Science Study (SISS) conducted in 1981and 1984, the Trends in International Mathematics and Science Study (TIMSS) conducted in 1995, 1999, and 2003, and the Program for International Student Assessment (PISA) conducted in 2000. Gender gap estimates for these studies are reported in Harnisch et al. (1986), Keeves (1973), Hanna, Kundiger, and Larouche (1990), Postlethwaite and Wiley (1992), Mullis et al. (2000), and OECD (2001), respectively.[3] In general, these studies find a small gender gap favoring boys across most participating countries (see Appendix Table 2).

[2] They use Project Talent (1960), National Longitudinal Study (1972), National Longitudinal Study of Youth (1979), High School and Beyond (1980), National Educational Longitudinal Study (1988), and National Assessment of Educational Progress (1977-1992) data.
[3] Notice that most of these studies come directly from the testing agency’s user guide.

Our work diverges from other examinations of math and/or science gender test score gaps in several ways. First, we use more recent data from the 1995, 1999, and 2003 Trends in International Mathematics and Science Study (TIMSS). Second, we use a representative sample of developed countries that includes all OECD participants who report the required information (see Section 3 for more detail). Three, we examine test score gaps among fourth and eighth graders. This avoids important selection differences countries during later grades. Finally, we explore the possibility that educational institutions may contribute to observed differences in gender test score gaps across countries.

3. Data

The data used in this study come from the 1995, 1999, and 2003 Trends in International Mathematics and Science Study (TIMSS). TIMSS provides information about math and science test scores and students' educational and socioeconomic background. TIMSS surveys two groups of students, third and fourth graders in 26 countries in 1995 and 2003 and seventh and eighth graders in 41, 38, and 47 countries in 1995, 1999, and 2003, respectively. We restrict the sample to OECD countries with close to universal school participation in grade eight. Turkey is eliminated because a sizable minority of girls leave school before grade eight. The only other exclusion is Korea in 1995. This exclusion is necessary because the data appear to be flawed; the male-female ratio is unbelievably different in the grade seven and eight samples in 1995. These exclusions leave us with a sample of 18 countries for third and fourth graders and 26 countries for seventh and eighth graders, and a sample of 445,835 observations across all ages and countries.[4] Table 1 reports the country and grade specific sample sizes.

TIMSS tests two groups of students. The 1995 and 2003 TIMSS includes test scores for two different grade groups. The first set of scores is for students enrolled in the two adjacent grades that contain the largest proportion of nine year olds – third and fourth graders in most countries. For expositional ease, we refer to these students as fourth graders. The second set of scores is for students enrolled in the two adjacent grades that contain the largest proportion of thirteen year olds – seventh and eighth graders in most countries. We refer to these students as eighth graders. In contrast, the 1999 TIMSS includes only one age group in a single grade. While the 1999 TIMSS uses the 1995 definition to target the two adjacent grades containing the

[4] The 38,195 students who do not report their sex, test scores, and age are also excluded.

most thirteen year olds, only students in the upper of the two grades were tested – eighth graders in most countries. We again refer to these students as eighth graders.

The TIMSS test scores used in all analyses are standardized within test book across all TIMSS participants to mean 50 and a standard deviation 10. Summary statistics are reported in Table 1 by country. As one would expect, the country-specific internationally standardized mean scores are generally above 50 because we are focusing on OECD countries.

All test score models include a basic set of socioeconomic controls. These include indicator variables for sex, grade, test year, native-born mother, native-born father, child living with both parents, child has a calculator, child has a computer, child has more than 100 books, and parental education[5] (in eighth grade models only),[6] and a continuous measure for the number of people residing in the child's household. Unfortunately, some of the socioeconomic controls suffer from substantial non-reporting. As we do not want to lose observations due to missing socioeconomic information, we replace the missing control variable observations with zeros and include a set of missing data indicators. In addition to the basic set of control variables that are included in all models, the class fixed effects specification includes teacher/class indicators. More specifically, students are defined as being in a specific math (science) class if they have the same set of math (science) classes with the same teachers in the same class periods. In most countries this is fairly simple because most students in a specified homeroom are with the same set of students for math and science, but in some countries students from a single homeroom class are in several different math and science classes, the U.S. is a good example.

4. The gender gap in math and science

We begin with a simple descriptive model of the relationship between gender and test scores.

$$S_{cgi} = \alpha_{cg} + \beta_{cg} F_{cgi} + X_{cgi}\gamma_{cg} + \varepsilon_{cgi} \qquad (1)$$

where S_{cgi} denotes the test score, for student *i* in country *c* in grade *g*, *F* is a female indicator, *X* is the vector of controls described in Section 3, and ε is the usual error term.[7] All models are estimated separately for each grade, subject, and country.

[5] We have collapsed maternal and paternal education into three categories in order to make them comparable across test years. The collapsed groups are: high school dropouts, college graduates, and all other education levels.
[6] Parental education is not reported for fourth graders in any country or eighth graders in England and Japan.
[7] Alternatively, we could allow all coefficients to vary by gender and then use an Oaxaca (1973) decomposition to isolate the unexplained part of the gender gap. However, we prefer the simpler approach described by equation (1)

The gender gap (female-male) estimate from equation (1) is only an unbiased estimate of the innate gender difference if all omitted factors are uncorrelated with gender. At a minimum, this implies that the gender gap estimate obtained from equation (1) is a combination of innate gender-specific ability differences, and parental, teacher, and peer interaction differences across boys and girls. In order for these to be the only factors included in β_{cg}, educational opportunities must be uncorrelated with gender. In particular, the assignment rules used to place children in classes or streams must be gender neutral. On the surface this seems like a reasonable assumption, but reality may be quite different. In countries that sort students into ability-based streams using teacher evaluations, gender-biased ability assessments may lead to gender-specific streaming rules – even if teachers themselves do not realize that they are doing so. This is, of course, in addition to any overt gender-biases that might exist in program placement decisions in some countries.

In contrast to the usual practice of discussing the results in ascending grade order, we first discuss the eighth grade results and then come back to the fourth grade results. The reason for the peculiar order will become clear shortly.

4.1. OLS results for grade eight

Columns 1 and 3 in Table 2 report the OLS estimates for equation (1) for math and science, respectively. For interpretive ease, columns 5 and 7 report the same results using the OECD percentile score.[8] Given the easier interpretation of the percentile scores, the text focuses on these results. Examination of columns 5 and 7 reveals three important facts. First, eighth grade boys outscore eighth grade girls in math and science in most OECD countries. The average gender test score gap (female-male) is –2.2 percentiles in math and –6.0 percentiles in science. These averages reveal the second fact: The gender test score gap is much bigger in science than in math. Third, the magnitude of the gender test score gap varies substantially across countries. In fact, the gender test score gap even differs across sub-sets of countries that one might have thought would be similar – Canada/U.S. and Finland/Norway/Sweden are good examples. This is an important finding as it suggests that educational structures may affect gender differentials.

because the male-female mean differences are so small that almost the entire gender gap is unexplained (due to coefficient differences rather than mean differences).

[8] These are approximated using the unweighted ranking (0 being the lowest and 100 being the highest) of standardized scores across the OECD sample used in the analysis.

4.2. Class fixed effects for grade eight

While it is impossible, given the available data, to purge the gender gap estimates of the bias induced by differential parental and teacher behavior towards girls and boys that encourages differential success rates in math and science, we can control for differential class assignment, at least to the extent that it is captured by current class assignment. However, as will become clear shortly, this may not be the right approach given the apparent wide spread use of gender-biased class sorting and its relationship with the observed (OLS or raw) gender gap. That being said, we proceed to estimate class fixed effects specifications in order to more fully understand this process. More specifically, we estimate the following fixed effects model:

$$S_{cgti} = \phi_{cgt} + \beta_{cg}^{FE} F_{cgti} + X_{cgti}\gamma_{cg}^{FE} + \nu_{cgti} \qquad (2)$$

where S_{cgti} denotes test score, for student *i* in country *c* in grade *g* in class (with teacher) *t* and ϕ_{cgt} is a vector of class indicators.

The class fixed effects results for the OECD math and science percentile scores are reported in columns 6 and 8. Focusing first on the math results, in all but five cases the fixed effects estimates are more negative than the OLS (non-fixed effects) results, and in eleven cases the difference is statistically significant at the 5 percent level. The most extreme examples are Flemish Belgium, Germany, and the Netherlands, all of which have fixed effects gender gap estimates that are more than 2 percentage points more negative than the corresponding OLS estimate. At the other end of the spectrum, seven countries have OLS and fixed effects estimates that are effectively identical – within 0.2 percentiles of each other. These countries include Denmark, Finland, Greece, Italy, Japan, Norway, and Spain.

