

The Gender Gap in Mathematics: Evidence from Chile

PRASHANT BHARADWAJ

University of California, San Diego

GIACOMO DE GIORGI

Federal Reserve Bank of New York, ICREA-MOVE, BREAD, CEPR, and IPA

DAVID HANSEN

University of Wisconsin–Madison

CHRISTOPHER NEILSON

Princeton University

I. Introduction

Throughout much of the developing world, women tend to be disadvantaged in terms of job opportunities and wages (Sen 1999). This is at least partially due to a significant gender gap in educational levels that remains large in many countries (World Development Report 2012), although it is reversed in the United States for recent cohorts (Goldin, Katz, and Kuziemko 2006). This gap is also potentially due to difference in the types of human capital women and men acquire, even conditional on the same level of education. For example, there is substantial evidence of a strong correlation between math test scores, math-based curriculum, mathematical majors in college, and future income earned, suggesting that observed differences in math skills across genders can explain part of the wage gap.¹ It is therefore important for research to address the differences in the development of math skills and the determinants of math-related specializations and the role played by myriad factors starting from early childhood, such as parental investments, preferences, expectations and innate ability. Understanding when and how differences between men and women begin to

q1

Giacomo De Giorgi acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075) and ECO2011-28822, and the European Union through the Marie Curie CIG grant FP7-631510. All views expressed in this article are those of the authors only and do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System in general.

¹ See Paglin and Rufolo (1990), Grogger and Eide (1995), Murnane, Willett, and Levy (1995), Altonji and Blank (1999), Weinberger (1999, 2001), and Murnane et al. (2000) for evidence on math scores and wages, and see Altonji, Blom, and Meghir (2012) for a review on math curriculum and college major.

develop in the process of human capital accumulation is crucial to understanding the gender gap in wages and job opportunities later in life.

We motivate our analysis by looking at a series of comparable data across countries that were part of the PISA (Programme for International Student Assessment) OECD (Organization for Economic Cooperation and Development) test score data, administered to children around age 15 in 65 countries (see Guiso et al. 2008). We show a substantial gap in math test scores in the pooled data across both developed and developing countries (although the gap appears to be larger in developed countries). While the PISA data serve as excellent motivation and evidence of external validity, they lack the detailed demographic data allowing for an in-depth exploration of the reasons behind the gender gap. The core of our analysis explores dynamics of the gender gap and more detailed explanations for the gap using administrative data from Chile. A similar gap to that identified in the PISA-OECD data in math scores is also present in the Chilean Sistema de Medición de la Calidad de la Educación (SIMCE) data (the SIMCE is a national test that is administered to all fourth and eighth graders in the country). In SIMCE scores at age 9 and 10 (grade 4), we see a sizable gap in the mean and median achievement. At the same time we note that these gaps expand substantially, nearly doubling by age 13 and 14 (grade 8). Further, for the top 5% of the performance distribution, the gap continuously increases across grades; that is, the ratio of boys to girls is 1.9 and 2.2 for grades 4 and 8, respectively. The detailed nature of the data from Chile allows us to explore some more possibilities along the broad categories of individual, parental, and school characteristics that are not available in the PISA data. A recent paper by Fryer and Levitt (2010) is closely related to our study. In that paper the authors establish that while the gender gap does not appear to be present at the beginning of school-going age, boys perform better in math by the end of the first 6 years of school. A host of explanatory variables are unable to account for the gap in their setting.

Our article expands on and complements the work of Guiso et al. (2008), Machin and Pekkarinen (2008), and Fryer and Levitt (2010). We are able to control for a rich set of individual and family characteristics (e.g., by examining differences within twins) and at the same time attempt to account for classroom composition, teacher gender, and other related factors. Our data on math ability perception and parental investments are unique in this setting and allow us to further pursue relevant hypotheses regarding the emergence of the gender gap in math skills.

In terms of individual-level characteristics, we examine factors such as birth weight and gestational age in the determination of test scores. Early childhood health has been shown to be an important determinant of a host of later life

outcomes like income, school performance, and health (Black, Devereux, and Salvanes 2007; Bharadwaj, Loken, and Neilson 2011). If there are differences in early childhood health across genders, then perhaps some of this gap is driven by these “health endowments.” We are also able to explore gender gaps within twin pairs, given the richness of the data from Chile. A twins fixed effect in this setting attempts to control for a large portion of unobserved family background and parental investment variables that twins share. Even within twins, we see a sizable gender gap in math.

We then examine the school and class environment. Given the population of school students observed in Chile, we can examine how gender composition of the classroom and classroom size affects performance in math. The motivation for examining this comes from an influential literature suggesting that girls and boys perform differently in competitive environments. We can also test the role that teacher gender plays in mitigating or exacerbating the gender gap in math. Dee (2007) and Carrell, Page, and West (2010) find that the gender of the teacher matters for female performance on science and math. While we cannot account for selection via random assignment of teachers, controlling for the gender of the math teacher does little to close the gap. We also examine the role played by schools in influencing the gender gap, by using school fixed effects.

Finally, we examine parental characteristics, which are a critical part of the hypothesis related to sibling competition and differential resources within the household. In addition, we study the role of parental investments and student perceptions regarding their own math abilities. We find that students have different perceptions (on the basis of gender) regarding their own math abilities. Even conditional on math score, we find that boys tend to be more optimistic about their math ability. For example, boys are much more likely to say that they are good at math than are girls. However, it is not the case that girls are pessimistic about everything—while answering questions about general ability in school, the differences between boys and girls is much smaller. These gaps in perceptions appear to arise specifically in the context of mathematics ability. While we do not use data on perceived ability as an explanatory variable, the differences appear to suggest that this would be an important area for future research to focus on.

Taken together with the results of Fryer and Levitt (2010), it is apparent that not only is there a well-established gender gap in math scores across many parts of the world but a host of potential explanations does little to explain this gap. The evidence we present suggests that a combination of differential parental investments and perceptions may play a role. Future research should focus on these aspects to further shed light on this issue.

II. Data

A. PISA Data

Table 1 uses data from the 2006 and 2009 PISA tests to document the existence of the gender gap in math across several countries. The PISA is designed by the OECD to produce student outcomes that are comparable across countries and to provide information about the characteristics of successful students, families, schools, and national educational systems. It is administered to students who are between age 15 years 3 months and 16 years 2 months. Students take a 2 hour test that includes both multiple choice and handwritten long-form answers. Each student is tested on only a portion of the available test modules, so the raw scores are scaled using the Rasch model of item response theory to reflect this fact. We focus on the results for mathematics. The math questions are in such areas as algebra, geometry, and interpretation of graphs, especially as they relate to solving real-world problems. We use the methods recommended by the OECD for analyzing these data, as set forth in OECD (2005), including the use of multiple plausible values for student test scores to calculate point estimates and employment of Balanced Repeated Replication on the results in order to calculate standard errors.²

B. Administrative Data from Chile

In addition to the introductory picture that the PISA data provide us, we use detailed Chilean administrative data, including the SIMCE. The SIMCE (loosely translated as System for Measuring Educational Quality) is a national test administered annually in Chile for all fourth graders. On alternate years, eighth and tenth grade students are also given this test. We have data on the fourth grade SIMCE from 2002 and subsequently on a yearly basis from 2005 on. We use the eighth grade SIMCE as administered in 2007 and 2009. While there were waves of the eighth grade SIMCE administered before 2007, we do not use them, as the cohort that took those tests was born before 1992 and hence was not available in the vital statistics database (as detailed below).

