NBER WORKING PAPER SERIES

AN EMPIRICAL ANALYSIS OF THE GENDER GAP IN MATHEMATICS

Roland G. Fryer, Jr
Steven D. Levitt

Working Paper 15430
http://www.nber.org/papers/w15430

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2009

We are grateful to Glenn Ellison, Claudia Goldin, Lawrence Katz, and Luigi Zingales for helpful comments and suggestions. Katherine Barghaus, Dimitrios Batzilis, Diana Chang, Corinne Espinoza, Kenneth Mirkin, Wonhee Park, Katherine Penner, Erin Robertson, Jörg Spenkuch, and David Toniatti provided exceptional research assistance. The authors are grateful to the Education Innovation Laboratory at Harvard University and the Becker Center for Chicago Price Theory for generous support. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

An Empirical Analysis of the Gender Gap in Mathematics
Roland G. Fryer, Jr and Steven D. Levitt
NBER Working Paper No. 15430
October 2009
JEL No. I20

ABSTRACT

We document and analyze the emergence of a substantial gender gap in mathematics in the early years of schooling using a large, recent, and nationally representative panel of children in the United States. There are no mean differences between boys and girls upon entry to school, but girls lose more than two-tenths of a standard deviation relative to boys over the first six years of school. The ground lost by girls relative to boys is roughly half as large as the black-white test score gap that appears over these same ages. We document the presence of this gender math gap across every strata of society. We explore a wide range of possible explanations in the U.S. data, including less investment by girls in math, low parental expectations, and biased tests, but find little support for any of these theories. Moving to cross-country comparisons, we find that earlier results linking the gender gap in math to measures of gender equality are sensitive to the inclusion of Muslim countries, where in spite of women's low status, there is little or no gender gap in math.

Roland G. Fryer, Jr
Department of Economics
Harvard University
Littauer Center 208
Cambridge, MA 02138
and NBER
rfryer@fas.harvard.edu

Steven D. Levitt
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
slevitt@midway.uchicago.edu

The gender gap in mathematics is an important and extremely divisive issue of academic debate (e.g. Sweeney 1953, Fennema and Sherman 1977, Goldin 1994, Hausmann et. al 2008.) Figure 1 plots the gender gap on the mathematics and verbal components of the Scholastic Aptitude Test (SAT) – over the past forty years. On the math section, female scores are on average 0.30 standard deviations lower than male scores; on the verbal portion there is no clear gender difference (College Board 2007).[1] An important shortcoming of the SAT data is that the population taking the test is not representative, and sample selection may occur differently across gender. For instance, since college attendance rates are presently higher for females, the female sample of SAT takers may be drawing more heavily from the middle or left tail of the ability distribution. Data from the National Assessment of Educational Progress (NAEP), a nationally representative sample that does not have sample selection problems, also shows boys consistently outperforming girls in fourth and eighth grade over the last two decades, though the magnitude of the gap is smaller (Lee, Grigg, and Dion 2007). The bulk of the evidence in the past 50 years suggests that the gender gap in mathematics does not exist before children enter school, but is large and significant in the middle school years and beyond. For instance, in a meta-analysis of 100 studies with a total sample of more than 3 million students, Hyde et al (1990) found a .29 standard deviation gender gap in math in high school.[2]

The patterns on math tests are especially striking when one considers that females either systematically outperform males or have made enormous gains on many educational dimensions. The high school dropout rate is 28% for females compared to 35% for males (Greene and

---

[1] Among elite achievers, these differences are even more pronounced. Men outnumber women by more than two to one above the 99th percentile in SAT mathematics scores (College Board 2007). Males also score four percent higher on AP calculus exams and 6 percent higher on AP science exams (Freeman 2004, College Board 2007).

[2] For surveys of the literature, see Fennema 1974, Hyde et. al 1990, Maccoby 1966, Maccoby and Jacklin 1974).Additional papers of note include Benbow and Stanley 1980, 1983, Benbow 1988, Dee 2005, Entwisle et al. 1994, Halpern et al., 2005, Hyde 2005, Hyde et al 1990, and Hyde and Mertz 2009.

Winters 2006). As noted by Goldin, Katz, & Kuziemko (2006), in 2003 there were 1.35 females graduating from four-year colleges for every male. In stark contrast, in 1960 there were 1.6 males graduating from 4-year colleges for each female. In 1970, women made up only 9% of combined Medicine, Dentistry, and Law degree recipients. Thirty years later, women accounted for 47% of full time, and 44% of part-time students pursuing such degrees (Freeman 2004). Women make up 45% of all doctorate degrees (Freeman 2004). A 2000 study, commissioned by the U.S. Congress, found that "[t]he large gaps in educational attainment that once existed between men and women have in most cases been eliminated" (Bae et al. 2000).

A wide range of theories have been explored to explain the gender gap in mathematics. One strand of the literature looks to biological differences. These biological theories argue that innate differences in spatial ability, higher order thinking, or brain development produce a gap in achievement (Wilder and Powell 1989).[3] Another group of researchers emphasize societal factors as the cause of the gap.[4] Societal explanations focus on how girls are socialized into believing that math is not important, useful, doable, or part of the identity of a girl (Wilder and Powell 1989).[5] Gaining a better understanding of the gender gap in mathematics is an issue of first-order importance. Paglin and Rufolo (1990) argue that the observed difference in

---

[3] Research in this domain includes work on differences in brain composition (Gur et al 1999; De Bellis et al 2001; Cahill 2005; Gallagher and Kaufman 2005), differences in hormone levels (Davison and Susman 2001), differences in strategy (Carr and Jessup 1997; Fennema and Carpenter 1998; Kucian et al 2005), and differences in spatial ability (Johnson and Meade 1987; Witelson 1976, Lawton and Hatcher 2005).

[4] Gneezy, Nierderle, and Rustichini (2001) argue then boys males are more competitive, leading to better test performance. Others argue that sex-typing in the wording of items plays a role (McLarty, Noble, and Huntley 1989; Kepner and Koehn 1977; Moss and Brown 1979). Voyer, Rodgers, and McCormick 2004 argue that males do better in timed settings. Math Anxiety on the part of females is another possibility (Hsiu-Zu et al 2000; Tobias 1976.)

[5] These theories take a number of forms: differential parental treatment or expectations (Eccles and Jacobs 1986; Parsons et al 1982; Muller 1998; Bouffard and Hill 2005; Bhanot and Jovanovic 2005), differential treatment by teachers (Dweck et al 1978; Heller and Parsons 1981; Leinhardt et al 1979; Parsons et al 1982), stereotype threat (Spencer et al 1999; Brown and Joseph 1999; O'Brien and Crandall 2003), and other environmental factors (Gneezy et al 2003; Levine et al 2005).

mathematical ability between women and men explains much of male-female differences in occupational choice and wages among recent college graduates.

In this paper, we utilize the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K) to shed new light on the gender gap in mathematics. ECLS-K is a data set administered by the Department of Education. The survey covers a sample of more than 20,000 children from roughly 1,000 schools entering kindergarten in the fall of 1998. An enormous amount of information is gathered for each individual, including family background, school and neighborhood characteristics, teacher and parent assessments and expectations, and test scores, which allows us to test several important theories for gender differences within a unified framework. The original sample of students has been subsequently re-interviewed in the spring of kindergarten, first grade, third grade, and fifth grade.

The results we obtain using these new data are informative and, in some cases, quite surprising. Consistent with the prior literature, when children enter kindergarten, girls and boys are observationally equivalent in both math and reading. By the end of fifth grade, however, girls have fallen more than 0.2 standard deviations behind their male counterparts in math.[6] The math gap is equivalent to 2.5 months of schooling. Girls are losing ground in math in every region of the country, every racial group, all levels of the socio-economic distribution, every family structure, and in both public and private schools. By the end of the sample, girls do significantly worse than boys on every math skill tested. Underperformance by girls is evident not just in mean test scores, but also in the upper tail of the math distribution. On entry to kindergarten, girls make up 45 percent of the top five-percentiles in math test scores; by the end of fifth grade

---

[6] Over that same time span there is little change in the reading gender gap, which consistently favors girls.

just 28 percent of the top five percent are female. Girls are underrepresented in the bottom tail of the math distribution in kindergarten, but overrepresented in the bottom tail by fifth grade.

Due to limitations of the data, we can test only a subset of the possible socialization theories for the divergent trajectory of girls' math scores in the early years of school, and none of the biological explanations. Among those hypotheses that we can test, we fail to uncover compelling support for any of them. Although teachers tend to rate girls more favorably than test scores would predict, girls lose nearly as much ground on subjective teacher ratings of math ability as they do on standardized tests, suggesting that the poor relative performance by girls is not simply an artifact of standardized testing. We attempt to test socialization hypotheses in a number of ways. Parental expectations regarding math are lower for girls than boys even after accounting for test scores, but controlling for these expectations does nothing to reduce the gender gap. We also find that girls with mothers working in math-related occupations lose just as much ground as those whose mothers are not in math-related occupations, making it unlikely that low familial expectations for girls in math lie at the root of the issue.[7] Parents report spending equal amounts of time with boys and girls doing math-related activities. As a result, including these variables has no effect on the gender gap.

If broader societal forces are working to undermine girls in math, then one might expect to see females fall further behind in states with greater levels of gender inequality in wages, employment, or education. Again, including these variables as covariates does little to alter the gender gap. An important caveat is that some socialization forces, such as the media, operate on an even broader level. Jacobs and Eccles (1985) argue that the mass media has the ability to be a

---

[7] We define a "mathematical occupation" as: engineers, surveyors, & architects; and natural scientists & mathematicians. Roughly one percent of the students in our sample have a mother with a math occupation and four percent of fathers are in math occupations.

socializing force that can both alter parental responses to stereotypes (sometimes reinforcing, sometimes counteracting) and affect the salience of various issues (gender stereotyping included).

Having exhausted our ability to explain the gender gap in math using ECLS and concerned about the appropriate level of aggregation to test socialization, we turn our attention to data from the Trends in International Mathematics and Science Study (TIMSS 2003) and Program for International Student Assessment (PISA 2003), the only two datasets with which one can make reasonable comparisons of math achievement across countries. PISA is a triennial world-wide test of academic achievement among 15 year olds. TIMMS is an international assessment of the mathematics and science knowledge of fourth and eighth grade students from around the world. Following Guiso et al. (2008), we investigate the relationship between the gender math gap and societal level female socialization as measured by the World Economic Forum's (WEF) gender gap index which reflects economic and political opportunities, education, and well-being for women.[8]

We are able to replicate the findings of Guiso et al. (2008) using PISA data: there is a strong positive association between the WEF measure of female opportunity and the relative performance of girls in math. In stark contrast, however, there is no such relationship between gender equality (as measured by the WEF index) and female math performance in TIMMS. The difference in results across these two datasets is driven by the samples of countries included; when one restricts TIMMS to the same countries as PISA, the positive relationship reemerges. The primary difference between the two data sets is that TIMMS includes a large number of

---

[8] The World Economic Forum constructs this index and makes it available at http://www.weforum.org/en/initiatives/gcp/Gender%20Gap/index.htm. In 2007 it ranged between .4510 (Yemen) and .8146 (Sweden), where bigger numbers indicate less gender inequality.

Middle Eastern countries that are absent in PISA. Surprisingly, although these Middle Eastern countries have a high degree of gender inequality, there is no gender gap in math on average in these places.

Our paper contributes to the gender gap literature in three important ways. First, our data are more recent, nationally representative, and remarkably rich relative to the previous literature – containing five assessments over the first six years of schooling. This not only provides the most accurate picture of the gender gap in early schooling, it allows us to test many of the socialization theories in a single unified framework. Second, contrary to the previous literature, we find gender gaps in mathematics that arise as early as third grade and are present in every strata of society. Third, our analysis of cross country data show the sensitivity of previous results to the inclusion of a larger set of countries. Fourth, we are the first to show that countries with high gender inequality and same-sex schools have no gender gap.

