ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

MESE 1

Measurement in STEM Education (MESE1)
Naples, 30-31 January, 1 February

DISP Dipartimento di Scienze Politiche
Università degli studi di Napoli
Federico II

Università degli Studi di Napoli Federico II - Dipartimento di
FISICA "ETTORE PANCINI"

Assessing The Gender Gaps In Maths Competence: An Overview Of What We Known From Invalsi Data

Stefania Mignani

Dipartimento di Scienze Statistiche

Università di Bologna

Outline

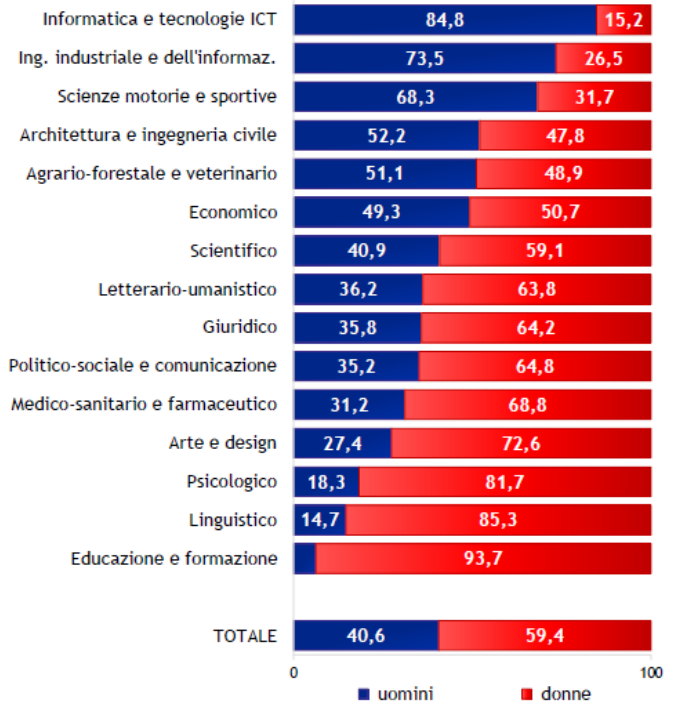
- Brief introduction on gender differences in STEM Degree enrolment
- Why dealing with mathematics competence
- The use of standardized test to assess the acquired competences
- An overview of main results for gender gaps from Invalsi Data
- Conclusions and suggestions for future studies

Introduction

- The issue of the differences in school results of girls and boys is very important from different points of view, educational, social, economic…
- As well-known from Literature, girls perform better in understanding the language while boys do better in Stem disciplines, in particular Mathematics
- There are many possible explanations of these differences.
- Any case they must be faced referring to knowledge fundamentals for the rights and duties of active citizenship.

The enrolment to STEM degree

Figura 4.1 Laureati dell'anno 2021: genere per gruppo disciplinare (valori percentuali)



Fonte: AlmaLaurea, Indagine sul Profilo dei Laureati.

DISPARITY in the choices made at the university level, which in turn has an impact on the choices made at the professional level.

- The percentage of students with a technical-scientific degree programmes (STEM degrees) in Italy amounts to **only 25%.**
- a.a. 2021/22 The females enrolled in STEM are 21,2%, while males 41,6% (MUR-USTAT, 2021)
- For females, the choice to undertake a STEM study path, is influenced by a plurality of factors including the different approach to competition, to non STEM skills and to job,…

Why considering mathematics competences

- Gender disparities in mathematics achievement have a remarkable impact on the enrolment into scientific degree courses and, consequently, on the job market
- Mathematical literacy is still a major challenge in basic education.
- The lack of adequate mathematical skills keeps the doors to possible careers closed

