Tackling the gender gap in mathematics with active learning methodologies

Daniela Piazzalunga[a]

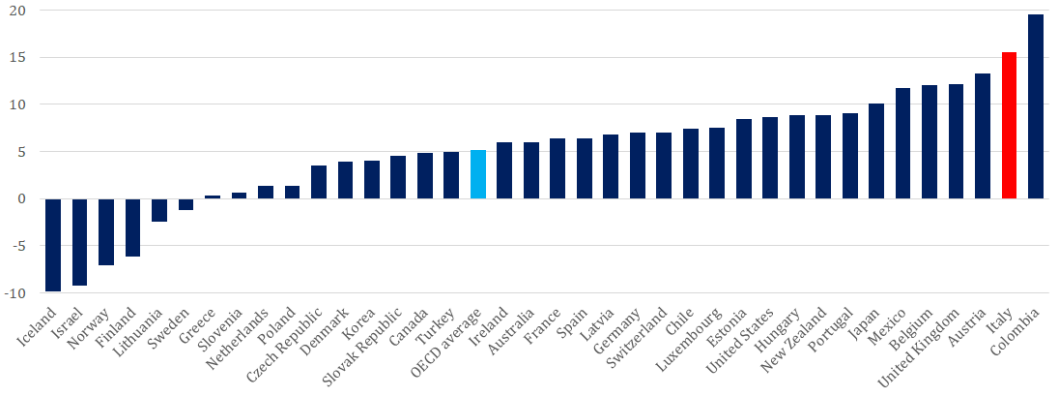
with M.L. Di Tommaso[b], D. Contini[b], D. De Rosa[c]
F. Ferrara[b], O. Robutti[b]

[a] University of Trento
[b] University of Torino
[c] Italian Ministry of Economy and Finance

September 24, 2021
Joint Workshop on Gender and Mathematics
INVALSI, Collegio Carlo Alberto, and University of Torino

Gender gap in mathematics across countries

Italy one of the countries with the highest gap



PISA 2018 (grade 9)

Boys' - girls' results (Positive value indicate boys advantage)

The role of teaching practices

- Existing research focuses on the role of teachers' and parents' stereotypes, beliefs, and expectations
- Surprisingly, to the best of our knowledge, no empirical study addressing the effect of math teaching practices on the gender gap with rigorous evaluation studies
- Group-work and mathematical discussion, investigative work and cognitive activation strategies seem to improve girls performances

  Boaler 2002a, 2002b; Boaler 2009, Zohar & Sela 2003
- Correlation studies, prone to endogeneity issues

Italian context

- The gap is particularly large in Italy
- Causes difficult to establish, out of the scope of our work
- Possible reasons:
  - more gender stereotypes/less gender equal society?
  - more traditional teaching (teacher-centred instruction)? Anecdotal evidence and TALIS results

Research question and our project

Could properly designed teaching practices
help reduce the gender gap in mathematics?

1. Implement a teaching practice (math lab) at an early stage of schooling, in Italy, potentially capable of reducing gender differences
2. Conduct a Randomized Controlled Trial to evaluate the impact of the lab

Trail registered in the AER RCT Registry (AEARCTR-0003651)

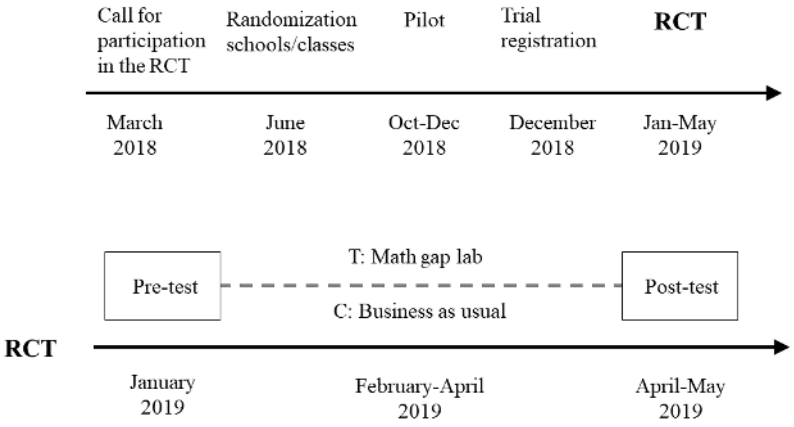
The RCT: treatment delivery

- Third grade pupils (8 years old)
- Treatment at the class level
- 5 laboratory meetings of 3 hours each, 5 consecutive weeks, **during school time**
  School year: 33 weeks; Math: about 6 hours pw
- All students take part to the activities (including student with disabilities or special needs)
- Children in the control group follow the usual curriculum
- Treatment delivered by 4 tutors, trained in math education (Master or Ph.D.)
- Teachers present with the role of observers