The remainder of the paper is structured as follows. Section II describes the data used in the analysis. Section III presents the basic facts and patterns in test scores in the first six years of school using these data. Section IV investigates the extent to which a variety of alternative hypotheses can account for the fact that girls are steadily losing ground in mathematics. Section V explores the patterns in two cross-country data sets. Section VI concludes.

II. The Data

The primary dataset used in this paper is the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K) – a nationally representative sample of over 20,000 children entering kindergarten in 1998. Information on these children has so far been gathered at six separate points in time. The full sample was interviewed in the fall and spring of kindergarten, spring of first grade, spring of third grade, and spring of fifth grade. A random sample of one-fourth of the respondents was also interviewed in the fall of first grade. The sample will eventually be followed through eighth grade. Roughly 1,000 schools are included in the sample, with an average of more than twenty children per school in the study.

A wide range of data is gathered on the children in the study, as described in detail at the ECLS website http://nces.ed.gov/ecls. We utilize just a small subset of the available information in our baseline specifications.

Summary statistics for the variables we use in our core specifications are displayed by gender in Table 1. Students who are missing data on test scores or gender are dropped from our sample. To keep the sample constant, we further restrict the sample to students who have valid test scores in all waves.[9] Our primary outcome variables are math and reading standardized test scores.[10] The math test evaluates number recognition, counting, comparing and ordering numbers, solving word problems, and interpreting picture graphs. The reading test includes questions designed to measure basic skills (print familiarity, letter recognition, beginning and ending sounds, rhyming sounds, and word recognition), vocabulary and comprehension, listening and reading comprehension, knowledge of the alphabet, and phonetics. The values reported in the table are item response theory (IRT) scores provided in ECLS-K, which we have transformed

---

[9] We have ensured that this sample restriction is of no consequence for our results by comparing our sample with regressions in which we only exclude students who don't have valid test scores in each wave. The results of this exercise are available from the authors upon request.

[10] These tests were developed especially for the ECLS, but are based on existing instruments including Children's Cognitive Battery (CCB); Peabody Individual Achievement Test-Revised (PIAT-R); Peabody Picture Vocabulary Test-3 (PPVT-3); Primary Test of Cognitive Skills (PTCS); and Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R).

to have mean zero and a standard deviation of one for the overall sample on each of the tests and time periods.[11] In all instances sample weights provided in ECLS-K are used.[12]

In the fall of kindergarten math test, male students score -.004 standard deviations below the mean in math and female students score .005 standard deviations above the mean, yielding a trivial achievement gap in favor of girls. By the spring of their 5th grade year, the math picture shifts dramatically. Male students score .096 above the mean in mathematics and girls score .101 below the mean – roughly a .2 standard deviation gap in achievement. In reading, girls start .15 standard deviations ahead of boys and maintain almost all of their initial advantage.

The remainder of Table 1 presents summary statistics for the other variables used in the analysis. The appendix contains precise definitions and details of the construction of these variables. In contrast to the test scores, for which we have observations at multiple points in time, most of the control variables are either collected only once (typically kindergarten fall, but in some cases kindergarten spring), or exhibit little variation over time for individual students. Thus we report only a single entry for these variables, rather than separate values for each wave.

For the purely demographic variables (e.g. socio-economic status of the family, whether the mother works in a math occupation, etc.), mean differences across gender are small because children's gender is more or less randomly assigned across families. This need not be the case for the covariates which have a behavioral component. For instance, boys in the sample are

---

[11] Because children were asked different questions depending on the answers they provided to the initial questions on the test, IRT-adjusted scores are preferable to simple test-score measures reflecting the number of correct answers a child provided. For more detail on the process used to generate the IRT scores, see chapter 3 of the ECLS-K Users Guide. Tests were administered orally to the full sample in the fall of kindergarten and in the spring of first grade. Tests in the third and fifth grade combined oral and written components. Tests in the fifth grade are more writing intensive.

[12] Because of the complex manner in which the ECLS-K sample is drawn, different weights are suggested by the providers of the data depending upon the set of variables used. We utilize the weights recommended for making longitudinal comparisons (i.e. C1_6FC0). None of our findings are sensitive to other choices of weights, or not weighting at all.

slightly older than girls, implying that parents are more likely to delay kindergarten entry for boys. Parents also report having lower expectations for their daughters when it comes to math. There is no evidence, however, of large differences across gender in the mean frequency that parents engage in math or reading activities with the child, although they believe girls will complete more years of schooling.

To compare the gender gap across countries, we use data from the Program for International Student Assessment (PISA) and Trends in Mathematics and Science Study (TIMMS). Countries included in either sample are described in Appendix Table 1. PISA is a triennial world-wide test of academic achievement among fifteen year olds (roughly ninth grade). Every period of assessment focuses on one particular subject, but also tests the other main areas studied. The implementation of PISA is coordinated by the Organization for Economic Cooperation and Development (OECD). The first PISA exam was administered in 2000 and focused on reading literacy. The 2003 assessment specialized in mathematics (testing real life situations in which mathematics is useful). Given our focus is on mathematics achievement, we focus on the 2003 PISA. Over 275,000 students from 41 countries took the two-hour handwritten PISA test in 2003, including all 30 OECD countries.[13]

TIMMS is an international assessment of the mathematics and science knowledge of fourth and eighth grade students from around the world. TIMMS was developed by the International Association for the Evaluation of Educational Achievement to allow participating nations to compare student's educational achievement across borders. TIMMS was first administered in 1995, and has been given every 4 years thereafter. To make comparisons with PISA, we use the 2003 TIMMS cohort which contains 47 countries.

---

[13] Britain's data collection did not meet PISA's quality standards and all UK data was subsequently omitted.

TIMMS consists of an assessment of mathematics and science, as well as student, teacher, and school questionnaires. The assessment includes topics that students have likely been exposed to up to and including grades 4 and 8. A key difference between PISA and TIMMS is that PISA asks students to apply their mathematical knowledge to solve problems in various real-world contexts. TIMMS, on the other hand, measures more traditional classroom context such as an understanding of fractions and decimals and the relationship between them. The TIMMS assessment divides mathematical domains into two dimensions: cognitive and content. The cognitive domains that the assessment covers are "Knowing Facts and Procedures, Using Concepts, Solving Routine Problems and Reasoning." The content domains are "Number, Algebra, Measurement, Geometry, and Data."

Following Guiso et al, (2008), to test the impact of country level socialization on the gender achievement gap, we link the above data on achievement with the World Economic Forum's Global Gender Index (WEF-GGI). The index is comprised of four sub-indices which measure economic participation and opportunity, educational attainment, political empowerment, health and survival. For example, the Economic Participation and Opportunity Sub-index is a weighted average of the ratio of female to male: (1) labor force participation, (2) income, (3) legislators, senior officials, and managers, (4) wage equality for similar work, and (5) professional and technical workers. More details of the index are described in Appendix Table 2.

III. The gender gap in test scores over the first six years of school

Table 2 presents a series of estimates of the gender gap in the ECLS data for the tests taken in the fall of kindergarten through the spring of fifth grade. The specifications estimated are of the form

(1)

where *i* indexes students. *Gender* is an indicator variable. In our most fully parameterized models, the vector of other covariates, denoted X_i, includes an exhaustive and mutually exclusive set of race indicators, age and age squared, socio-economic status, a categorical variable for the mother's age at birth, and whether or not the family participates in the WIC program – similar to the parsimonious specification estimated in Fryer and Levitt (2004). In all instances, the estimation is done using weighted least squares, with weights corresponding to the sampling weights provided in the data set.

The first five columns of Table 2 report the raw gender gap from the time students enter kindergarten through the spring of their fifth grade year. These numbers parallel those found in Table 1. Over the first year of school, the gender gap in math is statistically indistinguishable from zero. By the end of first grade the gap increases to .076 standard deviations and is marginally significant. In the spring of their third grade year, female students are doing .205 standard deviations worse than their male counterparts. Given the progress of a typical student over the course of a school year, this amounts to roughly 2.5 months of schooling. The gap remains roughly constant between third and fifth grade and is similar in magnitude to the gaps on the mathematics section of the SAT reported in the Introduction.

The last five columns in Table 2 include the full set of controls. The gender gap story is little changed when the covariates are added. Boys and girls remain indistinguishable upon entry to kindergarten, but steadily lose ground through the end of 3rd grade. The magnitude of the gender gap after controlling for other factors is slightly larger than the raw gap: .226 standard deviations.

The magnitude and sign of the other covariates appear plausible. Black and Hispanic students perform significantly worse than whites; Asians perform significantly better. Age, age of the child's mother at birth, and socioeconomic status are all positively related to achievement. WIC participation is negatively correlated with test scores, suggesting that it is capturing aspects of poverty not detected by our SES measures, rather than any true underlying benefits of the program.

Table 3 provides a parallel analysis on reading achievement. Female students enter school ahead of boys in reading and these differences persist through fifth grade in raw scores as well as after controlling for our set of controls. As was the case with math, including the controls slightly improves the performance of boys relative to girls.

Although not shown in tabular form, we have explored the sensitivity of our findings to different choices of weights (including giving all observations equal weight); the gender gap in math and reading are virtually unchanged. Similarly, employing an alternative test score measure (T-scores, which are norm-referenced measurements of achievement), has almost no impact on the results.

Thus far we have concentrated on the mean gender math gap that emerges over the first six years of life. It is also informative to investigate the distribution of math achievement and how it evolves as children age. The ratio of girls to boys in the full sample is .949. If girls were equally represented throughout the achievement distribution, we would expect any cut of the distribution to have a girl to boy ratio of approximately .949.

Table 4 reports the ratio of girls to boys in different parts of the raw test score distribution. Upon kindergarten entry, the ratio of girls to boys above and below the median is similar to the overall sample. By the spring semester of their fifth grade year, the distribution has shifted in favor of the boys: the ratio of girls to boys above the median is .80 and below is 1.125. The tails of the distribution are even more pronounced. The ratio of girls to boys in the top 5% of the distribution is .811 when they enter kindergarten and .381 in the end of fifth grade.[14]

IV. Why are girls losing ground in mathematics in ECLS-K?

As a first attempt at understanding the forces driving the divergence in test scores across genders, Table 5 examines how the gender gap in math varies across sub-samples of the data.[15] To the extent that either the absolute magnitude of the gap, or the degree of widening varies along observable dimensions, it may shed light on the sources of the gap. The first two columns of Table 5 report the estimated gender gap in math scores upon entry to kindergarten and at the end of fifth grade, respectively. Column 3 is the difference between those numbers, i.e. the amount of ground lost or gained by girls over the first six years of school. In all cases, the values reported in the table are the coefficients on the indicator variable for gender from regressions including the full set of controls. Each entry in columns 1 and 2 comes from a separate regression.

Each row of the table corresponds to a different subset of the data. The top row of the table presents the baseline results using the full sample. These numbers match the corresponding numbers presented in Table 2. The next five rows divide the sample by race of the child. Upon entry to kindergarten (column 1), we cannot reject gender equality of math scores for any of the racial groups. Girls of every race lose ground (column 3), with the biggest losses relative to boys observed for whites and the smallest losses for Asians. Asians are the only race for which we

---

[14] Ellison and Swanson (2009) use the American Math Competition tests to study the far right-hand tail of the math distribution, finding a large gender gap among very high math achievers.

[15] We have also analyzed differences across sub-samples in reading test scores, and these are available from the authors on request.

cannot reject equality of scores for boys and girls at the end of fifth grade, in part due to large standard errors because of the small number of Asians in the data.

The next five rows segment the sample into quintiles using the composite SES measure. Again we cannot reject equality of math scores across genders upon entry to school, but in every quintile of the SES distribution girls trail boys substantially six years later. The greatest decline for girls relative to boys is in the top quintile of SES, where girls lose almost 0.3 standard deviations.