Why using results of Standardized tests

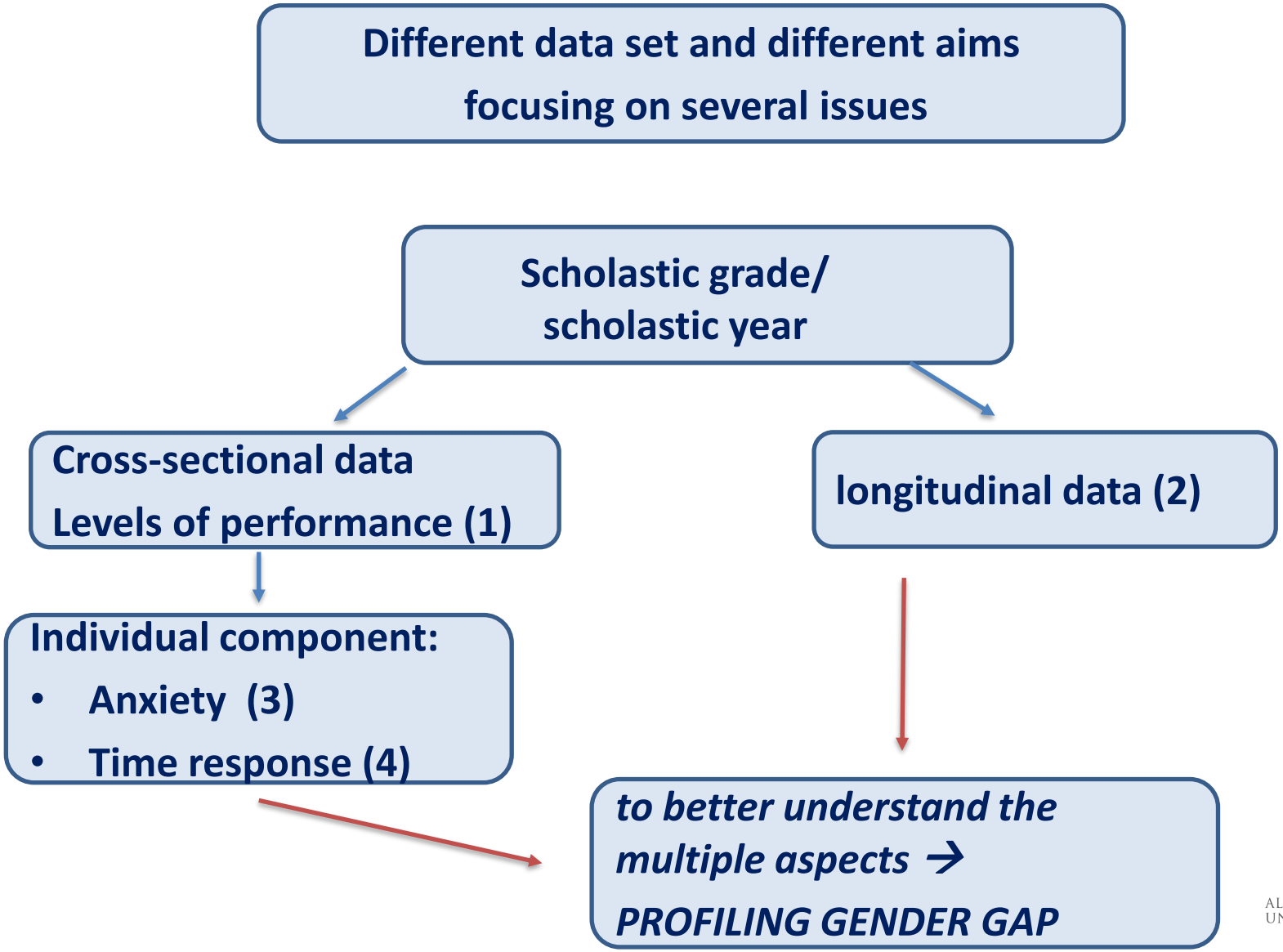
→ They allow to control the effect of the teacher or school, although there is a more hidden component that is less visible in the standardized evaluation

→ The problem of gender differences is much more complex and profound than what is commonly thought.

→ There is not only an average difference of acclaimed skill levels in all national and international research, but also a problem of visibility of the problem in school evaluative practice.

→ The phenomenon is complex: not only cognitive elements but also cultural, social, behavioural ,...