The SIMCE test for fourth graders consists of reading, math, and a social science component.³ These results are published at various levels to track performance of schools and (what would be the equivalent of) school districts.

The Chilean administrative data we use in this article allow us, in several ways, to go beyond what is possible with the PISA data. First, the SIMCE data cover almost all children in the relevant grades over multiple years, which al-

² An earlier version of this article included more detailed analysis of the PISA data. Please see Bhadraraj et al. (2012) for more on the PISA data and evidence on the gender gap in many low- and middle-income countries.

³ For detailed information on the SIMCE and sample questions, one can visit <http://www.simce.cl/>. Note that our data set does not contain the social science scores, just math and reading.

lows us to examine students who took the SIMCE in both fourth and eighth grade to see how the gender gap in math scores changes as students grow older while eliminating possible selection effects. At the same time, as we are dealing with population data, we can analyze the very top performers, such as the top 1% or the top 5%. According to UNESCO (2010), the survival rate to fifth grade in the 2007 school year was 96%, indicating that selection of students out of school is unlikely to bias the analysis for fourth grade. From the same source, the 2010 net enrollment rate for secondary school as a whole was 84% among boys and 87% among girls, indicating that differential dropout rates are similarly unlikely to bias the eighth grade results.

Because we are interested in understanding outcomes at the top end of the distribution, we should consider whether the SIMCE data do a better job than the PISA data of distinguishing students at the very top end of the distribution from one another. For example, can these tests tell us with some certainty that a student at the 99th percentile on a test is actually better at math than a student at the 95th percentile? Test design is an important issue when considering this question, as Ellison and Swanson (2010) show at length. Both the SIMCE and the PISA are designed to be broad-based tests that can be used at a national level (and an international level in the case of PISA) to compare regional performance and not to look at the performance of high achievers specifically. Both tests, however, do provide enough difficulty that very few students answer all the questions correctly, which increases our confidence that they can distinguish students at the top. Although both tests should be able to do a decent job, the SIMCE seems better able to address the gender gap at the top end of the distribution because of its structure and scope. The sample size of the SIMCE is much larger, so that we simply have more high achievers in the sample and can reduce the standard error of our estimates when dealing with them. In addition, the SIMCE is better able to help us understand individual performance because it is longer and more uniform across individuals, while the PISA is shorter and uses different modules with different students, which allows for comparability across countries but does not give as much information about an individual student's performance as we would like.

Second, because we have the data for almost the entire relevant population of Chile for children born over a period of several years, we have a large number of sets of twins and of siblings, which allows us to control for family background, as well as unobserved school and neighborhood characteristics that they share. We cannot entirely control for genetic ability using these data—unlike studies that use genetically identical homozygous twins—but the fact that twins of opposite gender share more genetic information, on top of identical prenatal care, than two randomly chosen students of opposite gender will help us understand the genetic component of the gender gap.

TABLE 1
LINEAR REGRESSION OF TOP SCORES FOR POOLED PISA DATA, FULL SAMPLE OF PISA COUNTRIES

	Above Quantile 50			Above Quantile 75			Above Quantile 90			Above Quantile 95		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)				
Male	.0755*** (00343)	.0618*** (00664)	.0737*** (00366)	.0584*** (00649)	.0465*** (00317)	.0389*** (00589)	.0284*** (00175)	.0241*** (00334)				
Age	-.0554*** (00650)	.00347 (00571)	-.0280*** (00494)	.0158*** (00445)	-.00920*** (00312)	.0135*** (00301)	-.00384 (00272)	.00898*** (00258)				
Year 2009	-.00101 (00467)	-.0294*** (00495)	-.00170 (00502)	-.0197*** (00508)	-.00247 (00430)	-.0104** (00478)	-.00131 (00272)	-.00470* (00281)				
Grade 8 or below	-.238*** (00790)	-.171*** (00706)	-.142*** (00522)	-.0916*** (00430)	-.0607*** (00342)	-.0346*** (00264)	-.0298*** (00234)	-.0148*** (00181)				
Grade 10	.263*** (00660)	.143*** (00540)	.198*** (00622)	.105*** (00491)	.103*** (00508)	.0529*** (00412)	.0574*** (00332)	.0291*** (00279)				
Grade 11 or above	.357*** (0131)	.155*** (00806)	.286*** (0119)	.121*** (00763)	.161*** (0105)	.0679*** (00652)	.0930*** (00602)	.0386*** (00366)				
In OECD country		-.139*** (00636)		-.110*** (00501)		-.0580*** (00362)		-.0328*** (00264)				
Male × OECD		.0141* (00824)		.0176*** (00624)		.00772 (00572)		.00411 (00361)				
Mother at ISCED 1		.0341*** (00659)		.00965* (00532)		-.000375 (00366)		-.000782 (00267)				
Mother at ISCED 2		.0359*** (00490)		.0173*** (00395)		.00512 (00324)		.00226 (00234)				
Mother at ISCED 3		.0579*** (00498)		.0399*** (00446)		.0203*** (00460)		.0104*** (00289)				
Mother at ISCED 4		.00116 (00379)		-.00507 (00426)		-.00460 (00297)		-.00375 (00237)				
Mother at ISCED 5		.0256*** (00415)		.0307*** (00455)		.0215*** (00352)		.0144*** (00276)				
Mother at ISCED 6		.0468*** (00514)		.0623*** (00595)		.0476*** (00426)		.0339*** (00432)				
Mother upper blue collar		.00233 (00555)		-.00424 (00580)		-.00505 (00445)		-.00290 (00313)				

000

Mother upper white collar	.0891*** (.00469)	.0669*** (.00375)	.0373*** (.00226)	.0226*** (.00223)
Mother lower white collar	.0214*** (.00395)	.00678* (.00362)	-3.62E-05 (.00329)	-.000653 (.00243)
Father upper blue collar	.0174*** (.00428)	.00798** (.00322)	.00361 (.00246)	.00177 (.00157)
Father upper white collar	.133*** (.00435)	.116*** (.00430)	.0661*** (.00325)	.0380*** (.00227)
Father lower white collar	.0740*** (.00500)	.0608*** (.00539)	.0323*** (.00443)	.0161*** (.00267)
Mother is immigrant	-.0640*** (.00450)	-.0401*** (.00462)	-.0157*** (.00365)	-.00730*** (.00246)
Wealth index	.00326** (.00162)	.00111 (.00191)	-5.18E-05 (.00141)	.000243 (.00108)
11-25 books	.0436*** (.00434)	.0226*** (.00423)	.00674*** (.00245)	.00188 (.00156)
26-100 books	.117*** (.00440)	.0719*** (.00423)	.0320*** (.00305)	.0165*** (.00234)
101-200 books	.169*** (.00506)	.122*** (.00490)	.0601*** (.00360)	.0325*** (.00278)
201-500 books	.246*** (.00548)	.198*** (.00545)	.109*** (.00451)	.0634*** (.00339)
More than 500 books	.225*** (.00716)	.203*** (.00585)	.120*** (.00548)	.0737*** (.00445)
Constant	1.347*** (.441)	.668** (.333)	.229 (.175)	.103 (.112)
Country dummies	Yes .086	Yes .063	Yes .036	Yes .022
R ²	50	75	90	95
Above quantile	1.135	1.312	1.447	1.751
Ratio male to female	1.104	1.234	1.566	1.596