Comparing the OLS and fixed effects results for math raises two important questions. First, why are the fixed effects estimates almost uniformly more negative than the OLS estimates? Second, why does the difference between the OLS and fixed effects estimates vary so much across countries? Gender-biased sorting across classes and/or academic programs appears to be an important part of the answer to both questions.

The easiest way to see this is to compare the degree of gender-biased sorting to the difference between the OLS and fixed effects estimates. We construct a simple measure of gender-biased sorting by regressing class rank on a female indicator.

$$R_{cgti} = \theta_0 + \theta_1 F_{cgti} + \theta_2 X_{cgti} + \nu_{cgti} \qquad (3)$$

where R_{cgti} denotes class rank for student *i* in country *c* in grade *g* in class *t*. Classes are ranked from 0 (the class with the lowest average score) to 1 (the class with the highest average score).[9] $\theta_1 = 0$ if, on average, male and female students are placed in equally ranked classes. If, on the other hand, girls are placed in lower than average classes $\theta_1 < 0$ and if girls are placed in better than average classes $\theta_1 > 0$.

Table 3 reports the equation (3) estimates. We begin by focusing on the mathematics results. What is, at first glance, somewhat surprising is the frequency of positive and statistically significant female coefficients (θ_1). Three countries have negative and significant female coefficients (girls are assigned to worse than average classes), thirteen countries have statistically insignificant female coefficients (gender-neutral class assignment), and twelve countries have positive and statistically significant female coefficients (girls are assigned to better than average classes). However, one should be cautious when interpreting these coefficients for countries with a sizable fraction of students in same-sex classes since sorting may be very different in nature in these cases. The most extreme examples are Ireland and Korea, where only 51 and 39 percent of students are in gender-mixed classes respectively.

The relationship between pro-female sorting and the difference between the fixed effects and OLS estimates is graphed in Figure 1. The x-axis is the differential female class assignment by class rank reported in Table 3. The y-axis is the difference between the fixed effects estimates and the OLS estimates reported in Table 2. Panel A plots the relationship for math and Panel B plots the relationship for science. Finally, to give the reader a sense of the precision of the gender-sorting measure, the circles in all graphs are an increasing function of the t-statistic on gender from equation (3).[10]

The negative slope depicted in Figure 1 indicates that countries that place a greater percentage of girls in higher scoring classes have more negative fixed effects estimates compared to their OLS estimates. In other words, the greater the degree of pro-female sorting (or class placement), the worse the within class relative performance of girls compared to boys. For example, the five most rightward circles in panel A are Flemish Belgium, Germany, Hungary,

[9] The results are similar if classes are ranked using average male scores instead of overall average scores.
[10] Circle size is a function of the gender sorting t-statistic rather than for the fixed effects or OLS gender coefficients from equations (1) or (2), since the OLS and fixed effects estimates are quite precise.

the Netherlands, and Portugal. A closer look at Panel A further reveals that most countries have both a bigger fixed effects gender gap estimate than an OLS estimate and pro-female sorting.[11]

In summary, countries that disproportionately place girls in better classes have smaller gender gaps. Stated somewhat differently, in the absence of pro-female class assignment the average gender gap in many countries might be substantially larger. This is a surprising finding for anyone who's intuition or casual observation of the world leads them think that the class-sorting in highly ability streamed countries favors boys, but it is consistent with Lavy's (2008) finding that Israeli teachers award higher grades to girls.

Thus far, we have focused on the eighth grade fixed effects gender gap in math. While the patterns that we have discussed are almost all equally applicable to science, there is one substantive difference between math and science: The science gap is generally much larger than the math gap. On average, the science gap is 3.7 percentiles more negative than the math gap. This is a large difference given an average math gap of -3.1 percentiles. As we will see in the next section, this is interesting in light of the fact that the math and science gaps are of a much more similar magnitude in grade 4.

4.3. Grade four

Table 4 replicates Table 2 for fourth graders. For interpretive ease we again focus on the results using the percentile scores reported in columns 5-8. Similar to the eighth grade results, fourth grade boys have higher math and science scores than fourth grade girls in almost all OECD countries whether we look at the OLS or fixed effects (FE) estimates. Also similar to the eighth grade results, the size of the gender gap varies substantially across countries, although to a lesser extent than in grade eight. In contrast to the eighth grade results, the science gender gap is only 1 percentage point larger than the math gap. Further, the similar gap size across math and science is entirely the result of a much smaller science gap at the fourth grade level. The average OLS (FE) math gap is -2.3 (-2.3) at the fourth grade level and -2.2 (-3.1) at the eight grade level compared to a -3.2 (-3.3) OLS (FE) science gap in grade four and -6.0 (-6.8) in grade eight.

The final, and perhaps most striking, feature of Table 4 is the fact that the OLS and fixed effects estimates are much more similar for grade four. More specifically, the difference

[11] In contrast, Hallinan and Sorensen (1987) find that boys are more often assigned to a high-ability group, but they find little evidence that this effects math achievement.

between the OLS and fixed effects estimates is less than one percentile in all but two countries: The Czech Republic and Ireland. The reason for the similar OLS and fixed effects estimates is easily seen by examining the pro-female sorting results for grade four reported in Table 3. Columns 5 and 6 in Table 3 report the coefficients for female indicator in equation (4) for math and science. In math, the female coefficient is negative (girls are assigned to worse than average classes) in three countries, positive (girls are assigned to better than average classes) in three countries, and statistically insignificant (gender-neutral class assignment) in all other countries. The results are similar for science: The female coefficient is negative in three countries, positive in one country, and statistically insignificant in all other countries. Overall, fourth grade class assignment appears to be gender neutral in the vast majority of OECD countries. This is easy to see in Panels C and D in Figure 1. In contrast to the eighth grade results graphed in Panels A and B, in the fourth grade panels most of the data points are located in close proximity to zero. The downward slope is preserved however because countries with gender-biased sorting follow the same pattern as before; positive female sorting is associated with a bigger FE-OLS gap and negative female sorting is associated with a smaller FE-OLS gap.

5. Understanding the gender gap

Even if one begins with the working hypothesis that boys are innately better at math and science, unless the underlying innate skill distributions differ substantially across OECD countries, which seems unlikely, other factors must be driving the observed variation in gender test score gaps across countries. In other words, innate gender differences can generate a female-male test score gap, but cannot explain the variation in gaps observed across OECD countries. As such, differences in the structure of the education systems, economies, or cultures across OECD countries must play important roles. However, attempts to explain differences across countries are plagued by omitted variables bias due to the impossibility of controlling for all cross-country differences. As a result, any exploration of the factors that may contribute to the observed variation in gender gaps across countries should be viewed as descriptive rather than causal. In other words, this is a descriptive exercise with the objective of revealing correlations between gender gaps and educational institutions across developed countries.

In this vein, we investigate the possibility that the OLS gender gaps ($\hat{\beta}_{cg}$), reported in columns 5 and 7 in Tables 2 and 4 are correlated with the structure of education systems using a simple descriptive model. While it is impossible to fully describe an education system using a small number of variables, we include several important characteristics that are publically available and comparable across countries.

$$\hat{\beta}_{cg} = \pi_g + \delta_1 E_{cg} + \delta_2 \hat{\theta}_{cg} + \delta_3 M_{cg} + \delta_4 \bar{S}^m_{cg} + \upsilon_{cg} \quad (4)$$

where E measures the degree of ability streaming, $\hat{\theta}$ is the degree of pro-female sorting from equation (3), M is the fraction of students in mixed gender classes, and $\bar{S}^m$ is the average male test score. As there is no perfect definition or measure for the ability streaming (or tracking), we use three alternative measures: the percent of tenth graders enrolled in the academic stream, the grade when streaming begins, and the percentage of people aged 25-34 who have a university/ tertiary degree (these variables are reported in Appendix Table 1). Including the average male test score allows for the possibility that higher scoring countries may have larger or smaller gender gaps. All equation (4) estimates are weighted by the inverse sampling variance of the left-hand side variable from equation (1).

The base specification estimates for eighth graders are reported in columns (1) and (5), for math and science respectively, in the top panel of Table 5. A 10 percentage point larger fraction of students enrolled in the academic stream in grade ten is associated with female-male test score gaps that are 0.18 smaller for math and 0.23 smaller for science. To put these numbers in perspective, given a streaming standard deviation of 0.3 and β_{cg} standard deviations of 1.4 for both math and science, a one standard deviation larger academic stream size is associated with a female-male test score gap that is 0.4 standard deviations smaller for math and 0.5 standard deviations smaller for science. Hanushek and Wobmann (2006) similarly find that countries that stream at early ages have greater educational inequality. In a similar vein, countries that place girls in better math classes also have smaller female-male gaps. More specifically, a one standard deviation higher $\hat{\theta}$ is associated with a math gap that is 0.5 standard deviations smaller. In contrast, pro-female sorting in science is not statistically or economically related to the science gender gap.