The framework of the illustrated studies



The main questions addressed in the studies

The research questions:

- ✓ *Cross-sectional data:*
  1. are there significant differences between males and females in their performance at the different performance levels?
  2. Is the gap Influenced by attitude toward the test (i.e. Anxiety and time response)?

- ✓ *Longitudinal data*:
  1. can these differences already be detected at the primary education stage?
  2. How do these differences evolve as students progress through school grades?

Cross-section data: first study

Aim of the analysis

Is gender gap larger at the upper and lower tail fo performance distribution?

- HIGH PERFORMANCES→ to understand why girls who, even though they obtained good results, decide not to enrol in a STEM degree

- LOW PERFOMANCES→ to promote learning activities to help girls who have shown problems with mathematics during their school life.

Data, measures and method

DATA

- ✓ Standardised mathematics tests Scholastic year 2017-2018
- ✓ Grade 5, 8 and 10
- ✓ 48-49% girls and 51-52% boys, depending on grade

MEASURE

Rasch scores

METHODOLOGICAL TOOL

A quantile regression analysis to explore the difference along the entire distribution of Rasch scores using gender as covariate.

(Matteucci, Mignani, 2021)

Main results

Quantile	Coefficient Grade 5	Coefficient Grade 8	Coefficient Grade 10
0.05	0.000	-0.168	0.000
0.10	-0.140 ***	-0.148	-0.001
0.25	-0.252 **	-0.252**	-0.001
0.50	-0.240	-0.240	-0.114 **
0.75	-0.252	-0.252	-0.356 ***
0.95	-0.215	-0.228	-0.208 **

- The slope is always negative: being female determines on average a reduction in the performance level
- For grade 5 and 8 the results are similar: the gender gap is just evident for low levels of performance and constant along the distribution
- The most notable impact is seen at the 25th and the 75th percentiles.
- For grade 10 the differences start to become significant at the medium performance levels and are particularly marked in the subsequent percentile (75th).
- The difference reduces at the higher-performing levels: among students with low ability, there is no significant difference between boys and girls.

Longitudinal data

- Studies are often used to answer questions about educational progress and obstacles to such progress.

- Monitoring progress in the mathematical knowledge of the same generation of students provides insights to support different stakeholders in detecting and reducing knowledge gaps.

- Learning processes are cumulative, and a single-year comparisons are inadequate to evaluate the school's ability to affect student academic progress.

- Measuring the achievements of the same generation of pupils in different years of schooling is a long-term process.

Longitudinal data: second study

Aims of the analysis

- discover, during the student school life, the different behavior of specific categories of students;
- identify when early differences between students appear and how they change over the schooling years in order to promote specific learning;
- what are the main determinants of the scoring, taking into account both individual and schools characteristics.

Data, measure and methods

- To facilitate the observation of students over time, the INVALSI assigns to each student a unique identification code: the student's *SIDI Code*.
- The selected cohort consists of **34545 students**, grouped in **2808 schools**.

INVALSI: grade 8
School year 2013/2014

INVALSI: grade 10
School year 2015/2016

INVALSI: grade 13
School year 2018/2019

- The response variable is the **Rasch score**
- Hierarchical approach with three nested levels

→Multilevel growth model

→Percentile approach

The information available are:

- **Characteristics of the students** (*Gender*, *Immigration status*, *ESCS*, *Oral mark in Mathematics in the 1st quarter*);
- **Characteristics of the schools** (*Type of school attended*, *geographical Macro-area in which the school is located*, *ESCS*, *Size of school*).