The next three breakdowns of the data are by family structure, region, and urbanicity, all of which confirm our original results. We observe little difference in the trends across marital status, or mother's age at the time of the child's birth. Looking regionally, girls lose the most ground in New England and fare the best in the South. The patterns look similar in cities and suburbs; girls in small towns are not losing quite as much ground.

Girls appear to fall further behind in private schools than in public schools, more when the mother is highly educated, and more when the mother is out of the labor force over the years the child goes from kindergarten to fifth grade. It does not appear to matter whether the mother is in a math-related occupation (engineer, architect, scientist, e.g.) or whether the mother is more highly educated than the father.

In summary, Table 5 demonstrates that girls are losing ground in every category we examine. If anything, the losses are greatest at the top of the SES/educational spectrum. Otherwise, there is little in Table 5 that points to a likely source for the phenomenon, suggesting instead that whatever forces are at work are quite widespread.[16]

---

[16] Ellison and Swanson (2009) come to a similar conclusion when analyzing the far right-tail of the distribution of math achievers.

Are the results simply an artifact of standardized testing?

Given the potential difficulties of evaluating student achievement using standardized tests in children so young, one possibility is that our results are simply an artifact of standardized testing. To assess this hypothesis, we examine the relative performance of children on an alternative measure of math skill: subjective teacher evaluations. Teachers were asked to answer 20 questions about the child's academic performance, ranking them on a scale of "Not Yet" to "Proficient." These answers were then transformed into IRT scores. As was done with test scores, these subjective assessments have been re-normalized to have mean zero and a standard deviation of one. We run regressions controlling for the same set of covariates as with test scores. In addition, because of concerns about heterogeneity across teachers in the way they may rate students on these subjective evaluations, we include teacher-fixed effects. Thus, the estimates are based on a student's evaluation relative to other students in the same classroom. The results, reported in Appendix Table 3, show that teachers rate girls .075 standard deviations (se=.05) higher than boys in math in the fall of kindergarten, but .172 standard deviations (se=.06) behind boys by the end of fifth grade.[17] The change in teacher assessments of roughly .25 standard deviations for girls relative to boys is remarkably similar in magnitude to the decline observed on standardized tests, suggesting that our findings are not simply an artifact of the particular tests in the ECLS.

Do girls spend less time studying math and thus fall behind?

---

[17] On teacher assessments of reading skills, girls go from having a .169 standard deviation advantage (se=.03) in fall of kindergarten to .261 standard deviations (se=.04) in spring of fifth grade. Thus girls gain roughly one-tenth of a standard deviation in reading relative to boys over the same time period that they lose .25 standard deviations in math.

While some of the gains that students experience as they age may be due purely to brain maturation, it seems likely that much of the improvement in mathematical performance is due to human capital investment. If girls are investing less than boys at these ages, it could explain why they are losing ground. Unfortunately, ECLS-K contains only a very crude proxy for the time and effort devoted to math – the parent's response to a question about how many times per week the child engages in math activities. We do not observe any clear gender differences on this variable. Parents report that their girls spend 2.27 days per week on math related activities and their boys spend 2.29 days per week.[18] While we cannot rule out this explanation, we have no affirmative support for it.

Can low parental expectations explain why girls fall behind in math?

As noted in the summary statistics in Table 1, parents have lower math expectations for their daughters, which could adversely affect female math trajectories. When we include parental expectations with respect to math as a covariate in our regressions, however, the gender gap decreases by at most .04 standard deviations. In some of our regressions it even increases slightly. This lack of impact on parental expectations is consistent with the absence of an impact of being in a family where the mother has more education than the father, or where the mother works in math-related professions.

---

[18] Even if we did observe gender differences, there is no systematic relationship between responses to this question and the child's math score in the spring of fifth grade. The question may simply not be capturing effectively human capital investment. Alternatively, it may be the case that parents work more with those students who are most in need of remedial help in math, so that reverse causality obscures the positive impact of the investment.

Does the material tested change as children age in a manner that lowers the relative performance of female students?

If gender differences vary across different mathematical subjects, then the measured performance of girls might decline over time simply due to changing composition of the test. [19] Table 6 reports the estimated gender gap in the probability of proficiency on a particular math skill by grade. These regressions control for the same set of covariates as in earlier tables. We report coefficients only for those grade/math skill combinations where there is substantial variation across children, e.g. by spring of first grade almost all students have mastered counting so no entries are made in the table only in kindergarten for counting. Standard errors are reported in parentheses. In kindergarten and first grade, girls outperform boys on some aspects of the test and do worse on other portions. By third grade, however, boys have higher proficiency rates on each of the math skills examined. The same is true for fifth grade. Gender differences in math performance on these tests are widespread; they are not concentrated in just one or two types of math.[20]

In summary, our search for sharp explanations for a gender gap in mathematics within ECLS-K must be judged a failure. Our tests of a wide range of a priori plausible stories generate little strong support for any particular causal pathway. Rather, it appears that girls are losing ground at a relatively similar magnitude more or less regardless of their circumstances. There is

---

[19] In personal correspondence, ECLS reports that the fraction of the exam devoted to each set of skills remains constant as children age. In the fall of kindergarten, few children are expected to correctly answer questions involving multiplication and division, and by the spring of first grade, few children are expected to miss questions involving counting. Even though the same share of the test is devoted to each set of skills, the portion of the test that provides the identifying variation changes over time. If almost all students (or, equivalently, virtually none of the students) have mastered a task, then there is little variation in scores on that part of the test, and consequently it does not contribute to measured gender differences.

[20] In contrast, girls outperform boys in every facet of reading in every grade.

one major caveat: we are unable to account for societal level socialization factors that operate at a macro level.

V. Cross-country analysis of the gender gap in mathematics

Having learned what we can from the ECLS-K, we conclude our analysis by studying cross-country data on gender differences in math. In a recent paper, Guiso et al. (2008) use the PISA exam of fifteen year olds to estimate gender gaps in math across forty countries. Largely consistent with the ECLS-K data for the United States, they find that boys outperform girls in math in the great majority of countries, whereas girls outperform boys in reading in every country in the sample. Guiso et al. (2008) further demonstrate that the gender gap in math scores is strongly negatively correlated with various measures of gender equality such as the World Economic Forum's Gender Gap Index (WEF-GGI) and an index of cultural attitudes toward women based on questions in the World Values Surveys.

In more gender-equal countries like Norway and Sweden, the gender gap disappears completely. The results from Guiso, et. al. (2008) suggest an important role for a country's culture influencing the relative performance of girls on both math and reading. Our inability to isolate strong determinants of the gender gap within a single country in the ECLS-K, while not direct support for this hypothesis, is consistent with their conclusion.

Figure 2 presents graphically the relationship in the raw PISA data between a country's gender gap in math (panel A), reading (panel B) and gender equality as measured by the WEF-GGI. As noted by Guiso et al. (2008) and as evidenced by the regression line superimposed on the data, a positive correlation (ρ=.424 for math and ρ=.407 for reading exists.

PISA is not the only cross-country test of math achievement. A second data set capturing international math patterns is TIMMS – developed by the International Association for the Evaluation of Educational Achievement to measure trends in students' mathematics and science achievement. There are 17 countries included in both PISA and TIMMS, and the WEF survey of gender equality. The correlation in math scores across PISA and TIMMS for that subset of countries is .89. Thus, it is not surprising that when we plot the TIMMS scores for these overlapping countries against gender equality in panel A of Figure 3, the observed patterns are similar to those presented earlier using PISA data.

What is surprising, however, is what happens to the relationship between the gender gap in math and gender equality when we use *all* the TIMMS sample for which WEF gender gap index data are available (41 out of 47 countries), as opposed to only the subset of countries that are also included in PISA.[21] A graph of the full data is also shown in panel A of figure 3. To facilitate comparison, the countries that are in both datasets are shown in, as is the regression line corresponding to those countries. The additional countries, along with the regression line that fits the whole data set, are shown in gray. Using the broader set of nations, any relationship between gender equality and the gender gap in math disappears. Indeed, in countries like Bahrain and Iran, which are among the worst in terms of gender equality, girls are actually outperforming boys on math, and this is due to relatively strong performance by girls, not an unusually bad showing among the boys. Consequently, the gender gap in math scores over this full sample actually disappears.[22]

---

[21] To be included in the WEF Survey of Gender Equality a country must have data on 12 out of 14 sub-indices of the WEF-GGI.

[22] The strong performance by girls in the set of countries with very low WEF scores does not appear to be due to selection, i.e. a smaller share of girls in school to be tested. See Appendix Table 1

Table 7 adds more structure to the discussion above by estimating a series of equations similar to (1) for the full sample of PISA (columns 1 through 3), the full sample of TIMMS (columns 4 through 6) and the set of 17 countries that are in both PISA and TIMMS (columns 7 through 9). For each sample, we estimate three equations. The first regresses the gender gap on theWEF-GGI measure. The second equation estimated adds GDP, and the third adds the gender ratio of test takers to the second equation to account for the possibility of sex-specific selection as to who is taking the test. The coefficient on WEF-GGI is positive and significant in PISA and in the set of countries where TIMMS and PISA overlap. In the full TIMMS sample, however, the coefficient on gender is small and insignificant,

There are a number of possible explanations for the stark difference between these samples. One possibility is that the results based on the PISA countries are spurious, and introducing more data reveals this to be the case. An intriguing alternative hypothesis emerges from another cut of the data. Using information provided in TIMMS, we identify those countries in which same-sex classrooms or schools are prevalent. For most of the world, almost all education is done in mixed-gender settings. In a handful of countries in TIMMS (Bahrain, Iran, Jordan, Palestine, and Saudi Arabia), virtually all secondary schooling is sex-segregated. In a few other countries there is a mix of same and mixed gender classrooms, but the majority of students are in same-sex classrooms (Egypt, Korea, Syria). The fact that the set of countries that have gender-segregated classrooms are largely Muslim countries confounds the analysis. In Figure 3B, these countries with a high-degree of sex-segregated education are shown in gray, whereas other countries are in black.[23] Regression lines for those two subsets of the data are also presented. When the countries with sex-segregated classrooms are not included, a positive

---

[23] One worry is that there is selection on gender in these countries as to who takes the test. Controlling for the gender ratio of test takers does not alter the results.

relationship between gender equality and relative female test scores once again emerges. The regression line for countries with single-sex education has a similar slope, but is shifted upwards by roughly 0.1 standard deviations .[24] Controlling for other country level covariates such as GDP does not alter the results. Although imprecisely estimated, it appears both that girls are doing better in the single-sex countries, and that the boys are doing worse.

While of course highly speculative, these cross-country data are consistent with the hypothesis that mixed-gender classrooms are a necessary component for gender inequality to translate into poor female math performance, although it is difficult to distinguish single-sex classrooms from Islamic religion in the data.[25]

VI. Conclusion

In this paper, we analyze the emergence of the gender gap in math using ECLS-K, a nationally representative data set from the United States. There are no mean differences between boys and girls upon entry to school, but girls lose one-fourth of a standard deviation relative to boys over the first six years of school. The ground lost by girls relative to boys is half the

---

[24] Because of the small sample of countries with sex-segregated classrooms and the fact that there is a limited amount of variation in the WEF gender equality index for these countries, the slope of the line for this subset of the data is imprecisely estimated. The difference in intercept relative to the broader set of countries, however, is highly statistically significant. When we include a dummy for countries having the majority of students in same sex classrooms in a regression on the whole TIMMS sample, the coefficient on that dummy is .16 with a standard error of .07.

