The Gender Gap in Mathematics: The Role of the Testing Environment

José Montalbán (SOFI at Stockholm University)
**Almudena Sevilla** (UCL-IoE)

SRI Gender Equality Workshop Program

Thursday 21st April, 2022

What is the gender gap in mathematics?

- On average, girls' math results are 9 percent of a standard deviation lower than boys' at 15 years of age

  ▸ Math and Reading Gender Gap PISA 2015

- Highest at the top of the score distribution (Machin and Pekkarinen 2008, Ellison and Swanson 2010)

- Non-existent in childhood, more than triples from about 3 percent of a standard deviation at age 9/10 to about 9 percent at age 15/16 (Borgonovi et al. 2018; Bharadwaj et al. 2016; Contini et al 2017; Ellison and Swanson 2018; Borra, Iacovou, and Sevilla, 2021)

Gender Differences in Math and Reading. PISA 2015
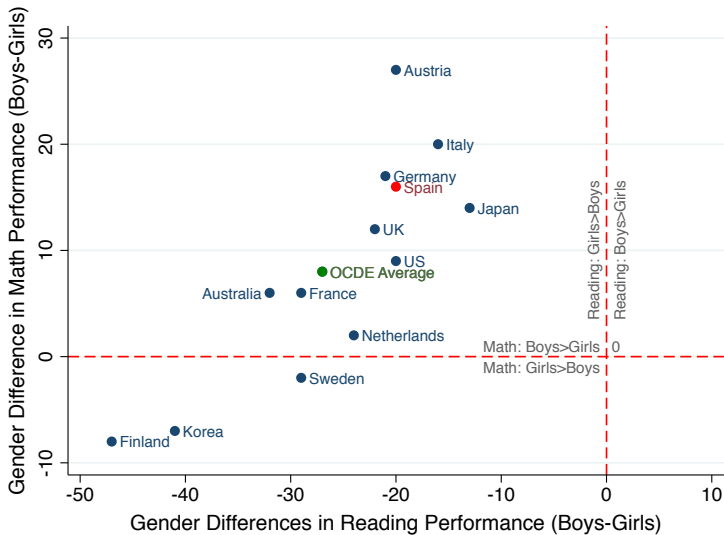
Why we care about the gender gap in mathematics?

- Gender differences in the labor market and academic attainment and achievement have reversed over the past 60 years (Goldin et al. 2006, Gimenez-Nadal and Sevilla, 2011)

- Positive effect of math performance on future income (for examples and discussion, see Paglin and Rufolo, 1990; Murnane, Willet, and Levy, 1995; Grogger and Eide, 1995; Weinberger, 1999, 2001; Murnane, Willett, Duhaldeborde, and Tyler, 2000; Altonjii and Blank, 1999).

Why we care about the gender gap in mathematics?

- Under the assumption than men and women share at birth similar distributions of mathematics-relevant talent, the mathematics gender gap is likely to lead to a waste of talent and lower economic growth, e.g., Cuberes and Teignier 2016, Hsieh et al. 2019)

- If one were to choose two people at random, the differences across a multitude of traits and abilities between a man and a woman (or a boy and a girl) will be much smaller than the differences between two women (girls) or the differences between two men (boys) (Hyde, 2005, Bertrand, 2020) ▸ Gender Differences in Preferences

- The plastic brain- higher boy exposure to spatially complex environments leads to brain changes (Berenbaum, Martin, Hanish, Briggs, and Fabes, 2008; Rippon, 2019)

Gender Differences in Preferences

HEIGHT

WOMEN'S AVERAGE

MEN'S AVERAGE

SHORT   TALL

EFFECT SIZE: 1.72

SELF-ESTEEM/CONFIDENCE

WOMEN'S AVERAGE

MEN'S AVERAGE

LOW   HIGH

EFFECT SIZE: 0.10

NOTE STATISTICIANS CONSIDER AN EFFECT SIZE OF LESS THAN 0.20 TO BE "TRIVIAL,"
0.20–0.49 TO BE "SMALL," 0.50–0.79 TO BE "MEDIUM," AND 0.80 OR MORE TO BE "LARGE."
FROM "WHAT MOST PEOPLE GET WRONG ABOUT MEN AND WOMEN,"
BY CATHERINE H. TINSLEY AND ROBIN J. ELY, MAY–JUNE 2018

© HBR.ORG

◂ Back

Stereotypes and Gender Gap in Mathematics

- Gender stereotypes are fixed ideas about what someone of a particular gender is like (descriptive) or should be like (prescriptive), simply for the mere fact of belonging to a certain gender group

- Prescriptive stereotypes: Gender Roles (Guiso et al 2008; Nollenberger, Rodriguez, and Sevilla, 2016; Dossi et al., 2021)

- Descriptive stereotype: "boys are better at math"
  1. Family: Mothers who endorsed a male-math stereotype underestimated their daughters' ability in math, child's confidence (Jacobs, 1991)
  2. Teachers: More stereotypical teachers 1- advise girls to pursue less maths-intensive subject choices, and 2- girls taught by these teachers display a lower confidence in the subject (Carlana, 2019).

- Women/Girls have lower expected returns, have lower self-confidence (Niederle and Vesterlund, 2011), choke under pressure (stereotype threat- Steel, 1997; Cai et al., 2019) in male-dominated tasks

Stereotypes and Test-taking Environments

- **Research Question:** Does the familiarity with the testing environment impact the relative performance of boys and girls in mathematics?