Bianconcini, Mignani, Mingozzi (2023)

Multilevel growth models for nested longitudinal data

- three-level growth t (time), j (student) i (school)
- the student performance is assumed to follow a linear trajectory:

where $$y_{ijt} = \pi_{0ij} + \pi_{1ij} \cdot (t - 8) + \pi_2 \cdot Z_{ijt} + \varepsilon_{ijt}$$

π_{0ij} and π_{1ij} intercept and slope of the growth trajectory of the j-th student attending the i-th school and Z_{ijt} student covariate

$$\pi_{0ij} = \beta_{00i} + \boldsymbol{\beta}_{01}\mathbf{w}_j + r_{0ij}$$
$$\pi_{1ij} = \beta_{10i} + \boldsymbol{\beta}_{11}\mathbf{w}_j + r_{1ij}.$$

At school level

β_{00i} and β_{10i} intercept and slope of the specific linear trajectory of the school i

$$\beta_{00i} = \gamma_{000} + \boldsymbol{\gamma}_{001}\mathbf{w}_i + u_{00i}$$
$$\beta_{10i} = \gamma_{100} + \boldsymbol{\gamma}_{101}\mathbf{w}_i + u_{10i}$$

γ_{000} and γ_{100} the expected intercept and rate of growth of the trajectory defined for the overall population, common to all the students and schools.

Main results

	Parameter	Estimate	St. err.	t
Intercept	γ_{000}	193.80	1.280	151.340***
Time	γ_{100}	1.428	0.237	6.022***
Gender	β_{01}	9.068	0.396	22.896***
1G	β_{02}	−6.349	1.082	−5.867***
2G	β_{03}	−5.172	0.980	−5.281***
ESCS student	β_{04}	1.158	0.208	7.592***
ESCS school	γ_{001}	8.251	0.983	8.391***
North West	γ_{002}	3.958	0.925	4.280***
North East	γ_{003}	3.808	0.948	4.019***
South	γ_{004}	0.152	0.951	0.160
South and Islands	γ_{005}	0.817	1.049	0.779
OtLy	γ_{006}	−20.01	0.663	−30.191***
TeIn	γ_{007}	−21.93	0.787	−27.882***
VoIn	γ_{008}	−35.79	1.027	−34.833***
Time: Gender	β_{11}	−0.032	0.071	−0.455
Time: 1G	β_{12}	0.493	0.193	2.558**
Time: 2G	β_{13}	0.299	0.175	1.713*
Time: ESCS student	β_{14}	−0.042	0.037	−1.126
Time: ESCS school	γ_{101}	−0.001	0.194	−0.005
Time: North West	γ_{102}	1.588	0.193	8.250***
Time: North East	γ_{103}	2.081	0.199	10.422***
Time: South	γ_{104}	−1.124	0.198	−5.671***
Time: South and Islands	γ_{105}	−2.146	0.219	−9.802***
Time: OtLy	γ_{106}	−3.749	0.123	−30.496***
Time: TeIn.	γ_{107}	−1.946	0.148	−13.181***
Time:VoIn.	γ_{108}	−4.001	0.192	−20.792***

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

- Male students have significantly better performance in the maths test in the eighth grade than females.
- Their rates of growth are not significantly different implying a constant gap in the performance of male and female students, defined by the discrepancy in the average maths scores observed in the eighth grade.

Student growth percentile (SPG)

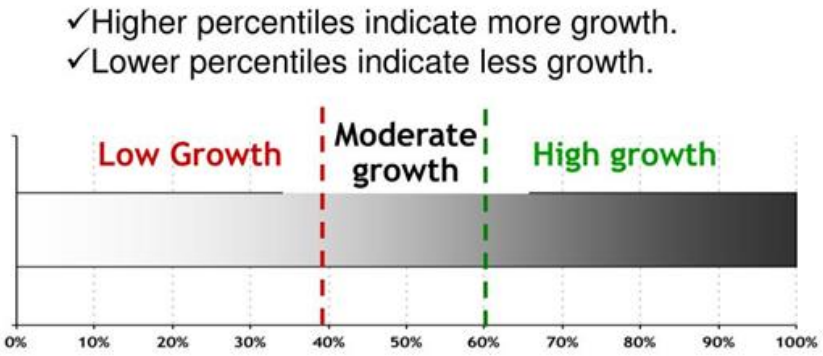
- The individual student performance over time is compared with that of other students with similar score histories, known as academic peers.

- The rate of change is expressed in percentile, a commonly understood manner of comparing things to one another.