Columns 2 and 6 add other educational structure and economic variables to check the robustness of the results.[12] Public expenditures on education as a fraction of GDP and private

[12] Public expenditures on education and the female-male university enrollment ratio are from *Education at a Glance* (2004), private school enrollment rates and the percentage of female teachers at the secondary level from the *Global*

school enrollment at the secondary level are included to isolate streaming from other aspects of educational 'quality' or structure. The fraction of secondary teachers who are female is intended to capture the impact of differential school performance by girls taught by women versus men. However, it is also possible that this variable also measures the fraction of women in math and science professions. The female-male university enrollment ratio is included to control for the impact of differences in female-male expectations about the probability that they will go onto university. The female labor force participation rate and GDP per capita are included to control for economic and labor market differences across countries. While the magnitudes of the point estimates of interest are similar whether or not these additional variables are included, the science point estimates become less precise.

The remaining columns in Table 5 include the complete set of regressors used in columns 2 and 5, but use alternate streaming measures. Columns 3 and 7 replace the percent of students in the academic stream in grade 10 with the grade at which formal streaming first occurs (this ranges from grade 4 to 12). While the magnitude of the coefficient differs, this simply reflects a difference in the scale of the streaming measure. Similar to previous columns, a one standard deviation older age at which streaming occurs is associated with a 0.6 standard deviation smaller female-male test score gap in math and a 0.3 standard deviation smaller gap in science. Columns 4 and 8 measure streaming by the percentage of the population who complete university or tertiary training (this ranges from 12 to 51 percent). Again, a one standard deviation larger percentage of people completing tertiary training is associated with a 0.6 standard deviation smaller gender gap in math and a 0.5 standard deviation smaller gap in science.

Perhaps more interesting than the finding that more heavily streamed countries tend to have bigger female-male test score gaps at the eighth grade level, is the finding that the same is true in grade four, long before streaming occurs in most countries. The bottom panel in Table 5 reports the same set of results for grade four. The primary finding at the fourth grade level is that the relationship between streaming and the gender gap is precisely estimated and of a similar magnitude to grade eight. Keep in mind, however, that the included countries differ across grades. Based on the results reported in columns 2-4, a streaming level that is one standard deviation higher is associated with a female-male math test score gap that is 0.6, 0.8, and 1.1

Education Digest (2003), female labor market participation rates from the *Yearbook of Labour Statistics* (2001), and Gross Domestic Product (GDP) per capita from the *Human Development Report* (1993-2000).

standard deviations smaller for the three streaming measures, respectively. Similarly for science, a one standard deviation reduction in streaming is associated with gender gaps that are 0.8, 0.7, and 1.0 standard deviations smaller for the three streaming measures. The fourth grade results may mean that if girls (or the parents of girls) believe that they are unlikely to participate in advanced math and science classes, or a career requiring advanced math or science skills, they may invest (or encourage) less effort in math and science even before formal streaming occurs (Catsambis 1994). Further, teachers may disproportionately encourage boys to take advanced math and science classes, which similarly reduces girls' expectations about their need for math and science and hence leads to reduced effort prior to formal streaming. While it is impossible to sort out the specific aspects of streaming that might cause or exacerbate a gender gap even before streaming occurs the reported results clearly show that such a relationship exists.

Overall, the results point to a substantial correlation between streaming and the gender test score gap at young ages. Further, since the relationship between streaming and the female-male gap arises before formal streaming occurs it likely works through indirect channels, such as family/teacher/peer interactions or student perceptions about the importance of math and science. There is also some evidence that pro-female sorting reduces the gender gap, at least in math.

6. Conclusion

Despite the fact that boys out score girls in math and science in almost every OECD country, uncovering the mechanisms behind the gap has proven incredibly difficult. The results reported in this paper contribute to the debate about gender skill gaps in at least three ways. First, we document the large differences in early gender gaps across developed countries using a large scale, modern, representative data source. Second, we show that countries with pro-female sorting, countries that place girls in better than average classes, have smaller gender test score gaps, at least in math. Third, we show that streaming/tracking is also correlated with observed gender gaps across developed countries.

References

Brown, Charles, and Mary Corcoran, “Content and the Male-Female Wage Gap,” *Journal of Labor Economics*, Vol.15, No.3 (1997), 431-465.

Catsambis, Sophia, “The Path to Math: Gender and Racial-Ethnic Differences in Mathematics Participation from Middle School to High School,” *Sociology of Education*, Vol.67, No.3 (1994), 199-215.

Coley, Richard J, *Differences in the Gender Gap: Comparisons across Racial/Ethnic Groups in Education and Work*, Educational Testing Service, Princeton (2001).

Dee, Thomas S, “Teachers and the Gender Gaps in Student Achievement,” *Journal of Human Resources*, Vol. 42, No. 3 (2007), 528-554.

Eide, Eric, “Accounting for Race and Gender Differences in College Wage Premium Changes,” *Southern Journal of Economics*, Vol. 63, No. 4 (1997), 1039-1050

Engelhard, George, “Gender Differences in Performance on Mathematics Items: Evidence from USA and Thailand,” *Contemporary Educational Psychology*, Vol.15 (1990), 13-16.

EURYDICE, European *Glossary on Education: Examinations, Qualifications, and Titles*, Eurydice, Brussels (1999), http://www.eurydice.org.

Freeman, Catherine E, *Trends in Educational Equity of Girls and Women: 2004* (NCES 2005-016), U.S. Department of Education, National Center for Education Statistics, Washington, DC: U.S. Government Printing Office (2004).

Hallinan, Maureen T., and Aage B. Sorensen, “Ability Grouping and Sex Differences in Mathematics Achievement,” *Sociology of Education*, Vol.60, No.2 (1987), 63-72.

Hanna, Gila, Erika Kundiger, and Christine Larouche, “Mathematical Achievement of Grade 12 Girls in Fifteen Countries,” *Gender and Mathematics: An International Perspective*, Edited by Leone Burton, Cassell Educational Limited (1990).

Hanushek, Eric A, and Ludger Wobmann, “Does Educational Tracking Affect Performance and Inequality? Difference-In-Difference Evidence Across Countries,” *Economic Journal*, Vol. 116, No. March (2006): C63-C76.

Harnisch, Delwyn L., Marjorie W. Steinkamp, Shiow-Ling Tsai, and Herbert J. Walberg, “Cross-national Differences in Mathematics Attitude and Achievement among Seventeen-year-olds,” *International Journal of Educational development*, Vol.6, No.4 (1986), 233-244.

Hedges, Larry V., and Amy Nowell, “Sex Differences in Mental Test Scores, Variability, and Number of High-Scoring Individuals,” *Science*, Vol.269, No.5220 (1995), 41-45.

ILO, *Yearbook of Labour Statistics: 2001*, International Labour Organization, Geneva, Switzerland (2001).

Jacob, Brian A., "Where the Boys Aren't: Non-Cognitive Skills, Returns to School and the Gender Gap in Higher Education," *Economics of Education Review*, Vol.21 (2002), 589-598.

Kaur, Berinderjeet, "Girls and Mathematics in Singapore: The Case of GCE "O" Level Mathematics," *Gender and Mathematics: An International Perspective*, Edited by Leone Burton, Cassell Educational Limited (1990).

Keeves, John, "Differences between the Sexes in Mathematics and Science Courses," *International Review of Education*, Vol.19, No.1 (1973), 47-63.

Lavy, Victor, "Do Gender Stereotypes Reduce Girls' and Boys' Human Capital Outcomes? Evidence from a Natural Experiment," *Journal of Public Economics*, Vol.91, No.10-11 (2008), 2083-2105.

Lummis Max and Harold W. Stevenson, "Gender Differences in Beliefs and Achievement: A Cross-Cultural Study," *Developmental Psychology*, Vol. 26 (1990), 254-263.

Mullis, Ina V.S., Michael O. Martin, Edward G. Fierros, Amie L. Goldberg, and Steven E. Stemler, *Gender Differences in Achievement: IEA's Third International Mathematics and Science Study (TIMSS)*, Chestnut Hill, MA: Boston College (2000).

National Center for Education Statistics, *The Nation's Report Card: Science Highlights 2000*, U.S. Department of Education, Office of Education Research and Improvement, National Center for Education Statistics, Washington, DC (2001).

OECD, *Knowledge and Skills for Life: First Results from the OECD Programme for International Student Assessment (PISA) 2000*, OECD, Paris (2001).