Note. Standard errors (in parentheses) are computed using the method suggested in OECD (2005). $N = 862,283$.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

Third, these administrative data include information on birth weight and gestational age (the time from when a child is conceived to when he or she is born). These data allow us to control for differential (prenatal) parental investment in boy children relative to girl children. It may also relate to some components of ability, if ability is considered to be the potential skill or genetic endowment a person has at birth. Electronic birth records were established in 1992 in Chile; detailed information is collected about every birth, such as birth weight, birth length, and gestational age, as well as basic parental characteristics, such as mother's age, education, and occupational status. The vital statistics used in this article stem from the same database used in Bharadwaj et al. (2011). That paper shows that the vital statistics records in the data set match rather well to nationally published records on births and deaths by year. Under the guidance of the Ministries of Health and Education, the vital statistics data were matched with the educational records for each child. Again, Bharadwaj et al. (2011) show that a large fraction of the population is observed with valid schooling data.

Fourth, some parental time investment variables are available in these data. These come from questions asked of parents, such as whether they do homework with their child or read to their child. These data will help us understand the role of differential postnatal parental investment in boys and girls in the gender gap.

Fifth, the SIMCE data include the school and classroom of each student in the sample, so that we can include school and classroom fixed effects to control for the influence of these environments on the student. We can also examine the effect of classroom size or the gender composition of the classroom on the size of the gender gap, which will be important if boys and girls perform differently in different classroom environments, perhaps because of different attitudes toward competition or because of the need for different teaching styles.

The descriptive statistics for the SIMCE and other Chilean administrative data used are in table A1. Some characteristics to note in the data are that the modal level of education for mothers that have children in the data set is high school, that most mothers do not work (and the majority were unmarried), and that about a quarter of twin pairs are mixed sex.

III. Empirical Analysis

We conduct a simple empirical exercise in which we regress math scores on a host of explanatory variables, including a dummy for student gender. The gender dummy will capture the effect of being male on test performance.

The main specification is as follows:

$$y_{ibst} = \alpha + \beta \text{male}_i + \gamma X_{ibst} + \varepsilon_{ibst}, \quad (1)$$

where y is the standardized (at the country level) test score in math, and the X variables are controls for characteristics at the level of the individual, family, school, and so on. Controls are introduced to be consistent with the hypotheses we test (one at a time in various subsections of this article): i indexes individuals, h households, s schools, and t is for the year or, more precisely, the cohort. Our main parameter of interest, β , is meant to capture the gender gap.

In addition to examining the standardized test score, we estimate linear models for the probability of falling above various percentiles of the test score distribution. In this case, y is a dummy variable for whether the standardized (at the country level) test score in math is in the top 50%, 25%, 10%, or 5%. In table A2, we show that results are consistent while using a logit model for this purpose.

A. PISA-OECD Evidence for Existence of the Gender Gap

Our results for the PISA-OECD data are presented in table 1. As we move from left to right in the table, we move across quantiles.⁴ At the bottom of the table we present the male-female ratio above the given quantile. For each quantile we have two columns, the first with the simple regression (including country fixed effects) and the second with additional controls for household and school characteristics (as well as a variable for OECD membership and an interaction of this with the “male” variable). This analysis always includes year fixed effects plus age and grade of the student.

The regression for the probability of being in the top half of the distribution shows a substantial gap in mathematical test score: table 1 column 1 shows a male-female ratio of 1.125 in this group, despite the fact that only slightly less than half (49.4%) of test-takers are male. This ratio increases monotonically along the distribution, and for the top 5% the ratio is as high as 1.878.

The even numbered columns of table 1 allow us to characterize the math gender gap separately for OECD countries (which make up about half of our sample) and non-OECD countries, a division that corresponds roughly to high income versus low and middle income. We find that the gender gap for the OECD countries is actually somewhat larger on average than the gap for the non-OECD countries. These regressions also present a pattern in the coeffi-

q3

⁴ In the PISA-OECD analysis we do not present the results for the top percentile (1%), as in this case we would be dealing with a very small number of observations.

lients of the control variables that is mostly consistent with our expectations: higher levels of maternal education, higher status maternal and paternal occupations, and more books at home generally increase the probability of being in the upper portion of the test score distribution for all four quantiles studied, and one's mother being an immigrant decreases this probability. Somewhat surprisingly, having a higher wealth index slightly reduces the probability of performing well.

B. Analysis from Chilean Administrative Data

The addition of our Chilean administrative data allows us to go well beyond what is possible using the PISA-OECD data alone. As mentioned in Section II, the Chilean data are extremely rich in many respects, including (i) population data (among other things we are now able to look at the top 1% of achievement), (ii) family background variables, (iii) a large sample of twins, (iv) birth outcomes like birth weight and gestational age, (v) postbirth parental investment measures, and (vi) self-assessed math ability.

1. Existence of the Gender Gap

Table 2 has simple ordinary least squares (OLS) regressions that use the standardized SIMCE math test scores as the dependent variable, with fourth grade

TABLE 2
MAIN REGRESSION SPECIFICATION USING SIMCE RESULTS

	Standardized Fourth Grade SIMCE			Standardized Eighth Grade SIMCE		
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy for male	.0826*** (.00164)	.0874*** (.00162)	.0927*** (.00162)	.196*** (.00305)	.205*** (.00267)	.211*** (.00267)
Log birth weight		.243*** (.00488)	.238*** (.00482)		.245*** (.00801)	.244*** (.00795)
Full-term birth		.00185 (.00205)	-.000208 (.00203)		-.000399 (.00329)	-.00222 (.00326)
Mother's age at birth		.00218*** (.000132)	.00193*** (.000131)		.00138*** (.000218)	.00121*** (.000217)
Unmarried mother		-.0405*** (.00170)	-.0360*** (.00168)		-.0218*** (.00277)	-.0175*** (.00275)
Mother attended high school		.228*** (.00208)	.209*** (.00206)		.161*** (.00330)	.144*** (.00328)
Mother attended college		.422*** (.00332)	.397*** (.00330)		.346*** (.00562)	.322*** (.00559)
Mother employed		.0797*** (.00200)	.0736*** (.00199)		.0601*** (.00337)	.0556*** (.00335)
Additional controls		School FE	Class FE		School FE	Class FE
Observations	1,444,735	1,225,761	1,225,761	422,319	418,017	418,017
R ²	.003	.255	.345	.010	.353	.400

Note. Robust standard errors in parentheses. All specifications have year of test fixed effects.