Familiar vs. Unfamiliar Testing Environments

- **In 2015 - Unfamiliar (high pressure/high perceived stakes):**
Externally Administered Testing. School teachers did not talk to
students about the tests and the testing procedure, and did not
emphasize the low-stakes of the test. Invigilators (and markers)
were allocated centrally (by the Madrid Regional Government)
from other schools

- **2016 onwards- Familiar (low pressure/low perceived stakes):**
Internally Administered Testing. School teachers talk to students
about the tests and the testing procedure, and emphasized the
low-stakes of the test. Invigilators (and markers) were allocated
by the school (they cannot be the student's main teacher)

Contributions I

- Competititive pressure (e.g., Niederle and Vesterlund 2007, 2010; Buser, Niederle and Oosterbeek, 2014; Buser and Yuan, 2019; Iriberri and Rey-Biel, 2019)

- Pressure from Stakes (e.g., Azmat, Calsamiglia and Iriberri, 2016; Cai, Lu, Pan and Zhong, 2019)

- Pressure from time constraints (e.g., Shurchkov, 2012; De Paola and Gioia, 2016; Galasso and Profeta, 2018)

This paper: **Pressure from unfamiliar test-taking environments**

Contributions II

Mechanisms: We can test directly the mechanism at play (e.g., stress, self-confidence, level of effort)

Contributions III

External Validity: Population-level data for the entire region of Madrid

Other factors constant (e.g., competitiveness, stakes, skills being tested)

Outline

Standardized testing in Spain

- Spanish organic law of education *LOMCE* (**Article 21.1**)

- Performed in the Region of Madrid since 2015/2016:
    1. 6th Grade: Mathematics, Science, Spanish and English
    2. 10th Grade: Mathematics, Spanish, English and "Social and Civic Competence"

- **Type of test**:
    1. Mathematics and Science: Multiple-choice
    2. English and Spanish: Linguistic competence (multiple-choice test) and written communication (written test).

- **Length**:
    1. 6th Grade: Math (60 minutes), Science (40 minutes), Spanish (80 minutes), and English (75 minutes)
    2. 10th Grade: 1 hour

Randomization: Familiar vs. Unfamiliar Testing Environments

- **Randomization unit**: School.

- **School selection**: Random and stratified at the District and school ownership level (public, private, semi-public)

- Five districts: Capital, North, South, East, West

- **Timing** of Randomization: ▸ Map

  1. **2015**: No Randomization
  2. **2016**: Randomization in North, South, East and West (Capital excluded)
  3. **2017**: Randomization in Capital, East and West (North and South excluded)
  4. **2018**: Randomization in Capital, North and South (East and West excluded)
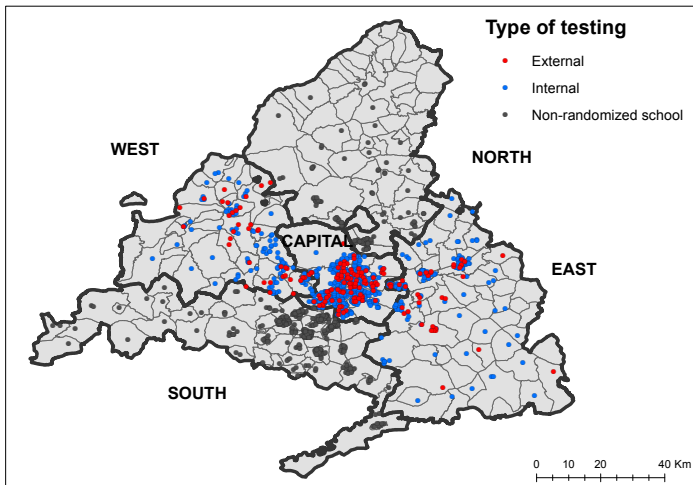
- **Why this policy?** Costs
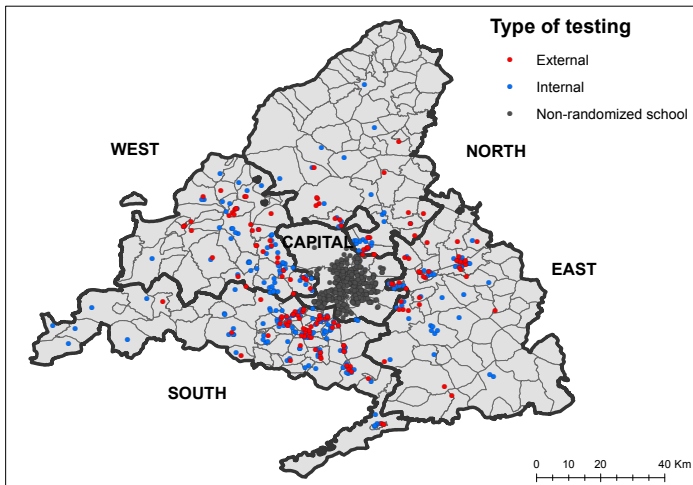
Distribution of schools in Grade 6 in 2016/17
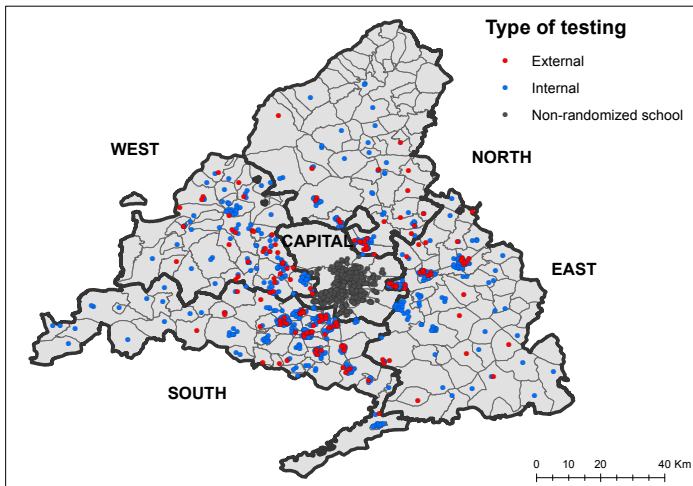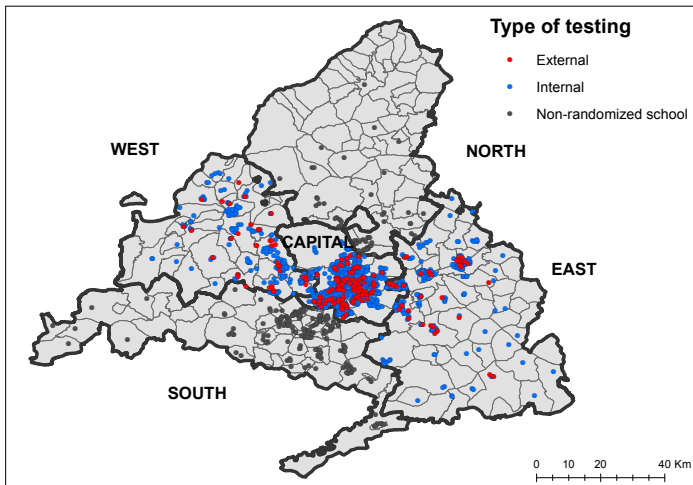
Distribution of schools in Grade 6 in 2017/18

