The Gender Test Score Gap Across OECD Countries

Kelly Bedard and Insook Cho

Department of Economics
University of California, Santa Barbara

January 2007
Preliminary Draft

Abstract

This paper begins with three facts. One, math and science skills are scarce resources that are highly rewarded in the labor market. Two, women are underrepresented in most university math/science programs and are less likely to be employed in math/science related fields. Three, isolating the important reasons why women are less likely to enter math/science fields is difficult, if not impossible, using data from a single country since biology, family, educational institutions, and the labor market interact in ways that are almost impossible to disentangle. We circumvent some of these problems using international math and science test score data from the Trends in International Mathematics and Science Study (TIMSS) to identify some of the educational institutions that are important determinants of early gender gaps in math and science. We contribute to the stock of knowledge about gender skill gaps in two ways. First, we show that countries with education systems that use highly selective academic streaming have larger gender gaps, even before streaming occurs. Second, we further show that pro-female biased class/program assignment substantially reduces the observed gender gap in many countries.

We thank Heather Antecol, Bill Fischel, Peter Kuhn, and Cathy Weinberger for helpful comments.

1. Introduction

In a January 2005 speech at the NBER Lawrence Summers suggested that part of the reason that women are under-represented in the sciences at elite universities may stem in from "innate" differences between men and women. His comments triggered a minor media frenzy and reignited a long-standing academic interest in gender skill differences. 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 excellent overview) and its impact on the gender pay gap among college graduates (Eide 1997; Brown and Corcoran 1997; Weinberger 1998, 1999, 2001). Backing up a step to childhood and adolescence, when many basic math and science skills are acquired, if girls accumulate fewer math and science skills they will make different higher education choices and earn lower wages because of their skill deficits.

Since math and science skills are highly valued in the economy, it seems important to understand the origin, or cause, of any skill gap that might lead to differential education and labor market choices across genders. We know, for example, that individuals with more high school math courses earn higher wages later in life (Rose and Betts 2005). Unfortunately, the root cause of observed math and science skill gaps during adolescence and adulthood are likely a complex combination of biology and the environment from birth onward. While simple gender test score gaps at any particular age tell us the average observed gender test score difference at that age, they are only an unbiased estimate of the "innate" gender gap if all factors omitted from the estimating equation are uncorrelated with gender.

It is easiest to think about the problem in reverse. In most cases, researchers simply report that girls score X percent lower than boys in math or science. How should one interpret this gap? While it is tempting to infer that the average boy is X percent "better" at math or science than the average girl, one must think very carefully about what "better" means in this context. Since the test score of interest is for a given grade, every factor that has influenced human capital accumulation up to that point in embedded in the score. In particular, innate math/science ability, family environment, teacher interactions and evaluations, peer interactions, student expectations about the current and future importance of math and science skills, and class/program/stream assignments are all embedded in test scores. If the ultimate goal is to estimate the true underlying difference in innate math or science ability, one needs to find an experiment that separates innate ability from the other aforementioned influences or have a set of variables that are sufficiently rich to control for all of these factors. Not surprisingly, neither of these approaches seems likely to succeed. As such, estimating the causal effect of gender is an incredibly difficult, if not impossible, task.

In contrast, if one starts with the understanding that what we observe at a point in time is a composite measure of innate ability, parent/teacher/peer interactions, student expectations about the future importance of math and science, and institutional structures, then comparing gender gaps across countries may give us important insight into the relative importance of various contributing factors. For example, the existence of internationally comparable data allows us to ask whether the gender test score gap varies systematically across school system structures. In particular, do countries with more elitist streaming rules have systematically bigger or smaller gender gaps in math and science, even among students in identical classes or programs? Further, does streaming impact the gender gap exclusively through class placement or through other avenues, such as student expectations about the future usefulness of math and science and teacher interactions, as well?

2. Math, science, gender, and the importance of international comparisons

There seems to be little dispute that boys outperform girls in math and science in most developed countries.[1] 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). Instead, the dispute is over the cause (nature versus nurture) and size of the gap. At one extreme, some authors almost dismiss the math and science gender score gaps as insignificant given their relatively small size compared to other socioeconomic gaps. For example, Dee (2005) points out that the gender gap is only 10 to 20 percent of black-white gap. At the other extreme, some authors argue that even small early gaps may have crucially important long run consequences (Marsh and Yeung 1998; Ethington

[1] Recent examples include, Freeman (2004), Coley (2001), Hedges and Nowell (1995), Jensen (1988), Chipman et al (1991), Moore et al. (1987), and Marshall et al. (1987).

and Wolfle 1986). In other words, even a small female deficiency might propagate itself through a multiplier process if skills accumulated in childhood and adolescence are complementary to later learning (see Cuhna, Heckman, Lochner, and Masterov 2006).

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 out score girls by three percentage points. In contrast, using 1999 National Assessment of Educational Progress (NAEP) data, Dee (2005) 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 (2004) 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. While their results are somewhat difficult to summarize because they report many subcategories, 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)

[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 (1971-1992) data.

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 (1991), Mullis et al. (2000), and PISA-OECD (2000), respectively.[3] In general, these studies find a small gender gap favoring boys across most participating countries (see Appendix Table 2).

Baker and Jones (1993) and Wiseman, Baker, and Ramirez (1999) attempt to explain the male math premium in SIMS and TIMSS, respectively. Both studies report the correlation between the male-female math gap and a variety of higher education and labor market opportunity measures. Baker and Jones (1993) report a negative correlation between the percentage of women in higher education and the eighth grade male-female math gap and a similarly negative correlation between female labor force participation and the eighth grade gender test score gap. In contrast, Wiseman, Baker, and Ramirez (1999) do not find these statistically significant correlations in TIMSS.

Keeves (1973) and Hanna, Kundiger, and Larouche (1990) further examine the gender gap in FIMS/FISS and SIMS, respectively, among students in their final year of secondary school. Keeves (1973) makes several important observations. First, terminal year math and science classes have two boys for every girl, except in Finland. Second, conditional on enrollment, boys out score girls. Third, boys are more interested in math and science, conditional on enrollment. Hanna, Kundiger, and Larouche (1990) further attempt to uncover the factors that lead to worse performance in mathematics by girls in the final year of secondary school in SIMS. Although they find statistically significant gender gap differences across participating countries, they are unable to identify family background or school structure factors capable of explaining any of the observed international variation in the gender math score gap at the end of secondary school.

Since all of the studies discussed so far report either raw female-male test score gaps or regression adjusted gaps, it is impossible to determine the underlying cause of the

[3] Notice that most of these studies come directly from the testing agency's user guide.

gender gap unless all unobservables are uncorrelated with gender. While it may be reasonable to assume that family background and other socioeconomic factors are uncorrelated with child gender, unobserved parent and teacher interactions related to math and science and sorting across academic programs are unlikely to be independent of gender. As a result, researchers have had difficulty determining the importance of biological factors relative to environmental influences.

Despite the difficulty associated with determining the relative importance of biology and the environment for the gender test score gap, it is nonetheless important to think carefully about the potential major biological and environmental influences because their relative importance and interactions determine the correct interpretation of OLS gap estimates. For example, parental gender stereotypes can influence children's attitudes towards mathematics and hence their competency levels (Tiedeman 2000; Carr et al. 1999; Jacob and Eccles 1992). In a similar vein, teachers may interact differently with boys and girls in ways that lead to differential cognitive development (Tiedemann 2000; Carr et al. 1999). It is also possible that subjective teacher evaluations favor either boys or girls in ways that reinforce or partially counteract gender stereotypical human capital investment (Lavy 2004; Hilderbrand 1996; Sadker and Sadker 1994; AAUW 1992).