[25] There are at least three reasons why sex segregated schools might have a positive influence on math achievement for girls. First, as discussed in Wilder and Powell (1989), differential treatment of male and female students by teachers may perpetuate stereotypes of gender roles (girls are more verbal, boys are more cognitive, e.g.). Second, single-sex education has, in some instances, been able to increase levels of self-esteem in girls (Haag 2000). Third, as demonstrated in Lee and Marks (1990), girls that underwent single-sex schooling were less likely to hold stereotypical views of gender roles even after single-sex schooling had concluded. There is a substantial literature examining the impact of same-sex schools and classrooms at the micro level (e.g. Lee and Bryk 1986, Lepore and Warren 1997, Younger and Warrington 2006), yielding mixed results. The identification strategies used in many of these studies is questionable, and those with any form of plausibly exogenous variation have had small samples that limit inference.

magnitude of the black-white test score gap that appears over these same ages. We document the presence of this gender gap across every strata of society. We explore a wide range of possible explanations in the U.S. data, including less investment by girls in math, low parental expectations, and biased tests, but find little support for any of these theories.

Although we are not successful at isolating the root causes of the gap in the ECLS-K data, we are able to test a number of potentially important prior explanations, finding scant support for any of these. For instance, our evidence suggests that the gender math gap is especially large among children who attend private schools, have highly-educated mothers, and have mothers working in math-related occupations -- all factors that one might think under some theories would be conducive to girls' success in math.

Data Appendix

We describe below how we combined and recoded some of the ECLS variables used in our analysis.

Socio-economic Composite Measure.

The socioeconomic scale variable (SES) was computed by ECLS at the household level for the set of parents who completed the parent interview in Fall Kindergarten or Spring Kindergarten. The SES variable reflects the socioeconomic status of the household at the time of data collection for spring kindergarten. The components used for the creation of SES were: Father/male guardian's education; Mother/female guardian's education; Father/male guardian's occupation; Mother/female guardian's occupation; and Household income.

Child's Age.

We used the Child's Age at Assessment Composite variable provided by ECLS. The Child's age was calculated by determining the number of days between the child assessment date and the child's date of birth. The value was then divided by 30 to calculate the age in months.

Mother's Age at First Birth.

Mothers were asked how old they were at the birth of their first child.

WIC Participant

Parents were asked whether their child received any benefits from the WIC program. WIC is a nutrition program aimed at low income mothers and children.

Mother/Father in Math Intensive Occupation

Parents were asked for their respective occupations.
We consider the following occupational categories in ECLS to be math intensive:"engineers, surveyors, & architects"; and "natural scientists & mathematicians."
The occupational categories not considered to be math intensive are: "executive, admin, managerial"; "social scientist/workers, lawyers"; "physicians, dentists, veterinarians"; "registered nurses, pharmacists"; "health technologists & technicians"; "technologists, except health"; "university teachers, postsecondary counselors"; "teacher, except postsecondary"; "writers, artists, entertainers, athlete";
"marketing & sales occupation"; "administrative support, including clerk"; "service"; "agriculture, forestry, fishing"; "mechanics & repairs"; "construction & extractive occupations"; "precision production occupation"; "production working occupation"; "transportation, material moving"; "handler, equip, cleaner, helpers, labor"; "unemployed or retired".

Mother's/Father's Occupational Prestige Score

ECLS provides a prestige score for the occupations of a child's parents. The raw score ranges from 29-78. We use this information to create two variables indicating the quintile into which a child's mother and father fall.

Teacher Evaluations

Teachers were asked 20 questions in which they rated a student's proficiency in several skills based on the teacher's past observation and experience with the child. The scale on which each item is rated ranges from 1 ("net yet [proficient]") to 5 ("proficient"). Item Response Theory is then used to combine the teacher's answers into three continuous scores, which reflect a child' achievement in mathematical thinking, language and literacy, and general knowledge. We use the math and language scores, after normalizing them to each have mean 0 and a standard deviation of 1.

Parental Expectations Educational Attainment
Parents were asked how far in school they expected their child to go. Answer choices were 1--"receive less than a high school degree", 2--"graduate from high school", 3--"attend two or more years of college", 4--"finish a 4-or-5-year college degree", 5--"earn a master's degree or equivalent", 6--"get Ph.D., or MD, or higher degree". We use this information to construct six mutually exclusive indicator variables.

Parental Expectations Relative Performance
Parents were asked how well they think their child is doing in math and reading, respectively, compared to other children in the same school. Answer choices ranged from 1 ("much worse") to 5 ("much better"). Based on this we construct five mutually exclusive indicator variables.

Frequency of Math Activities
Parents were asked how many times they, or any other family member, did math related activities with their child in the week after July 4th. Examples of math related activities were "learning numbers", "adding", "subtracting", and "measuring". Answer choices were "never", "once or twice", "three to six times", and "everyday".

Frequency of Reading Activities
Parents were asked how many times they, or any other family member, read books to their child in the week after July 4th. Answer choices were "never", "once or twice", "three to six times", and "everyday".

Wage Inequality Index
We use the location of the school each child attends (state), from the unrestricted ECLS-K data file, to merge the data with state-level information from the 2000 Census. Our measure of wage inequality is the ratio of full-time year-round female workers' median income in a particular state to that of their male counterparts.

Employment Inequality Index
We use the location of the school each child attends (state), from the unrestricted ECLS-K data file, to merge the data with state-level information from the 2000 Census. Our measure of gender inequality in employment is the ratio of employed females relative to employed males.

Educational Attainment Inequality Index

We use the location of the school each child attends (state), from the unrestricted ECLS-K data file, to merge the data with state-level information from the 2000 Census. Our measure of educational inequality is the ratio of females with a graduate education relative to males.

References

Bae, Y., S. Choy, C. Geddes, J. Sable, and T. Snyder. (2000). Trends in Educational Equity of Girls and Women. Washington, DC: National Center for Education Statistics.

Benbow, C.P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral & Brain Sciences, 11,* 11-62.

Benbow, C. P., & Stanley, J. C. (1980). "Sex differences in mathematical ability: Fact or artifact?" *Science*, 210, 1262-1264.

Benbow, C. P., & Stanley, J.C., "Sex Differences in Mathematical Reasoning Ability: More Facts," *Science,* 222, 1029-1031.

Bhanot, Ruchi, and Jasna Jovanovic. "Do Parents' Academic Gender Stereotypes Influence Whether They Intrude on their Children's Homework?" *Sex Roles,* Vol. 52, Nos. 9/10, May 2005.

Bouffard, Suzanne M. & Nancy E. Hill. (2005). "Maternal Perceptions of Competence and Children's Academic Adjustment: Longitudinal Relations across Early Elementary School". *Social Psychology of Education*. Nov 2005.

Brown, F. G. "Sex Bias in Achievement Test Items: Do They Have Any Effect on Performance?," Unpublished manuscript, Iowa State University, 1975.

Brown, R.P., & R.A. Joseph. (1999) "A burden of proof: Stereotype relevance and gender difference in math performance". *Journal of Personality and Social Psychology.* 76, 246-257.

Cahill, Larry. "HIS BRAIN, HER BRAIN". *Scientific American*: May 2005, Vol 292 Issue 5, pg. 40-47.

Carr, Martha, & Donna L. Jessup. (1997). "Gender Differences in First-Grade Mathematics Strategy Use: Social and Metacognitive Influences". *Journal of Educational Psychology.* 89, 318-328.

Davison, Kirsten Krahnstover, & Elizabeth J. Susman. (2001). "Are Hormone Levels and Cognitive Ability Related During Early Adolescence?". *International Journal of Behavioral Development.* 25, 416-428.

Dee, Thomas S. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?," *American Economic Review*, 95, 2 (2005), 158-165.

De Bellis, Michael D., Matcheri S. Keshavan, Sue R. Beers, Julie Hall, Karin Frustaci, Azadeh Masalehdan, Jessica Noll, and Amy M. Boring. (2001). "Sex Differences in Brain Maturation during childhood and Adolescence". *Cerebral Cortex.* 11, 552-557.

Dweck, Carol S., William Davidson, Sharon Nelson, and Bradley Enna, "Sex Differences in Learned Helplessness: II. The Contingencies of Evaluative Feedback in the Classroom and III. An Experimental Analysis," *Developmental Psychology*, 14, No. 3 (1978), 268-276.

Eccles, Jacquelynne S., & Janis E. Jacobs. (1986). " Social Forces Shape Math Attitudes and Performance". *Journal of Women in Culture and Society.* 11, 367-380.

Ellison, Glenn, and Ashley Swanson. (2009). "The Gender Gap in School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions," Unpublished Manuscript.

Fennema, Elizabeth, "Mathematics Learning and the Sexes: A Review," *Journal for Research in Mathematics Education*, 5, No. 3(May 1974), 126-139.

Fennema, Elizabeth & Julia Sherman (1977). "Sex-Related Differences in Mathematics Achievement, Spatial Visualization and Affective Factors" *American Educational Research Journal,* 14, No. 1, 51-71.

Fennema, Elizabeth, & Thomas P. Carpenter. (1998) "New Perspectives on Gender Differences in Mathematics: An Introduction" *Educational Researcher.* 27, 4-5.

Freeman, Catherine E. (2004). "Trends in Educational Equity of Girls & Women: 2004". U.S. Department of Education: National Center for Education Statistics. 2005-016.

Fryer, Roland G. & Steven D. Levitt, "Understanding the Black-White Test Score Gap in the First Two Years of School," *The Review of Economics and Statistics*, 86, No. 2(2004), 447-464.

Fryer, Roland G., Steven D. Levitt, and John A. List, "Exploring the Impact of Financial Incentives on Stereotype Threat: Evidence from a Pilot Study," *American Economic Review*, 98, No. 2 (2008), 370-375.

Gallagher, Ann M., and James C. Kaufman. (Eds). Gender Differences in Mathematics. Cambridge University Press, Cambridge: 2005.

Gneezy, Uri, Kenneth Leonard, and John A. List. forthcoming. "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society," *Econometrica*.

Gneezy, Uri, Murial Niederle, and Aldo Rustichini. (2003). "Performance in competitive Environments: Gender Differences". *Quarterly Journal of Economics,* 118, No. 3, 1049-1074.

Goldin, Claudia. 1994. Understanding the Gender Gap: An Economic History of American Women In *Equal Employment Opportunity: Labor Market Discrimination and Public Policy* edited by Paul Burstein, 17-26 (Piscataway, New Jersey: Aldine Transaction, 1994).

Goldin, Claudia, Larry Katz, & Ilyana Kuziemko (2006). “The Homecoming of American College Women: the reversal of the gender gap”. NBER working paper.

Goldstein, David, Diane Haldane, and Carolyn Mitchell, “Sex Differences in Visual-Spatial Ability: The Role of Performance Factors,” *Memory and Cognition*, 18, No. 5(1990), 546-550.

Greene, Jay P. and Marcus A. Winters (2006) “Leaving boys behind: public high school graduation rates” Manhattan Institute for Policy Research Civic Report No. 46.

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales, “Culture, Gender, and Math,” *Science*, 320, No. 5880 (2008), 1164-1165.

Gur, Ruben C., Bruce I. Turetsky, Mie Matsui, Michelle Yan, Warren Bilker, Paul Hughett, and Raquel E. Gur. "Sex Differences in Brain Gray and White Matter in Healthy Young Adults: Correlations with Cognitive Performance". *The Journal of Neuroscience*. 19, 4065-4072.

Haag, Pamela (2000) “K-12 single-sex education: what does the research say?” *ERIC Clearinghouse on Elementary and Early Childhood Education.*

Halpern, Diane F., Jonathan Wai, and Amanda Saw (2005). A psychobiosocial model: Why females are sometimes greater than and sometimes less than males in math achievement. In A.M. Gallagher & J.C. Kaufman (Eds.), *Gender differences in mathematics* (pp. 48-72). New York: Cambridge University Press.

Hausmann, Ricardo, Laura D. Tyson, and Saadia Zahidi, “The Global Gender Gap Report 2008,” World Economic Forum (2008).

Heller, Kirby A., Jacquelynne Eccles Parsons. “Sex Differences in Teacher’s Evaluative Feedback and Students’ Expectancies for Success in Mathematics”. *Child Development.* 52, 1015-1019.