- This approach helps in understanding the heterogeneity in the performances of different groups and supports the identification of effective practices that could help students attain higher academic performances

Student growth percentile

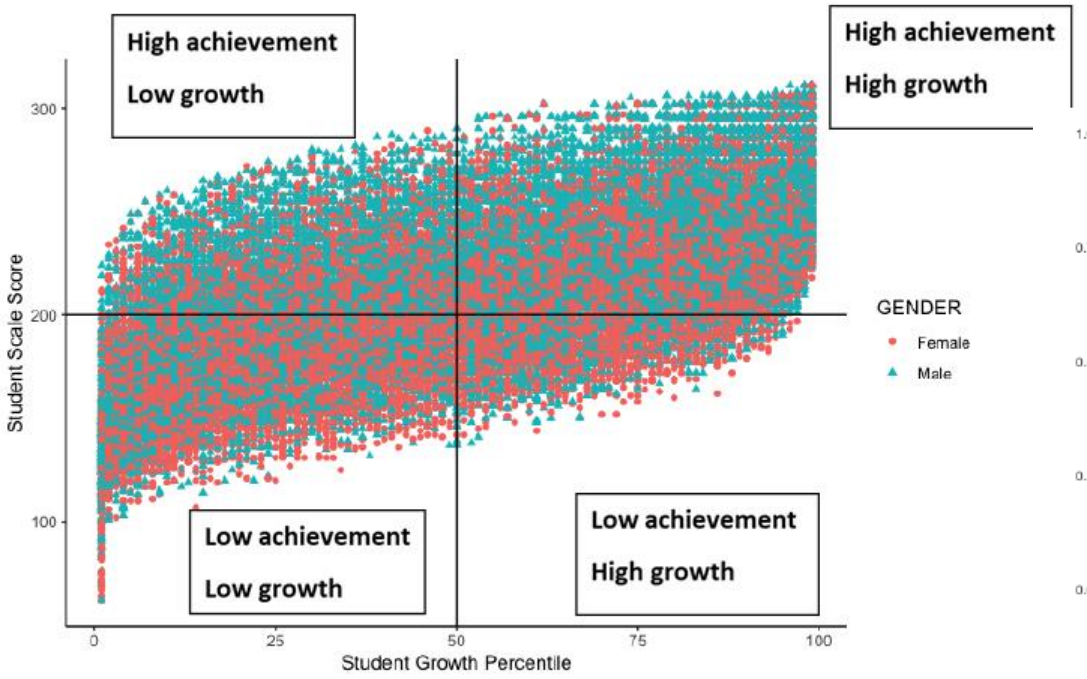
- Percentiles: percentage of cases that fall below a certain score.
- SGPs communicate the degree to which a student has learned in a particular domain, compared to a group of academic peers who had a comparable score on the previous test (or multiple previous tests) in that subject.
- SGP is a number between 1 and 99. Receiving an SGP score of 70 indicates that you demonstrated more growth than 70 percent of your academic peers.
- A student with a low raw score can show high growth, and a student with a high raw score can demonstrate low growth. Similarly, two students with very different test scores can have the same SGP.