At the same time, gender specific brain structures and hormones may play an important role in determining mental aptitude (Benbow and Stanley 1980). Fitch et al (1997) and Kimura (1987, 1999) argue that superior visual-spatial skills among males and superior verbal skills among females are at least partly the result of male brains being more developed in the right hemisphere and female brains being more focally organized in the left hemisphere.[4] In contrast, Silverman et al. (1999) and Phillips and Silverman (1997) attribute part of this difference to higher levels of testosterone in boys and higher levels of estrogen in girls. These arguments notwithstanding, researchers are reluctant to conclude that biological differences are the sole cause of the existing gender test score gap, since biological factors interact with environmental factors (Neisser et al. 1996).

Our work diverges from previous attempts to estimate the math and science gender gap in at least two important ways. First, our use of internationally comparable TIMSS tests facilitates cross-country comparisons that highlight differences in

[4] Also see Cahill (2005), Blum (1997), and Kimura (1992).

parent/teacher/peer interactions, student expectations, and institutional structures that lead to differential test score gaps across countries, to the extent that it is reasonable to assume that the distribution of innate math and science skills are the same across OECD countries. For example, TIMSS allows us to gain important insight into some of the education system differences that appear to exacerbate the gender test score gap. In particular, we find that countries that formally stream students into academic and vocational programs at early ages have greater gender test score gaps in math and science – even before formal streaming has occurred.

Second, we document the substantial degree of gender-biased class assignment in many OECD countries at or before grade eight. We begin by comparing standard OLS estimates with class fixed effects estimates. We find that the standard OLS gender gap understates the gender gap in many countries because girls are more likely to be in "better" classes. More specifically, once we control for female-biased class sorting the eighth grade math gender gap estimate increases by at least 35 percent in eleven countries.[5] This might occur because girls are better behaved and easier to deal with, which may increase the probability that they are seen as more able.[6] For example, Lavy (2005) finds that Israeli teachers give girls higher grades, holding ability constant. We further show that countries with higher degrees of female-biased class sorting have lower observed gender gaps. In other words, pro-female class sorting reduces the gender gap.

3. Econometric framework

3.1. Estimating the gender test score gap

We begin with a simple model of the relationship between gender and student outcomes.

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

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 4, and ε is the usual error term.[7] All models are estimated separately for each grade, subject, and country.

---

[5] This count reflects statistically significant OLS and FE differences results reported in Table 2.
[6] At the other end of the spectrum, teachers who are predisposed to see boys as more academically talented may over assess male ability levels.
[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, for simplicity we prefer the

A gender gap (female-male) estimate using equation (1) will only be 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 the 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.

One of the strengths of TIMSS is that it includes enough information about teachers to allow us to construct class indicators (see Section 4). We therefore estimate the following fixed effects model:

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

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. Section 5.2. discusses the gender-biased allocation of students across classes in many countries in more detail.

3.2. Factors contributing to the gender test score gap

The second part of the analysis tries to uncover the education system features that contribute to the wide range of gender gaps observed across developed countries. More specifically, we hypothesize that β_{cg} depends on the structure of the education system: The degree of streaming, the amount of gender-biased class/stream sorting, and the use of single-sex classes.

$$\beta_{cg} = \pi_g + \delta_1 E_{cg} + \delta_2 FB_{cg} + \delta_3 M_{cg} + \upsilon_{cg} \qquad (3)$$

where E measures the degree of ability streaming, FB is the degree of female-biased sorting (discussed in detail in Section 5.2), and M is the fraction of students in mixed

---

simpler approach described by equation (1) 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).

gender classes (also discussed in detail in Section 5.2).[8] All equation (3) estimates are weighted by the inverse sampling variance for the left-hand side variable from equation (1). Alternatively, the effect of a particular school system structure could be estimated in a single step by augmenting equation (1) to include interactions between the female indicator and the school system structure measure and pooling the data across countries. The primary advantage of the two-stage procedure is that it reduces the data to the country level, which makes it easy to see the relationship between the gender test score gap and the structure of the education system.

4. 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 484,030 observations across all ages and countries. However, we exclude students who do not report their sex, test scores, and age. This reduces the sample by 38,195 students. 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 theses students as

[8] Section 6 further discusses the importance of including controls for female labor market conditions that may confound the relationship between the test score gap and academic structureseaming.

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 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 socio-economic 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[9] (in eighth grade models only),[10] and a continuous measure for the number of people residing in the child's household. Unfortunately, there is fair amount of non-reporting for some of the socioeconomic controls, and 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 dummy variable indicating missing data. 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 United States is a good example.

[9] We have collapsed the maternal and paternal education categories 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.
[10] Parental education is not reported for fourth graders in any country or eighth graders in England and Japan.

In Section 6 the TIMSS data are supplemented with education system data from *Education at a Glance* (2004) and *Eurydice* (2002). In particular, we measure the degree of streaming in an education system in three ways: The grade when students are streamed for the first time, the percentage of students in the academic stream in grade 10, and the percentage of people aged 25-34 who are university graduates. These variables are reported in Appendix Table 1. Some specifications also include average math or science test scores calculated from the TIMSS, public expenditures on education and the female-male university enrollment ratio from the *Education at a Glance* (2004), private school enrollment rates and the percentage of female teachers at the secondary level from the *Global 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).

5. The gender gap in math and science

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.

5.1. Baseline 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.[11] 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

[11] 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.

subsets of countries that one might have thought would be similar – Canada/U.S. and Finland/Norway/Sweden are good examples.

One must be careful, however, when interpreting the gender coefficients reported in columns 5 and 7. In particular, while these point estimates tell us the average observed female-male test score difference, they are only an unbiased estimate of the "gender" gap if all factors omitted from equation (1) are uncorrelated with gender. In other words, unobservable family characteristics and educational opportunities must be uncorrelated with gender for the female coefficient to be interpreted as the gender gap. As discussed earlier, this is unlikely to be true.

5.2. 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 purge the estimates of the impact of differential class assignment, to the extent that it is captured by current class assignment. More specifically, we add a vector of class indicators to the model (see equation (2)). Including class indicators controls for differences in educational opportunities and human capital accumulation across classes thereby purging the gender gap estimates of bias induced by non-gender neutral class assignment.

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 is the distance between the OLS and fixed effects estimates so different 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. More specifically, we construct a simple measure of gender-biased sorting by regressing class rank on a female indicator.

$$R_{cgti} = \theta_0 + \theta_1 F_{cgti} + \theta_2 X_{cgti} + \nu_{cgti} \tag{4}$$

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).[12] $\theta_1 = 0$ if, on average, male and female students are placed in equal ranking 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 (4) estimates. 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 gender-biased 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

[12] The results are similar if classes are ranked using average male scores instead of overall average scores.

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 (4).[13]

The negative slope depicted in Figure 1 indicates that countries that place more girls in better classes have more negative fixed effects estimates compared to their OLS estimates. In other words, once class placement is held constant, the increasingly poor performance of girls compared to boys in the same classes as class placement becomes increasingly pro female-biased is revealed. For example, the five most rightward circles in panel A are Flemish Belgium, Germany, Hungary, 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-biased class placement.[14]

In summary, it appears that the positive sorting of girls into better classes in many countries masks the severity of the gender gap. Stated somewhat differently, in the absence of positive female-biased 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 (2005) 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 much more similar magnitude in grade 4. At this stage, we do not have a good explanation for the difference in the size of the eighth grade math and science gaps.