Hsiu-Zu Ho, Deniz Senturk, Amy G. Lam, Jules M. Zimmer, Sehee Hong, Yukari Okamoto, Sou-Yung Chiu, Yasuo Nakazawa, and Chang-Pei Wang, “The Affective and Cognitive Dimensions of Math Anxiety: A Cross-National Study,” *Journal for Research in Mathematics Education*, 31, No. 3 (May 2000), 362-379.

Hyde, Janet Shibley (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581-592.

Hyde, Janet Shibley, Elizabeth Fennema, and Susan J. Lamon, “Gender Differences in Mathematics Performance: a Meta-Analysis,” *Psychological Bulletin*, 107, No. 2 (1990), 139-155.

Hyde, Janet. and Mertz, Janet, “Gender, Culture, and Mathematics Performance,” *Proceedings of the National Academy of Sciences*, 106 (22), p. 8801-8807.

Jacobs, Janice E. and Jacquelynn S. Eccles (1985). Gender difference in math ability: the impact of media reports on parents. *Educational Researcher, 14, No. 3*, 20-25.

Jimenez, Emmanuel and Marlaine E. Lockheed (1989). Enhancing girls’ learning through single-sex education: Evidence and a policy conundrum. *Educational Evaluation and Policy Analysis, 11, No. 2*, 117–142.

Johnson, Edward S., Ann C. Meade. (1987). "Developmental Patterns of Spatial Ability". *Child Development.* 58, 725-740.

Kepner, Henry S. and Lilane R. Koehn, “Sex Roles in Mathematics: a Study of the Status of Sex Stereotypes in Elementary Mathematics Texts,” *Arithmetic Teacher*, 24, No. 5 (May 1977), 379-385.

Kucian, Karin, Thomas Loenneker, Thomas Dietrich, Ernst Martin, Michael von Aster. (2005). "Gender Differences in brain activation patterns during mental rotation and number related cognitive tasks". *Psychology Science.* 47, 112-131.

Lawton, Carol A. & David W. Hatcher. (2005) "Gender Differences in Integration of Images in Visuospatial Memory". *Sex Roles,* 53, 717-726.

Lee, Jihyun, Wendy Grigg, and Gloria Dion, 2007, The Nation’s Report Card: Mathematics 2007, Washington DC: National Center for Education Statistics.

Lee, Valerie E. and Anthony S. Bryk (1986). Effects of single-sex secondary schools on student achievement and attitudes. *Journal of Educational Psychology, 73, No. 5*, 381–395.

Lee, Valerie E. and Helen M. Marks (1990). Sustained effects of the single-sex secondary experience on attitudes, behaviors, and sex differences. *Journal of Educational Psychology, 82* (3), 578-592.

Lepore, Paul C. and John Robert Warren (1997). A comparison of single-sex and coeducational Catholic secondary schooling: Evidence from the National Educational Longitudinal Survey of 1988. *American Educational Research Journal, 34, No. 3*, 485–511.

Levine, Susan C., Marina Vasilyeva, Stella F. Lourenco, Nora S. Newcombe, & Janellen Huttenlocher. (2005) "Socioeconomic Status Modifies the Sex Difference in Spatial Skill". *Psychological Science.* 16, 841.

Maccoby, E. 1966. *The Development of Sex Differences*. Stanford University Press, Stanford CA.

Maccoby, E., and Jacklin C. 1974. *The Psychology of Sex Differences*. Stanford University Press, Stanford CA

Marsh, Herbert W. and Kenneth J. Rowe (1996). The effects of single-sex and mixed-sex mathematics classes within a coeducational school: A reanalysis and comment. *Australian Journal of Education, 40, No. 2*, 147–162.

McLarty, Joyce R., A. Candace Noble, & Renee M. Huntley. (1989) "Effects of Item Wording on Sex Bias". *Journal of Educational Measurement.* 26, 285-293.

Moss, Jacque D., F.G. Brown. (1979) "Sex Bias and Academic Performance: an Empirical Study" *Journal of Educational Measurement*. 16, 197-201.

Muller, Chandra. (1998) "Gender Differences in Parental Involvement and Adolescents' Mathematics Achievement". *Sociology of Education.* 17, 336-356.

National Science Foundation, Division of Science Resources Statistics, Survey of Earned Doctorates, 1966-2003.

O'Brien, Laurie T., & Christian S. Crandall. (2003) "Stereotype Threat and Arousal: Effects on Women's Math Performance". Personality and Social Psychology Bulletin. 29, 782-789.

Organisation for Economic Cooperation and Development (OECD). (2004). *Education at a Glance* (using PISA 2000).

Paglin, Morton & Anthony M. Rufolo, "Heterogenous Human Capital, Occupational Choice, and Male-Female Earnings Differences," *The Journal of Labor Economics*, 8, No. 1 (1990), 123-144.

Parsons, Jacquelynne Eccles., Terry Adler, Caroline Kaczala. (1982) "Socialization of Achievement Attitudes and Beliefs: Parental Influences". *Child Development*. 53, 310-321.

Riordan, Cornelius (1985). Public and Catholic schooling: The effects of gender context policy. *American Journal of Education, 93, No. 4*, 518–540.

Singh, Kusum, Clare Vaught, and Ethel W. Mitchell (1998). Single-sex classes and academic achievement in two inner-city schools. *The Journal of Negro Education, 67, No. 2*, 157–167.

Spencer, Steven J., Claude M. Steele, and Diane M. Quinn (1999) Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, No. 1, 4-28.

Sweeney, Edward J., "Sex Differences in Problem-Solving," Department of Psychology, Report No. 1, Stanford University (1953).

Tobias, Sheila, "Math Anxiety," *Ms Magazine*, 5 (1976), 56-59.

Voyer, Daniel, Marquerite A. Rodgers, & Peter McCormick. (2004). "Timing Conditions and the magnitude of gender differences on the Metal Rotation Tests". *Memory & Cognition.* 32, 72-82.

Wei, Thomas, “Stereotype Threat, Gender, and Math Performance: Evidence from the National Assessment of Educational Progress,” Working Paper, Harvard University (2008).

Wilder, Gita Z. and Kristin Powell (1989). Sex differences in test performance: a survey of the literature. *College Board Report, No. 89-3.*

Witelson, Sandra F. (1976) "Sex and the Single Hemisphere: Specialization of the Right Hemisphere for Spatial Processing". *Science.* 193, 425-427.

Woodward, Lianne J., David M. Fergusson, and L. John Horwood (1999). Effects of single-sex and coeducational secondary schooling in children’s academic achievement. *Australian Journal of Education, 43, No. 2*, 142–156.

Younger, Michael Robert and Molly Warrington (2003). “We decided to give it a twirl”: single-sex teaching in English comprehensive schools. *Gender and Education, 14, No. 4*, 339–350.