◄ Back

Distribution of schools in Grade 10 in 2016/17

Distribution of schools in Grade 10 in 2017/18

Summary Statistics

- 1300 primary schools (1285 in 2016, 1293 in 2017, and 1298 in 2018)
- 800 secondary schools (780 in 2016, 787 in 2017, and 800 in 2018)
- Not all schools underwent the randomization in any given year
- The percentage of schools that underwent the randomization gradually increased from about 60 (52) per cent of schools to over 70 per cent of schools
- Schools that did not undergo randomization continued with the externally administered environment.
- Among the schools which underwent the randomization, between 20 per cent and 25 (40) per cent of primary (secondary) schools had an externally administered testing environment

Summary Statistics: Schools 2016/2017

A. 6th Grade:			
District	**Internal**	**External**	**Frac. External**
Capital	0	542	1.00
North	80	25	0.24
South	260	58	0.18
East	137	36	0.21
West	111	36	0.24
Total	588	697	
A. 10th Grade:			
District	**Internal**	**External**	**Frac. External**
Capital	0	368	1.00
North	29	23	0.44
South	122	58	0.34
East	49	35	0.42
West	61	35	0.36
Total	261	519	

Summary Statistics: Schools 2017/2018

A. 6th Grade:			
District	**Internal**	**External**	**Frac. External**
Capital	453	95	0.17
North	0	105	1.00
South	0	322	1.00
East	137	36	0.21
West	109	36	0.25
Total	699	594	

A. 10th Grade:			
District	**Internal**	**External**	**Frac. External**
Capital	270	96	0.26
North	0	53	1.00
South	0	180	1.00
East	50	36	0.42
West	66	36	0.35
Total	386	401	

Summary Statistics: Schools 2018/2019

A. 6th Grade:			
District	**Internal**	textbfExternal	**Frac. External**
Capital	457	90	0.16
North	80	25	0.24
South	264	60	0.19
East	0	174	1.00
West	0	148	1.00
Total	801	497	
A. 10th Grade:			
District	**Internal**	textbfExternal	**Frac. External**
Capital	279	94	0.25
North	30	23	0.43
South	122	60	0.33
East	0	88	1.00
West	0	104	1.00
Total	431	369	

Outline

Data

1. **Administrative Student Register**: Test scores and individual question scores for the population of 6th and 10th graders in academic years 2015, 2016, 2017 & 2018.

2. **Administrative School Register**: Internally or externally administered testing, socioeconomic composition of student body, private or state.

3. **Family Questionnaire**: parental educational attainment, occupation, investments at home, immigrant status (Very high non-response rate after 2015).

4. **2016 & 2018 Main Teacher Questionnaire**: Gender, experience, age

5. **2017 & 2018 Subject-specific student Questionnaire**: After each test- Self-perceived stress, self-confidence, and effort

Overview of Main Registers and Questionnaires

Register/Questionnaire Name (Period)	Unit of Obs	Data Description
Student Register (15/16–18/19)	Student/ question	Standardized scores by grade and subject
School Register (15/16–18/19)	School	School Characteristics (Internally/Externally test (dummy), school size, ownership status, type of instruction, socioeconomic composition of student body)
Family Questionnaire (15/16–18/19)	Student	Family Characteristics (Socio-economic measures, repeater dummy)
Main Teacher Questionnaire (16/17; 18/19)	Teacher	Teacher Background Variables (Age, gender, type of contract (permanent dummy, full-time dummy), work experience)
Self-perception Student Questionnaire (17/18; 18/19)	Student	Self-perceived stress, self-confidence and effort by subject)

Outline

Empirical Strategy

- The impact of external versus internal testing on gender relative student performance:

$$Y_{isdt} = \alpha + \beta Fem_i + \gamma Ext_{sdt} + \delta Fem_i * Ext_s + \theta_{dt} + \epsilon_{isdt}$$

- Y_{isd} is the standardized score in a given subject for student i in school s located in administrative district d at year t

- Fem_i takes the value 1 if the student i is female, and 0 otherwise

- Ext_{sdt} has value 1 if student attended a school that was assigned to administer the test externally, and 0 otherwise

- θ_{dt} are year-administrative district fixed effects

- Standard errors are **clustered at the school-class level**(Abadie, Athey, Imbens, Wooldridge, 2017)

Sample

- Our sample spans over three academic years (from 2016 to 2019)
- 72,381 students in 6th Grade enrolled in 1,292 primary schools
- 77,973 students in 10th Grade enrolled in 799 secondary high-schools

Balance Test in Pre-treatment Student Characteristics 6th Grade (I)