OECD, *OECD Handbook for Internationally Comparative Education Statistics: Concepts, Standards, Definitions and Classifications*, OECD, Paris (2004).

OECD, *Education at a Glance: OECD Indicators*, OECD, Paris (2004).

Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, Vol.14, No.3 (1973), 693-709.

Postlethwaite, T.N. and D.E. Wiley, *Science Achievement in Twenty-Three Countries: The IEA Study of Science II*, Pergamon Press, Oxford (1992).

Santapau, Shari L., *The Nation's Report Card: Mathematics Highlights 2000*, U.S. Department of Education, Office of Education Research and Improvement, National Center for Education Statistics, Washington, DC (2001).

UNESCO Institute for Statistics, *Global Education Digest 2003*, UNESCO Institute for Statistics, Montreal, Canada (2003).

UNDP, *Human Development Reports 1993-2000*, Oxford University Press, USA (1993-2000).

Weinberger, Catherine J., “Race and Gender Wage Gaps in the Market for Recent College Graduates,” *Industrial Relations*, Vol.37, No.1 (1998), 67-84.

Weinberger, Catherine J., “Is Teaching More Girls More Math the Key to Higher Wages?” in *Squaring Up: Policy Strategies to Raise Women’s Incomes in the U.S.*, edited by Mary C. King: University of Michigan Press (2001).

Weinberger, Catherine J., “Is the Science and Engineering Workforce Drawn from the Far Tail of the Math Ability Distribution?” UCSB working paper (2005).

Weinberger, Catherine J., and Marlene Kim, “The Working Poor- A Statistical Artifact?” *Eastern Economic Journal*, Vol.25, No.2 (1999), 155-167.

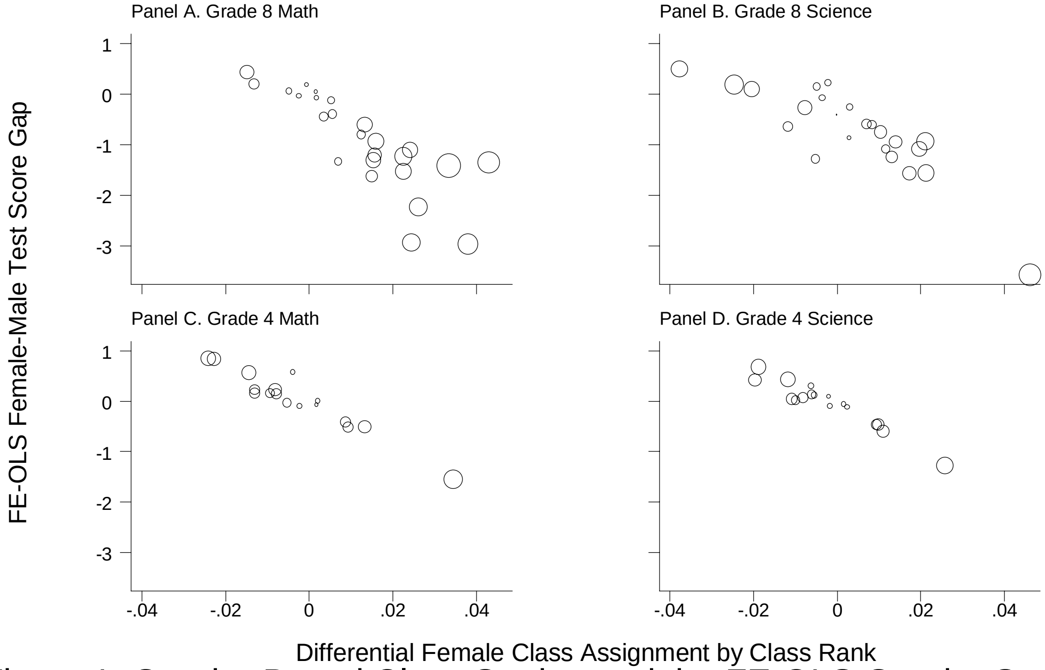

Figure 1. Gender-Based Class Sorting and the FE-OLS Gender Score Gap

Table 1. Summary Statistics

	Fourth Grade					Eighth Grade				
	Math (1)	Science (2)	Female (3)	Math Sample (4)	Science Sample (5)	Math (6)	Science (7)	Female (8)	Math Sample (9)	Science Sample (10)
Australia	51.62	52.99	0.50	15,237	15,045	52.96	53.89	0.51	20,967	17,956
	(9.22)	(8.90)	(0.50)			(9.20)	(9.46)	(0.50)		
Austria	52.36	52.58	0.49	5,047	5,047	53.34	53.28	0.53	5,002	5,578
	(9.27)	(8.55)	(0.50)			(8.70)	(9.19)	(0.50)		
Belgium - Flemish	56.83	53.29	0.50	4,712	4,712	56.82	53.49	0.51	15,845	13,356
	(6.83)	(6.42)	(0.50)			(8.21)	(7.91)	(0.50)		
Belgium - French						53.17	47.25	0.53	4,502	4,502
						(7.97)	(8.33)	(0.50)		
Canada	50.23	51.37	0.50	15,523	15,533	52.72	52.63	0.50	24,871	24,660
	(9.30)	(8.88)	(0.50)			(8.56)	(8.83)	(0.50)		
Czech Republic	53.18	51.80	0.52	6,523	6,523	54.30	54.78	0.50	10,119	10,119
	(9.31)	(8.53)	(0.50)			(8.65)	(8.17)	(0.50)		
Denmark						49.64	46.62	0.51	3,079	3,046
						(8.55)	(8.89)	(0.50)		
England	50.39	52.67	0.50	9,644	9,644	51.88	55.58	0.49	6,982	6,879
	(10.04)	(9.44)	(0.50)			(9.07)	(9.52)	(0.50)		
Finland						53.97	54.77	0.50	2,896	2,905
						(7.33)	(7.65)	(0.50)		
France						51.90	47.76	0.50	5,616	5,616
						(8.08)	(8.31)	(0.50)		
Germany						50.24	51.38	0.51	5,294	5,117
						(8.81)	(9.37)	(0.50)		
Greece	46.40	47.17	0.50	5,759	5,759	47.08	47.78	0.48	7,310	7,568
	(9.90)	(8.90)	(0.50)			(9.12)	(9.03)	(0.50)		
Hungary	52.33	50.99	0.50	9,020	9,020	53.89	54.71	0.51	12,158	12,158
	(9.24)	(8.81)	(0.50)			(9.12)	(8.84)	(0.50)		
Iceland	44.33	46.19	0.51	3,408	3,422	48.06	48.09	0.49	3,713	3,719
	(8.70)	(9.07)	(0.50)			(7.99)	(8.43)	(0.50)		
Ireland	51.32	50.54	0.49	5,753	5,753	51.72	51.73	0.52	6,201	5,686
	(9.52)	(9.00)	(0.50)			(9.07)	(9.22)	(0.50)		
Italy	52.17	52.99	0.48	4,282	4,282	49.83	50.32	0.51	12,439	12,439
	(8.56)	(8.15)	(0.50)			(8.87)	(8.93)	(0.50)		
Japan	56.25	54.40	0.50	12,731	12,731	58.56	55.42	0.49	19,670	19,670
	(8.05)	(7.47)	(0.50)			(8.45)	(8.46)	(0.50)		
Korea	57.05	55.82	0.49	5,586	5,586	61.48	56.59	0.49	11,422	11,422
	(7.57)	(6.83)	(0.50)			(8.62)	(8.62)	(0.50)		
Netherlands	54.26	52.89	0.49	7,636	7,636	54.98	54.69	0.51	9,963	9,963
	(7.76)	(6.91)	(0.50)			(8.55)	(8.25)	(0.50)		
New Zealand	48.77	51.13	0.51	9,211	9,174	50.80	51.83	0.49	14,219	14,099
	(9.50)	(9.44)	(0.50)			(9.04)	(9.46)	(0.50)		
Norway	46.68	48.75	0.48	8,703	8,703	48.91	50.71	0.49	9,864	9,849
	(8.61)	(8.96)	(0.50)			(8.06)	(8.52)	(0.50)		
Portugal	45.53	45.60	0.49	5,447	5,447	44.54	46.00	0.50	6,745	6,746
	(9.28)	(9.55)	(0.50)			(7.10)	(8.24)	(0.50)		
Scotland	49.58	50.77	0.50	10,329	10,329	49.89	50.44	0.49	9,272	9,152
	(9.17)	(8.95)	(0.50)			(8.99)	(9.63)	(0.50)		
Slovak Republic						53.76	53.09	0.50	14,791	14,761
						(8.80)	(8.60)	(0.50)		
Spain						47.48	49.95	0.50	7,595	7,595
						(8.07)	(8.32)	(0.50)		
Sweden						52.21	53.18	0.49	12,939	12,943
						(8.77)	(9.12)	(0.50)		
Switzerland						54.60	51.80	0.50	10,132	10,131
						(8.57)	(8.98)	(0.50)		
United States	51.94	53.35	0.50	20,885	20,813	51.01	52.88	0.50	28,634	28,188
	(9.02)	(8.84)	(0.50)			(9.28)	(9.79)	(0.50)		

Test scores are internationally standardized to mean 50 and standard deviation 10. Sample means are population weighted.