[13] 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.
[14] In contrast, using longitudinal data for 1477 students (grades 4-7) in Northern California, 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.

5.3. 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 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 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 gender-biased 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, 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 fixed FE-OLS gap and negative female sorting is associated with a smaller FE-OLS gap.

6. Explaining 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 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 wide range of gap magnitudes observed across OECD countries. As such, differences in the structure of the education systems, economies, or cultures across OECD countries must play an important role. In terms of educational structure differences, we hypothesize that streaming and gender-biased class sorting might play important roles. In fact, they may even have an impact before they formally occur. For example, if girls view the probability that they will participate in advanced math and science classes, or a career requiring advanced math or science skills, as low they may invest less effort in studying 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. It is also possible that girls placed in higher ranked classes in heavily streamed education systems may be at a disadvantage because females tend to perform less well in competitive environments (Gneezy, Niederle, and Rustichini 2001). 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 first step is to ask whether such a relationship exists, and whether or not it persists once other cultural and labor market conditions are accounted for.

We assess the extent to which the degree of streaming and gender-biased sorting impact the gender test score gap using the β_{cg} estimates reported in Tables 2 and 4 combined with the degree of gender-biased sorting and the percentage of students educated in mixed-gender classes reported in Table 3 and the educational structure data from *Education at a Glance* (2004) and *Eurydice* (2002) to estimate equation (3). In particular, we hypothesize that β_{cg} depends heavily streamed the education system is, measured by the percentage of students in the academic stream in grade 10,[15] as well as the extent to which girls are placed in better classes relative to similarly skilled boys. The results for grade eight are reported in the top panel of Table 5. The base model includes the percent of tenth grade students enrolled in the academic stream, the degree of female-biased class placement (from Table 3), the percent of students educated in mixed-gender classes (reported in Table 3), the mean male test score for each country, and a constant. The estimates for this specification are reported in columns (1) and (5), for math and science respectively. The average male test score is included to allow for the possibility that higher scoring countries may have larger or smaller gender gaps. As can be seen in Table 5, a 10 percentile increase in the average male test score decreases the female-male test score gap by 0.2 percentiles for math and 0.6 for science. If anything, higher scoring countries have smaller gender gaps.

The percent of students in the academic stream and the degree of female-biased class placement are the primary variables of interested. For the base specification, a 10 percentage point increase in the fraction of students enrolled in the academic stream in grade 10 reduces the female-male test score gap by 0.20 for math and 0.24 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 increase in the academic stream size decreases the female-male test score gap by 0.4 standard deviation in math and 0.5 standard deviations in science. In a similar vein, countries that place girls in better classes also have smaller female-male gaps. More specifically, a one standard deviation increase in *FM* decreases the female-male gap by 0.42 standard deviations for math and 0.26 standard deviations for science.

Columns 2 and 5 add other educational structure and economic variables to check the robustness of the results. Public expenditures on education as a fraction of GDP and private 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

[15] We use also use the age at which formal streaming occurs and the percentage of people completing university or tertiary education to measure streaming (see columns 6,7,13, and 14 in Tables 6 and 7).

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 with or without these additional variables are included, the science results become more imprecise.

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 that 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 increase in the grade at which streaming occurs decreases the female-male test score gap by 0.47 standard deviation in math and 0.38 standard deviations in science. Columns 4 and 8 replace the percent of students in the academic stream in grade 10 with percent of people who complete university or tertiary training (this ranges from 12 to 51 percent). Again, a one standard deviation increase in the percentage of people completing tertiary training decreases the female-male test score gap by 0.58 standard deviation in math and 0.72 standard deviations in science.

Perhaps even more interesting than the finding that more heavily streamed countries have bigger female-male test score gaps at the eighth grade level, is the finding 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 impact of streaming is precisely estimated and of a similar magnitude to grade eight. More specifically, based on the results reported in columns 2-4, a one standard reduction in streaming leads to a 0.75, 0.68, and 1.24 standard deviation decrease in the female-male math test score gap for the three streaming measures respectively. Similarly for science, one standard reduction in streaming leads to a 0.88, 0.61, and 0.95 standard deviation decrease in the female-male test score gap for the three streaming measures.

Overall, the results point to a large impact of streaming on the female-male test score gap at young ages. The results also suggest that the gender gap is substantially reduced by female-biased class/stream placement in some countries. Further, since the relationship between streaming and the female-male gap arises before formal streaming occurs it must partly be working through indirect channels, such as family/teacher/peer interactions or student perceptions about the importance of math and science.

7. The gender gap distribution

Up to now, we have focused on the average gender gap. However, one might also wonder how the gap distribution differs across countries. One of Summers's points is that there may be fewer women in the far right tail of the math/science distribution, where Ph.D.'s are likely drawn from. In order to examine this, Tables 6a and 6b report the estimates for equation (1) at the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} quantiles, and Table 7 replicates equation (3) at each of the same quantiles.

Focusing first on the quantile gender gap estimates for grade eight reported in Table 6a, at least two aspects of the Table warrant comment. First, countries with small average gaps tend to have small gaps at all points in the gap distribution while countries with large mean gaps have large gaps at all points in the distribution. For example, the Swedish math gap is less than -1.3 at all quantiles while the Czech math gap ranges from -2.3 to –5.8. This is an important finding because it shows that it is not just at the top end that girls lag behind boys. Second, there are two general gap shapes across quantiles: flat and downward sloping (with a bit of upward slope at the 90^{th} quantile in some cases). While we have already established that countries with more selective streaming structures have bigger gaps, Table 6a raises the question of whether countries with certain types of educational institutions tend to have steeper gap gradient across deciles.

Table 6b reports the same set of estimates for fourth graders. As we know from section 5, the main difference between the fourth and eighth grades is the slightly smaller average math gap and the substantially smaller mean science gap among fourth graders. While the gap magnitudes differ, the pattern across deciles is very similar across age groups. Countries with flat gradients across deciles in grade eight have flat gradients in grade four and countries with negatively sloped gradients in grade eight typically have negatively sloped gradients in grade four.

Finally, to assess the extent to which the degree of streaming and gender-biased sorting affects the gender test score gap across quantiles, Table 7 reports the estimates for equation (3) at the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} quantiles. All models include the set of regressors included in columns 2 and 5 in Table 5. The standout feature of Table 7 is the consistency of the impact of educational structures across quantiles in most cases. The one exception is eighth grade science. In this case, streaming has a much larger impact ofn the female-male test score gap at high quantiles than at low quantiles. The consistency across quantiles is somewhat surprising as one might have expected the impact to be generally stronger at one end of the distribution or the other. In contrast, female-biased class sorting has a similar impact across quantiles for eighth grade math, has no impact on fourth grade math at any quantile, and only has a positive impact on science at high quantiles at both grade levels.

8. Conclusion

While there seems to be a general consensus that boys score better than girls in math and science, estimating the magnitude of the gap and uncovering the mechanisms causing the gap have proven incredibly difficult. The results reported in this paper contribute the stock of knowledge about gender skill gaps in two ways. First, we show that countries with education systems that use very selective academic streams have larger gender gaps, even before streaming occurs. Second, while it is true that the biggest test score gender gaps are at high test score deciles in some countries, the gender gap is nearly constant across deciles in other countries. Combined, these results clearly suggest that institutional structures are capable of manipulating gender-specific skill accumulation patterns, and hence observed gender skill gaps.