*** $p < .01$.

scores used in the columns 1–3 and eighth grade scores in columns 4–6. Columns 1 and 4 report the results of the most basic, “unconditional” regression of test score on a dummy variable for whether a student is male. We discuss the remaining columns in the next sections.

The most important result of this exercise is that the coefficient on “male,” which measures the gender gap, remains substantially the same as these control variables are added, at around 0.08 of a standard deviation for fourth grade and around 0.2 of a standard deviation for eighth grade, suggesting that the gender gap is not the result of boys’ and girls’ having systematically different family, classroom, or school environments.⁵ It is also striking to observe that the gender gap more than doubles between fourth and eighth grade.⁶

This doubling also occurs when the sample of students is limited to those for whom we have both fourth grade and eighth grade scores, so that selection is not a plausible cause of this phenomenon. Fryer and Levitt (2010) find a similar pattern among US schoolchildren, noting that “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” (210). In the Chilean data, for computational reasons given to the very large sample sizes, we do not run quantile regressions, but we still investigate the distributional gap through the a linear probability model. The main findings from this model are that the unconditional male-female ratios increase from about 1.3 to 1.9 for fourth graders between the top 10% and the top 1% (table 3). What is again striking is a large jump in the gaps for older children: for eighth graders the male-female ratios jump to 1.4 and 2.6 for the top 10% and 1%, respectively (table 3).⁷ In table A2, we show that these results are robust to a logit fixed effects specification (estimated on a random 20% sample to ease the computational burden).

Another way to examine the dynamics of the gender gap is to examine eighth grade test scores while controlling for fourth grade scores. In this case, if fourth grade scores are reflective of parental investments and other aspects that are driving differences between boys and girls up to grade 4, then controlling for these should lower the main effect of the gender gap in grade 8. Table 4

q4

⁵ One reassuring outcome is that the results of this regression with regard to the eighth grade gender gap are very similar to what the regression using the PISA data for Chile produces for students who are only a year or so older.

⁶ It is important to note that this increase in the gender gap across grades is also replicable in the PISA-OECD data (see a previous version of this article in Bharadwaj et al. [2012]). However, some studies focusing on individual low-income countries find evidence to the contrary; see Suryadarma (2015) for evidence from Indonesia.

⁷ One would also have to consider the possibility that the fourth grade test is not able to capture the very top performances as well as the eighth grade test.

TABLE 3
LINEAR REGRESSION OF TOP SCORES FROM SIMCE DATA

	Probability of Being in the Top 10%		Probability of Being in the Top 5%		Probability of Being in the Top 1%				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Standardized fourth grade SIMCE:									
Dummy for male	.0251*** (.000520)	.0246*** (.000548)	.0253*** (.000563)	.0156*** (.000394)	.0150*** (.000422)	.0155*** (.000433)	.00591*** (.000226)	.00567*** (.000246)	.00588*** (.000253)
Controls		School FE	Class FE		School FE	Class FE		School FE	Class FE
Male/female	1.27	1.26	1.27	1.35	1.33	1.35	1.86	1.67	1.86
Observations	1,444,735	1,225,761	1,225,761	1,444,735	1,225,761	1,225,761	1,444,735	1,225,761	1,225,761
R ²	.002	.119	.173	.001	.082	.133	.001	.038	.085
Standardized eighth grade SIMCE:									
Dummy for male	.0353*** (.000961)	.0337*** (.000879)	.0339*** (.000887)	.0221*** (.000730)	.0208*** (.000680)	.0209*** (.000684)	.00896*** (.000430)	.00812*** (.000414)	.00786*** (.000414)
Controls		School FE	Class FE		School FE	Class FE			
Male/female	1.42	1.40	1.41	1.56	1.50	1.50	2.64	2.36	2.29
Observations	422,319	418,017	418,017	422,319	418,017	418,017	422,319	418,017	418,017
R ²	.003	.230	.265	.002	.175	.210	.001	.089	.122

Note. Robust standard errors in parentheses. All specifications have year of test fixed effects and controls for log birth weight, gestational age, mother's education, age, and marital and employment status.
*** $p < .01$.

TABLE 4
LINEAR REGRESSION OF EIGHTH GRADE TOP SCORES FROM SIMCE DATA CONTROLLING FOR PERFORMANCE IN FOURTH GRADE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Probability of Being in the Top 10%									
Without male interaction term:									
Dummy for male	.0258*** (.00135)	.0267*** (.00131)	.0272*** (.00137)	.0191*** (.00108)	.0189*** (.00105)	.0197*** (.00109)	.00838*** (.000665)	.00769*** (.000657)	.00763*** (.000681)
Top X percentile status in fourth grade (dummy variable)	.438*** (.00206)	.331*** (.00344)	.328*** (.00351)	.349*** (.00215)	.265*** (.00434)	.262*** (.00442)	.207*** (.00223)	.165*** (.00636)	.163*** (.00643)
Controls		School FE	Class FE		School FE	Class FE		School FE	Class FE
Observations	179,785	178,015	178,015	179,785	178,015	178,015	179,785	178,015	178,015
R ²	.210	.351	.379	.133	.261	.289	.048	.133	.163
With male interaction term:									
Dummy for male	.0225*** (.00144)	.0231*** (.00120)	.0234*** (.00127)	.0158*** (.00111)	.0154*** (.000961)	.0161*** (.00101)	.00758*** (.000673)	.00691*** (.000619)	.00689*** (.000646)
Top X percentile status in fourth grade (dummy variable)	.423*** (.00304)	.314*** (.00490)	.310*** (.00501)	.321*** (.00323)	.233*** (.00629)	.229*** (.00641)	.186*** (.00346)	.142*** (.00938)	.141*** (.00949)
Top X percentile status in fourth grade (dummy variable) × male	.0273*** (.00412)	.0312*** (.00647)	.0330*** (.00664)	.0502*** (.00432)	.0577*** (.00851)	.0581*** (.00867)	.0361*** (.00453)	.0397*** (.0127)	.0372*** (.0129)
Controls		School FE	Class FE		School FE	Class FE		School FE	Class FE
Observations	179,785	178,015	178,015	179,785	178,015	178,015	179,785	178,015	178,015
R ²	.210	.352	.379	.134	.262	.290	.048	.133	.163

Note. Other controls: see table 3. Standard errors in parentheses.

*** p < .01.

presents two versions of this test, with the first panel just controlling for fourth grade test scores, and the second panel controlling for an interaction between fourth grade test scores and the male dummy. Focusing on the coefficient for the male dummy, we find that despite controlling for fourth grade performance, males perform better in eighth grade, but these coefficients are smaller (as expected) than the ones in table 3. The interaction term in the second panel shows that males who do well in fourth grade do even better by eighth grade.