Table 2. Grade 8 Math and Science

	Standardized Score				OECD Percentile Score			
	Math		Science		Math		Science	
	OLS (1)	FE (2)	OLS (3)	FE (4)	OLS (5)	FE (6)	OLS (7)	FE (8)
Australia	**-0.29**	**-0.71**	**-1.61**	**-1.76**	-0.9	**-2.2**	**-5.1**	**-5.4**
	(0.14)	(0.14)	(0.16)	(0.19)	(0.4)	(0.4)	(0.5)	(0.6)
Austria	**-0.91**	**-1.34**	**-1.86**	**-2.26**	**-2.8**	**-4.2**	**-5.9**	**-7.1**
	(0.26)	(0.23)	(0.26)	(0.26)	(0.8)	(0.7)	(0.8)	(0.9)
Belgium - Flemish	-0.10	**-1.05**	**-1.92**	**-2.52**	-0.3	**-3.3**	**-6.6**	**-8.5**
	(0.18)	(0.14)	(0.17)	(0.18)	(0.6)	(0.4)	(0.6)	(0.6)
Belgium - French	**-0.93**	**-1.43**	**-2.30**	**-2.48**	**-3.0**	**-4.6**	**-7.0**	**-7.6**
	(0.24)	(0.23)	(0.25)	(0.26)	(0.8)	(0.8)	(0.7)	(0.8)
Canada	-0.10	**-0.49**	**-1.49**	**-1.73**	-0.4	**-1.6**	**-4.9**	**-5.6**
	(0.16)	(0.15)	(0.16)	(0.16)	(0.5)	(0.4)	(0.5)	(0.5)
Czech Republic	**-1.39**	**-1.88**	**-2.92**	**-3.39**	**-4.4**	**-6.0**	**-9.4**	**-11.0**
	(0.20)	(0.19)	(0.18)	(0.19)	(0.6)	(0.6)	(0.6)	(0.6)
Denmark	**-1.32**	**-1.25**	**-3.11**	**-2.91**	**-4.1**	**-3.9**	**-9.2**	**-8.7**
	(0.32)	(0.32)	(0.33)	(0.36)	(1.0)	(1.0)	(0.9)	(1.0)
England	**-0.98**	**-1.09**	**-2.05**	**-2.14**	**-2.9**	**-3.3**	**-6.3**	**-6.7**
	(0.21)	(0.22)	(0.22)	(0.27)	(0.7)	(0.7)	(0.7)	(0.9)
Finland	-0.46	-0.44	**-1.08**	**-1.32**	-1.5	**-1.4**	**-3.5**	**-4.4**
	(0.28)	(0.32)	(0.30)	(0.38)	(0.9)	(1.0)	(1.0)	(1.3)
France	**-0.75**	**-1.10**	**-2.19**	**-2.39**	**-2.4**	**-3.5**	**-6.6**	**-7.3**
	(0.21)	(0.21)	(0.21)	(0.23)	(0.7)	(0.7)	(0.6)	(0.7)
Germany	**-0.70**	**-1.69**	**-1.88**	**-3.02**	**-2.2**	**-5.2**	**-6.2**	**-9.7**
	(0.24)	(0.23)	(0.26)	(0.25)	(0.7)	(0.7)	(0.8)	(0.7)
Greece	**-1.00**	**-1.05**	**-1.67**	**-1.62**	**-3.2**	**-3.3**	**-5.1**	**-4.9**
	(0.20)	(0.21)	(0.20)	(0.24)	(0.6)	(0.6)	(0.6)	(0.7)
Hungary	**-0.28**	**-0.72**	**-2.13**	**-2.41**	-0.7	**-2.1**	**-6.5**	**-7.5**
	(0.15)	(0.16)	(0.15)	(0.15)	(0.5)	(0.5)	(0.5)	(0.5)
Iceland	-0.13	-0.37	**-1.87**	**-2.19**	-0.6	-1.4	**-6.4**	**-7.5**
	(0.29)	(0.25)	(0.31)	(0.30)	(0.9)	(0.8)	(0.9)	(0.9)
Ireland	**-1.74**	**-2.10**	**-1.76**	**-2.71**	**-5.6**	**-6.5**	**-5.6**	**-8.6**
	(0.22)	(0.28)	(0.23)	(0.33)	(0.7)	(0.9)	(0.7)	(1.1)
Italy	**-0.95**	**-0.97**	**-1.45**	**-1.51**	**-3.0**	**-3.1**	**-4.9**	**-5.1**
	(0.16)	(0.17)	(0.16)	(0.17)	(0.5)	(0.5)	(0.5)	(0.5)
Japan	**-0.49**	**-0.43**	**-1.09**	**-1.01**	**-1.4**	**-1.2**	**-3.5**	**-3.3**
	(0.12)	(0.10)	(0.12)	(0.11)	(0.4)	(0.3)	(0.4)	(0.4)
Korea	**-0.60**	**-0.72**	**-1.72**	**-1.59**	**-1.5**	**-1.9**	**-5.4**	**-4.9**
	(0.15)	(0.23)	(0.15)	(0.25)	(0.4)	(0.7)	(0.5)	(0.8)
Netherlands	**-0.78**	**-1.50**	**-1.93**	**-2.43**	**-2.5**	**-4.7**	**-6.3**	**-7.9**
	(0.20)	(0.13)	(0.19)	(0.15)	(0.6)	(0.4)	(0.6)	(0.5)
New Zealand	0.03	**-0.27**	**-1.34**	**-1.70**	**0.0**	**-0.9**	**-4.4**	**-5.7**
	(0.15)	(0.16)	(0.16)	(0.17)	(0.5)	(0.5)	(0.5)	(0.5)
Norway	**-0.48**	**-0.49**	**-1.56**	**-1.59**	**-1.5**	**-1.5**	**-5.1**	**-5.2**
	(0.18)	(0.18)	(0.18)	(0.18)	(0.5)	(0.5)	(0.6)	(0.6)
Portugal	**-0.76**	**-1.24**	**-2.12**	**-2.50**	**-2.2**	**-3.6**	**-6.1**	**-7.2**
	(0.17)	(0.18)	(0.19)	(0.20)	(0.5)	(0.5)	(0.6)	(0.6)
Scotland	**-1.08**	**-0.94**	**-2.29**	**-2.21**	**-3.4**	**-3.0**	**-7.2**	**-7.0**
	(0.17)	(0.16)	(0.18)	(0.20)	(0.5)	(0.5)	(0.5)	(0.6)
Slovak Republic	**-0.59**	**-0.79**	**-2.19**	**-2.37**	**-1.8**	**-2.4**	**-6.9**	**-7.5**
	(0.15)	(0.16)	(0.14)	(0.16)	(0.5)	(0.5)	(0.5)	(0.5)
Spain	**-0.79**	**-0.79**	**-2.32**	**-2.28**	**-2.3**	**-2.3**	**-7.2**	**-7.1**
	(0.18)	(0.21)	(0.19)	(0.21)	(0.5)	(0.6)	(0.6)	(0.7)
Sweden	-0.11	-0.23	**-1.20**	**-1.32**	-0.2	**-0.6**	**-3.7**	**-4.1**
	(0.15)	(0.15)	(0.16)	(0.16)	(0.5)	(0.5)	(0.5)	(0.5)
Switzerland	**-1.07**	**-1.46**	**-2.24**	**-2.57**	**-3.3**	**-4.5**	**-7.1**	**-8.1**
	(0.18)	(0.14)	(0.18)	(0.17)	(0.6)	(0.5)	(0.6)	(0.5)
United States	**-0.78**	**-0.92**	**-1.81**	**-1.87**	**-2.4**	**-2.9**	**-5.7**	**-6.0**
	(0.12)	(0.09)	(0.12)	(0.13)	(0.4)	(0.3)	(0.4)	(0.4)

Population weighted. Fixed effects clustered at the class level. All models inlcude the variables listed in Section 4.

Bold coefficients are significant at the 5% level. Shaded OLS and FE coefficients are statistically different at the 5% level.