References

American Association of University Women, *How Schools Shortchange Girls*, Washington, D.C.: AAUW Educational Foundation (1996).

Baker, David P., and Deborah Perkins Jones, "Creating Gender Equality: Cross-national Gender Stratification and Mathematical Performance," *Sociology of Education*, Vol.66, No.2 (1993), 91-103.

Benbow, Camilla Persson, and Julian C. Stanley, "Sex Differences in Mathematical Ability: Fact or Artifact?" *Science*, Vol.210, No.4475 (1980), 1262-1264.

Blum, Deborah, *Sex on the Brain: The Biological Differences between Men and Women*, Viking Press (1997).

Brandell, Gerd, Peter Nystrom, and Christina Sundqvist, "Mathematics-A Male Domain?" Unpublished paper submitted to ICME 10 (2006).

Braswell, James S., Anthony D. Lutkus, Wendy S. Griggs, Shari L. Santapau, Brenda S.-H. Tay-Lim, and Matthew S. Johnson, *The Nation's Report Card: Mathematics 2000*, U.S. Department of Education, Office of Educational Research and Improvement (NCES 2001-517), National Center of Education Statistics Washington, DC (2001).

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

Cahill, Larry, "His Brain, Her Brain," *Scientific American*, Vol.292, No.5 (2005), 40-47.

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.

Carr, Martha, Jessup L. Donna, and Diana Fuller "Gender Differences in First-Grade Mathematics Strategy Use: Parent and Teacher Contributions," *Journal for Research in Mathematics Education*, Vol.30, No.1 (1999), 20-46.

Casey, M.Beth, Ronald L. Nuttall, and Elizabeth Pezaris, "Spatial-Mechanical Reasoning Skills Versus Mathematics Self-Confidence as Mediators of Gender Differences on Mathematics Subtests Using Cross-National Gender-Based Items," *Journal for Research in Mathematics Education,* Vol.32, No.1 (2000), 28-57.

Chipman, Susan. F., Sandra P. Marshall, and Patricia A. Scott, "Content Effects on Word Problem Performance: A Possible Source of Test Bias?" *American Educational Research Journal*, Vol.28 (1991) 897-915.

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

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov, “Interpreting the Evidence on Life Cycle Skill Formation,” NBER Working Paper No. W11331, (2005).

Dee, Thomas S, “Teachers and the Gender Gaps in Student Achievement,” NBER Working Paper 11660 (2005).

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.

Ethington, Corinna A., and Lee M. Wolfle, “A Structural Model of Mathematics Achievement for Men and Women,” *American Educational Research Journal*, Vol.23, No.1 (1986), 65-75.

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

Fitch, R.Holly., Steve Miller, and Paula Tallal, “Neurobiology of Speech Perception,” *Annual Review of Neuroscience*, Vol.20 (1997), 331-353.

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).

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini, “Performance in Competitive Environments: Gender Differences,” *Quarterly Journal of Economics*, Vol.118, No.3 (2003), 1049-1074

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).

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.

Hilderbrand, Gaell M, "Redefining Achievement," in *Equity in the Classroom: Towards Effective Pedagogy for Girls and Boys*, Edited by Patricia F. Murphy and Caroline V. Gipps, Falmer and UNESCO, Washington, D.C., (1996), 149-171.

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.

Jacobs, Janis and Jacquelynne S. Eccles, "The Impact of Mothers' Gender-Role Stereotypic Beliefs on Mothers' and Children's Ability Perceptions," *Journal of Personality and Social Psychology*, Vol.63, No.6 (1992), 932-944.

Jensen, Arthur R., "Sex Differences in Arithmetic Computation and Reasoning in Pre-pubertal Boys and Girls," *Behavioral and brain sciences*, Vol.11 (1988), 198-199.

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.

Keller, Johannes, and Dirk Dauenheimer, "Stereotype Threat in the Classroom: Dejection Mediates the Disrupting Threat Effect on Women's Math Performance," *Personality and social Psychology Bulletin*, Vol.29, No.3 (2003), 371-381.

Kimura, Doreen, "Are Men's and Women's Brains Really Different?" *Canadian Psychology*, Vol.28 (1987), 133-147.

Kimura, Doreen, "Sex Differences in the Brain," *Scientific American*, Vol. 268, No.3 (1992), 118-125.

Kimura, Doreen, *Sex and Cognition*, Cambridge, MA: MIT Press (1999).

Kuhn, Peter, and Catherine J. Weinberger, "Leadership Skills and Wages," *Journal of Labor Economics*, Vol.23, No.3 (2005), 395-436.

Lauzon, Darren, "Gender Differences in Large Scale, Quantitative Assessments of Mathematics and Science Achievement," Unpublished paper presented in the Statistics Canada-John Deutsch Institute-WRNET conference, Canada, Ottawa (2001).

Lavy, Victor, "Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment," NBER Working Paper 10678 (2004).

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

Marsh, Herbert W., and Alexander S. Yeung, "Longitudinal Structural Equation Models of Academic Self-Concept and Achievement: Gender Differences in the Development of Math and English Constructs," *American Educational Research Journal*, Vol.35, No.4 (1998), 705-738.

Marshall, Sandra P., and Julie D. Smith, "Sex Differences in Learning Mathematics: A Longitudinal Study with Item and Error Analyses," *Journal of Educational Psychology*, Vol.79 (1987), 372-383.

Moore E.G. J., and A. W. Smith, "Sex and Ethnic Group Differences in Mathematical Achievement: Results from the National Longitudinal Study," *Journal for Research in Mathematics Education*, Vol.18 (1987), 25-36.

Mullis, Ina V.S., Michael O. Martin, Albert E. Beaton, Eugenio J. Gonzalez, Dana L. Kelly, and Teresa A. Smith, *Mathematics and Science Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Study*, Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, Chestnut Hill, MA, (1998).

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).

Neisser, Ulric, Gwyneth Boodoo, Thomas J. Bouchard, A.Wade Boykin, Nathan Brody, Stephen J. Ceci, Diane F. Halpern, John C. Loehlin, Robert Perloff, Robert J. Sternberg, and Susana Urbina, " Intelligence: Knowns and Unknowns," *American Psychologist*, Vol.51 (1996) 77-101.

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).

O'Sullivan, Christine Y., Mary A. Lauko, Wendy S. Grigg, Jjahe Qian, and Jinming Zhang, *The Nation's Report Card: Science 2000*, U.S. Department of Education, Institute of Education Sciences (NCES 2003-453), National Center of Education Statistics, Washington, DC (2003).

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

Phillips, Krista, and Irwin Silverman, "Differences in the Relationship of Menstrual Cycle Phase to Spatial Performance on Two- and Three-Dimensional Tasks," *Hormones and Behaviors*, Vol.32 (1997), 167-175.

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

Rose, Heather and Julian R. Betts, "The Effect of High School Courses on Earnings," *Review of Economics and Statistics*, Vol. 86, No. 2 (2004): 497-513.

Sadker, Myra and David Sadker, *Failing at Fairness: How Our Schools Cheat Girls*, New York: Touchstone (1994).

Silverman, Irwin, Don Kastuk, Jean Choi, and Krista Phillips, "Testosterone and Spatial Ability in Men," *Psychoneuroendocrinology*, Vol.24 (1999), 813-822.

Tiedemann, Joachim, "Parents' Gender Stereotypes and Teachers' Beliefs as Predictors of Children's Concept of Their Mathematical Ability in Elementary School," *Journal of Educational psychology*, Vol.92, No.1 (2000), 144-151.