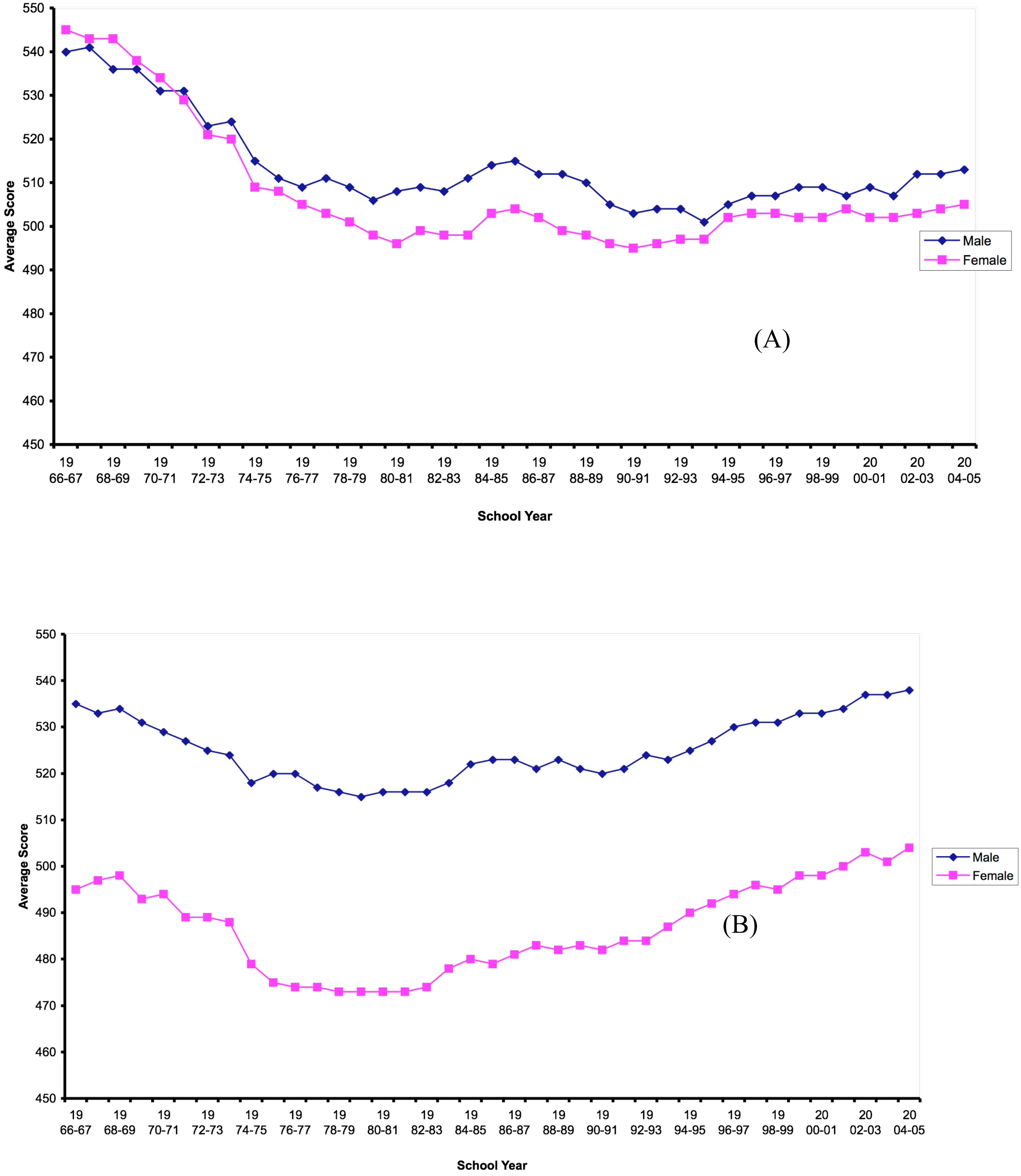

Figure 1: Verbal (A) and Mathematics (B) Achievement on the SAT, by Gender

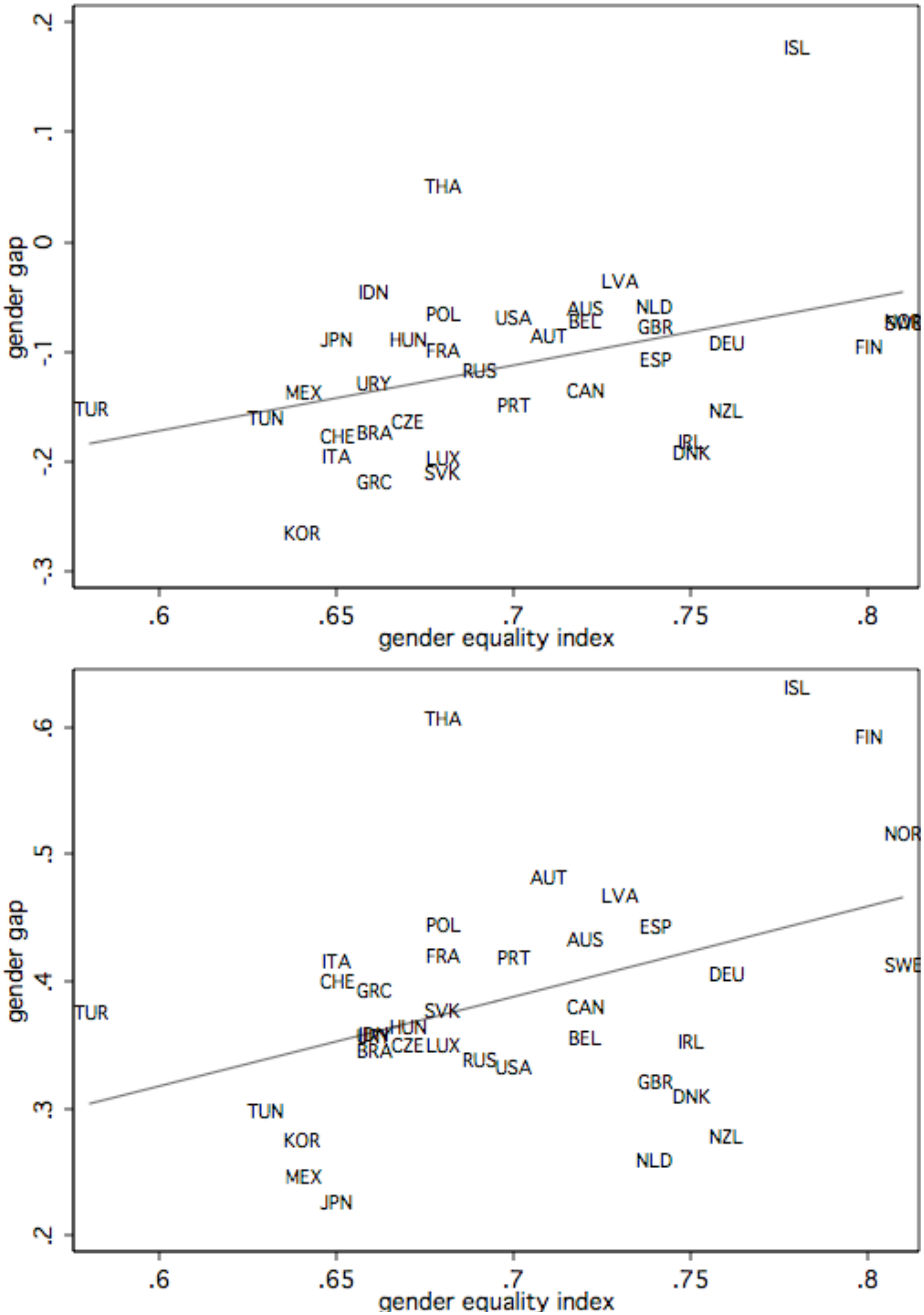

Figure 2: The Relationship Between the Gender Gap in Mathematics (A) and Reading (B) and Gender Equality, PISA 2003

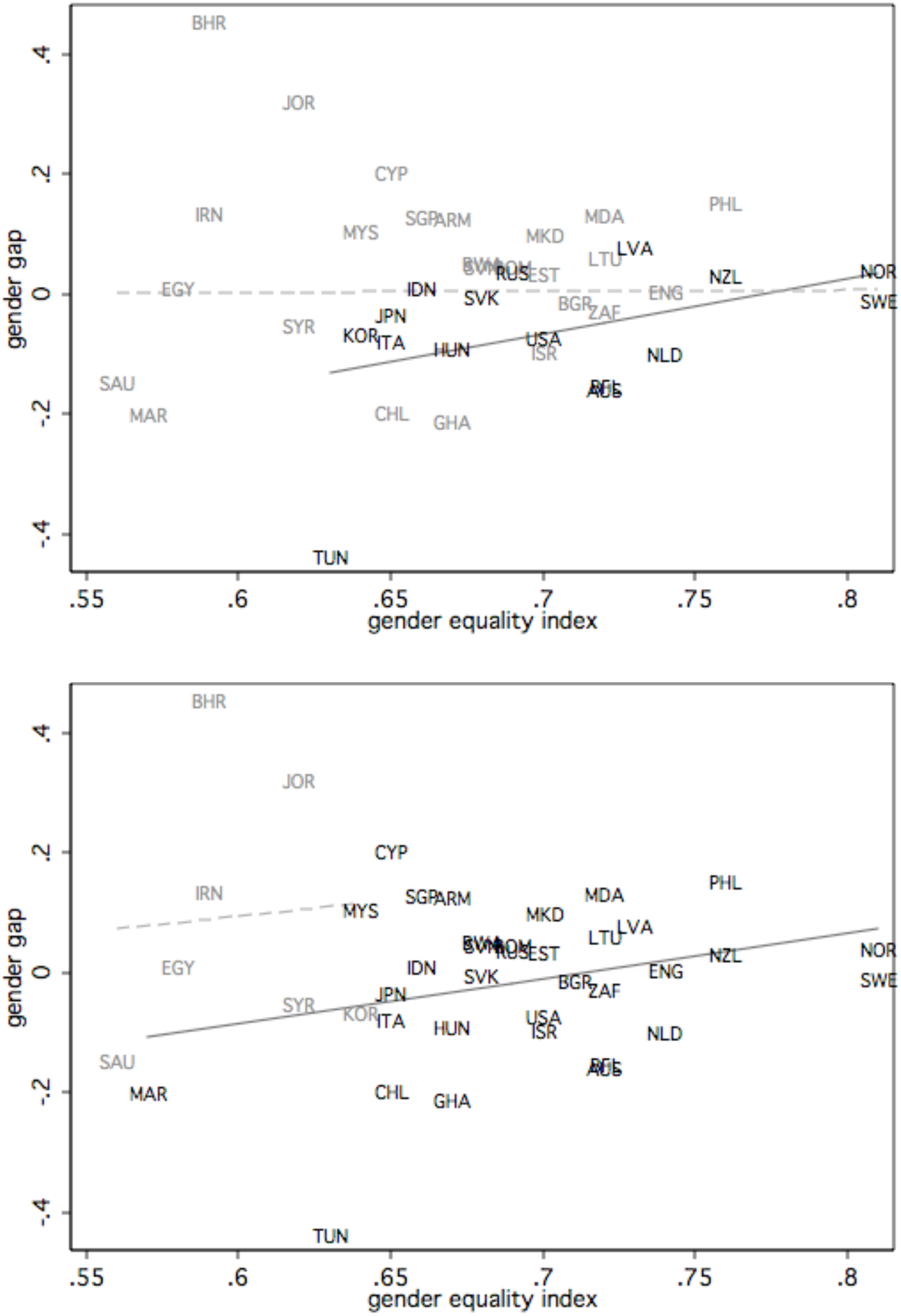

Figure 3: The Relationship Between the Gender Gap in Mathematics and Gender Equality; (a) Full Sample of TIMMS 2003 and intersection of TIMMS 2003 and PISA 2003 and (b) Countries with predominantly Single Sex versus Mixed Sex Schools in TIMMS 2003

Table 1 - Summary Statistics by Gender

Variable	Full Sample	Male	Female	Mean Difference Significant
Math Test Scores:				
Fall kindergarten	0.000	-0.004	0.005	
	(1.000)	(1.028)	(0.970)	
Spring kindergarten	0.000	0.014	-0.015	
	(1.000)	(1.038)	(0.959)	
Spring first grade	0.000	0.037	-0.039	
	(1.000)	(1.041)	(0.954)	
Spring third grade	0.000	0.100	-0.105	**
	(1.000)	(1.015)	(0.974)	
Spring fifth grade	0.000	0.096	-0.101	**
	(1.000)	(0.998)	(0.992)	
Reading Test Scores:				
Fall kindergarten	0.000	-0.073	0.076	**
	(1.000)	(0.984)	(1.011)	
Spring kindergarten	0.000	-0.082	0.085	**
	(1.000)	(0.998)	(0.995)	
Spring first grade	0.000	-0.086	0.089	**
	(1.000)	(1.019)	(0.972)	
Spring third grade	0.000	-0.073	0.075	**
	(1.000)	(1.018)	(0.975)	
Spring fifth grade	0.000	-0.063	0.065	**
	(1.000)	(1.038)	(0.955)	
Parental Expectations:				
Educational Attainment	4.057	3.996	4.122	**
	(1.120)	(1.130)	(1.105)	
Relative Performance, Math, 1st-Spring	3.949	4.009	3.886	**
	(0.957)	(0.939)	(0.972)	
Relative Performance, Math, 3rd-Spring	3.849	3.943	3.749	**
	(1.023)	(1.028)	(1.007)	
Relative Performance, Reading, 1st-Spring	3.980	3.889	4.077	**
	(1.071)	(1.088)	(1.044)	
Relative Performance, Reading, 3rd-Spring	3.880	3.809	3.956	**
	(1.062)	(1.066)	(1.053)	
Female	0.487	0.000	1.000	**
	(0.500)	(0.000)	(1.000)	
Race:				
White	0.583	0.594	0.571	
	(0.493)	(0.491)	(0.495)	
Black	0.164	0.164	0.163	
	(0.370)	(0.370)	(0.369)	
Hispanic	0.188	0.183	0.193	
	(0.391)	(0.387)	(0.395)	
Asian	0.022	0.019	0.026	*
	(0.147)	(0.135)	(0.158)	
Other	0.044	0.040	0.047	
	(0.204)	(0.196)	(0.213)	
Other Controls:				
Age (in months)	56.999	57.225	56.760	**
	(4.318)	(4.466)	(4.143)	
SES composite measure	0.000	-0.031	0.033	
	(1.000)	(1.024)	(0.974)	
Mother's age at time of first birth	23.436	23.329	23.547	
	(5.468)	(5.350)	(5.588)	
WIC participant	0.469	0.480	0.458	
	(0.499)	(0.500)	(0.498)	

Socialization Variables:				**
Mother in math occupation	0.007	0.004	0.011	
	(0.084)	(0.059)	(0.103)	
Father in math occupation	0.038	0.041	0.035	
	(0.192)	(0.198)	(0.184)	
Frequency of math activities (per wk)	2.277	2.288	2.267	
	(0.876)	(0.848)	(0.901)	
Frequency of reading activities (per wk)	3.148	3.117	3.177	
	(0.862)	(0.883)	(0.842)	
Indices of Gender Inequality:				
Wage	0.641	0.640	0.642	
	(0.049)	(0.050)	(0.047)	
Percent employment	0.879	0.878	0.880	
	(0.035)	(0.035)	(0.035)	
Educational attainment	0.866	0.863	0.868	
	(0.098)	(0.098)	(0.097)	
Frequency of Missing Values:				
SES composite measure	0.030	0.036	0.024	
	(0.172)	(0.187)	(0.154)	
Mother's age at Birth	0.082	0.086	0.078	
	(0.275)	(0.281)	(0.268)	
WIC	0.079	0.086	0.071	
	(0.269)	(0.281)	(0.257)	
Mother in math occupation	0.348	0.343	0.352	
	(0.476)	(0.475)	(0.478)	
Father in math occupation	0.311	0.299	0.323	
	(0.463)	(0.458)	(0.468)	
Frequency of math activities (per wk)	0.691	0.711	0.669	*
	(0.462)	(0.453)	(0.471)	
Frequency of reading activities (per wk)	0.691	0.711	0.669	*
	(0.462)	(0.453)	(0.471)	
Wage	0.025	0.032	0.017	
	(0.156)	(0.177)	(0.130)	
Employment	0.025	0.032	0.017	
	(0.156)	(0.177)	(0.130)	
Educational attainment	0.025	0.032	0.017	
	(0.156)	(0.177)	(0.130)	

NOTES: The entries are means and standard deviations of student-level data for those students in ECLS-K who do not have missing values for gender, age, and test scores. Test scores are IRT scores, normalized to have a mean of 0 and a standard deviation of 1 in the full, weighted sample. Parental Expectations are categorical variables. Educational Attainment indicates the highest level of education the parents expect their child to attain: 1 - HS dropout; 2 - HS grad; 3 - some college; 4 - Bachelor; 5 - Masters; 6 - Ph.D. The category white includes only non-Hispanic whites. Relative Performance indicates how well, compared to other children, parents expect their child to do: 1 - much worse; 2 - a little worse; 3 - about the same; 4 - a little better; 5 - much better. The SES composite measure incorporates information on parental education, occupational status, and family income. The SES measure ranges from -5.88 to 3.46 in the sample, with larger numbers indicating higher SES. Indices of gender inequality are means of state level data on: full-time year-round female workers' median income relative to male's; ratio of employed females relative to employed males;and the ratio of females with a graduate education relative to males. Precise definitions of the variables are provided in the data appendix. The total number of students in the sample who receive a positive weight in the estimation is 9500. The bottom panel of the table reports the number of missing values for the covariates. Sample weights are used.