The gender gap in test scores may influence later wage differentials by gender. At the very least, the description of the math test score gender gap from the previous paragraph is consistent with the form that the wage gender gap among adults takes on. Ñopo (2006), for example, finds that the unexplained gender gap in wages in Chile is at around 25% of average female wages with an increasing gap by wage percentile: “While for the lowest percentiles of the wage distribution, males tend to earn an unexplained premium of 10 percent to 20 percent over comparable females, at the top of the distribution this premium increases to 40 percent to 80 percent, depending on the set of matching characteristics” (31).

Despite the gender gap in test scores, women receive more years of school than men on average in Chile, which one could hypothesize would compensate for some of the test score difference. Ñopo (2006) shows that, over 1992–2003, women had 1.6 more years of schooling on average than men in rural areas and 0.5 years more in urban areas. Interestingly, the gender gap in wages was nonexistent in rural areas and quite large in urban areas, and the largest gender gap in wages by far was among college graduates. Education levels were increasing throughout this period. Nevertheless, women had relatively low, although increasing, rates of labor force participation during this period, of around 35%.

2. Family Background and School and Class Characteristics

In order to test whether family background and school and class characteristics are crucial in explaining the gap, we first add control variables able to capture such confounders, such as birth weight and gestational age of the student and the education, employment, marital status, and age of the student’s mother. Table 2 columns 2 and 5 also include school fixed effects, while columns 3 and 6 have classroom fixed effects instead.

Once again, columns 2, 3, 5, and 6 of table 2 confirm the previous “unconditional” findings reported in columns 1 and 4. The average gender gap cannot be explained on the basis of the maternal characteristics included or classroom and school selection. Looking at the control variables in table 2, the children of more educated women had much better test scores, while the chil-

dren of older mothers and married mothers did somewhat better. The log of birth weight also has a statistically significant impact of about 0.25 standard deviations on both fourth grade and eighth grade test scores, indicating that a student who had a birth weight that was 10% higher could expect a math test score 0.025 standard deviations higher.

We can look at the right tail of the distribution, with the results shown in table 3. We examine how the probability of being in the top 1%, 5%, or 10% relates to the dummy variable for being male and the same set of control variables used in table 2. In order to have a consistent basis of comparison as the base percentage changes, as for the PISA-OECD analysis, we also calculate the predicted male-to-female ratio. We find that, as in the PISA data, the gender gap increases at higher scores: for the fourth grade specification with class fixed effects, there are predicted to be 1.27 male students for every female in the top 10%, 1.35 males per female in the top 5%, and 1.86 males per female in the top 1%, while—consistent with the increase in coefficient in the OLS regression above—these ratios are 1.41, 1.50, and 2.33 among eighth graders. These results are not sensitive to the use of fixed effects at the school instead of the class level.⁸

We performed the same analysis for the twins sample, both on the continuous (mean) test outcome as well as at the top end of the distribution. While twins-based comparisons are useful from a research perspective, we present the twins results with the usual caveat that twins studies might not be generalizable to the broader population (see also Bharadwaj et al. [2011] and Figlio et al. [2014] for more on this). Nevertheless, it helps us control for key unobservables at the household level. The findings of this analysis are presented in table 5. While the twins analysis is meant to address several of the proposed explanations for the gender gap, in this section the focus is on parental background, which twins typically share. As is clear from the table, if anything, the gap is larger across the board for twins.

3. Parental Investment

Prenatal Investment

In this section we examine whether differences in parental investments can help explain some of the gender gap in math test scores. Our motivation for examining this channel begins with the observation that studies have shown that boys typically tend to perform worse on early assessments of health such

⁸ Once again because of standard concerns with predictions from a linear probability model like this (in the tails), we also estimated a logit model with similar results (table A2).

TABLE 5
REGRESSIONS ON TWINS SAMPLE

	Mean	Probability of Being in the Top Percentile			
		25	10	5	1
Standardized fourth grade SIMCE:					
Dummy for male	.148*** (.0167)	.0643*** (.00893)	.0485*** (.00698)	.0305*** (.00568)	.00774** (.00349)
Constant	-.137*** (.00870)	.202*** (.00466)	.0741*** (.00364)	.0388*** (.00296)	.0129*** (.00182)
Male to female ratio		1.26	1.67	1.99	2.08
Observations	22,074	22,074	22,074	22,074	22,074
Number of twin groups	12,461	12,461	12,461	12,461	12,461
Standardized eighth grade SIMCE:					
Dummy for male	.269*** (.0327)	.115*** (.0173)	.0728*** (.0136)	.0405*** (.0110)	.0113* (.00674)
Constant	-.187*** (.0166)	.187*** (.00881)	.0668*** (.00693)	.0379*** (.00561)	.0120*** (.00342)
Male to female ratio		1.57	1.90	1.94	3.44
Observations	6,293	6,293	6,293	6,293	6,293
Number of twin groups	3,834	3,834	3,834	3,834	3,834

Note. Robust standard errors in parentheses.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

as Apgar scores (Gissler et al. 1999), even though they are typically born with higher birth weight. Moreover, in terms of early childhood cognitive achievement, females tend to perform better. Early studies such as Willerman, Naylor, and Myriantopoulos (1970), using data from the United States, show that females perform significantly better on tests such as the Bayley Motor Test (administered at age 8 months). According to the same study, females also perform better at the Binet IQ test administered at age 4. Simple correlations using data from the children of the National Longitudinal Survey of Youth 1979 sample suggest that females do indeed perform better on early motor and social development skills tests (administered to children between age 0 and 3). However, in tests such as the Peabody Picture and Vocabulary Test, which is administered to children age 3 and up, gender gaps appear in later ages, with males performing slightly better than females. Hence, it appears that girls do better than boys along various health and cognitive measures in very early childhood.

The most convincing evidence we provide relies on the twins sample, since twins are subjected to the same investment or care in the womb. The richness of the data allows us to fully difference out prenatal investment and family background characteristics through the use of data on twins. This is not to say that twins are a representative sample of the population; in fact, they are quite

clearly at the bottom end of the distribution in many respects, including test scores.

Table 5 contains the results of OLS and probability regressions similar to those above, with the sample restricted to twins. These results display the same trend of a gender gap that increases with age and at higher percentiles of the score distribution seen in previous tables. In fact, the twins results display an even larger gender gap than that in the overall sample. Twins are, of course, not average individuals, but the persistence of the gender gap among them does allow us to conclude that differential prenatal parental investment perhaps does not play a large role in creating this gap. As we can see from table 5, girls fare worse than boys in math test scores even between twins.