	Internal (C) (1)	External (T) (2)	Diff (T − C) (3)	> p-value (4)
Panel A. Randomization 2016/2017				
Average Score in Spanish	-0.020	-0.057	-0.037	**0.506**
Average Score in Math	0.015	0.001	-0.013	**0.770**
Average Score in English	0.199	0.134	-0.065	**0.401**
Average Score in Science	0.006	-0.032	-0.038	**0.409**
Repeater	0.081	0.074	-0.008	**0.404**
Start School After 3 years-old	0.250	0.251	0.002	**0.918**
Less than 50 books at home	0.206	0.196	-0.010	**0.572**
College Educated Parent	0.576	0.579	0.003	**0.913**
Blue Collar Parent	0.222	0.220	-0.002	**0.939**
Business Owner Parent	0.067	0.063	-0.004	**0.559**
Self-Employer Parent	0.407	0.402	-0.005	**0.782**
Non-Spanish Parents	0.119	0.113	-0.006	**0.731**
N Students	14,467	3,490		
Test of joint significance	**F-stat: 0.676 (p-value: 0.775)**			>

Balance Test in Pre-treatment Student Characteristics 6th Grade (II)

	Internal (C) (1)	External (T) (2)	Diff (T − C) (3)	> p-value (4)
Panel B. Randomization 2017/2018				
Average Score in Spanish	-0.053	-0.152	-0.099	**0.150**
Average Score in Math	0.003	-0.048	-0.050	**0.312**
Average Score in English	-0.086	-0.076	0.010	**0.909**
Average Score in Science	-0.021	-0.014	0.008	**0.888**
Repeater	0.100	0.096	-0.004	**0.617**
Start School After 3 years-old	0.315	0.335	0.020	**0.188**
Less than 50 books at home	0.251	0.238	-0.013	**0.478**
College Educated Parent	0.375	0.380	0.005	**0.873**
Blue Collar Parent	0.370	0.352	-0.018	**0.511**
Bussiness Owner Parent	0.075	0.081	0.006	**0.305**
Self-Employer Parent	0.294	0.305	0.011	**0.616**
Non-Spanish Parents	0.130	0.115	-0.015	**0.301**
N Students	11,651	3,447		
Test of joint significance	*F*-stat: 1.054 (*p*-value> 0.399)			

Balance Test in Pre-treatment Student Characteristics 6th Grade (III)

	Internal (C) (1)	External (T) (2)	Diff (T − C) (3)	> p-value (4)
Panel C. Randomization 2018/2019				
Average Score in Spanish	0.097	0.039	-0.058	**0.189**
Average Score in Math	0.057	0.016	-0.041	**0.292**
Average Score in English	0.049	0.044	-0.005	**0.939**
Average Score in Science	0.072	0.028	-0.044	**0.238**
Repeater	0.083	0.083	0.000	**0.989**
Start School After 3 years-old	0.330	0.335	0.005	**0.750**
Less than 50 books at home	0.228	0.217	-0.011	**0.505**
College Educated Parent	0.578	0.605	0.028	**0.284**
Blue Collar Parent	0.266	0.250	-0.016	**0.491**
Business Owner Parent	0.052	0.050	-0.002	**0.642**
Self-Employer Parent	0.430	0.446	0.016	**0.390**
Non-Spanish Parents	0.164	0.156	-0.007	0.628
N Students	30,246	6,681		
Test of joint significance	F-stat: 0.906 (p-value: 0.541)			
N Total Schools	1,273			

Balance Test in 2016-2018 School Characteristics. 6th Grade

	Internally (C) Administered (1)	Externally (T) Administered (2)	Difference T − C (3)	> p-value (4)
Panel A. 6th Grade				
School Characteristics				
Public or Semi-Public School	0.827	0.801	-0.026	**0.134**
Private School	0.173	0.199	0.026	**0.134**
Bilingual School	0.413	0.360	-0.053	**0.060**
Avg. Pre-Intervention School Score in Spanish	-0.057	-0.107	-0.050	**0.113**
Avg. Pre-Intervention School Score in Math	-0.037	-0.073	-0.036	**0.209**
Avg. Pre-Intervention School Score in English	-0.014	-0.036	-0.022	**0.618**
Avg. Pre-Intervention School Score in Science	-0.045	-0.081	-0.036	**0.220**
Number of Students	45.928	46.317	0.390	**0.819**
Class Size	21.796	21.841	0.045	**0.885**
Number of Clases	2.419	2.427	0.008	**0.900**
ISEC Index	0.158	0.203	0.045	**0.160**
Frac. Students Started After 3yo	0.033	0.035	0.002	**0.754**
Frac. Non-Spanish Parents	0.169	0.161	-0.008	**0.562**
Frac. Less than 50 Books at Home	0.275	0.275	-0.000	**0.977**
Frac. College Parents	0.630	0.650	0.020	**0.177**
Frac. Basic Occupation	0.056	0.049	-0.007	**0.368**
Teacher Characteristics				
Frac. Female Teachers	0.766	0.766	-0.000	**0.988**
Average Age of Teachers	44.261	44.126	-0.135	**0.849**
Average Experience of Teachers	16.245	16.256	0.011	**0.982**
N Schools-Year	1,305	303		
Test of joint significance	F**-stat: 0.885 (**p**-value: 0.597)**			

Robust standard errors are clustered at the school level and displayed in parenthesis.