The RCT: invitation and criteria

- Primary schools of Torino province (180 primary schools)
- Schools participate **voluntarily** with at least two classes: one randomized to treatment and one to the control group
- The two classes should
  - different math teachers, to avoid spillover effects
  - not involved in other math laboratories in the same school year
- 50 classes, 25 schools, approx. 1000 pupils
- Random selection of the 25 participating schools and of the 2 participating classes per school

Timeline


Call for participation in the RCT
Randomization schools/classes
Pilot
Trial registration
RCT
March 2018
June 2018
Oct-Dec 2018
December 2018
Jan-May 2019
T: Math gap lab
Pre-test
Post-test
C: Business as usual
RCT
January 2019
February-April 2019
April-May 2019

The math lab I

Theoretical framework

- **Social constructivism** (post-Vygotskji):
  theories of active learning that emphasize the need for students to construct their own understanding
- ***Laboratorio di matematica*** (Anichini et al. 2004):
  Italian didactics developed at the beginning of 2000s
  (well known in the international math education community)

The math lab II

Fundamental elements

- Doing instead of 'listening'
- Problem solving
- Small-group and peer work
- Sharing and comparison of ideas, arguing
- Mistakes as opportunities for learning
- Use of tools and materials
- Role of the teacher: orchestrate class activities

Characteristics that could help girls.
Focus on *Numeracy*

The story of the forest trolls

C'era una volta una casa nel bosco dove abitava una famiglia di folletti dei boschi. La famiglia era composta da mamma folletta, papà folletto e dai due loro figli.

Era autunno e bisognava iniziare a fare le provviste per la brutta stagione.

La prima a uscire fu folletta figlia, uscì con il suo cesto, prese il sentiero, fece venti passi in direzione delle montagne e arrivò a un albero di mele. Riempì il cesto di mele e tornò a casa.

Poi uscì di casa il figlio, anche lui con il suo cesto, prese il sentiero in direzione delle montagne, fece venti passi e arrivò a un albero di castagne. Raccolse castagne fino a riempire il cesto e ritornò.

Più tardi uscì la mamma e portò con sé un secchio vuoto, prese il sentiero in direzione del lago, fece venti passi e arrivò alla fontana. Riempì d'acqua il suo secchio e tornò indietro.

Per ultimo uscì papà folletto, camminò sul sentiero in direzione del lago, fece venti passi e arrivò al mercato. Qui comprò dei pesci, poi ritornò a casa.

The story of the town to be enlarged

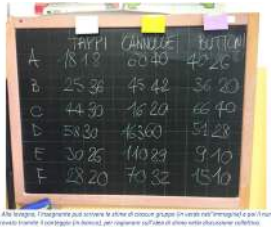


Figura 2. Alla lavagna, l'insegnante può scrivere le stime di ciascun gruppo. Un verde nell'immagine) e poi il numero di oggetti trovati (rosso). Il confronto (in basso), per ragionare sull'idea di stima e sulla discussione collettiva.

Riciclò, sindaco di Contamille, vuole ingrandire la sua città. Per fare questo deve costruire un plastico con il progetto della zona nuova di Contamille. Il plastico sarà molto grande e sarà fatto di tappi, cannucce e bottoni. Riciclò ha bisogno di molti aiutanti per realizzarlo.

"Da solo non posso farcela. Ragazzi: ho bisogno del vostro aiuto! Raccogliete tappi di plastica, cannucce e bottoni. Cercate questi oggetti attorno a voi per i prossimi 3 minuti

Data

- Children's math outcomes (Pre-test and Post-test scores)
  - Designed by math scholars participating to the project
  - Similar to national standardized assessments (INVALSI)
  - 20 items
- Children's attitudes towards math (after the post-test)
- Additional background data provided by school boards (parental education and migratory background)
- Aggregate data in grade 2 at the class level on family background and test scores from Invalsi (external validity?)