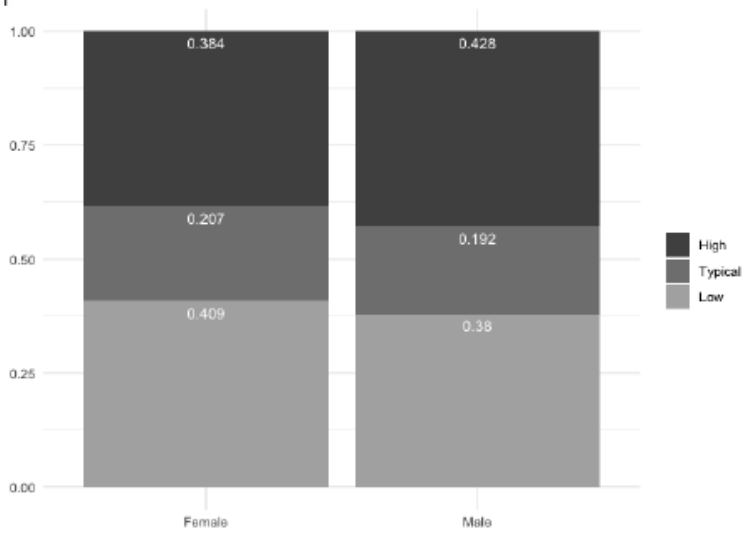


Student growth percentile

- SGP is based upon the estimation of the conditional density associated with a student's score at time t using the student's prior scores at previous occasions, as the conditioning variables.
- SGP is then defined as the percentile of the score in this conditional density.
- The percentile reflects the likelihood of such an outcome given the student's prior achievements.
- Quantile regressions are used to estimate features of the conditional distribution of student achievement

Main results





- Male students mainly obtain scores above the national average in grade 13 and generally higher than those obtained by female students.
- In terms of growth, both groups are almost equally distributed on the left and right of the vertical line.
- This distribution by the three clusters are very similar between males and females for typical and low growth.
- A slightly different in proportion for high growth .

Cross sectional data: third study

Aim of the analysis

Investigate the relationship between the emotional component and the student performance, taking into account the interactions with other background variables, such as socio-demographic (gender, geographical area, immigrant status), economic, and those more related to the educational path.

Data, measure and method

DATA

- ✓ INVALSI 2018 test data for grades 5, 8 and 10
- ✓ Mathematics item responses + background variables + student questionnaire items
  - Grade 5: P&P test with 41 items
  - Grade 8: CBT test, totally 206 items
  - Grade 10: CBT test, totally 143 items
- ✓ Info from student questionnaire: anxiety items
- ✓ National sample data, no missing data: n_5=23817, n_8=25377, n_{10}=36342

MEASURES

- ✓ Rasch model for achievement score
- ✓ Graded model for anxiety score

METHODS

- ✓ Linear regression models

Matteucci, Mignani, Spaccapanico Proietti (2023)

Main results

Rasch model item difficulty parameters							
Grade	**Min**	**25th Qu.**	**Median**	**Mean**	**75th Qu.**	**Max**	**Sd**
5	-2.627	-0.679	-0.091	-0.148	0.578	1.846	0.996
8	-2.916	-0.726	0.082	0.047	0.776	3.372	1.140
10	-3.564	-1.543	-1.013	-1.040	-0.477	1.765	0.942

Math ability scores							
Grade	**Min**	**25th Qu.**	**Median**	**Mean**	**75th Qu.**	**Max**	**Sd**
5	-3.306	-0.885	-0.131	-0.101	0.652	3.339	1.094
8	-3.416	-1.004	-0.251	-0.194	0.572	3.416	1.112
10	-4.303	-1.873	-1.093	-0.994	-0.187	2.418	1.153

- Anxiety has a negative significant effect for all grades
- A negative and significant effect for gender in favor of male, by controlling for the other variables.

	Grade					
	5		**8**		**10**	
Intercept	-0.001		0.000		-0.001	
	(0.019)		(0.018)		(0.015)	
Gender (female)	-0.188	***	0.004		-0.285	***
[ref: male]	(0.015)		(0.014)		(0.011)	
Enrollment (late)	-0.376	***	-0.640	***	-0.432	***
[ref: regular or early]	(0.054)		(0.030)		(0.016)	
Immigrant status (first gen)	-0.290	***	-0.244	***	-0.030	
[ref: native]	(0.054)		(0.041)		(0.027)	
Immigrant status (second gen)	-0.139	***	-0.138	***	-0.168	***
[ref: native]	(0.028)		(0.029)		(0.021)	
Geographical area (North-East)	0.052	**	0.134	***	0.060	***
[ref: North-West]	(0.023)		(0.022)		(0.016)	
Geographical area (Center)	0.052	*	-0.027		-0.297	***
[ref: North-West]	(0.024)		(0.022)		(0.017)	
Geographical area (South)	-0.066	***	-0.475	***	-0.715	***
[ref: North-West]	(0.021)		(0.019)		(0.015)	
ESCS	0.270	***	0.338	***	0.091	***
	(0.008)		(0.008)		(0.006)	
Anxiety	**-0.325**	***	**-0.400**	***	**-0.298**	***
	(0.008)		(0.008)		(0.006)	
School Type (technical)	-		-		-0.539	***
[ref: lyceum]					(0.013)	
School Type (professional)	-		-		-1.191	***
[ref: lyceum]					(0.016)	