Balance Test in 2016-2018 School Characteristics. 10th Grade

	Internally (C) Administered (1)	Externally (T) Administered (2)	Difference T − C (3)	> p-value (4)
Panel B. 10th Grade				
School Characteristics				
Public or Semi-Public School	0.956	0.946	-0.010	**0.603**
Private School	0.044	0.054	0.010	**0.603**
Bilingual School	0.119	0.093	-0.025	**0.192**
Number of Students	80.683	83.936	3.253	**0.164**
Class Size	23.261	22.747	-0.514	**0.178**
Number of Clases	3.146	3.288	0.143	**0.070**
ISEC Index	-0.419	-0.428	0.009	**0.781**
Frac. Students Started After 3yo	0.039	0.038	-0.000	**0.955**
Frac. Non-Spanish Parents	0.109	0.114	0.006	**0.555**
Frac. Less than 50 Books at Home	0.170	0.170	-0.000	**0.977**
Frac. College Parents	0.588	0.587	-0.001	**0.967**
Frac. Basic Occupation	0.050	0.052	0.002	**0.778**
Teacher Characteristics				
Frac. Female Teachers	0.673	0.649	-0.024	**0.031**
Average Age of Teachers	42.431	42.820	0.389	**0.345**
Average Experience of Teachers	14.487	14.296	-0.191	**0.656**
N School-Year	927	414		
Test of joint significance	F-stat: 1.024 (p-value: 0.427)			

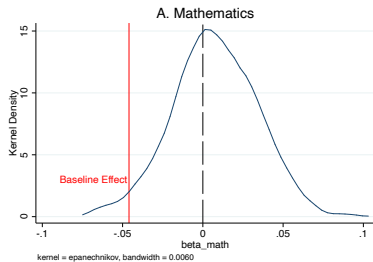Robust standard errors are clustered at the school level and displayed in parenthesis.

Outline

Baseline Estimates

Dependent Variable: Score in:

A. 6th Grade	Mathematics	Science	Spanish	English
	(1)	(2)	(3)	(4)
Female*External	**-0.046****	-0.045**	-0.025	0.001
	(0.021)	(0.020)	(0.019)	(0.021)
Female	**-0.148*****	0.050***	0.410***	0.274***
	(0.010)	(0.009)	(0.009)	(0.009)
External	-0.029	-0.028	-0.061***	-0.086***
	(0.024)	(0.021)	(0.023)	(0.027)
Constant	0.173***	0.022	-0.015	0.120**
	(0.043)	(0.035)	(0.036)	(0.047)
DAT*Year FE	Yes	Yes	Yes	Yes
Observations	72,381	72,381	72,381	72,381
B. 10th Grade	**Mathematics**	**Social**	**Spanish**	**English**
	(5)	(6)	(7)	(8)
Female*External	**-0.058*****	-0.031	-0.030	-0.026
	(0.020)	(0.019)	(0.018)	(0.019)
Female	**-0.200*****	-0.159***	0.163***	0.141***
	(0.012)	(0.012)	(0.011)	(0.012)
External	0.002	-0.056**	-0.019	-0.060**
	(0.023)	(0.023)	(0.022)	(0.024)
Constant	0.151***	0.116**	-0.018	0.159***
	(0.049)	(0.048)	(0.047)	(0.054)
DAT*Year FE	Yes	Yes	Yes	Yes
Observations	77,973	77,973	77,973	77,973

Standard errors are clustered at by-year-school-class level and displayed in parenthesis.
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Baseline Estimates

- In line with the literature, girls outperform boys in traditionally female-dominated subjects such as Spanish and English (REFS) and underperform in traditionally male-dominated subjects such as mathematics

- Gender gap in Mathematics (20%) and Spanish(16%) are similar to 2015 PISA scores for Spanish students in 10th Grade in Mathematics (16%) and reading (21%)

- The gender gap in mathematics raises by one third from 6th to 10th Grade

- Given that learning gains over one year are between one quarter and one third of a standard deviation (Woessman, 2016), the increase in the math gender gap that we find represents one and a half month of learning loss for students under externally administered testing environments

Baseline Estimates

- The interaction coefficient is particularly precisely estimated for Mathematics: 0.05 (0.06) for students in 6th (10th) Grade
- Gender gaps in mathematics are 30 per cent higher in schools with externally administered testing
- Boys performance in mathematics is the same in internally and externally administered testing environments
- Boys seem to perform worse under the external condition in subjects where their comparative advantage is lower (i.e., in Spanish and English), although the estimates are less efficiently estimated

Permutation Test in 6th Grade
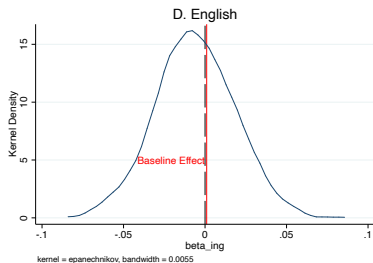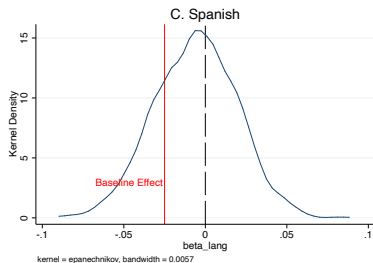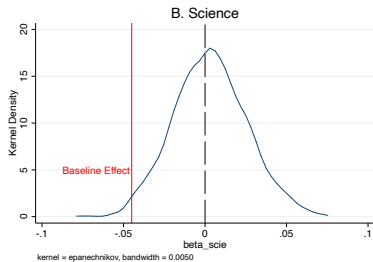
Permutation Test in 10th Grade
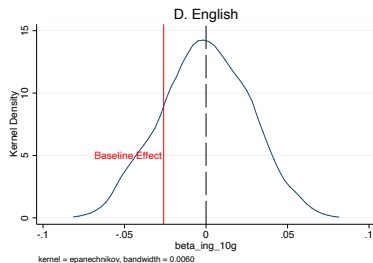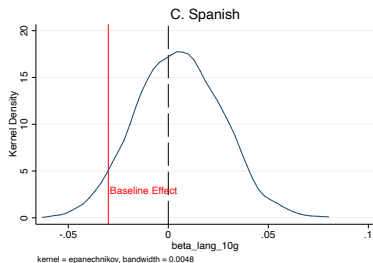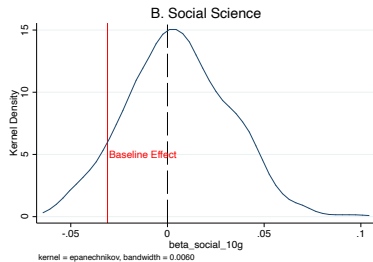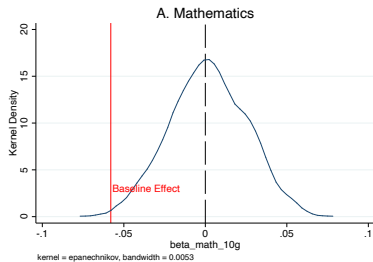
A. Mathematics