Table 2 - Estimated Gender Gap over the First Six Years of School, Math

	Fall-K	Spring-K	Spring-1st	Spring-3rd	Spring-5th	Fall-K	Spring-K	Spring-1st	Spring-3rd	Spring-5th
Female	0.009	-0.029	-0.076	-0.205**	-0.197**	0.007	-0.035	-0.085*	-0.225**	-0.226***
	(.04)	(.04)	(.04)	(.04)	(.04)	(.03)	(.03)	(.04)	(.03)	(.03)
Controls:										
Black						-0.179**	-0.301**	-0.376**	-0.449**	-0.515**
						(.05)	(.05)	(.06)	(.06)	(.06)
Hispanic						-0.290**	-0.279**	-0.239**	-0.187**	-0.128**
						(.04)	(.04)	(.05)	(.05)	(.05)
Asian						0.223**	0.196*	0.076	0.089	0.242**
						(.08)	(.08)	(.08)	(.09)	(.07)
Other						-0.281**	-0.237**	-0.300**	-0.310**	-0.313**
						(.07)	(.07)	(.07)	(.06)	(.07)
Age in months						0.248*	0.357**	0.436**	0.419**	0.352**
						(.11)	(.09)	(.09)	(.09)	(.10)
Age in months, squared						-0.002	-0.003**	-0.003**	-0.003**	-0.003**
						(.00)	(.00)	(.00)	(.00)	(.00)
Teenage mother at time of first birth						0.269**	0.253**	0.233**	0.254**	0.263**
						(.03)	(.03)	(.03)	(.03)	(.03)
Mother at least 30 at age of first birth						-0.202**	-0.187**	-0.190**	-0.172**	-0.189**
						(.04)	(.04)	(.05)	(.05)	(.05)
Socioeconomic status						0.272**	0.179**	0.117**	0.170**	0.176**
						(.05)	(.05)	(.05)	(.05)	(.04)
WIC participant						-0.264**	-0.244**	-0.231**	-0.242**	-0.212**
						(.04)	(.05)	(.04)	(.05)	(.05)
Constant	-0.004	0.014	0.037	0.100**	0.096**	-8.575**	-11.528**	-13.271**	-12.226**	-9.854**
	(.03)	(.03)	(.03)	(.03)	(.03)	(3.01)	(2.54)	(2.72)	(2.66)	(2.93)
Observations	9481	9481	9481	9481	9481	9481	9481	9481	9481	9481
R-squared	0	0	0.001	0.01	0.01	0.321	0.291	0.242	0.263	0.267

NOTES: The dependent variables vary by column, but contain the math test score in Fall Kindergarten, Spring Kindergarten, Spring First Grade, and Spring Fifth Grade. Test Scores are IRT scores, normalized to have a mean of 0 and a standard deviation of 1 in the weighted sample. Non-Hispanic whites are the omitted category. The unit of observation is a student. Standard errors are in parentheses. Estimation is done using weighted least squares, using sample weights provided in the data set. In addition to the variables in the table, indicator variables for students with missing values on each covariate are also included in the regression. * denotes significance at 5%-level; ** denotes significance at 1%-level.

Table 3 - Estimated Gender Gap over the First Six Years of School, Reading

	Fall-K	Spring-K	Spring-1st	Spring-3rd	Spring-5th	Fall-K	Spring-K	Spring-1st	Spring-3rd	Spring-5th
Female	0.149**	0.166**	0.175**	0.148**	0.128**	0.152**	0.163**	0.169**	0.136**	0.112**
	(.04)	(.04)	(.04)	(.04)	(.04)	(.04)	(.04)	(.04)	(.04)	(.04)
Controls:										
Black						0.025	-0.044	-0.12	-0.322**	-0.358**
						(.05)	(.06)	(.06)	(.07)	(.06)
Hispanic						-0.159**	-0.045	-0.061	-0.119	-0.093
						(.05)	(.05)	(.06)	(.06)	(.06)
Asian						0.282**	0.338**	0.298**	0.051	0.091
						(.08)	(.08)	(.08)	(.07)	(.06)
Other						-0.232**	-0.173**	-0.234**	-0.299**	-0.231**
						(.07)	(.06)	(.08)	(.07)	(.08)
Age in months						-0.073	0.159	0.463**	0.319**	0.284**
						(.13)	(.12)	(.12)	(.11)	(.10)
Age in months, squared						0.001	-0.001	-0.004**	-0.003**	-0.002**
						(.00)	(.00)	(.00)	(.00)	(.00)
Teenage mother at time of first birth						0.277**	0.262**	0.238**	0.260**	0.270**
						(.03)	(.03)	(.03)	(.04)	(.03)
Mother at least 30 at age of first birth						-0.208**	-0.190**	-0.215**	-0.165**	-0.232**
						(.05)	(.05)	(.06)	(.06)	(.06)
Socioeconomic status						0.299**	0.245**	0.119*	0.150**	0.142**
						(.06)	(.06)	(.06)	(.05)	(.05)
WIC participant						-0.213**	-0.177**	-0.180**	-0.269**	-0.232**
						(.05)	(.05)	(.05)	(.05)	(.05)
Constant	-0.073*	-0.082**	-0.086**	-0.073*	-0.063*	0.679	-5.765	-14.160**	-9.563**	-8.351**
	(.03)	(.03)	(.03)	(.03)	(.03)	(3.53)	(3.39)	(3.39)	(3.06)	(2.92)
Observations	8867	8867	8867	8867	8867	8867	8867	8867	8867	8867
R-squared	0.006	0.007	0.008	0.006	0.004	0.257	0.208	0.187	0.23	0.242

NOTES: The dependent variables vary by column, but contain the reading test score in Fall Kindergarten, Spring Kindergarten, Spring First Grade, and Spring Fifth Grade. Test Scores are IRT scores, normalized to have a mean of 0 and a standard deviation of 1 in the weighted sample. Non-Hispanic whites are the omitted category. The unit of observation is a student. Standard errors are in parentheses. Estimation is done using weighted least squares, using sample weights provided in the data set. In addition to the variables in the table, indicator variables for students with missing values on each covariate are also included in the regression. * denotes significance at 5%-level; ** denotes significance at 1%-level.

Table 4 – Analysis of the Tails, Math

Ratio of girls to boys:	0.949					
	Bottom 1%	Bottom 5%	Below Median	Above Median	Top 5%	Top 1%
Fall of Kindergarten	.914	.793	.954	.944	.811	.629
Spring of 1st Grade	.370	.781	1.011	.891	.497	.288
Spring of 3rd Grade	1.211	1.243	1.134	.793	.455	.234
Spring of 5th Grade	.508	1.249	1.125	.800	.381	.428

NOTES: Math scores are standardized scores with no controls. The distribution is partitioned within each year.

Table 5 – Sensitivity Analysis for Losing Ground in Math

	Fall-K	5th Grade	Lost Ground
Baseline	.007	-.226	-.233
	(.03)	(.03)	(.03)
By Race:			
White	.033	-.239	-.272
	(.05)	(.05)	(.04)
Black	-.051	-.212	-.160
	(.06)	(.09)	(.08)
Hispanic	.087	-.134	-.221
	(.05)	(.07)	(.06)
Asian	-.060	-.167	-.107
	(.14)	(.12)	(.11)
Other	-.252	-.394	-.142
	(.11)	(.13)	(.13)
By SES Quintile:			
Bottom	-.061	-.232	-.171
	(.05)	(.09)	(.07)
Second	.127	-.141	-.268
	(.07)	(.07)	(.07)
Third	.031	-.193	-.224
	(.07)	(.08)	(.06)
Fourth	-.120	-.317	-.197
	(.07)	(.07)	(.07)
Fifth	.060	-.234	-.294
	(.08)	(.06)	(.07)
By Family Structure:			
Two Biological Parents	.024	-.221	-.245
	(.04)	(.04)	(.04)
Single Mother	.000	-.283	-.282
	(.06)	(.08)	(.07)
Teen Mother at Birth	.014	-.196	-.209
	(.06)	(.07)	(.06)
Mother in Her 20s at Birth	-.030	-.265	-.235
	(.04)	(.05)	(.04)
Mother over 30 at Birth	.179	-.196	-.375
	(.09)	(.07)	(.08)
By Region:			
Northeast	-.028	-.373	-.344
	(.07)	(.07)	(.07)
Midwest	.038	-.222	-.260
	(.06)	(.06)	(.06)
South	-.014	-.156	-.142
	(.05)	(.06)	(.05)
West	.035	-.227	-.262
	(.06)	(.08)	(.06)
By Location Type:			
Central City	-.025	-.288	-.263
	(.05)	(.05)	(.05)
Suburban	.078	-.163	-.241
	(.06)	(.05)	(.05)
Town	-.033	-.149	-.117
	(.09)	(.10)	(.08)
Rural	.011	-.193	-.204
	(.07)	(.08)	(.08)
By School Type:			

Public	-.003	-.216	-.213
	(.03)	(.04)	(.03)
Private	.074	-.264	-.337
	(.08)	(.06)	(.07)
By Parents in Math Occupations:			
Mother not in Math Occupation	.007	-.227	-.234
	(.03)	(.03)	(.03)
Mother in Math Occupation	.180	-.021	-.201
	(.29)	(.17)	(.25)
Higher Education than Father	.006	-.259	-.265
	(.06)	(.07)	(.06)
By Mother's Education Level			
High School Drop Out	.047	-.135	-.182
	(.06)	(.10)	(.09)
High School Grad	.027	-.221	-.248
	(.05)	(.06)	(.05)
Some College	-.074	-.244	-.170
	(.06)	(.06)	(.06)
Bachelors Degree	.047	-.274	-.321
	(.09)	(.07)	(.07)
Graduate Degree	.150	-.199	-.349
	(.13)	(.09)	(.10)
By Mom's Labor Force Participation:			
Not in LF between birth and 5th Grade	.156	-.169	-.325
	(.15)	(.13)	(.13)
Not in LF between Birth and K	.071	-.139	-.202
	(.07)	(.07)	(.05)
Always in LF	-.025	-.237	-.213
	(.18)	(.17)	(.14)

NOTES: Specifications are variations on those reported in table 2, columns 6 and 10. Only the gender coefficients are reported. The top row simply reproduces the baseline results in columns 6 and 10 of table 2. The remaining rows correspond to separate regressions run using different weights, test score measures, or particular subsets of the data. For further details of the baseline specification, see the notes to table 2.