Postnatal Investment

It is still possible that parents invest more into boys than girls because of son preference (although this is unlikely in the case of Chile) or because of higher returns on the investment (wage differentials). We have some measures of parental investment in the Chilean data, which can be summarized as mathematical or reading investments. For example, for the 2002 data on fourth graders, the only year of data used in this part of the analysis, parents were asked: “How often do you pose math problems to your child?” They could reply in five categories from 1 (never) to 5 (very often). Similar questions were asked for reading investment.⁹

We introduce such variables in our empirical model for the fourth grade test as continuous controls, the intuition being that if postnatal parental investment is the fundamental cause of the gender gap in math, once one controls for that (and other correlated confounders) the gap should shrink or disappear.¹⁰ Clearly, parental investment is not a random event, so the analysis here is not able to make causal statements about the role of parental investments in the education production function. For our purposes, however, what we need to understand is whether parental investments are correlated with the gender of the child. When we control for parental investments as shown in table 6, the gender gap remains relatively unchanged; note that this table also includes the interaction of parental investments with the male dummy. As expected, the direct effect of parental investments appears to have a large and positive effect on test scores, and the results are also indicative of a positive correlation between math investments and boys (the interaction term between

⁹ The comparable reading-related question is how often the parent reads to the child.

¹⁰ The large majority of the students, 93% of the fourth graders with valid SIMCE scores, have valid parental investment information.

TABLE 6
FOURTH GRADE SIMCE MATH SCORES AND PARENTAL INVESTMENTS

	(1)	(2)	(3)
Dummy for male	.0796*** (.00770)	.0885*** (.00682)	.0914*** (.00677)
Investments in math	.168*** (.00693)	.110*** (.00602)	.104*** (.00598)
Investments in reading	.103*** (.00710)	.0148** (.00621)	.0161*** (.00618)
Male × investments in reading	−.00327 (.0102)	−.00184 (.00897)	−.00519 (.00890)
Male × investments in math	.0196** (.00988)	.0112 (.00861)	.0122 (.00853)
Controls		School FE	Class FE
Observations	184,144	182,189	182,189
R ²	.015	.301	.339

Note. Other controls: see table 3. Robust standard errors in parentheses.

** $p < .05$.

*** $p < .01$.

math investments and the male dummy is positive throughout and just shy of significance in the last two columns). Even if our measures of parental investments do not capture the entire range of possible investments, we can turn to the evidence from the twins analysis, under the assumption that within twins, the difference in investment is minimal.

The central message is that parental investment, measured as time spent challenging children in math and reading, has a positive and significant impact on the overall performance of boys and girls but does not explain the gender gap, while at the same time we do find some indication for differential investment between boys and girls. Hence, we believe that further research along this measure would be important. One drawback of our parental investment measure is that it only measures investments as a flow at a certain point in time, rather than a stock that has accumulated since birth. Perhaps differences in the stock of parental investments are much more important for explaining the gender gap.

4. Classroom Environment

Tables 7 and 8 explore the role of the class environment in explaining the gender gap. These tables explore various aspects of the classroom environment already captured by the classroom fixed effects regressions seen earlier. Hence, while it is useful to understand the extent to which individual classroom-level variables matter for the gender gap, we do not expect these variables to explain the gap since the classroom fixed effects regressions did not appear to matter much for changing the coefficient of interest.

Table 7 shows that adding teacher gender as an explanatory variable does not change the gender gap much. Approximately 20% of the fourth grade math

TABLE 7
TEACHER GENDER AND MATH SCORES

	(1)	(2)	(3)
Dummy for male	.0914*** (.00209)	.0914*** (.00209)	.0896*** (.00226)
Math teacher male		-.0635*** (.00571)	-.0703*** (.00648)
Male × math teacher male			.0127** (.00602)
R ²	.256	.257	.257

Note. OLS: standardized fourth grade SIMCE score. Standard errors clustered at the classroom level in parentheses. Other controls include class size, fraction male in class, and school fixed effects. All regressions also include log birth weight, gestational age, mother's age, education, and marital and employment status. Teacher gender information not available for 2009 and 2010. $N = 797,102$.

** $p < .05$.

*** $p < .01$.

teachers in Chile are male. Typically, a teacher who teaches math in this grade also teaches other subjects like language and reading. While having a male teacher in general implies lower math scores, boys do slightly better in the presence of a male teacher (col. 3). However, the main message from these regressions is that (selection issues aside) a teacher's gender does not appear to matter much in terms of explaining the gender gap.

A very influential experimental literature (Gneezy, Niederle, and Rustichini 2003; Niederle and Vesterlund 2007; for a summary, see Niederle and Vesterlund 2010) finds that females tend to shy away from competition and that such behavior can explain a substantial part of the gender gap in performance in many realms that involve a significant competitive element, such as taking tests in school. It would be very hard, however, to square these results in particular experiments with experiments that involve verbal competition tasks, where it appears that girls do better than boys. In order to investigate such an issue, we introduce among the control variables the classroom composition in terms of share of male students. The thought experiment is as follows: as the fraction of males in the classroom increases, females would feel a higher competitive pressure, which might hamper their performance. If this is a principal cause of the gender gap, female students should perform the same as males in predominantly female classrooms.

To investigate the influence of the competitive environment in the classroom on boys and girls, we regressed fourth and eighth grade test scores on different combinations of the male dummy, the proportion of males in the student's fourth grade class, the number of students in the student's fourth grade class, and interactions of "male" with proportion male and class size. Table 8 reports the results of these regressions. We find that test scores go down for both boys

TABLE 8
SIMCE SCORES REGRESSED ON CLASS SIZE AND COMPOSITION

	Full Sample			Excluding Single-Sex Schools
	(1)	(2)	(3)	(4)
Fourth grade SIMCE scores:				
Male × class size	.00109*** (.000121)		.000930*** (.000122)	.000989*** (.000121)
Male	.0514*** (.00427)	.0869*** (.00891)	.0545*** (.00998)	.0622*** (.0101)
Class size	-.000729*** (.000180)		-.000676*** (.000180)	-.000717*** (.000188)
Male × fraction male		.0103 (.0168)	.0133 (.0168)	-.00549 (.0172)
Fraction male		-.109*** (.0137)	-.103*** (.0138)	-.127*** (.0167)
Observations	1,225,761	1,225,761	1,225,761	1,125,046
R ²	.255	.255	.255	.245
Eighth grade SIMCE scores:				
Male × class size	.000983*** (.000335)		.000561 (.000519)	.000495 (.000524)
Male	.175*** (.0105)	.246*** (.0119)	.213*** (.0247)	.203*** (.0257)
Class size	.00728*** (.000570)		.0108*** (.00102)	.0109*** (.00108)
Male × fraction male		-.0641*** (.0222)	-.0394 (.0328)	-.0153 (.0357)
Fraction male		-.0821*** (.00905)	-.0428*** (.0149)	-.0391** (.0177)
Observations	418,017	418,017	181,029	162,548
R ²	.354	.354	.393	.366

Note. Standard errors clustered at the classroom level in parentheses. Controls include birth weight, gestational age, mother's age and education, mother employment status, and school fixed effects.

** $p < .05$.

*** $p < .01$.

and girls in fourth grade as the proportion of the class that is male increases, and that the effect is not significantly different by sex. For the eighth grade, we find that both sexes again do worse in the presence of more boys and that, in this case, boys experience this effect even more strongly than girls. Class size has a small negative effect on girls and a small positive effect on boys in fourth grade and a small positive effect on both sexes in eighth grade, even controlling for school fixed effects. We might be concerned that a few outliers, classes that had almost all boys or almost all girls, are inordinately driving these results, so we also include the results of the regression excluding schools in which all students are of the same gender. This does not materially change the results. Thus, because we do not see a strong positive coefficient on the interaction of “male” and “fraction male,” we learn from these class composition regressions that competition be-

tween boys and girls does not seem to be the main fact explaining the gender gap in math test scores either.