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

Wiseman, Alexander W., David P. Baker, and Francisco O. Ramirez, “Opportunity Structures, Gender Differences and Achievement: Further Exploration of Cross-national Mathematics Achievement and Gender in TIMSS,” Unpublished paper presented at the annual meeting of the comparative and international education society, Toronto (1999).

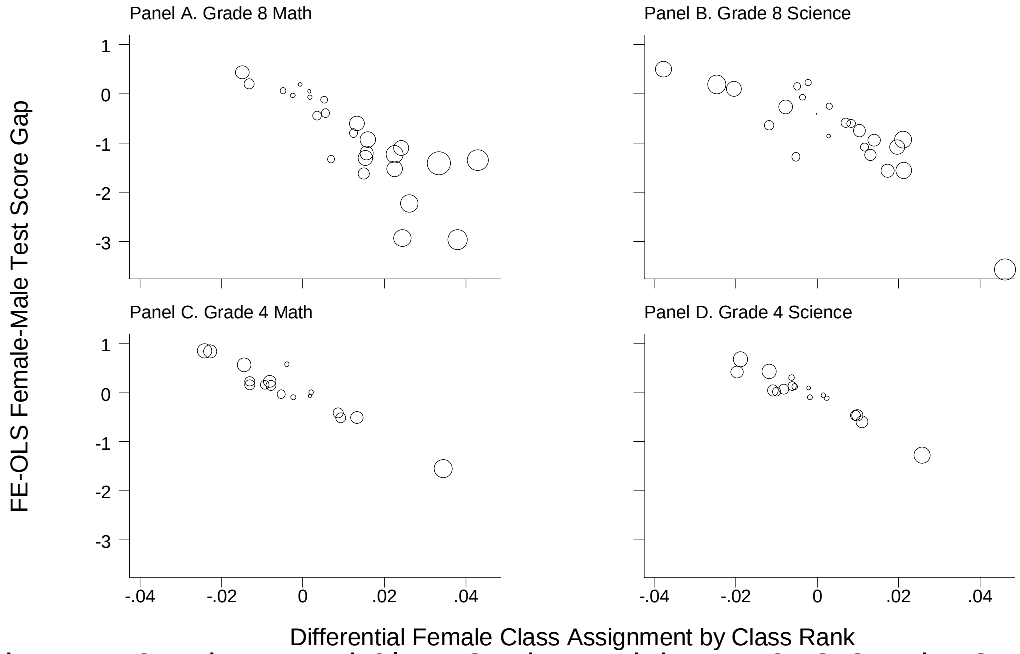

Figure 1. Gender-Based Class Sorting and the FE-OLS Gender Score Gap

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.

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

Population weighted. All models inlcude the variables listed in Section 4.

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.

Table 5. Explaining the Gender Test Score Gap

	Math				Science			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grade 8								
Percent Academic	***1.97***	**2.55**	***0.26***	**7.21**	**2.43**	1.69	0.22	**9.40**
	(1.02)	(0.96)	(0.13)	(3.18)	(1.18)	(1.07)	(0.14)	(3.06)
Pro-Female Sort	**33.95**	**45.61**	**57.45**	**47.38**	13.71	25.48	***34.52***	**36.38**
	(16.56)	(16.95)	(20.28)	(17.90)	(16.58)	(16.14)	(17.82)	(14.43)
% Mixed Classes	2.93	3.47	1.18	2.05	1.58	2.17	0.51	1.50
	(2.24)	(2.34)	(2.42)	(2.35)	(2.83)	(2.72)	(2.69)	(2.26)
National Male Score	0.02	**0.08**	***0.07***	0.04	***0.06***	**0.12**	**0.12**	**0.11**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Public Educ Expend		**132.57**	***100.11***	**136.93**		100.27	94.54	**144.45**
		(50.78)	(54.64)	(53.50)		(59.01)	(58.93)	(52.84)
Priv School Enroll Rate		-2.54	-2.23	-2.11		-2.57	-2.29	**-2.87**
		(1.65)	(1.77)	(1.72)		(1.62)	(1.64)	(1.37)
% Female Teachers		-2.85	-1.13	-0.89		-3.04	-2.10	-1.41
		(3.58)	(3.91)	(3.80)		(4.11)	(4.13)	(3.54)
F/M Univ Enroll Ratio		1.45	1.49	0.42		**3.67**	**3.62**	**2.60**
		(1.18)	(1.27)	(1.38)		(1.29)	(1.30)	(1.18)
Female LFP Rate		-2.76	-1.96	-2.68		**-11.01**	**-11.00**	**-12.37**
		(3.65)	(3.96)	(3.82)		(4.10)	(4.10)	(3.50)
GDP		-0.07	-0.07	-0.13		0.05	0.03	-0.05
		(0.07)	(0.07)	(0.08)		(0.08)	(0.08)	(0.07)
Adjusted R-Squared	0.22	0.39	0.30	0.34	0.13	0.38	0.38	0.55
Sample Size	28	27	27	27	28	27	27	27
Grade 4								
Percent Academic	**2.35**	**2.59**	**0.29**	**11.75**	**2.55**	**3.83**	**0.33**	**11.43**
	(0.62)	(0.69)	(0.10)	(3.20)	(1.06)	(0.93)	(0.11)	(3.39)
Pro-Female Sort	**37.50**	-11.31	-13.17	***66.12***	9.22	28.33	-70.01	-75.13
	(13.87)	(26.88)	(31.41)	(32.77)	(31.64)	(60.30)	(64.57)	(57.93)
% Mixed Classes	-1.83	***-8.86***	***-10.94***	3.81	1.59	4.54	-6.79	-2.70
	(2.53)	(4.00)	(4.66)	(5.54)	(4.04)	(5.89)	(5.90)	(5.74)
National Male Score	0.01	***0.09***	***0.11***	-0.04	0.04	0.05	***0.24***	***0.22***
	(0.02)	(0.04)	(0.05)	(0.05)	(0.04)	(0.10)	(0.10)	(0.09)
Public Educ Expend		***97.58***	93.98	**119.51**		**174.95**	**198.84**	**295.10**
		(45.14)	(52.80)	(48.24)		(45.65)	(58.43)	(62.94)
Priv School Enroll Rate		-2.08	-2.26	1.23		-1.25	-2.84	***-3.61***
		(1.60)	(1.87)	(1.84)		(1.67)	(1.95)	(1.69)
% Female Teachers		-5.58	-6.27	1.27		-0.38	-3.48	-1.51
		(3.00)	(3.46)	(3.89)		(3.44)	(4.02)	(3.85)
F/M Univ Enroll Ratio		0.63	1.81	-1.02		-1.61	3.29	1.56
		(1.24)	(1.31)	(1.54)		(2.56)	(2.40)	(2.40)
Female LFP Rate		1.29	3.99	***-10.61***		***-13.72***	-1.06	-7.87
		(3.31)	(3.76)	(5.16)		(6.52)	(6.38)	(6.62)
GDP		***-0.23***	**-0.26**	-0.16		-0.07	***-0.28***	**-0.38**
		(0.09)	(0.11)	(0.08)		(0.12)	(0.14)	(0.13)
Adjusted R-Squared	0.63	0.75	0.66	0.74	0.15	0.77	0.62	0.69
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.