Table 6 – Adjusted Means of Skill Components of the Tests

	Fall-K	Spring 1st	Spring 3rd	Spring 5th
Skill Tested	Coefficient on Female:			
Math:				
Count, number, shapes	0.013*	-	-	-
	(.01)	-	-	-
	[0.918]	[1.000]	[1.000]	[1.000]
Relative size	0.004	-	-	-
	(.01)	-	-	-
	[0.538]	[0.990]	[1.000]	[1.000]
Ordinality, sequence	-0.004	0.003	0	-
	(.01)	(.01)	(.00)	-
	[0.192]	[0.951]	[0.999]	[1.000]
Add, subtract	-0.005	-0.011	-0.007*	-
	(.00)	(.01)	(.00)	-
	[0.034]	[0.713]	[0.970]	[0.994]
Multiply, divide	-	-0.046**	-0.050**	-0.020**
	-	(.01)	(.01)	(.01)
	[0.002]	[0.226]	[0.763]	[0.925]
Place value	-	-0.019**	-0.089**	-0.070**
	-	(.00)	(.01)	(.01)
	[0.000]	[0.029]	[0.416]	[0.737]
Rate and measurement	-	-0.002**	-0.057**	-0.085**
	-	(.00)	(.01)	(.01)
	[0.000]	[0.003]	[0.131]	[0.431]
Fractions	-	0	-0.008**	-0.065**
	-	(.00)	(.00)	(.01)
	[0.000]	[0.000]	[0.008]	[0.133]
Area and volume	-	-	-0.001*	-0.013**
	-	-	(.00)	(.00)
	[0.000]	[0.000]	[0.001]	[0.018]
Reading:				
Letter recognition	0.059**	-	-	-
	(.01)	-	-	-
	[0.670]	[0.997]	[1.000]	[1.000]
Beginning sounds	0.058**	0.006	-	-
	(.01)	(.01)	-	-
	[0.295]	[0.977]	[1.000]	[1.000]
Ending sounds	0.036**	0.018**	-	-
	(.01)	(.01)	-	-
	[0.166]	[0.938]	[0.998]	[1.000]
Sight words	0.007	0.058**	0.004*	-
	(.01)	(.01)	(.00)	-
	[0.031]	[0.772]	[0.990]	[1.000]
Words in context	0.003	0.054**	0.018**	0.007**
	(.00)	(.01)	(.01)	(.00)
	[0.015]	[0.486]	[0.917]	[0.975]
Literal inference	-	0.024**	0.035**	0.024**
	-	(.01)	(.01)	(.01)
	[0.004]	[0.174]	[0.706]	[0.880]
Extrapolation	-	0.003	0.046**	0.036**
	-	(.01)	(.01)	(.01)
	[0.001]	[0.037]	[0.448]	[0.731]
Evaluation	-	0.004	0.024**	0.025*
	-	(.00)	(.01)	(.01)
	[0.001]	[0.036]	[0.255]	[0.463]

Notes: Entries are adjusted mean proficient probability score differences on specific areas of questions. Standard errors are given in parentheses. Mean probability scores for the whole sample are reported in square brackets. These scores are constructed using IRT scores and provide the probability of mastery of a specific set of skills. We adjust for our parsimonious set of controls and refrain from reporting minuscule differences. ** denotes significance at 5%-level; *** denotes significance at 1%-level.

Revise and Resubmit Table 7 - Estimated Gender Gap in Math Scores

	PISA			TIMSS- full sample			TIMSS-PISA overlap		
WEF_GGI	0.599*	0.824**	0.789*	.019	.006	.069	.916	1.131*	1.204*
	(.25)	(.28)	(.30)	(.47)	(.56)	(.75)	(.58)	(.51)	(.45)
Controls									
GDP		-0.001	-0.001		0.000	0.000		-0.001	-0.001
		(.00)	(.00)		(.00)	(.00)		(.00)	(.00)
Gender Ratio			.119			-.059			-1.065*
			(.16)			(.29)			(.48)
Constant	-0.531**	-0.647**	-0.746***	-0.008	-0.002	0.015	-0.707	-0.818	0.205
	(.17)	(.18)	(.20)	(.34)	(.38)	(.36)	(.43)	(.38)	(.39)
Observations	41	41	41	47	47	47	17	17	17
R-squared	0.148	0.196	0.211	0.003	0.005	0.007	0.207	0.224	0.541

NOTES: Estimation is done using ordinary least squares. Test scores are standardized by country to have mean 0 and standard deviation over the full weighted sample. Gender gap is calculated as the weighted mean score of females minus the weighted mean score of males. Gender ratio was constructed at the country level as the weighted ratio of female to male among test takers. GDP was divided by 1000. Standard errors of coefficients are reported in parentheses. * denotes significance at 5% level; **denotes significance at 1% level.

Appendix Table 1 – Countries Represented in PISA and TIMMS

Country	ISO Code	PISA	TIMMS	Gender ratio of test takers	Gender ratio at birth
Armenia	ARM		X	1.12	1.15
Australia	AUS	X	X	1.05	1.05
Austria	AUT	X			
Bahrain	BHR		X	1.02	1.03
Belgium (Flemish)	BFL	X	X	1.16	1.04
Botswana	BWA		X	1.06	1.03
Brazil	BRA	X			
Bulgaria	BGR		X	0.93	1.06
Canada	CAN	X			
Chile	CHL		X	0.92	1.05
Czech Republic	CZE	X			
Cyprus	CYP		X	0.95	1.05
Denmark	DNK	X			
Egypt	EGY		X	0.87	1.05
England	ENG		X	1.00	1.05
Estonia	EST		X	1.00	1.06
Finland	FIN	X			
France	FRA	X			
Germany	DEU	X			
Ghana	GHA		X	0.82	1.03
Greece	GRC	X			
Hong Kong (China)	HKG	X	X	1.00	1.08
Hungary	HUN	X	X	1.02	1.06
Iceland	ISL	X			
Indonesia	IDN	X	X	1.02	1.05
Iran, Islamic Republic of	IRN		X	0.68	1.05
Ireland	IRL	X			
Israel	ISR		X	1.08	1.05
Italy	ITA	X	X	0.99	1.07
Japan	JPN	X	X	0.98	1.06
Jordan	JOR		X	0.95	1.06
Korea	KOR	X	X	0.93	1.08
Latvia	LVA	X	X	0.97	1.05
Lebanon	LBN		X	1.34	1.05
Liechtenstein	LIE	X			
Lithuania	LTU		X	0.98	1.06
Luxembourg	LUX	X			
Macedonia	MKD		X	0.97	1.08
Macao (China)	MAC	X			
Malaysia	MYS		X	1.00	1.07
Mexico	MEX	X			
Moldova	MDA		X	1.05	1.06
Morocco	MAR		X	0.99	1.05
Netherlands	NLD	X	X	0.98	1.05
New Zealand	NZL	X	X	1.07	1.05
Norway	NOR	X	X	0.98	1.05
Palestinian National Authority	PSE		X	1.21	1.06
Philippines	PHL		X	1.36	1.05
Poland	POL	X			
Portugal	PRT	X			
Romania	ROM		X	1.07	1.06
Russian Federation	RUS	X	X	0.97	1.06
Saudi Arabia	SAU		X	0.76	1.05
Scotland	SCO		X	1.00	-

Serbia and Montenegro	SCG		X	0.95	-
Singapore	SGP		X	0.95	1.08
Slovakia	SVK	X	X	0.91	1.05
Slovenia	SVN		X	1.01	1.07
South Africa	ZAF		X	1.06	1.02
Spain	ESP	X			
Sweden	SWE	X	X	1.03	1.06
Switzerland	CHE	X			
Syria	SYR		X	1.19	1.06
Taiwan	TWN		X	0.94	1.09
Thailand	THA	X			
Tunisia	TUN	X	X	1.11	1.07
Turkey	TUR	X			
United Kingdom	GBR	X			
Uruguay	URY	X			
United States	USA	X	X	1.07	1.05
Yugoslavia	YUG	X			
Total		41	47		

NOTES: Data were obtained from http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html (PISA) and http://timss.bc.edu/ (TIMMS). Gender ratio of test takers is calculated as the ratio of the number of females to the number of males for each country using the TIMSS data set. Gender ratio at birth is taken from 2007 CIA World Factbook.

Appendix Table 2 - Structure of the Global Gender Gap Index

Subindex	Variables	Sources	Weights
Economic Participation and Opportunity	Ratio: female labor force participation over male value	International Labor Organization, Key Indicators of the Labor Market, 2005	0.199
	Wage equality between women and men for similar work (converted to female-over-male ratio)	World Economic Forum, Executive Opinion Survey 2007	0.310
	Ratio: estimated female earned income over male value	United Nations Development Program, *Human Development Report 2006,* 2004 or latest available data	0.221
	Ratio: female legislators, senior officials and managers over male value	International Labor Organization, *LABORSTA Internet*, online database, 2006, or latest year available	0.149
	Ratio: female professional and technical workers over male value	International Labor Organization, *LABORSTA Internet,* online database, 2006, or latest year available	0.121
Educational Attainment	Ratio: female literacy rate over male value	UNESCO Statistics Division, Education Indicators, 2006; CIA World Factbook, 2004-2005 estimates	0.191
	Ratio: female net primary level enrollment over male value	*Worldbank, World Development Indicators Online*, accessed June 2007; 2005 data or latest year available	0.459
	Ratio: female net secondary level enrollment over male value	*Worldbank, World Development Indicators Online*, accessed June 2007; 2005 data or latest year available	0.230
	Ratio: female gross tertiary level enrollment over male value	*Worldbank, World Development Indicators Online,* accessed June 2007, 2005 data or latest year available	0.121
Political Empowerment	Ratio: females with seats in parliament over male value	International Parliamentary Union, April 2007	0.310
	Ratio: females at ministerial level over male value	United Nations Development Program, *Human Development Report 2006*	0.247
	Ratio: number of years of a female head of state (last 50 years) over male value	own calculations, as of June 2007	0.443
Health and Survival	Ratio: female healthy life expectancy over male value	World Health Organization, *'World Health Statistics 2007*' and *'The World Health Report 2007'*	0.307
	Sex ratio at birth (converted to female-over-male ratio)	CIA World Factbook, U.S. Census Bureau, International Data Base (IDB), retrieved May 2007	0.693

Source: Global Gender Gap Report 2007, http://www.weforum.org/en/initiatives/gcp/Gender%20Gap/index.htm. All variables within each sub-index sum to one. The World Economic Forum's Gender Gap Index is the average of the subindices, ranging from 0 to 1 with a max value of .81 (Norway and Sweden) and a minimum value of .56 (Saudi Arabia).

Appendix Table 3: The Evolution of the Performance Gap on Subjective Teacher Assessments

	Teacher's Subjective Assessment of Student Ability		Test Scores	
	Raw Data	Including Controls	Raw Data	Including Controls
Math:				
Fall Kindergarten	0.098	0.075	-0.068	-0.019
	(.08)	(.05)	(.08)	(.06)
Spring Kindergarten	0.196*	0.118	-0.092	-0.047
	(.08)	(.08)	(.08)	(.07)
Spring 1st Grade	-0.01	-0.015	-0.064	-0.077
	(.08)	(.07)	(.07)	(.07)
Spring 3rd Grade	-0.03	-0.093	-0.252**	-0.190*
	(.08)	(.08)	(.08)	(.08)
Spring 5th Grade	0.001	-0.172**	-0.210**	-0.312**
	(.08)	(.06)	(.08)	(.06)
Reading:				
Fall Kindergarten	0.151**	0.169**	0.094	0.157**
	(.05)	(.03)	(.05)	(.03)
Spring Kindergarten	0.230**	0.213**	0.109*	0.128**
	(.05)	(.03)	(.05)	(.03)
Spring 1st Grade	0.172**	0.189**	0.135**	0.101**
	(.05)	(.03)	(.05)	(.03)
Spring 3rd Grade	0.235**	0.186**	0.125*	0.093**
	(.05)	(.04)	(.05)	(.04)
Spring 5th Grade	0.319**	0.261**	0.106*	0.068*
	(.05)	(.04)	(.05)	(.03)

NOTES: The table entries are estimated gaps in IRT scores and subjective teacher assessment of student achievement provided by teachers in fall kindergarten, spring first grade, spring 3rd grade, and spring 5th grade. The odd columns are raw gaps; the even columns are residual gaps after controlling for the parsimonious set of controls and teacher fixed effects. Test scores and teacher assessments are normalized to have a mean of 0 and a standard deviation of 1 in the weighted sample of students with valid test scores. The method of estimation is weighted least squares using sample weights provided by ECLS. The number of observations is 6096 for reading and 2445 for math, which is the set of children for whom teacher assessments and test scores were available at every point in the time. Standard errors are given in parentheses. * denotes significance at 5% level; ** denotes significance at 1% level.