Model

$$Y_{iks} = \beta_0 + \beta_1 T_{ks} + \beta_2 X_{iks} + \beta_3 Y_{0iks} + \gamma_s + \epsilon_{iks}$$

Y and Y_0 standardized (Pre and post-test designed by the team)

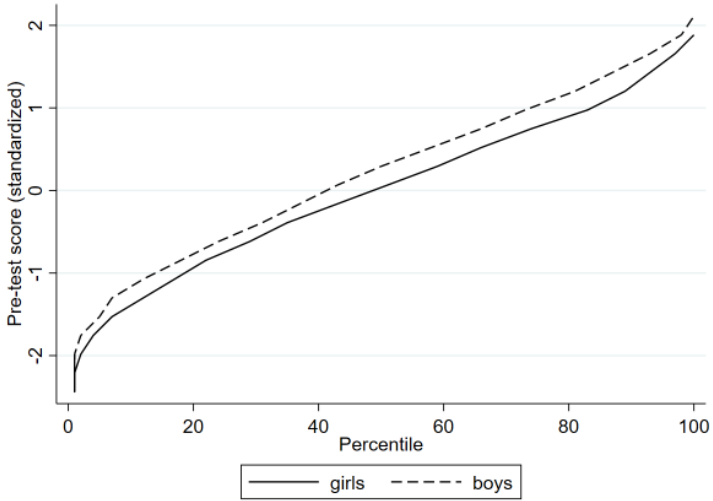
Control variables:

- math pre-test score Y_0
- individual and school characteristics X
  gender, age, migratory background, parental education, fulltime, class size
- School fixed effects γ_s

SE clustered at the class level

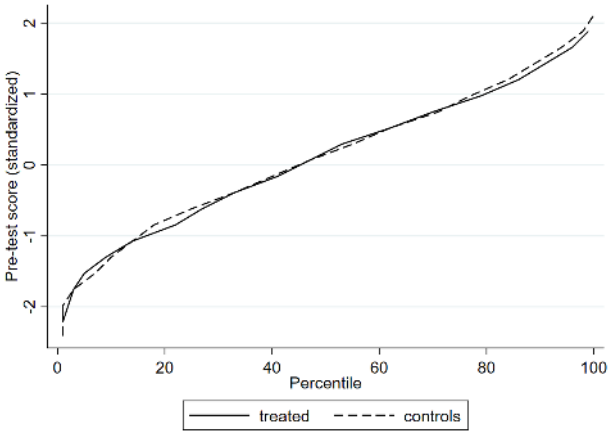
Balance, attrition, and compliance

Math gender gap in the pre-test



Pre-test scores

Balance:
confirmed on most characteristics and on the pre-test



Overall	Mean	Sd
Controls	10.78	4.45
Treated	10.70	4.33
Girls		
Controls	10.39	4.29
Treated	10.15	4.32
Boys		
Controls	11.18	4.60
Treated	11.27	4.27

Balance after attrition - individual level

	Controls	Treated	Diff
Pre-test score	10.77	10.86	
Girl	0.50	0.51	
Special needs	0.14	0.15	
Parents low educ.	0.67	0.74	**
Parents high educ.	0.33	0.26	**
Native child	0.88	0.85	*
By gender			
Pre-test score (F)	10.36	10.23	
Pre-test score (M)	11.19	11.50	

$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Attrition

Balance - class level

	Control Classes	Treated Classes	Diff
Size of class	21.0	20.8	
Pre-test score (mean)	10.8	10.7	
Pre-test score (s.d.)	4.3	4.2	
Permanent contract teachers %	100.0	92.0	
Teaching experience (years)	21.4	22.6	
Teaching math in class (years)	2.8	2.4	*

$^{***}p < 0.01,^{**}\, p < 0.05,^{*}\, p < 0.1$

Compliance and attendance

- Perfect compliance
- Very large attendance rate

Share of labs attended	Children	Boys	Girls
0%	0.0%	0.0%	0.8%
$\geq 50\%$	99.3%	100.0%	98.6%
$\geq 70\%$	95.8%	97.2%	94.5%
$\geq 80\%$	94.2%	95.8%	92.7%
100%	73.8%	75.9%	71.7%