Table 3. Differential Assignment to Class Rank for Females

	Grade 8				Grade 4			
	Pro-Female Sort		Gender Mixed Class		Pro-Female Sort		Gender Mixed Class	
	Math (1)	Science (2)	Math (3)	Science (4)	Math (5)	Science (6)	Math (7)	Science (8)
Australia	**0.015** (0.005)	**-0.017** (0.005)	0.78	0.63	0.009 (0.006)	0.009 (0.006)	0.97	0.95
Austria	0.007 (0.008)	0.013 (0.007)	0.81	0.93	-0.004 (0.010)	-0.006 (0.010)	1.00	1.00
Belgium - Flemish	**0.024** (0.006)	**-0.034** (0.006)	0.75	0.62	-0.009 (0.009)	-0.002 (0.009)	0.99	0.99
Belgium - French	0.015 (0.009)	-0.012 (0.009)	0.90	0.90				
Canada	**0.023** (0.006)	**0.010** (0.005)	0.98	0.97	**-0.024** (0.008)	-0.008 (0.006)	0.99	0.99
Czech Republic	**0.022** (0.007)	**0.021** (0.006)	0.99	0.99	**0.034** (0.007)	**0.026** (0.007)	0.99	0.99
Denmark	-0.013 (0.011)	**-0.038** (0.011)	0.79	0.78				
England	-0.006 (0.006)	**-0.027** (0.006)	0.66	0.62	-0.005 (0.006)	-0.006 (0.006)	0.98	0.98
Finland	-0.005 (0.011)	0.003 (0.011)	0.98	0.93				
France	**0.024** (0.008)	0.008 (0.008)	0.98	0.98				
Germany	**0.038** (0.008)	**0.046** (0.008)	0.93	0.89				
Greece	0.005 (0.007)	-0.005 (0.007)	0.93	0.96	**-0.023** (0.010)	**-0.020** (0.009)	1.00	1.00
Hungary	**0.033** (0.005)	**0.021** (0.005)	0.99	0.99	**0.013** (0.006)	0.011 (0.006)	0.99	0.99
Iceland	0.012 (0.011)	0.012 (0.011)	0.99	0.99	-0.013 (0.010)	-0.010 (0.010)	0.97	0.98
Ireland	**-0.035** (0.007)	**-0.037** (0.008)	0.51	0.44	**0.021** (0.008)	0.002 (0.008)	0.64	0.64
Italy	0.002 (0.005)	0.003 (0.005)	1.00	1.00	-0.013 (0.009)	-0.005 (0.009)	1.00	1.00
Japan	-0.001 (0.004)	-0.002 (0.004)	0.98	0.98	0.002 (0.005)	0.002 (0.005)	1.00	1.00
Korea	**-0.017** (0.005)	**-0.096** (0.004)	0.39	0.39			1.00	1.00
Netherlands	**0.026** (0.006)	**0.017** (0.007)	0.97	0.97	-0.008 (0.006)	-0.011 (0.007)	1.00	1.00
New Zealand	**0.016** (0.005)	-0.005 (0.005)	0.73	0.72	0.009 (0.006)	0.010 (0.006)	0.98	0.97
Norway	-0.002 (0.007)	-0.004 (0.007)	1.00	1.00	-0.002 (0.006)	-0.002 (0.006)	1.00	1.00
Portugal	**0.043** (0.007)	**0.020** (0.007)	1.00	1.00	0.002 (0.008)	0.002 (0.008)	0.99	0.99
Scotland	**-0.015** (0.006)	**-0.025** (0.005)	0.98	0.91	**-0.014** (0.006)	**-0.019** (0.006)	0.99	0.99
Slovak Republic	**0.013** (0.005)	0.007 (0.005)	0.98	0.98				
Spain	0.002 (0.007)	**-0.020** (0.007)	0.93	0.93				
Sweden	0.006 (0.005)	0.000 (0.005)	0.99	0.96				
Switzerland	**0.016** (0.007)	**0.014** (0.007)	0.98	0.98				
United States	0.004 (0.003)	**-0.008** (0.003)	0.99	0.92	-0.008 (0.004)	**-0.012** (0.004)	1.00	1.00

Population weighted. All models inlcude the variables listed in Section 4. Bold coefficients are significant at the 5% level.

Table 4. Grade 4 Math and Science

	Standardized Score				OECD Percentile Score			
	Math		Science		Math		Science	
	OLS (1)	FE (2)	OLS (3)	FE (4)	OLS (5)	FE (6)	OLS (7)	FE (8)
Australia	**-0.58**	**-0.73**	**-0.43**	**-0.53**	**-1.9**	**-2.3**	**-2.1**	**-2.5**
	(0.18)	(0.19)	(0.19)	(0.20)	(0.6)	(0.6)	(0.6)	(0.6)
Austria	**-1.30**	**-1.11**	**-1.17**	**-1.08**	**-4.2**	**-3.6**	**-4.4**	**-4.0**
	(0.30)	(0.26)	(0.28)	(0.25)	(0.9)	(0.8)	(0.9)	(0.8)
Belgium - Flemish	**-0.54**	**-0.48**	**-0.60**	**-0.58**	**-1.7**	**-1.5**	**-2.3**	**-2.2**
	(0.21)	(0.22)	(0.20)	(0.22)	(0.7)	(0.7)	(0.7)	(0.8)
Canada	**-0.96**	**-0.69**	**-0.82**	**-0.81**	**-3.0**	**-2.2**	**-3.0**	**-3.0**
	(0.22)	(0.22)	(0.21)	(0.21)	(0.7)	(0.7)	(0.7)	(0.7)
Czech Republic	**-0.81**	**-1.30**	**-1.57**	**-1.96**	**-2.5**	**-4.1**	**-5.6**	**-6.9**
	(0.22)	(0.22)	(0.20)	(0.21)	(0.7)	(0.7)	(0.7)	(0.7)
England	**-0.70**	**-0.70**	-0.26	-0.22	**-2.4**	**-2.5**	**-1.3**	-1.2
	(0.19)	(0.20)	(0.18)	(0.19)	(0.6)	(0.6)	(0.6)	(0.6)
Greece	**-1.02**	**-0.70**	**-1.34**	**-1.14**	**-2.8**	**-1.9**	**-4.2**	**-3.8**
	(0.27)	(0.25)	(0.23)	(0.23)	(0.8)	(0.7)	(0.7)	(0.7)
Hungary	**-0.38**	**-0.55**	**-1.21**	**-1.38**	**-1.1**	**-1.6**	**-4.1**	**-4.7**
	(0.18)	(0.19)	(0.17)	(0.18)	(0.6)	(0.6)	(0.5)	(0.6)
Iceland	**-1.19**	**-1.12**	**-1.30**	**-1.28**	**-3.1**	**-3.0**	**-4.1**	**-4.1**
	(0.30)	(0.33)	(0.32)	(0.32)	(0.8)	(0.9)	(0.9)	(0.9)
Ireland	0.11	-0.31	**-0.62**	**-1.05**	0.1	-1.2	**-2.8**	**-4.1**
	(0.23)	(0.30)	(0.22)	(0.28)	(0.7)	(0.9)	(0.7)	(0.9)
Italy	**-1.03**	**-0.97**	**-0.54**	**-0.48**	**-3.6**	**-3.3**	**-2.0**	**-1.8**
	(0.27)	(0.25)	(0.26)	(0.24)	(0.9)	(0.8)	(0.9)	(0.8)
Japan	**-0.40**	**-0.39**	**-0.52**	**-0.52**	**-1.2**	**-1.2**	**-2.2**	**-2.2**
	(0.14)	(0.14)	(0.13)	(0.14)	(0.4)	(0.4)	(0.5)	(0.5)
Korea	**-1.25**	**-1.29**	**-1.23**	**-1.31**	**-4.1**	**-4.2**	**-4.5**	**-4.8**
	(0.19)	(0.18)	(0.17)	(0.17)	(0.6)	(0.6)	(0.6)	(0.6)
Netherlands	**-0.96**	**-0.92**	**-1.38**	**-1.37**	**-3.1**	**-2.9**	**-5.0**	**-5.0**
	(0.17)	(0.17)	(0.16)	(0.18)	(0.5)	(0.6)	(0.6)	(0.7)
New Zealand	0.06	-0.10	0.28	0.10	-0.2	-0.5	0.3	-0.1
	(0.20)	(0.21)	(0.20)	(0.21)	(0.6)	(0.6)	(0.6)	(0.7)
Norway	**-0.92**	**-0.96**	**-0.72**	**-0.73**	**-2.8**	**-2.8**	**-2.5**	**-2.6**
	(0.18)	(0.19)	(0.20)	(0.21)	(0.5)	(0.6)	(0.6)	(0.7)
Portugal	**-0.74**	**-0.76**	**-1.13**	**-1.16**	**-2.1**	**-2.1**	**-3.4**	**-3.5**
	(0.24)	(0.23)	(0.25)	(0.26)	(0.7)	(0.6)	(0.7)	(0.7)
Scotland	**-0.75**	**-0.56**	**-0.92**	**-0.72**	**-2.4**	**-1.8**	**-3.5**	**-2.8**
	(0.18)	(0.18)	(0.17)	(0.18)	(0.5)	(0.6)	(0.6)	(0.6)
United States	**-0.45**	**-0.40**	**-0.83**	**-0.71**	**-1.4**	**-1.2**	**-3.2**	**-2.8**
	(0.14)	(0.15)	(0.14)	(0.16)	(0.4)	(0.5)	(0.5)	(0.5)

Population weighted. Fixed effects clustered at the class level. All models inlcude the variables listed in Section 4.