5. Self-Assessed Ability

A unique feature of the data from Chile is the ability to measure self-assessed ability in math in boys and girls in fourth grade. In table 9 we examine whether boys and girls differ in various aspects of how difficult they perceive math to be, conditional on their math score. What is perhaps surprising is that even conditional on math scores, across the wide range of questions, girls are much more likely to be pessimistic about their math abilities. For example, boys are 10% more likely to say that they are “good at math” and 8% more likely to say that they get good grades in math without studying much. Conditioning on math score matters significantly for the coefficient on gender, suggesting that self-assessed perceptions and attitudes might go a long way toward explaining some of the gender gap in math (an excellent, related paper on this topic is González de San Román and De La Rica [2012]).

The main empirical problem here is that it is unclear whether perceptions affect math scores or vice versa, and hence, our analysis here is largely suggestive. However, we are able to provide some evidence that boys are not always more confident of their abilities than girls. In table A3 we examine whether boys and girls differ in other attitudes toward school not related to math. While boys are still more likely to say that it is “important for me to get good grades,” relative to the mean response for this variable, the magnitude is quite small. Boys are less

q7

TABLE 9
GENDER AND PERCEPTIONS ABOUT MATH ABILITY

	In General I Am Quite Good at Math	I like Math Classes in My School	Math Is Harder for Me than for the Rest of My Peers	I Learn Math Quickly and Easily	I Am Not Good at Math	I like to Study Math	I Get Good Grades in Math without Studying
Dummy for male	.104*** (.00164)	.0272*** (.00163)	-.0168*** (.00136)	.116*** (.00163)	-.0139*** (.00124)	.0454*** (.00169)	.0885*** (.00154)
SIMCE math score	.164*** (.000907)	.0184*** (.000940)	-.132*** (.000766)	.134*** (.000916)	-.0951*** (.000715)	.0619*** (.000960)	.119*** (.000901)
Observations	364,337	389,653	387,355	386,128	383,740	372,307	368,427
R ²	.152	.086	.114	.106	.086	.079	.104
Mean of dependent variable	.477	.443	.223	.448	.168	.459	.301

Note. Robust standard errors in parentheses, clustered at the classroom level. Controls are year of SIMCE test, full-term birth, mother’s education, marital status, and age. School fixed effects in all regressions.

*** $p < .01$.

likely to agree that they “understand very little of what happens in class,” but the magnitudes are small.

IV. Concluding Remarks

Motivated by the existence of a gender gap in math in many countries, we investigated the dynamics and underlying causes behind the gender gap using detailed microdata in Chile. We explored the roles played by parents (background characteristics as well as investment behavior), classroom environments including teacher’s gender and class composition, and individual characteristics of the students. None of those seems to be able to account for a substantial portion of such a gap. Similarly, controlling for school, and even class, fixed effects does very little to the gender gap, suggesting that sorting across schools or classes is not what causes this gap.

There are two avenues for future research in this area. One promising path is to better explore the role of gender attitudes and self-perceptions toward math and the extent to which changing perceptions about math ability can influence performance on math tests. A second avenue is to examine, in a comprehensive manner, a host of education-based interventions that have already been examined in a developing-country context and explore heterogeneity in the impacts of those interventions by child gender. While these individual results likely already exist, a meta-analysis approach to understanding the heterogeneous impacts by gender of various education-based RCT programs would shine light on whether such interventions in turn can help alleviate the gender gap in mathematics.

q8

Appendix

TABLE A1
DESCRIPTIVE STATISTICS FOR CHILEAN ADMINISTRATIVE DATA

	Observations	Mean	SD
Birth weight (in grams)	1,687,269	3,357.58	512.08
Fraction full-term births (38–40 weeks gestation)	1,687,269	.804	.396
Fraction of mothers married	1,481,512	.41	
Mother’s education: primary school	1,679,693	.27	
Mother’s education: high school	1,679,693	.58	
Mother’s education: college	1,679,693	.14	
Fraction of mothers employed	1,686,582	.25	
Fraction twins both male	25,619	.34	
Fraction twins of mixed sex	25,619	.24	
Fraction twins both female	25,619	.42	
Fraction of students male in sample	1,687,269	.502	.50
Class size in fourth grade	1,675,350	31.82	13.88
Fraction male in fourth grade	1,675,350	.51	.18
Class size in eighth grade	784,973	34.33	16.3
Fraction male in eighth grade	784,973	.50	.18

TABLE A2
FIXED EFFECTS LOGISTICAL REGRESSION: TOP SCORES FROM SIMCE DATA

	Probability of Being in the Top 10%			Probability of Being in the Top 5%			Probability of Being in the Top 1%		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Fourth grade:									
Dummy for male	.266*** (.0120)	.293*** (.0150)	.316*** (.0170)	.285*** (.0159)	.292*** (.0195)	.316*** (.0222)	.306*** (.0279)	.281*** (.0336)	.336*** (.0381)
Controls		School FE	Class FE		School FE	Class FE		School FE	Class FE
Observations	289,033	212,085	98,752	289,033	183,889	64,360	289,033	119,577	25,728
Eighth grade:									
Dummy for male	.329*** (.0224)	.434*** (.0289)	.438*** (.0317)	.349*** (.0295)	.434*** (.0373)	.433*** (.0413)	.433*** (.0500)	.484*** (.0613)	.488*** (.0680)
Controls		School FE	Class FE		School FE	Class FE		School FE	Class FE
Observations	84,205	50,790	29,971	84,205	37,606	19,604	84,205	21,497	8,605

Note. Other controls: see table 3. Both panels use a random 20% sample of the data. Standard errors in parentheses.
 *** $p < .01$.

TABLE A3
GENERAL ATTITUDES TOWARD CLASSES AND GENDER

	I Am as Capable of Learning as the Rest of My Peers	Generally Understand Very Little of What Happens in Class	Have Trouble Concentrating in Class	I Do Certain Tasks Even Though I Find Them Difficult	My Notebooks Are Usually Incomplete	I like to Study for Tests	It Is Important for Me to Get Good Grades	I Take Good Notes in Class
Male	-.0191*** (.00108)	-.0168*** (.00136)	.000815 (.00129)	.116*** (.00163)	-.0139*** (.00124)	.0454*** (.00169)	.0885*** (.00154)	.0610*** (.00160)
SIMCE math score	.0398*** (.000609)	-.132*** (.000766)	-.0887*** (.000740)	.134*** (.000916)	-.0951*** (.000715)	.0619*** (.000960)	.119*** (.000901)	.0582*** (.000931)
Log birth weight	-.00315 (.00334)	-.0233*** (.00420)	-.00641 (.00397)	.0506*** (.00487)	-.0236*** (.00386)	.0241*** (.00507)	.0385*** (.00456)	.0321*** (.00480)
Constant	.977*** (.0273)	.470*** (.0349)	.260*** (.0330)	-.0471 (.0401)	.449*** (.0322)	.0826** (.0415)	-.0121 (.0376)	.0355 (.0395)
Observations	391,258	387,355	385,156	386,128	383,740	372,307	368,427	371,059
R ²	.067	.114	.111	.106	.086	.079	.104	.081
Mean of dependent variable	.878	.200	.303	.700	.150	.433	.951	.467

Note. Controls are year of SIMCE test, full-term birth, mother's education, marital status, and age. School fixed effects in all regressions. Robust standard errors in parentheses.