Main results

Anxiety as dependent variable

	Grade 5	Grade 8	Grade 10
Intercept	-0.19***	-0.23***	-0.18 ***
female	0.39 ***	0.46 ***	0.38 ***

- The female show a positive significative relation with anxiety at all grades

Rasch score as dependent variable

	Grade 5	Grade 8	Grade 10
Intercept	204.67 ***	200.81 ***	199.98 ***
female	-5.60 ***	2.53 ***	-0.78 *
anxiety	-13.48 ***	-17.00 ***	-14.00 ***
anxiety:generefemale	3.15 ***	1.65 **	0.63

- The male perform better than female especially at grade 5. At grade 8 the coefficient change the sign in favour of female. At grade 10 the gender gap is slightly significant.
- The impact of anxiety on the performance is negative and significative at all grades
- There is an interaction effect between anxiety and gender on the performance in favour of female.

Cross-sectional data: fourth study

Aim of the analysis

- To estimate the ability and speed of students and to evaluate the impact of some students' characteristics both to the performance and to the response time behaviour.

Data, measure and method

Data

- ✓ INVALSI 2018 test data for grades 10
- ✓ Mathematics item responses + background variables + response time

Measure and methods

- ✓ First step: IRT model for estimating ability and speed
- ✓ Second step: bivariate multilevel model, where the first-level units are represented by students and the second-level units are represented by classes. Covariates such as gender, school type, immigrant status, economic, social, and cultural status, prior achievement, grade retention, student anxiety, class compositional variables, and geographical area are included in the model.

Bungaro, Desimoni, Matteucci, Mignani (2023)

Main results: first step

	Item Difficulty	Time Intensity	Time Discrimination
Item Difficulty	1.000	0.370 (0.000)	0.234 (0.004)
Time Intensity	0.370 (0.000)	1.000	-0.014 (0.436)
Time Discrimination	0.234 (0.004)	-0.014 (0.436)	1.000

	Person Ability	Person Speed
Person Ability	1.000	-0.574 (0.000)
Person Speed	-0.574 (0.000)	1.000

Main results: second step step

Final model parameter estimates

	Ability			Speed		
	Estimate	S.E.	*p*-value	Estimate	S.E.	*p*-value
Intercept	0.520	0.050	0.000	-0.330	0.069	0.000
male	0.110	0.008	0.000	0.100	0.009	0.000
student's ESCS	0.002	0.004	0.708	0.018	0.005	0.000
student_repeating_one_or_more_grades	-0.149	0.011	0.000	0.225	0.012	0.000
low prior achievement vs average and high	-0.442	0.008	0.000	0.248	0.010	0.000
math test anxiety	-0.162	0.004	0.000	-0.030	0.005	0.000
second generation immigrant vs native	-0.085	0.016	0.000	0.010	0.018	0.593
first_generation_immigrant vs native	-0.090	0.016	0.000	-0.052	0.019	0.006
Class % of stud. with low prior achievement	-0.007	0.001	0.000	0.004	0.001	0.000
Class % of immigrants	-0.005	0.001	0.000	0.006	0.001	0.000
Class average ESCS	0.211	0.029	0.000	-0.164	0.040	0.000
Class % of students repeating grades	-0.001	0.001	0.203	0.003	0.001	0.008
Class average math test anxiety	-0.046	0.026	0.075	-0.289	0.035	0.000
North West vs Center	0.210	0.028	0.000	-0.169	0.039	0.000
North East vs Center	0.251	0.028	0.000	-0.233	0.039	0.000
South vs Center	-0.259	0.027	0.000	0.159	0.038	0.000
South Islands vs Center	-0.504	0.034	0.000	0.356	0.047	0.000
Liceum vs Vocational	0.106	0.038	0.005	-0.251	0.052	0.000
Technical Inst vs Vocational	0.177	0.027	0.000	-0.371	0.037	0.000

- As for gender, the associations with speed and ability are both positive and very similar in size: males are slightly perform better and work slightly faster than females.
- though being faster leads to worse results males seem to use time in a better way than females.