Bold coefficients are significant at the 5% level. Shaded OLS and FE coefficients are statistically different at the 5% level.

Table 5. Explaining the Gender Test Score Gap

	Math				Science			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grade 8								
Percent Academic	**1.75**	**2.85**	**0.33**	**8.02**	**2.30**	1.72	0.14	***6.76***
	(0.80)	(1.03)	(0.14)	(3.14)	(0.90)	(1.09)	(0.15)	(3.46)
Pro-Female Sort	**41.43**	**37.61**	**41.24**	**36.82**	-2.07	5.78	5.57	15.14
	(15.02)	(16.64)	(18.29)	(17.02)	(16.59)	(16.00)	(17.59)	(17.05)
% Mixed Classes	-0.18	2.40	1.28	1.63	1.63	2.21	1.51	1.53
	(1.62)	(2.07)	(2.05)	(2.02)	(2.64)	(2.40)	(2.46)	(2.27)
National Male Score	***0.04***	**0.09**	**0.09**	0.05	0.03	**0.07**	***0.08***	**0.07**
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Public Educ Expend		**102.22**	77.48	**114.65**		79.57	68.71	***112.30***
		(48.93)	(50.16)	(51.65)		(53.31)	(54.95)	(56.71)
Priv School Enroll Rate		-1.41	-1.61	-1.26		**-3.33**	**-3.60**	**-3.32**
		(1.34)	(1.41)	(1.38)		(1.45)	(1.50)	(1.38)
% Female Teachers		-1.28	-2.02	0.11		**-6.36**	**-6.94**	-3.88
		(2.93)	(3.03)	(3.22)		(3.00)	(3.10)	(3.36)
F/M Univ Enroll Ratio		1.00	***2.31***	0.94		***2.57***	**3.51**	***2.49***
		(1.37)	(1.29)	(1.43)		(1.48)	(1.36)	(1.39)
Female LFP Rate		-1.90	0.94	-1.35		***-6.44***	-4.95	***-7.82***
		(3.51)	(3.44)	(3.54)		(3.74)	(3.72)	(3.81)
GDP		***-0.10***	***-0.12***	**-0.13**		-0.02	-0.02	-0.06
		(0.05)	(0.06)	(0.06)		(0.06)	(0.07)	(0.06)
Adjusted R-Squared	0.20	0.39	0.31	0.36	0.09	0.37	0.31	0.41
Sample Size	28	27	27	27	28	27	27	27
Grade 4								
Percent Academic	**1.84**	***2.27***	***0.31***	***10.40***	**3.13**	**3.68**	**0.37**	**11.81**
	(0.62)	(0.98)	(0.16)	(4.58)	(1.09)	(1.05)	(0.15)	(3.82)
Pro-Female Sort	**40.45**	7.63	2.52	70.47	40.46	31.81	-36.43	-67.20
	(15.48)	(34.32)	(37.12)	(42.37)	(29.85)	(64.04)	(68.58)	(57.30)
% Mixed Classes	-1.57	-5.43	-7.26	5.46	2.14	4.35	-3.76	-2.35
	(2.89)	(5.45)	(5.90)	(7.34)	(4.44)	(6.62)	(6.91)	(6.21)
National Male Score	0.02	0.07	0.10	-0.04	0.00	0.04	0.17	0.18
	(0.02)	(0.05)	(0.06)	(0.06)	(0.04)	(0.10)	(0.11)	(0.09)
Public Educ Expend		72.42	81.13	92.93		**189.70**	**218.33**	**325.98**
		(59.43)	(64.67)	(62.03)		(54.97)	(65.67)	(63.94)
Priv School Enroll Rate		-1.62	-1.97	1.03		-1.84	-3.34	-3.10
		(1.57)	(1.67)	(2.06)		(1.72)	(1.90)	(1.69)
% Female Teachers		-2.95	-3.37	2.87		-2.34	-5.36	-1.66
		(3.19)	(3.39)	(4.41)		(3.58)	(4.05)	(4.04)
F/M Univ Enroll Ratio		0.84	1.82	-0.81		-2.11	1.46	-0.12
		(1.58)	(1.68)	(1.86)		(2.54)	(2.62)	(2.47)
Female LFP Rate		0.03	2.14	-10.08		***-14.69***	-5.82	-8.91
		(5.69)	(6.15)	(7.36)		(7.54)	(8.05)	(7.40)
GDP		-0.12	-0.18	-0.07		-0.11	-0.27	**-0.38**
		(0.10)	(0.12)	(0.09)		(0.13)	(0.14)	(0.13)
Adjusted R-Squared	0.48	0.33	0.23	0.32	0.24	0.71	0.57	0.66
Sample Size	18	17	17	17	18	17	17	17

Weighted by the inverse sampling variance from the first stage. Bold coefficients are significant at the 5% level and bold italics are significant at the 10% level.

Appendix Table 1. Differences in International Educational Systems

	Age at Start of Compulsory Education	First Grade with Formal Streaming	Percent Academic at Grade 10	Age at End of Compulsory Education	Population at least Upper Secondary Educaiton, Males	Population at least Upper Secondary Educaiton, Females	Population at least Tertiary Education, Males	Population at least Tertiary Education, Females
Austrailia	5	11	100	15	73	68	29	38
Austria	6	4	13	15	86	81	16	14
Belgium-Flemish	6	8	38	15-18	74	77	33	39
Belgium-French	6	7	53	15-18	74	77	33	39
Canada	5	none	100	16-18	88	91	45	56
Czech Republic	6	5	19	15	93	92	12	11
Denmark	7	9	48	16	85	88	25	34
England	5	11	100	16	70	65	30	29
Finland	7	9	43	16	84	90	30	46
France	6	9	48	16	78	78	32	37
Germany	6	4	26	16-19	87	84	23	20
Greece	6	9	61	15	69	76	21	27
Hungary	5	4	28	18	81	80	13	16
Iceland	6	10	56	16	64	59	25	29
Ireland	6	11	100	15	71	76	45	50
Italy	6	8	33	15	55	60	10	13
Japan	6	9	75	15	92	95	46	49
Korea	6	9	58	15	95	91	42	35
Netherlands	5	8	38	16-17	73	75	27	26
New Zealand	5	11	100	16	82	82	26	31
Norway	6	10	31	16	93	94	30	40
Portugal	6	6	71	15	28	37	10	17
Scotland	5	11	100	16	70	65	30	29
Slovak Republic	6	4	24	16	95	93	11	12
Spain	6	10	66	16	55	59	32	39
Sweden	7	9	87	16	90	91	34	39
Switzerland	6	9	23	15	93	91	35	17
United States	6	none	100	16-18	87	89	36	42

Note: Age at start of compulsory education and first grade with formal streaming data from EURYDICE (1999), www.euroeducation.net, and www.en.wikipedia.org. Percent academic at grade 10 data from OECD (2004).First grade with formal streaming indicates the grade level in which explicit academic or vocational tracks are offered. Percent academic at grade 10 is the percentage of students enrolled in an academic track. Population at least upper secondary and tertiary educaiton data from OECD (2002). Population at least upper secondary or tertial education are percentages of the population that has attained at least upper secondary education or at least tertiary educaiton among 25 to34-year-olds.