** $p < .05$.

*** $p < .01$.

References

- Altonji, J. G., and R. Blank. 1999. "Race and Gender in the Labor Market." In *Handbook of Labor Economics*, vol. 3c. Amsterdam: Elsevier.
- Altonji, J. G., E. Blom, and C. Meghir. 2012. "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers." Discussion paper, National Bureau of Economic Research, Cambridge, MA.
- Bharadwaj, P., G. De Giorgi, D. Hansen, and C. Neilson. 2012. "The Gender Gap in Mathematics: Evidence from Low- and Middle-Income Countries." Discussion paper, National Bureau of Economic Research, Cambridge, MA.
- Bharadwaj, P., K. Loken, and C. Neilson. 2011. "Early Life Health Interventions and Academic Achievement." Working paper.
- Black, S., P. Devereux, and K. Salvanes. 2007. "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *Quarterly Journal of Economics* 122, no. 1:409–39.
- Carrell, S., M. Page, and J. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics* 125, no. 3:1101–44.
- Dee, T. 2007. "Teachers and the Gender Gaps in Student Achievement." *Journal of Human Resources* 42, no. 3:528–54.
- Ellison, G., and A. Swanson. 2010. "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions." *Journal of Economic Perspectives* 24, no. 2:109–28.
- Fortin, N., P. Oreopoulos, and S. Phipps. 2012. "Leaving Boys Behind: Gender Disparities in High Academic Achievement." Working paper.
- Fryer, R. G., and S. D. Levitt. 2010. "An Empirical Analysis of the Gender Gap in Mathematics." *American Economic Journal: Applied Economics* 2:210–40.
- Gissler, M., M.-R. Järvelin, P. Louhiala, O. Rahkonen, and E. Hemminki. 1999. "Can Children's Health Be Predicted by Perinatal Health?" *International Journal of Epidemiology* 28:276–80.
- Gneezy, U., M. Niederle, and A. Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics* 118, no. 3:1049–74.
- Goldin, C., L. F. Katz, and I. Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives* 20, no. 4:133–56.
- González de San Román, A., and S. De La Rica. 2012. "Gender Gaps in PISA Test Scores: The Impact of Social Norms and the Mother's Transmission of Role Attitudes." IZA Discussion Paper no. 6338, Institute for the Study of Labor, Bonn.
- Grogger, J., and E. Eide. 1995. "Changes in College Skills and the Rise in the College Wage Premium." *Journal of Human Resources* 30, no. 2:280–310.
- Guiso, L., F. Monte, P. Sapienza, and L. Zingales. 2008. "Culture, Gender, and Math." *Science* 320, no. 5880:1164–65.
- Machin, S., and T. Pekkarinen. 2008. "Global Sex Differences in Test Score Variability." *Science* 322, no. 5906:1331–32.
- Murnane, R. J., J. B. Willett, Y. Duhaldeborde, and J. H. Tyler. 2000. "How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?" *Journal of Policy Analysis and Management* 19, no. 4:547–68.

q9

q10

- Murnane, R. J., J. B. Willett, and F. Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." NBER Discussion Paper w5076, National Bureau of Economic Research, Cambridge, MA.
- Niederle, M., and L. Vesterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics* 122, no. 3:1067–1101.
- . 2010. "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives* 24, no. 2:129–44.
- Ñopo, H. 2006. "The Gender Wage Gap in Chile, 1992–2003, from a Matching Comparisons Perspective." Discussion paper, Inter-American Development Bank.
- OECD. 2005. *PISA 2003 Data Analysis Manual*. Paris: OECD.
- Paglin, M., and A. Rufolo. 1990. "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences." *Journal of Labor Economics* 8:123–44.
- Sen, A. 1999. *Development as Freedom*. New York: Knopf.
- Suryadarma, D. 2015. "Gender Differences in Numeracy in Indonesia: Evidence from a Longitudinal Dataset." *Education Economics* 23, no. 2:180–98.
- UNESCO. 2010. "Global Education Digest 2010: Comparing Education Statistics across the World." Discussion paper, UNESCO Institute for Statistics.
- Weinberger, C. J. 1999. "Mathematical College Majors and the Gender Gap in Wages." *Industrial Relations* 38:407–13.
- . 2001. "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.* Ann Arbor: University of Michigan Press.
- Willerman, L., A. Naylor, and N. Myrianthopoulos. 1970. "Intellectual Development of Children from Interracial Matings." *Science* 170, no. 964:1329–31.

Abstract

Using a large administrative data set from Chile, we find that on average boys perform better than girls in math. In this article, we document several features of boys versus girls' relative performance in math. First, we note that the gender gap appears to increase with age (doubles between fourth grade and eighth grade). Second, we test whether commonly proposed explanations such as parental background and investments, unobserved ability, and classroom environment (including teacher gender) help explain a substantial portion of the gap. While none of these explanations help in explaining a large portion of the gender gap, we show that boys and girls differ significantly in perceptions about their own ability in math. Conditional on math scores, compared to boys, girls are much more likely to state that they dislike math or find math difficult. We highlight differences in self-assessed ability as areas for future research that might lead to a better understanding of the gender gap in math.

QUERIES TO THE AUTHOR

Q1. AU: World Development Report 2012 not in reference list; please add, or omit citation here. (See also query 10.)

Q2. AU: Defined SIMCE as Sistema de Medición de la Calidad de la Educación here, OK? If not, please define (i.e., spell out) here as first use.

Q3. AU: (1) The male-female ratio in column 1 is given as 1.135, not 1.125 as stated here. Please revise text or table for agreement. Also, (2) the ratio of 1.878 (the “high” mentioned in the last sentence of this paragraph) is not shown in table 1. Is this OK?

Q4. AU: Table 3: Are school FE and class FE intentionally omitted from the 8th grade controls row in columns 8 and 9, respectively?

Q5. AU: Is 2.33 (here) or 2.29 (table 3) correct? (See also query 4.)

Q6. AU: Figlio et al. 2014 not in reference list; please add, or omit citation here.

Q7. AU: Should 8% instead be 9% here to better approximate rounding 0.0885 from table 9?

Q8. AU: Please define RCT here.

Q9. AU: Please provide institution for this working paper.

Q10. AU: (1) Fortin et al. 2012 not cited; please cite, or omit from References. (Is this related to the uncited report from query 1?) Also, (2) please provide institution for this working paper.