B. Social Science

C. Spanish

D. English

kernel = epanechnikov, bandwidth = 0.0053

kernel = epanechnikov, bandwidth = 0.0060

kernel = epanechnikov, bandwidth = 0.0048

kernel = epanechnikov, bandwidth = 0.0060

Outline

Ruling Out Alternative Explanations

1. Gender-biased marking ▸ Non-Multiple Choice

2. Gender-biased help during the test ▸ Heterogeneous Effects

3. Manipulation of the Duration of the Test ▸ Manipulation of the Duration

Non-Multiple Choice

Dependent Variable: Fraction of Correct Questions (Standardized):		
A. 6 Grade	**Spanish** **Non-Multiple Choice** **(1)**	**English** **Non-Multiple Choice** **(2)**
Female*External	**-0.007**	**-0.002**
	(0.021)	(0.023)
Female	0.418***	0.265***
	(0.009)	(0.009)
External	-0.291***	-0.247***
	(0.028)	(0.029)
Constant	0.077***	0.298***
	(0.030)	(0.033)
District*Year FE	Yes	Yes
N Schools	1,288	1,288
Observations	72,414	72,414
B. 6 Grade	**Spanish** **Non-Multiple Choice** **(3)**	**English** **Non-Multiple Choice** **(4)**
Female*External	**-0.016**	**-0.028**
	(0.018)	(0.019)
Female	0.205***	0.182***
	(0.011)	(0.012)
External	-0.156***	-0.088***
	(0.020)	(0.025)
Constant	0.112***	0.162***
	(0.037)	(0.041)
District*Year FE	Yes	Yes
N Schools	799	799
Observations	77,958	77,958

► Back

Heterogeneous Effects by Above vs. Below Median of School Size for 6th Grade.
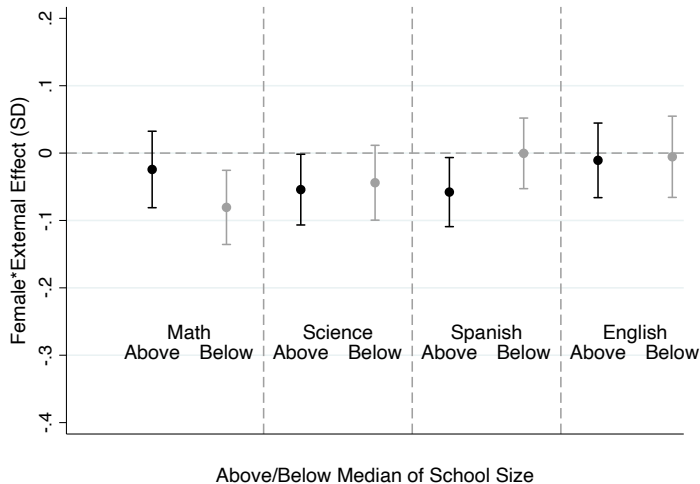
Heterogeneous Effects by Above vs. Below Median of School Size for 10th Grade.
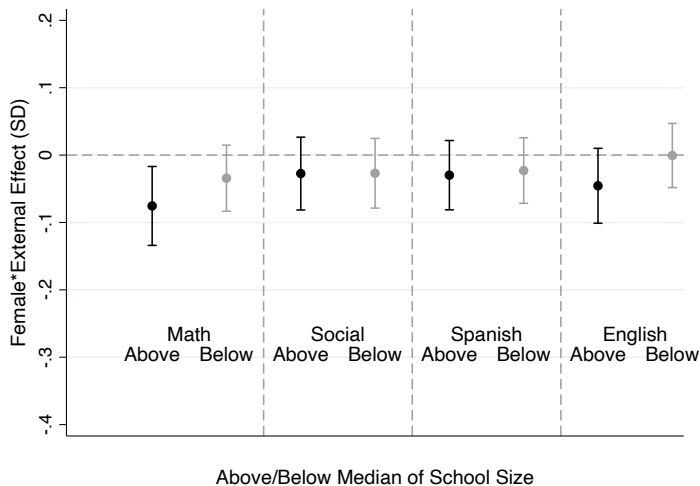
Manipulation of the Duration of the Test

Dependent Variable: Fraction of Questions in Blank (Standardized):				
6th Grade	**Mathematics** (1)	**Science** (2)	**Spanish** (3)	**English** (4)
Female*External	0.009	0.006	-0.028	-0.018
	(0.030)	(0.022)	(0.029)	(0.029)
Female	**0.107*****	-0.047***	-0.067***	-0.110***
	(0.012)	(0.011)	(0.012)	(0.012)
External	0.063**	0.015	0.050*	-0.065**
	(0.026)	(0.021)	(0.027)	(0.029)
Constant	-0.133***	-0.042***	-0.037***	-0.073***
	(0.018)	(0.017)	(0.011)	(0.018)
District FE	Yes	Yes	Yes	Yes
Observations	34,646	34,646	34,646	34,646
10th Grade	**Mathematics** (1)	**Science** (2)	**Spanish** (3)	**English** (4)
Female*External	0.081***	0.054***	-0.034**	-0.022
	(0.022)	(0.017)	(0.017)	(0.022)
Female	**0.040*****	-0.062***	-0.128***	-0.108***
	(0.014)	(0.012)	(0.011)	(0.013)
External	-0.061**	-0.096***	-0.081***	-0.021
	(0.024)	(0.022)	(0.022)	(0.028)
Constant	0.076	0.149**	0.194***	0.248***
	(0.058)	(0.071)	(0.062)	(0.046)
District FE	Yes	Yes	Yes	Yes
Observations	43,255	43,255	43,255	43,255