Appendix Table 2. Previous Gender Test Score Gaps Estimates

Author	Data set	Testyear	Country	Age	Subject	Test score		Gender gap (female-male)		Score range	Remarks
						Boys	Girls				
Freeman (2004)	ECLS-K	1998 (Fall)	United States	5	Math	22.3	21.5	-0.8		0-123	
		1999 (Spring)		5.5		32.5	31.7	-0.8		0-123	
		1999 (Fall)		6		39.6	38.6	-1.0		0-123	
		2000 (Spring)		6.5		56.8	54.9	-1.9		0-123	
		2002 (Spring)		8		87.4	83.2	-4.2		0-123	
		1998 (Fall)	United States	5	Math, Addition and	4.7	3.2	-1.5		0-123	
		1999 (Spring)		5.5	Subtraction	19.1	17.1	-2.0		0-123	
		1999 (Fall)		6		36.1	32.7	-3.4		0-123	
		2000 (Spring)		6.5		73.1	73.2	0.1		0-123	
		2002 (Spring)		8		97.3	96.8	-0.5		0-123	
	AP	2002	United States	16-17	Calculus	3.5	3.3	-0.2		1-5	
				16-17	Comp.Science	3.2	2.9	-0.3		1-5	
				16-17	Science	3.1	2.8	-0.3		1-5	
Coley (2001)	NAEP	1996	United States	9	Math			-3.2	*	0-500	(1), (2)
				13				1.0		0-500	
				17				-2.0		0-500	
				9	Science			-3.0		0-500	
				13				-9.9	*	0-500	
				17				-8.6	*	0-500	
Dee (2007)	NAEP	1999	United States	9	Math	232.9	231.2	-1.7		0-500	
				13		277.2	274.5	-2.7	*	0-500	(1)
				17		309.8	306.8	-3.0	*	0-500	
				9	Science	230.9	227.9	-3.0	*	0-500	
				13		258.7	252.9	-5.8	*	0-500	
				17		300.4	290.6	-9.8	*	0-500	
NCES (2001)	NAEP	2000	United States	9	Math	229.0	226.0	-3.0		0-500	
				13		277.0	274.0	-3.0		0-500	
				17		303.0	299.0	-4.0		0-500	
				9	Science	153.0	147.0	-6.0		0-300	
				13		154.0	147.0	-7.0		0-300	
				17		148.0	145.0	-3.0		0-300	
Hedges and Nowell (1995)	Project Talent	1960	United States	15	Math			-0.1			(3)
					Physics			-0.5			
					Biology			-0.3			
	NLS-72	1972	United States	17	Math			-0.2			
	NLSY	1980	United States	15-22	Arithmatic reasoning			-0.3			
					Mathematical knowledge			-0.1			
					Science			-0.4			
	HS&B	1980	United States	17	Math			-0.2			
	NELS: 88	1992	United States	13-17	Math			0.0			
					Science			-0.1			

Note: (1) The gender difference with * is statistically significant at 5% level. (2) The gender gap in this table is for white students only. (3) Hedges et al reported d-value, instead of raw score gaps. According to Cohen (1977), we can interprete the gap is small if d<0.2; medium if 0.2<d<0.5; and large if d>0.8. (4) B indicates blind tests or state-level tests and NB indicates non-blind tests or school-level tests.

Appendix Table 2. Previous Gender Test Score Gaps Estimates

Author	Data set	Testyear	Country	Age	Subject	Test score		Gender gap	Score range	Remarks
						Boys	Girls			
Hedges and Nowell (1995)	NAEP	1978	United States	17	Math			-0.2		
		1982		17				-0.2		
		1986		17				-0.2		
		1990		17				-0.1		
		1992		17				-0.2		
		1977		17	Science			-0.3		
		1982		17				-0.4		
		1986		17				-0.3		
		1990		17				-0.2		
		1992		17				-0.2		
Jacob (2002)	NELS: 88	1988-1992	United States	17	Math	50.1	47.8	-2.4	Mean 50	
Kaur (1990)	GCE "O" level	1986	Singapore	16	Math, Paper I	54.1	50.9	-3.2	N/A	
				16	, Paper II	47.3	46.5	-0.8	N/A	
				16	, Paper II-A	26.8	26.5	-0.3	N/A	
				16	, Paper II_B	20.5	20.0	-0.5	N/A	
				16	, Spatial ability	39.3	36.6	-2.7	N/A	
Lummis and Stevenson (1990)	Curriculum-based Independent Achievement Test	1979-1980	United States	6	Math	38.3	38.0	-0.3	N/A	
			Taiwan	6	Math	39.6	38.7	-0.9	N/A	
			Japan	6	Math	42.4	42.4	0.0	N/A	
		1985-1986	United States	7	Math	16.6	17.6	1.0	N/A	
			Taiwan	7	Math	21.2	21.1	-0.1	N/A	
			Japan	7	Math	20.7	19.5	-1.2	N/A	
			United States	11	Math	45.0	43.8	-1.2	N/A	
			Taiwan	11	Math	50.5	51.0	0.5	N/A	
			Japan	11	Math	53.0	53.5	0.5	N/A	
Lavy (2008)	Ministry of Education, Israel	2000-2002	Israel	15-16	Biology	79.7	80.8	1.1	0-100	B (4)
				15-16	Chemistry	76.8	78.8	2.0	0-100	
				15-16	Comp. Science	73.0	72.7	-0.3	0-100	
				15-16	Math	77.3	79.5	2.2	0-100	
				15-16	Physics	81.2	81.0	-0.2	0-100	
				15-16	Biology	81.6	84.8	3.2	0-100	NB (4)
				15-16	Chemistry	84.2	86.4	2.2	0-100	
				15-16	Comp. Science	83.0	85.0	2.0	0-100	
				15-16	Math	79.1	82.1	3.0	0-100	
				15-16	Physics	85.2	86.9	1.7	0-100	

Note: (1) The gender difference with * is statistically significant at 5% level. (2) The gender gap in this table is for white students only. (3) Hedges et al reported d-value, instead of raw score gaps. According to Cohen (1977), we can interprete the gap is small if d<0.2; medium if 0.2<d<0.5; and large if d>0.8. (4) B indicates blind tests or state-level tests and NB indicates non-blind tests or school-level tests.

Appendix Table 2. Previous Gender Test Score Gaps Estimates

Author	Data set	Testyear	Country	Age	Subject	Test score		Gender gap		Score range	Remarks
						Boys	Girls				
Hanna el al (1990)	SIMS	1977-1979	Average	15	Math	47.2	42.3	-4.9			
OECD (2001)	PISA	2000	OECD average	15	Mathematical literacy	506.3	495.0	-11.3		Mean 500	(1)
			Australia			539.3	527.3	-12.0		Mean 500	
			Austria			530.1	503.0	-27.1	*	Mean 500	
			Belgium			523.7	517.5	-6.2		Mean 500	
			Canada			538.8	528.6	-10.3	*	Mean 500	
			Czech Republic			503.8	492.1	-11.7	*	Mean 500	
			Denmark			522.1	507.3	-14.8	*	Mean 500	
			Finland			536.7	535.7	-1.0		Mean 500	
			France			524.8	510.7	-14.1	*	Mean 500	
			Germany			497.6	483.0	-14.6	*	Mean 500	
			Greece			450.8	444.3	-6.5		Mean 500	
			Hungary			491.7	484.7	-7.0		Mean 500	
			Iceland			513.5	518.0	4.6		Mean 500	
			Ireland			510.1	497.3	-12.9	*	Mean 500	
			Italy			462.1	453.7	-8.4		Mean 500	
			Japan			560.7	552.6	-8.2		Mean 500	
			Korea			558.6	532.1	-26.6	*	Mean 500	
			Luxembourg			454.1	439.2	-15.0	*	Mean 500	
			Mexico			392.7	382.0	-10.6		Mean 500	
			New Zealand			536.4	539.1	2.7		Mean 500	
			Norway			505.9	495.4	-10.5	*	Mean 500	
			Poland			472.5	467.7	-4.8		Mean 500	
			Portugal			464.3	445.8	-18.5	*	Mean 500	
			Spain			486.8	468.6	-18.2	*	Mean 500	
			Sweden			514.2	506.7	-7.5		Mean 500	
			Switzerland			537.0	522.8	-14.2	*	Mean 500	
			United Kingdom			534.3	526.2	-8.0		Mean 500	
			United States			496.8	489.6	-7.1		Mean 500	
			OECD average	15	Scientific literacy	500.5	500.7	0.2		Mean 500	
Mullis et al (2000)	TIMSS	1995	OECD average	9	Math	535.0	532.9	-2.1		Mean 500	
			OECD average	13	Math	518.8	512.4	-6.4		Mean 500	
			OECD average	17	Math	517.5	484.6	-33.0		Mean 500	
			OECD average	9	Science	534.0	524.9	-9.0		Mean 500	
			OECD average	13	Science	525.4	508.8	-16.6		Mean 500	
			OECD average	17	Science	521.0	481.6	-39.5		Mean 500	

Note: (1) The gender difference with * is statistically significant at 5% level. (2) The gender gap in this table is for white students only. (3) Hedges et al reported d-value, instead of raw score gaps. According to Cohen (1977), we can interprete the gap is small if d<0.2; medium if 0.2<d<0.5; and large if d>0.8. (4) B indicates blind tests or state-level tests and NB indicates non-blind tests or school-level tests.