Results

Main results: the intervention improves girls' achievements

	Post-test scores			Post-test scores with additional controls		
Variable	Overall	Girls	Boys	Overall	Girls	Boys
Treatment	0.076*	0.152***	-0.028	0.083**	0.142**	-0.009
	(0.030)	(0.053)	(0.045)	(0.033)	(0.055)	(0.046)
Pre-test	0.763***	0.744***	0.784***	0.739***	0.737***	0.748***
	(0.023)	(0.037)	(0.024)	(0.025)	(0.035)	0.033)
Constant	0.007	-0.132**	0.048	0.163	-0.194	0.290
	(0.040)	(0.058)	(0.045)	(0.157)	(0.225)	(0.249)
R-sq.	0.592	0.572	0.601	0.616	0.603	0.641
Obs.	888	448	440	888	448	440
Pre-test	X	X	X	X	X	X
School FE	X	X	X	X	X	X
Add. contr.				X	X	X

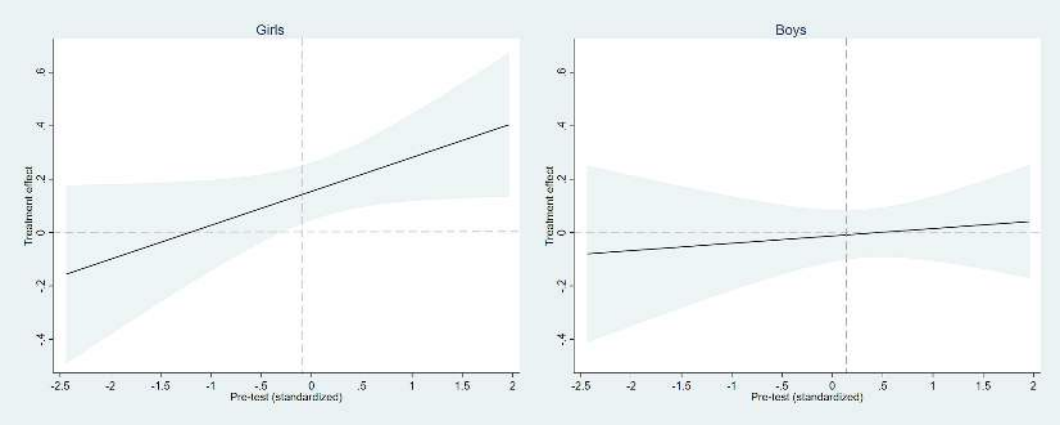
$^{***}p < 0.01$,** $p < 0.05$,* $p < 0.1$.
Standardized test scores. S.E. clustered at the class level.

Heterogeneous effects by initial math level I

	Girls	Boys
Treatment	0.155***	-0.013
	(0.053)	(0.048)
Pre-test scores	0.679***	0.735***
	(0.049)	(0.041)
Treatment*pre-test	0.127*	0.028
	(0.064)	(0.058)
Constant	-0.159	-0.292
	(0.224)	(0.251)
Observations	448	440
R-squared	0.611	0.656
School FE	X	X
Add. Contr.	X	X

$^{***}p < 0.01,^{**}\, p < 0.05,^{*}\, p < 0.1$

Heterogeneous effects by initial math level II



Additional heterogeneity

Heterogeneity by parental education and migratory background

- Migratory background
  - Significant effect among native girls
  - Larger effect for migrants girls
- Parental education
  - Larger effect for girls with low educated parents

Table

Exploring possible mechanisms

1. Does the treatment improve math competences
   only for some types of questions
   (multiple choice vs. open answer, different dimensions)?
   - No, overall improvement (in Numbers)

2. Does the treatment improve children's attitudes towards math?
   - No
   - Positive impact on skills not mediated by attitudes

   ▸ Questions ▸ Results

3. Does the treatment reduce item non-response?
   - Yes, but slightly and with same magnitude for boys and girls ▸
   - → Positive impact on the probability of answering
   - but the main impact is the direct impact on math competences

Limited external validity

	Classes		
Variable	Experimental	Piedmont	Italy
Invalsi score in Italian	0.393	0.067	0.000
Invalsi score in Math	0.559	0.023	0.000
Invalsi score Italian Female	0.389	0.113	0.017
Invalsi score Italian Male	0.407	0.021	-0.044
Invalsi score Math Female	0.439	-0.052	-0.070
Invalsi score Math Male	0.681	0.086	0.029
Gender Gap Math	-0.241	-0.139	-0.099
Kindergarden attendance	42.00	32.72	38.09
Mother tertiary education	31.61	22.28	24.23
Father tertiary education	22.01	16.20	16.39

All differences are significant $p < 0.01$

Conclusions

- The program improves girls' math skills (+**0.14 s.d.**)
- The effect is large and policy-relevant
  - One full year of primary school attendance: +0.89 s.d.
  - Similar interventions, lasting 12 weeks: +0.25-0.33 s.d.