Ruling out Alternative Explanations

Gender-biased Marking

- No difference in girls' scores in Spanish/English essay-based questions between internally and externally administered testing environments

- Girls were not marked relatively more favourably in internally administered testing environments in essay-based components

Ruling out Alternative Explanations

Gender-Biased Help During the Test

- Main results do not vary depending on year/class size- maths score penalty experienced by girls in externally administered testing environments

- Thus, the maths score penalty experienced by girls in externally administered testing environments is not is not likely to be driven by girls receiving greater help relative to boys in internal environments during the mathematics test

Ruling out Alternative Explanations

Manipulation of the Duration of the Test

- The gender difference in the number of blank questions is not consistently different depending on the testing environment

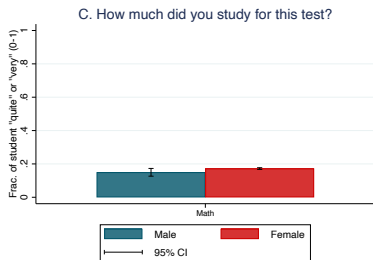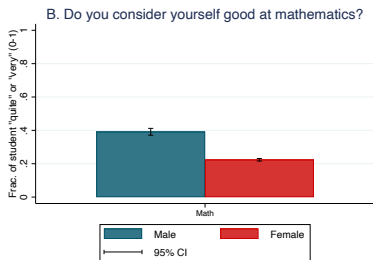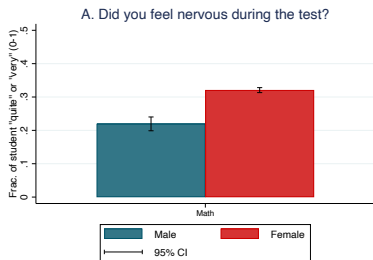- Thus, main results are not driven by girls answering fewer questions because they were more pressed with time in external environments

Outline

Student Questionnaire

- Student Questionnaire filled by all test-takers at the end of each exam of each subject in 2017 & 2018

- Students answered **4 questions**:

  1. Self-perceived Stress: Have you felt nervous during the test? (Not at all/A little/Somewhat/A considerable degree)

  2. Self-confidence: I consider myself to be good in Subject X (Not at all/A little/Somewhat/A considerable degree)

  3. Effort before the test: How much have you prepared (studied) for the test? (Not at all/A little/Somewhat/A considerable degree)

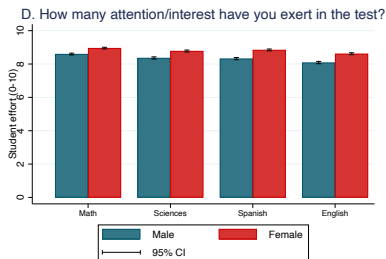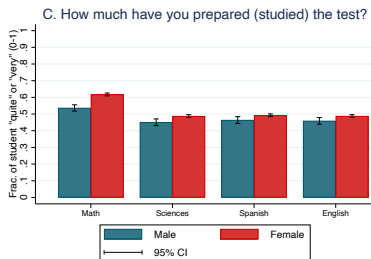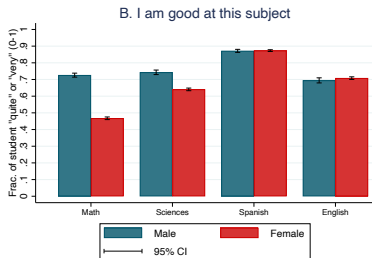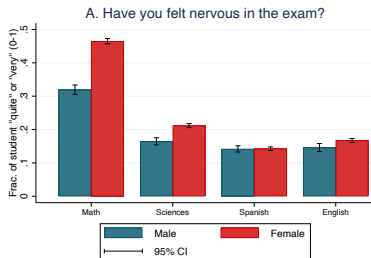 4. Effort during the test: How carefully did you consider each question during the test? [0-10]

Survey Questions by gender in 6th Grade (Mathematics)
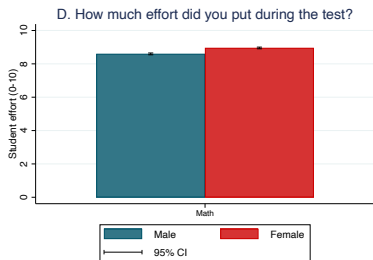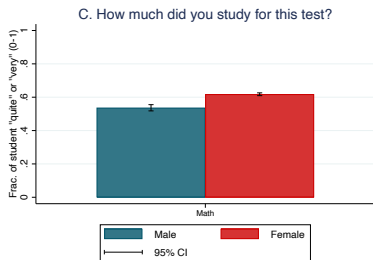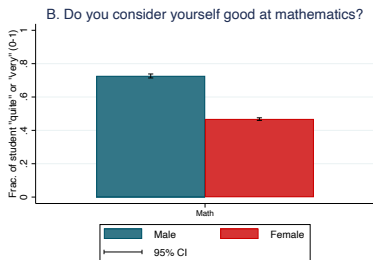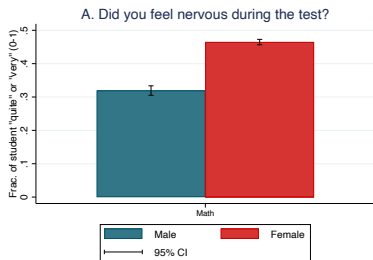
Survey Questions by gender in 10th Grade (Mathematics)
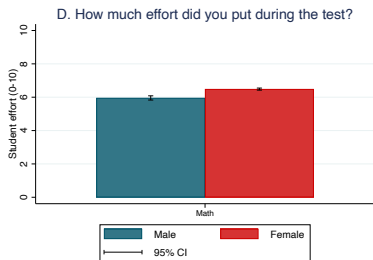
Survey Questions by gender and subject in 6th