Table 6a. Grade 8 Math and Science Quantile Gender Estimates

	Math					Science				
	10th (1)	25th (2)	50th (3)	75th (4)	90th (5)	10th (6)	25th (7)	50th (8)	75th (9)	90th (10)
Australia	**1.5**	-0.2	**-1.6**	**-2.8**	**-1.9**	**-1.2**	**-3.9**	**-1.6**	**-2.8**	**-1.9**
	(0.5)	(0.5)	(0.8)	(0.5)	(0.5)	(0.6)	(0.7)	(0.8)	(0.5)	(0.5)
Austria	-0.6	**-2.6**	**-2.7**	**-4.6**	**-4.1**	**-2.6**	**-5.4**	**-2.7**	**-4.6**	**-4.1**
	(1.2)	(1.2)	(1.3)	(1.1)	(0.7)	(1.2)	(1.3)	(1.3)	(1.1)	(0.7)
Belgium - Flemish	**2.0**	-0.6	-0.6	-0.4	-0.3	**-2.2**	**-6.0**	**-0.6**	-0.4	-0.3
	(0.8)	(1.0)	(0.6)	(0.4)	(0.3)	(0.7)	(0.8)	(0.6)	(0.4)	(0.3)
Belgium - French	-1.1	**-3.5**	**-3.7**	**-2.9**	**-2.8**	**-1.8**	**-4.2**	**-3.7**	**-2.9**	**-2.8**
	(1.0)	(1.2)	(1.2)	(0.9)	(0.9)	(0.4)	(0.6)	(1.2)	(0.9)	(0.9)
Canada	0.9	-0.3	-0.4	-0.1	**-1.2**	**-1.6**	**-3.3**	-0.4	-0.1	**-1.2**
	(0.5)	(0.6)	(0.8)	(0.5)	(0.6)	(0.5)	(0.6)	(0.8)	(0.5)	(0.6)
Czech Republic	**-2.3**	**-3.9**	**-5.8**	**-5.5**	**-3.0**	**-8.0**	**-9.6**	**-5.8**	**-5.5**	**-3.0**
	(0.8)	(0.7)	(0.9)	(0.6)	(0.5)	(0.8)	(0.9)	(0.9)	(0.6)	(0.5)
Denmark	-1.5	**-3.4**	**-4.6**	**-4.5**	**-6.3**	**-1.7**	**-4.3**	**-4.6**	**-4.5**	**-6.3**
	(0.8)	(1.3)	(1.8)	(1.5)	(1.5)	(0.5)	(0.7)	(1.8)	(1.5)	(1.5)
England	-1.2	-1.1	**-3.4**	**-5.3**	**-3.4**	**-2.7**	**-6.2**	**-3.4**	**-5.3**	**-3.4**
	(0.6)	(0.8)	(1.0)	(0.9)	(0.7)	(0.8)	(1.1)	(1.0)	(0.9)	(0.7)
Finland	-0.8	1.0	-1.1	**-4.0**	-1.6	-1.2	**-4.1**	-1.1	**-4.0**	-1.6
	(1.4)	(1.6)	(1.6)	(1.1)	(1.1)	(1.1)	(1.6)	(1.6)	(1.1)	(1.1)
France	-1.0	-1.5	**-2.4**	**-2.5**	**-3.0**	**-1.4**	**-4.0**	**-2.4**	**-2.5**	**-3.0**
	(0.8)	(1.0)	(1.0)	(0.9)	(0.9)	(0.4)	(0.6)	(1.0)	(0.9)	(0.9)
Germany	0.4	-0.6	**-2.5**	**-4.2**	**-4.8**	-1.3	**-5.0**	**-2.5**	**-4.2**	**-4.8**
	(0.7)	(0.9)	(1.2)	(1.2)	(1.1)	(0.7)	(1.0)	(1.2)	(1.2)	(1.1)
Greece	0.2	-0.5	**-2.7**	**-5.1**	**-5.1**	**-1.6**	**-3.2**	**-2.7**	**-5.1**	**-5.1**
	(0.3)	(0.5)	(0.8)	(1.0)	(1.0)	(0.4)	(0.5)	(0.8)	(1.0)	(1.0)
Hungary	**1.2**	-0.2	-0.9	**-1.5**	**-1.5**	**-4.1**	**-6.9**	-0.9	**-1.5**	**-1.5**
	(0.6)	(0.7)	(0.6)	(0.5)	(0.4)	(0.7)	(0.7)	(0.6)	(0.5)	(0.4)
Iceland	0.5	-0.1	-0.9	**-2.2**	**-1.4**	-1.2	**-3.7**	-0.9	**-2.2**	-1.4
	(0.7)	(0.9)	(1.3)	(1.2)	(1.6)	(0.6)	(0.8)	(1.3)	(1.2)	(1.6)
Ireland	-0.5	**-3.8**	**-7.5**	**-7.5**	**-4.3**	**-1.2**	**-4.3**	**-7.5**	**-7.5**	**-4.3**
	(0.7)	(0.8)	(1.1)	(0.8)	(0.9)	(0.6)	(1.0)	(1.1)	(0.8)	(0.9)
Italy	-0.2	**-1.3**	**-4.0**	**-4.3**	**-4.2**	**-0.9**	**-2.9**	**-4.0**	**-4.3**	**-4.2**
	(0.3)	(0.5)	(0.6)	(0.8)	(0.9)	(0.3)	(0.6)	(0.6)	(0.8)	(0.9)
Japan	**3.3**	0.0	**-3.5**	**-2.2**	**-1.2**	0.0	0.0	**-3.5**	**-2.2**	**-1.2**
	(0.9)	(0.8)	(0.4)	(0.2)	(0.1)	(0.6)	(0.8)	(0.4)	(0.2)	(0.1)
Korea	-0.2	**-2.2**	**-1.8**	**-0.5**	**-0.1**	**-2.7**	**-2.2**	**-1.8**	**-0.5**	**-0.1**
	(1.1)	(0.8)	(0.4)	(0.1)	(0.1)	(0.8)	(0.8)	(0.4)	(0.1)	(0.1)
Netherlands	-0.9	**-2.2**	**-2.8**	**-3.5**	**-1.7**	**-2.4**	**-2.2**	**-2.8**	**-3.5**	**-1.7**
	(0.9)	(0.9)	(0.9)	(0.6)	(0.5)	(1.0)	(0.9)	(0.9)	(0.6)	(0.5)
New Zealand	**1.1**	**1.8**	0.9	**-1.3**	**-2.7**	-0.8	**1.8**	0.9	**-1.3**	**-2.7**
	(0.4)	(0.5)	(0.8)	(0.7)	(0.7)	(0.4)	(0.5)	(0.8)	(0.7)	(0.7)
Norway	-0.3	-0.8	**-1.7**	**-2.6**	**-2.7**	**-1.6**	-0.8	**-1.7**	**-2.6**	**-2.7**
	(0.4)	(0.6)	(0.7)	(0.9)	(1.0)	(0.4)	(0.6)	(0.7)	(0.9)	(1.0)
Portugal	**-0.6**	**-1.2**	**-1.8**	**-2.7**	**-4.9**	**-1.5**	**-1.2**	**-1.8**	**-2.7**	**-4.9**
	(0.2)	(0.3)	(0.6)	(0.8)	(1.1)	(0.2)	(0.3)	(0.6)	(0.8)	(1.1)
Scotland	-0.5	**-1.5**	**-3.9**	**-5.6**	**-4.7**	**-2.2**	**-1.5**	**-3.9**	**-5.6**	**-4.7**
	(0.4)	(0.5)	(0.6)	(0.9)	(0.8)	(0.4)	(0.5)	(0.6)	(0.9)	(0.8)
Slovak Republic	0.1	-0.9	**-2.2**	**-2.9**	**-2.1**	**-3.8**	-0.9	**-2.2**	**-2.9**	**-2.1**
	(0.6)	(0.6)	(0.6)	(0.6)	(0.4)	(0.5)	(0.6)	(0.6)	(0.6)	(0.4)
Spain	-0.5	**-1.1**	**-2.2**	**-3.4**	**-4.9**	**-3.1**	**-1.1**	**-2.2**	**-3.4**	**-4.9**
	(0.3)	(0.5)	(0.7)	(0.9)	(1.2)	(0.5)	(0.5)	(0.7)	(0.9)	(1.2)
Sweden	0.4	-0.4	-0.2	-0.6	**-1.3**	**-1.1**	-0.4	-0.2	-0.6	**-1.3**
	(0.5)	(0.7)	(0.7)	(0.6)	(0.6)	(0.5)	(0.7)	(0.7)	(0.6)	(0.6)
Switzerland	-1.2	**-3.0**	**-3.6**	**-3.3**	**-2.6**	**-2.7**	**-3.0**	**-3.6**	**-3.3**	**-2.6**
	(1.0)	(0.9)	(0.6)	(0.5)	(0.4)	(0.6)	(0.9)	(0.6)	(0.5)	(0.4)
United States	0.1	**-0.9**	**-3.1**	**-4.2**	**-3.4**	-0.1	**-0.9**	**-3.1**	**-4.2**	**-3.4**
	(0.3)	(0.4)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)	(0.5)	(0.5)	(0.4)