  Bloom et al. 2008; Slavin and Lake 2008; Pellegrini et al. 2018
- No benefit no harm for boys
- Math gender gap (0.21 s.d. before) reduced by **40.1 - 47.5%**
- Girls with high pre-test scores benefit the most
- Properly designed teaching practices have the potential to reduce the gender gap in math in primary school

Thank you

daniela.piazzalunga@unitn.it

Project website

Paper download

Appendix

Attrition

		Both	Girls	Boys
Post-test	Overall	0.054	0.052	0.056
	Control	0.055	0.049	0.061
	Treated	0.054	0.056	0.051
	Difference (T-C)	0.001	0.006	-0.009
Pre- and Post-test	Overall	0.149	0.153	0.138
	Control	0.124	0.125	0.123
	Treated	0.167	0.179	0.155
	Difference (T-C)	0.043**	0.053*	0.037

$^{**}p < 0.05;^{*}\ p < 0.10$

Back

Additional heterogeneity

	Parents' education			Migratory backgr.	
	Girls	Boys		Girls	Boys
Treatment	0.182**	-0.075	Treatment	0.104*	0.032
	(0.072)	(0.068)		(0.062)	(0.062)
T*high educ. par.	-0.099	0.119	T*migrant	0.295*	-0.317
	(0.133)	(0.148)		(0.154)	(0.204)
Obs.	448	440	Obs.	448	440
R-sq.	0.604	0.643	R-sq.	0.605	0.643
Pre-test scores	X	X		X	X
School FE	X	X		X	X
Add. controls	X	X		X	X

$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Back

Children's attitudes

- 5 questions to record attitudes towards math
  (after the post-test)
  1. Do you like math?
  2. Are you good at math?
  3. Are you afraid of making mistakes when you do math?
  4. Are you relaxed when you do math?
  5. Are you afraid of not finishing in time when you do math exercises in class?
- Likert-scale answers

Back

	Attitudes (1)	Attitudes (2)
Girls	-0.750*	-0.831**
	(0.388)	(0.375)
Treatment effect on boys	-0.474	-0.477
	(0.301)	(0.298)
Treatment effect on girls	-0.495	-0.486
	(0.358)	(0.350)
Observations	882	882
R-squared	0.053	0.072
School FE	YES	YES
Additional controls	NO	YES

Back

Treatment effect on item-non response

Number of blank items			
	Overall	Boys	Girls
Treatment	-0.146**	-0.142*	-0.137*
	(0.061)	(0.077)	(0.072)
Gender	0.008		
	(0.054)		
N. of blank items at pre-test	0.138***	0.146**	0.115***
	(0.041)	(0.057)	(0.039)
Observations	888	448	440
R-squared	0.159	0.191	0.212
Pre-test score	YES	YES	YES
School FE	YES	YES	YES
Additional Controls	YES	YES	YES

Back

Issues and Limitations

Limitations

- Short intervention and **short term** effects.
  Longer term effects?
- Limited **external validity**

- What would happen with **scaling up** of the intervention?
  - Teachers instead of experts in math education
  - In other areas of the country the GGM is smaller
  - In other contexts where girls (and boys) are less performing
    $\rightarrow$ smaller effect?

  But scaling up implies longer intervention
  (possibly 'business as usual')

Additional issues: Internal validity

1. Assessments made by developers of the program

   "They could be unfair to control groups because they are aligned with the content taught in the treatment but not in the control group"

   Pellegrini et al. 2018

'Ex-ante' line of defence:

- Test: same conceptual framework of Invalsi tests
- Developed with collaboration of teachers not involved in trial
- Focus on numeracy (standard curriculum in grade 3)
- Teachers typically work on all domains during the year.
  Treatment classes not overexposed to numeracy
- Qualitative questionnaires to teachers confirm

Additional issues: Internal validity

1. Assessments made by developers of the program

   "They could be unfair to control groups because they are aligned with the content taught in the treatment but not in the control group"

   Slavin et al 2018

Ex-post:

- No effect on boys
- If there was an over-exposure to the content of the test, it would be on both M and F

Additional issues - black box?

2. Treatment:
   Active learning methodologies & 'gender gap awareness' (tutors)

- Pro (policy/involved children)
  larger effect of teaching practices + awareness
- Cons (research)
  Difficult (impossible?) to unpack the effects
- Cons (policy)
  Cost-effectiveness?
  Only information package could be cheaper and easier to implement