A. Have you felt nervous in the exam?

B. I am good at this subject

C. How much have you prepared (studied) the test?

D. How many attention/interest have you exert in the test?

Survey Questions by gender and subject in 10th

A. Have you felt nervous in the exam?

B. I am good at this subject

C. How much have you prepared (studied) the test?

D. How many attention/interest have you exert in the test?
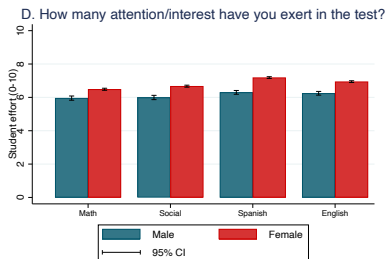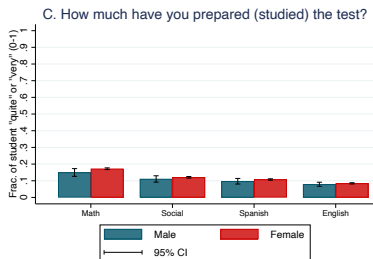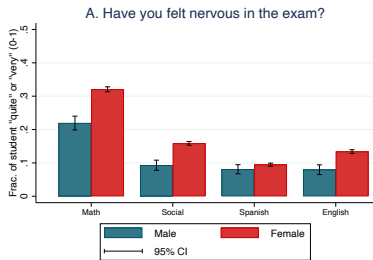
Self-reported stress, Self-confidence, and Effort before and during the Maths Test in 6th Grade

Dependent Variable: Four Survey Questions:				
A. 6th Grade	A. Student Self-Reported Stress (0-1) (1)	B. The Student considers herself good at this Subject (0-1) (2)	C. How much the student prepared the test (0-1) (3)	D. How much attention has the student exert (Standardized) (4)
Female*External	0.025**	-0.005	0.005	-0.015
	(0.012)	(0.012)	(0.013)	(0.026)
Female	0.140***	-0.258***	0.080***	0.202***
	(0.005)	(0.005)	(0.006)	(0.011)
External	-0.003	-0.003	0.016	0.029
	(0.009)	(0.009)	(0.013)	(0.022)
Constant	0.318**	-0.727	-0.533***	-0.033*
	(0.009)	(0.007)	(0.012)	(0.018)
District*Year FE	Yes	Yes	Yes	Yes
Observations	43,947	43,947	43,947	43,947

Self-reported stress, Self-confidence, and Effort before and during the Maths Test in 10th Grade

Dependent Variable: Four Survey Questions:				
B. 10th Grade	**A. Student Self-Reported Stress (0-1)**	**B. The Student considers herself good at this Subject (0-1)**	**C. How much the student prepared the test (0-1)**	**D. How much attention has the student exert (Standardized)**
	(5)	**(6)**	**(7)**	**(8)**
Female*External	-0.015	0.004	-0.002	-0.046**
	(0.010)	(0.010)	(0.006)	(0.023)
Female	**0.105***	**-0.169***	**0.023***	**0.201***
	(0.005)	(0.005)	(0.004)	(0.013)
External	-0.022***	-0.010	0.007	-0.029
	(0.007)	(0.008)	(0.005)	(0.021)
Constant	0.230***	0.395	0.153***	0.011
	(0.013)	(0.013)	(0.015)	(0.037)
District*Year FE	Yes	Yes	Yes	Yes
Observations	45,454	45,454	45,454	45,454

Summary of Results

- The coefficient on the female dummy indicates that differences between girls and boys in stress during the test, self-confidence in mathematics, study prior to the test, and effort exerted during the test are statistically significant and of the expected sign.

- 6th Grade: Doing the test in an externally administered environment seems to negatively affect girls' stress more so than boys - an 18% increment in the female-male stress differential for students in externally administered testing environments (from 14% to 16.5%).

- 10th Grade: Girls devote more effort during the test (about 20 percent of a standard deviation) in internally administered testing environments than boys. This differential is reduced to 15 percent of a standard deviation in externally administered testing environments - a reduction of the gender gap in exerted effort of 25%.

Outline

Conclusion

- Our results show that girls perform worse in mathematics in unfamiliar testing environments than boys. Not in other subjects
- This gender difference is explained by girls being more stressed during the test when younger, and by girls putting less effort during the test in adolescence.
- Results from stress: Unfamiliar, higher pressure, environments may exacerbate the stereotype threat (Steele, 1997) in male-dominated tasks
- Results from effort during the test: Unfamiliar environments, higher perceived stakes, environments may lead to boys investing more effort as the expected returns are higher
- No gender difference in self-confidence

Implications for Policy

- Standardized testing is commonplace to determine high school track (e.g., Cito exam in the Netherlands), and college access (e.g., SAT in the US, Baccalauréat in France, EVAU in Spain)

- Our results imply that standardized testing may not measure mathematical cognitive ability because of different responses to the testing environment

- Short run: Fix testing model

- Long run: Debunk (prescriptive and descriptive) stereotypes-Conscious vs unconscious (Sevilla, 2020)

Thank you very much!