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

Table 6b. Grade 4 Math and Science Quantile Gender Estimates

	Math					Science				
	10th (1)	25th (2)	50th (3)	75th (4)	90th (5)	10th (6)	25th (7)	50th (8)	75th (9)	90th (10)
Australia	-0.2	-1.1	**-2.5**	**-3.0**	**-1.7**	1.0	**-1.9**	**-3.6**	**-3.8**	**-2.3**
	(0.6)	(0.8)	(0.8)	(0.6)	(0.6)	(0.7)	(0.7)	(0.8)	(0.7)	(0.5)
Austria	-1.6	**-3.3**	**-5.2**	**-4.1**	**-3.5**	**-1.9**	**-2.8**	**-6.0**	**-6.8**	**-4.3**
	(1.1)	(1.1)	(1.2)	(1.0)	(0.8)	(1.0)	(1.3)	(1.4)	(0.7)	(0.8)
Belgium - Flemish	-0.3	-2.8	-1.8	**-2.3**	**-1.9**	-0.4	-1.2	**-1.6**	**-4.6**	**-3.2**
	(1.2)	(0.9)	(1.2)	(0.7)	(0.4)	(1.0)	(1.1)	(1.2)	(1.0)	(0.9)
Canada	**-1.8**	**-3.2**	**-3.7**	**-3.2**	**-2.0**	-0.2	**-2.6**	**-3.0**	**-4.8**	**-2.6**
	(0.3)	(0.7)	(1.1)	(0.8)	(0.8)	(0.7)	(0.9)	(0.9)	(1.0)	(0.7)
Czech Republic	**-2.6**	**-3.0**	**-3.1**	**-2.0**	**-0.9**	**-2.2**	**-5.3**	**-6.5**	**-5.7**	**-3.9**
	(1.1)	(1.0)	(1.0)	(0.7)	(0.5)	(0.7)	(1.0)	(1.0)	(1.0)	(0.7)
England	0.6	-0.2	**-2.9**	**-4.7**	**-2.7**	1.2	**1.0**	**-2.3**	**-3.2**	**-2.5**
	(0.5)	(0.8)	(0.9)	(0.6)	(0.5)	(0.7)	(0.9)	(0.9)	(0.6)	(0.4)
Greece	-0.7	**-2.1**	**-2.6**	**-3.8**	**-2.5**	**-1.0**	**-2.6**	**-4.8**	**-5.5**	**-4.6**
	(0.4)	(0.6)	(0.6)	(1.0)	(1.0)	(0.4)	(0.6)	(1.0)	(1.0)	(1.1)
Hungary	0.7	-1.4	-1.2	**-1.3**	**-1.5**	**-1.2**	**-2.3**	**-4.9**	**-6.0**	**-3.8**
	(0.7)	(0.8)	(0.8)	(0.6)	(0.5)	(0.6)	(0.7)	(0.8)	(0.6)	(0.6)
Iceland	**-1.0**	**-1.7**	**-2.9**	**-4.9**	**-5.7**	-0.8	**-1.5**	**-4.1**	**-7.6**	**-6.6**
	(0.3)	(0.6)	(1.0)	(1.4)	(2.1)	(0.5)	(0.7)	(1.1)	(1.2)	(1.9)
Ireland	**2.3**	**0.8**	-0.2	**-1.9**	-1.1	1.0	-1.6	**-3.4**	**-5.4**	**-4.3**
	(0.9)	(0.9)	(1.1)	(0.9)	(0.7)	(0.7)	(0.9)	(1.0)	(1.0)	(0.9)
Italy	0.1	-1.4	**-5.0**	**-6.7**	**-3.3**	-1.2	-1.2	**-2.7**	-2.0	**-1.5**
	(0.8)	(1.2)	(1.5)	(1.2)	(1.0)	(1.0)	(1.1)	(1.5)	(1.5)	(0.8)
Japan	**2.3**	-1.0	**-2.2**	**-2.2**	**-1.1**	0.4	**-1.8**	**-3.1**	**-3.4**	**-2.0**
	(1.1)	(0.8)	(0.6)	(0.4)	(0.3)	(0.9)	(0.8)	(0.7)	(0.5)	(0.3)
Korea	**-3.5**	**-5.6**	**-5.0**	**-2.8**	**-1.4**	**-4.1**	**-5.0**	**-6.1**	**-4.6**	**-2.5**
	(1.1)	(1.2)	(0.9)	(0.5)	(0.3)	(1.3)	(0.9)	(0.8)	(0.7)	(0.4)
Netherlands	**-2.2**	**-3.5**	**-4.5**	**-2.3**	**-1.7**	**-3.9**	**-5.4**	**-5.3**	**-5.9**	**-4.4**
	(0.9)	(0.7)	(0.7)	(0.7)	(0.5)	(0.7)	(0.7)	(0.8)	(0.8)	(0.8)
New Zealand	**1.5**	**1.5**	-1.0	-1.1	-1.5	0.0	**3.1**	-0.3	**-2.0**	**-1.8**
	(0.5)	(0.5)	(1.0)	(1.0)	(0.8)	(0.5)	(0.9)	(0.9)	(0.9)	(0.5)
Norway	-0.5	**-1.1**	**-2.4**	**-4.1**	**-4.7**	0.0	**-1.3**	**-3.0**	**-3.0**	**-3.8**
	(0.4)	(0.6)	(0.8)	(0.8)	(1.0)	(0.3)	(0.6)	(1.0)	(0.9)	(0.9)
Portugal	-0.4	-0.4	**-2.0**	**-3.9**	**-6.2**	0.2	**-1.2**	**-3.1**	**-6.1**	**-5.7**
	(0.4)	(0.6)	(0.8)	(1.2)	(1.3)	(0.5)	(0.6)	(0.6)	(1.4)	(1.5)
Scotland	0.3	-0.8	**-3.8**	**-4.7**	**-3.3**	-0.2	**-2.5**	**-4.7**	**-5.7**	**-3.6**
	(0.5)	(0.5)	(0.7)	(0.8)	(0.7)	(0.5)	(0.7)	(0.9)	(0.9)	(0.6)
United States	0.3	-0.9	**-2.0**	**-2.0**	**-1.6**	**-1.9**	**-2.5**	**-4.7**	**-2.3**	**-2.2**
	(0.5)	(0.6)	(0.6)	(0.6)	(0.3)	(0.7)	(0.7)	(0.6)	(0.5)	(0.4)

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

Table 7. Explaining the Gender Test Score Gap Across Quantiles

	Math					Science				
	10th (1)	25th (2)	50th (3)	75th (4)	90th (4)	10th (5)	25th (6)	50th (7)	75th (7)	90th (8)
<u>Grade 8</u>										
Percent Academic	**2.27**	***2.50***	***2.46***	**1.95**	**3.59**	0.46	1.31	***1.88***	***2.55***	**4.24**
	(0.94)	(1.25)	(1.19)	(0.83)	(1.26)	(0.90)	(0.84)	(1.02)	(1.41)	(1.84)
Pro-Female Sort	**46.71**	***42.03***	**56.21**	**43.53**	16.04	11.40	4.47	18.69	30.98	**77.84**
	(17.34)	(23.91)	(22.24)	(13.10)	(17.63)	(15.28)	(14.00)	(15.90)	(20.13)	(22.90)
% Mixed Classes	1.41	3.38	***5.94***	**5.11**	**8.71**	-0.36	2.78	3.79	2.51	1.26
	(2.12)	(2.84)	(3.36)	(2.24)	(3.28)	(2.30)	(2.10)	(2.62)	(3.67)	(4.28)
National Male Score	**0.08**	**0.08**	**0.10**	**0.09**	**0.17**	0.00	-0.03	**0.11**	**0.23**	**0.30**
	(0.03)	(0.04)	(0.04)	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)
Public Educ Expend	34.75	54.00	**158.16**	**182.77**	**171.60**	80.35	***97.86***	**146.04**	64.08	53.79
	(48.94)	(66.93)	(70.28)	(50.21)	(71.03)	(57.29)	(52.20)	(56.40)	(74.83)	(94.36)
Priv School Enroll Rate	**-3.77**	**-4.39**	***-4.27***	-2.17	-3.83	**-2.82**	***-3.35***	**-3.87**	-1.61	0.58
	(1.29)	(1.86)	(2.45)	(1.70)	(2.58)	(1.23)	(1.37)	(1.70)	(2.19)	(2.29)
% Female Teachers	-2.86	-2.86	-4.45	***-7.77***	-1.02	**-7.99**	-4.96	-4.96	1.91	7.54
	(2.81)	(3.91)	(5.34)	(4.01)	(6.16)	(3.11)	(3.09)	(4.02)	(5.80)	(7.23)
F/M Univ Enroll Ratio	**2.48**	**4.10**	2.17	1.16	-1.20	**4.05**	**3.74**	**3.74**	2.61	0.08
	(1.04)	(1.43)	(1.56)	(1.03)	(1.45)	(1.03)	(0.92)	(1.32)	(1.68)	(2.09)
Female LFP Rate	***-8.50***	-2.48	-3.44	-3.88	-4.42	***-7.62***	**-9.14**	**-8.67**	**-12.68**	**-17.34**
	(4.30)	(5.28)	(5.18)	(3.03)	(4.30)	(3.89)	(3.48)	(3.67)	(5.03)	(6.05)
GDP	0.04	-0.04	-0.11	***-0.13***	-0.05	**0.13**	0.08	-0.01	0.04	0.04
	(0.05)	(0.08)	(0.10)	(0.07)	(0.11)	(0.06)	(0.06)	(0.07)	(0.10)	(0.13)
Adjusted R-Squared	0.46	0.35	0.45	0.61	0.49	0.65	0.68	0.40	0.63	0.78
Sample Size	27	27	27	27	27	27	27	27	27	27
<u>Grade 4</u>										
Percent Academic	**2.98**	**3.67**	**2.42**	1.62	**2.76**	**4.05**	***3.49***	***3.56***	**3.82**	**3.40**
	(1.15)	(0.80)	(0.94)	(1.30)	(1.06)	(1.07)	(1.50)	(1.74)	(0.85)	(0.47)
Pro-Female Sort	-56.60	32.46	-47.24	-16.69	-23.98	-3.79	-104.03	93.48	**132.35**	**76.16**
	(35.05)	(26.12)	(35.85)	(62.05)	(55.90)	(63.47)	(90.40)	(111.50)	(53.26)	(31.87)
% Mixed Classes	-8.73	0.77	**-15.58**	-9.08	-3.36	-2.70	-6.34	8.41	**18.21**	**12.85**
	(5.07)	(3.82)	(5.23)	(9.20)	(8.10)	(6.45)	(9.19)	(11.07)	(4.95)	(2.90)
National Male Score	**0.11**	-0.01	***0.13***	0.14	0.16	0.09	0.20	-0.08	-0.07	0.04
	(0.04)	(0.04)	(0.06)	(0.10)	(0.09)	(0.11)	(0.15)	(0.18)	(0.09)	(0.05)
Public Educ Expend	**176.48**	**167.48**	***136.50***	37.64	17.20	**158.86**	**329.75**	148.93	73.51	-14.37
	(71.13)	(49.69)	(59.99)	(92.18)	(81.07)	(59.86)	(67.73)	(90.00)	(39.63)	(23.04)
Priv School Enroll Rate	**-6.43**	-2.40	-3.43	-3.53	-5.29	-2.89	***-4.55***	1.35	0.30	0.45
	(1.68)	(1.57)	(2.29)	(3.39)	(2.94)	(1.54)	(2.35)	(2.93)	(1.46)	(0.92)
% Female Teachers	**-9.93**	-3.72	***-8.45***	***-12.95***	-9.75	-2.72	-4.58	0.67	3.81	**6.17**
	(3.66)	(2.92)	(4.20)	(6.58)	(5.92)	(3.24)	(4.73)	(6.08)	(2.85)	(1.97)
F/M Univ Enroll Ratio	-1.76	-1.96	1.77	1.38	-2.62	-0.36	0.36	-1.37	***-4.72***	***-2.70***
	(2.28)	(1.43)	(1.76)	(2.34)	(1.91)	(2.95)	(4.21)	(4.99)	(2.03)	(1.22)
Female LFP Rate	-7.10	**-12.76**	7.95	11.99	-1.00	-3.37	-4.94	-16.94	**-26.30**	**-14.17**
	(5.70)	(3.44)	(4.37)	(6.77)	(5.69)	(7.11)	(10.42)	(12.24)	(5.54)	(3.40)
GDP	**-0.30**	-0.16	**-0.35**	-0.28	-0.19	-0.17	***-0.40***	0.08	0.16	***0.14***
	(0.12)	(0.09)	(0.12)	(0.19)	(0.17)	(0.13)	(0.18)	(0.22)	(0.10)	(0.06)
Adjusted R-Squared	0.75	0.81	0.67	0.41	0.81	0.76	0.66	0.35	0.91	0.97
Sample Size	17	17	17	17	17	17	17	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 (2002), 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 (2005)	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 (2000)	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 et al (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 et al (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 et al (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 (2004)	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 Boys	Test score Girls	Gender gap		Score range	Remarks
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.