Sum up: profiling the gender gap

MALES OUTPERFORM FEMALES

- Performances along the entire distribution:

→ at grade 5 and 8 the gender gap in favour of males just occurs at lowest level of performance and remain along the whole ditribution

→ At grade 10 the difference begin at medium level of performances

- Over time:

→just at 8 grade there is a disadvantage for female but the difference doesn't change over time

- Emotional components:

→ The females show higher levels of test anxiety at each grades but the anxiety effect on performance is less for female at 5 and 8 grade and not significant at grade 10

- Attitude toward the test:

→ males seem to use time in a better way than females.

Future remark

Policy implications:

- Promote actions just at primary school: at teaching level and in a more inclusive enviroment
- Support the girls with specific activities to ecourage the self-confidence toward mathematics→ peer activities in grops with boys
- Evaluate the teacher effect
- Control the role of *social persuasions*

Methodological developments:

- improve the models to address the complexity of the data and the multifaceted issues of gender gap.

References of the illustrated studies

- Silvia Bianconcini, Mignani; Jacopo Mingozzi*, Assessing maths learning gaps using Italian longitudinal data, «STATISTICAL METHODS & APPLICATIONS»,* in corso di stampa
- Luca Bungaro, Marta Desimoni, Mariagiulia Matteucci, Stefania Mignani,

  The joint estimation of accuracy and speed: An application to the INVALSI data, in "Proceeding of ASA Conference", in corso di stampa
- Mariagiulia Matteucci, Stefania Mignani (2021), *Investigating gender differences in mathematics by performance levels in the Italian school system, «STUDIES IN EDUCATIONAL EVALUATION»*
- Maria Giulia Matteucci, Stefania Mignani, Giada Spaccapanico Proietti, *Some insights on the relationship between student performanceand test related emotional aspects, in The school and its protagonists: the students. V Seminar "INVALSI data: a tool for teaching and scientific research",* Franco Angeli editore, in corso di stampa
- Roberto Ricci (2022), *Le differenze che non si vedono e la valutazione di scuola* «INDUZIONI», VOL. 60-61

Main references of the methodological tools

- Betebenner D.W. (2009). Norm- and Criterion-Referenced Student Growth. Educational Measurement: Issues and Practice. 28(4), 42-51.
- Curran, P. J., McGinley, J. S., Serrano, D., and Burfeind, C. (2012). A multivariate growth curve model for three-level data. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, and K. J. Sher (eds.). APA Handbook of Research Methods in Psychology. 3(17), 335-358. Washington, DC: American Psychological Association.
- Davino, F., Furno, M., Vistocco, D. (2013). *Quantile Regression: Theory and Applications*, Hoboken, NJ: John Wiley & Sons
- Goldschmidt P., Choi K. , Martinez F. and Novak J. (2010). Using growth models to monitor school performance: comparing the effect of the metric and the assessment. School Effectiveness and School Improvement. 21(3), 337-357
- Klein Entink, R. H., Fox, J.-P., van der Linden, W. J. (2009). A Multivariate Multilevel Approach to the Modeling of Accuracy and Speed of Test Takers. *Psychometrika*, **74**(1), pp. 21-48.
- Koenker, R. (2010), *Quantile Regression*, Cambridge: Cambridge University Press.
- Yen, W. M. (2007). Vertical scaling and No Child Left Behind. In N. J. Dorans, M. Pommerich, and P. W. Holland (Eds.). Linking and aligning scores and scales (pp. 273- 283). New York: Springer

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA