Pre-Analysis Plan

Tackling the Gender Gap in Mathematics in Piedmont

Dalit Contini[a], Maria Laura Di Tommaso[b], Daniela Piazzalunga[c]

[a] University of Torino

[b] University of Torino, Collegio Carlo Alberto CHILD, Frisch Center for Economic Research

[b] FBK-IRVAPP, CHILD-Collegio Carlo Alberto, and IZA

1. Introduction

According to the last available PISA data (OECD 2016), Italy is one of the countries with the highest GGM for 15-year-old students. While the Italian mean test scores in mathematics are similar to the OECD average (a mean score of 490), the gender differences in mathematics are much higher in Italy than in the OECD average (a 20-point difference in Italy against an average difference of 9 points in the OECD). This difference is the second highest among OECD countries, with only Austria having a higher difference. Also, 2015 TIMMS data (Trends in International Mathematics and Science Study) shows that Italy has the highest gender gap in mathematics for children in fourth grade among all the 57 countries included in the survey (Mullis et al. 2016). Contini et al. (2017), using the National Assessment for Italy (INVALSI) for year 2013, show that boys outperform girls in mathematics from 2nd to 10th grade. Moreover, the GGM is increasing with age even after controlling for some individual and family characteristics.

The general objective of the project is to devise a teaching method that aims to narrow the Gender Gap in Mathematics (GGM) and to measure the impact of this method on Piedmont children in primary school. If successful, this teaching method will be implemented in Piedmont schools.

The gender gap in mathematics in Piedmont (a region in the North-East of Italy) is even higher than the already high Italian average. It is slightly higher in 2nd and 5th grade and is much higher in 8th and 10th grade (Italian National Test, INVALSI, data for 2013). Therefore, Piedmont represents a good case study to test new methods devised at tackling the math gender gap. When focusing on the specific domains, the evidence from most assessments is that the gender gap in math is particularly high in the area of Numeracy, compared to “Data and previsions” and to “Space and figures” (own estimates with INVALSI data, years 2013-2017).

Many explanations have been proposed for the existence of a gender gap in mathematics. Among other factors affecting math performance, both how schools are organised and the educational methods and practices used in class matter and the project focuses on the way math is taught. Many studies show that when mathematics' teaching is centred upon problem solving, involving students in discussions and investigative work as opposed to traditional passive methods, the gender gap in math decreases and can even disappear (Boaler 2002, Zohar & Sela 2003, OECD 2016). These researchers frame the problems of GGM within the consolidated stream of constructivist and social methods in mathematical teaching/learning[1]. In a nutshell, according to these methods, mathematical learning involves activity on the part of the learners, leading to the idea that learners 'make things' together and 'communities of practice' are created (Lave & Wenger 1991). The focus is more on participation than on passive knowledge acquisition. In this perspective, it is the lack of constructivist and social methods in teaching mathematics that creates the GGM. In our project, we will instead use these teaching practices on both girls and boys to try to reduce the gender gap in mathematics.

We conduct a randomized control trial (RCT) to evaluate the impact of the implemented teaching methodology, designed incorporating recommendations from the existing literature, on the gender gap in mathematics in primary school. The RCT intervention covers about 50 third grade classes in 25 schools, evenly divided into treatment and control classes, involving about 1000 students. The intervention takes place in Piedmont, in the Northwest of Italy, and more specifically in the province of Torino. Due to type of intervention, we also evaluate the impact of the applied teaching practices on the math skills of all children, which is likely to be the first and strongest effect. The GGM will be reduced if math skills improvement is larger for girls, thanks to the specific design of the treatment.

2. The Italian schooling system

The Italian education system is organised in three stages. Students attend primary school from the age of 6 until the age of 11 years old. At the end of primary school, they enrol in lower secondary school, and remain within the same institution from the age of 11 until the age of 14 years old. All primary and lower secondary schools follow the same curriculum, defined at the national level. High school begins at the age of 14 and lasts for five years, but compulsory education terminates at 16 years old, so a share of children does not attain the upper secondary school qualifications.

At the end of middle school, students choose among different kinds of high schools, with significant differences in the curriculum. These educational programs are broadly classified into three main types: Lyceum, Technical High School, and Vocational High School. The curriculum

[1] For a survey, see Gutierrez & Boero (2006).

is generally organised at national level and all high schools have to offer some compulsory subjects (Italian, Mathematics, Sciences, History, one foreign language and Physical Education). However, there are significant differences in terms of the time allocated to each subject, and the specialised field of studies. Lyceums generally provide higher-level academic education, with a specialisation in the humanities, sciences, languages or arts. Technical institutes usually provide students with both a general education and a qualified technical specialization in a specific field (e.g.: business, accountancy, tourism, technology). Vocational institutes have specified structures for technical activities, with the objective of preparing students to enter the workforce.[2]

3. Focus of the Study

3.1. Purpose/Outcomes

This study will provide evidence of the impact of a new teaching methodology described below. The primary research questions are the following:

i. What is the impact of the treatment on children's math abilities?
ii. How does the impact differ by gender?
iii. How does the impact differ by children's prior achievements?

Due to the type of the intervention, the first and strongest effect we might expect is on children's math abilities in general. As the aim of the project is to reduce the gender gap, with a specific focus on the involvement of girls in class, we also evaluate if and how the impact differs by gender. Moreover, we may expect that different teaching methodologies will have differential impact on high and low achievement students, and thus explores heterogeneous effects by prior achievements.

Since the literature provides evidence that girls are generally more anxious than boys and exhibit less positive attitudes towards mathematics (Else-Quest, Hyde, & Linn, 2010; Lubiensky et al 2013; Mullis et al 2008; OECD 2016), additional research questions relate to the impact of the treatment on children's attitudes.

The identification of a causal effect relies on a randomized control trial.

3.2. Treatment methodology

The treatment consists in laboratories that implement a teaching practice aimed at improving children math skills and reducing the gender gap. Experts in mathematical education design and implement laboratory methods in which the learner is at the centre of activity instead of being subject to passive knowledge acquisition. The laboratories focus on peer interaction, sharing of

[2] For a general account of the Italian education system, see Ichino and Tabellini (2014).

ideas, students' engagement and tool use, problem solving and problem posing. These features are typical of the laboratory method, and tools can add value providing multiple representations and cognitive challenges for learners.

The treatment is applied homogenously across classes by tutors, trained in mathematical education (at Master or Ph.D. level), giving teachers the role of observers. It consists of different elements:

- Attention to the narrative context of the mathematical activity;
- Support to gender balanced interventions;
- Investigating the given problem with the children;
- Use of semiotic games (signs as words, gestures, glazes and uses of tools);
- Promote children interaction.

These strategies aims to activate children thinking, understanding, constructing mathematical meanings, and solving problems, through interaction and signs' production. We expect an increase in participation by children, in particular by the less self-confident ones, who are not always involved in standard activities.

Treatment delivery

According to the literature, the gender gap in math increases as children grow older, but starts developing at very young age (see for Italy, Contini et al. 2017). Hence, it is important to intervene on young children. However, since we want to assess the treatment's short-run impact, we need to implement it when the gap is already existing. For this reason, we have chosen to deliver the treatment in third grade, when the gap – although relatively small – can be already detected.

The treatment will be delivered at the class level, which on average have 20-25 pupils. All students in the class will take part to the activities, including student with disabilities or special needs to practice inclusion. However, their results will not be included in the statistical analyses for the impact evaluation. We plan to conduct 5 laboratory meetings of 3 hours each, once a week for 5 consecutive weeks, between January and May 2019. Children selected in the control group will follow the usual curricula.

4. Research Design

4.1. Setting, sampling procedure and randomization

The randomized control trial takes place in the primary schools of Torino province in the Piedmont region (North West of Italy), where there are about 180 primary schools.

The sampling procedure was set as follows. At the beginning of March 2018, the Regional Board of Education (*Ufficio Scolastico Regionale*) sent a letter informing all the principals of the primary schools in Torino and province about the project. In the same letter, principals and mathematics teachers were invited to a meeting that took place on the 10th of April. During the meeting, the project was presented by the Regional Board of Education, the University of Torino. During the meeting, the e-mail contact of the project administrator was provided for further information on the project.

The research team explained the following conditions for participation:

i) Each school must participate with at least 2 classes, one of which will be randomized to the control group and one to the treatment group.

ii) To avoid contamination and confounding effects, the two classes must have different math teachers and must not be involved in other math laboratories.

iii) Each school must sign a waiver allowing the research team to use the background information of pupils participating to the project and the pupils' individual results of the national test score INVALSI (see in section 5 for details about INVALSI data). The Principal Investigator of the project will sign a confidentiality agreement, in which she commits to not divulge to third parties the information collected during the project.

After the meeting, principals were able to enroll into the project signing a form that was sent to all the schools (both those who have participated to the meeting and those who have not) by the Regional Board of Education. The deadline for enrollment was the 31st of May 2018. The signed forms were sent to the project administrator.

Due to budget constraints, the plan was to enroll 50 classes and therefore at most 25 schools, implying approximately 1000-1250 pupils, half of which in the treated group. In the end, 31 schools applied. One was excluded because it was already participating to another math laboratory. Among the others, we randomly selected 25 participating schools; since some schools applied with more than two classes, we also randomly selected the two participating classes. Finally, within each school we randomly assigned one class to the treatment group and the other to the control one.[3]

[3] The sampling procedure plan was set before knowing how many schools and classes would have applied to participate to the project, and different rules were defined depending on the number of applications. Details can be found in the Appendix.

Both sampling of participating schools and randomization of classes to treatment or control group was public and took place at University of Torino on the 20th of June 2018. School principals received an e-mail inviting them to assist.

In case a class assigned to the treated group withdraws from the project, the entire school will be substituted with another school, chosen among the schools randomly excluded by the sampling procedure.

4.2. Identification

As explained above, we have included in the trial two classes per school, one randomized to the treatment group and the other to the control group. Since classes are much more similar within schools than across schools, this can be seen as a matching procedure, set up to improve the precision of the estimates and to increase the similarities between the treated and control group. Given the laboratorial nature of the intervention, performed by external professionals, and the fact that math teachers are different in the two classes, we do not expect contamination.

We identify the average treatment effect (ATE) by comparing the average test scores ($\bar{Y}$) after the intervention for the treated and control groups as in the following expressions:

$$ATE_{total} = \bar{Y}_T - \bar{Y}_C$$

We will also evaluate the heterogeneous effects by gender and by achievement levels

$$ATE_{Female} = \bar{Y}_{F,T} - \bar{Y}_{F,C}$$

$$ATE_{Male} = \bar{Y}_{M,T} - \bar{Y}_{M,C}$$

$$ATE_{High} = \bar{Y}_{H,T} - \bar{Y}_{H,C}$$

$$ATE_{Low} = \bar{Y}_{L,T} - \bar{Y}_{L,C}$$

Given the above effects, we can estimate the gender gap in math as follows:

$$ATE_{Gap} = \left(\bar{Y}_{M,T} - \bar{Y}_{M,C}\right) - (\bar{Y}_{F,T} - \bar{Y}_{F,C})$$

More specifically, the (total) treatment effect is identified using the following regression model:

$$Y_{iks} = \alpha + \beta T_{ks} + \gamma_s + \epsilon_{iks} \tag{1}$$

where *Y* is the post-intervention test score for the individual *i* in class *k* of school *s*; *T* is an indicator of whether the student belongs to a class in the treatment group when s/he is enrolled in school, γ is a vector of school fixed effects, and ϵ are random errors normally distributed. Due to randomization, β represents the treatment effect and is an unbiased estimate. Standard errors are adjusted for clustering at the class level because treatment is at the class level.

To increase the precision of the β estimate, we also include as control variables individual characteristics X (gender, age, migrant background, parental education and occupation) and math skills level (Y_0), from a pre-treatment test:

$$Y_{iks} = \alpha + \beta T_{ks} + \gamma X_{iks} + \delta Y_{0iks} + \gamma_s + \epsilon_{iks} \quad (2)$$

We standardize the pre- and post-test score, thus the effect of treatment β will represent by how many standard deviations test scores of the treated differ on average from those of the controls.

The math laboratories take place during school time and therefore we expect full participation to the project with possible absences due to sickness. In case of missing values for the pre-test or for individual characteristics, we assign a zero value and include a dummy variable for missing test-scores or for missing individual characteristics.

If children are absent to the post-test they will take the test in a different date, as close as possible to the original one.

If there are no-shows, which means that some pupils are absent during the treatment, the difference ($\bar{Y}_T - \bar{Y}_C$) will identify the intention-to-treat (ITT). ATE will be recovered by the Wald estimator, by dividing the ITT by the receipt rate for treatment group members.

We are exploring the possibility of utilizing children's attitudes towards mathematics as additional outcomes, using the same analysis just described. It will be defined in the next few months, before the starting of the intervention.

4.3. Power calculation

We set significance at α=0.05 and power (1-B) = 80%, following common convention; the multiplier of a two-tail test is thus 2.8 (see Bloom 2008). Under the assumption of individual randomization, the standardized minimum detectable effect MDES would be:

$$MDES(\bar{Y}_t - \bar{Y}_c) = M_{N-2}\sqrt{\frac{1}{NP(1-P)}} = 0.177$$

where N is the total number of individual in the RCT (1000 pupils) and P is the proportion assigned to the treatment (0.5).

We have, instead, group randomization (where groups are classes in our case) within blocks (schools); moreover, we will control for covariates, including a pre-test, to improve precision. The MDES becomes:

$$MDES(\overline{Y}_t - \overline{Y}_c) = M_{j-2}\sqrt{\frac{\rho(1-R_2^2)}{jP(1-P)} + \frac{(1-\rho)(1-R_1^2)}{jnP(1-P)}}$$

where j is the number of classes (50), n is the average number of children per class (20), R_2^2 is the proportion of group variance (at level two) predicted by covariates (included schools dummies), R_1^2 is the proportion of individual variance (at level one) predicted by covariates, and ρ is the intra-class correlation, defined as follows:

$$\rho = \frac{\tau^2}{\sigma^2 + \tau^2}$$

τ^2 is the between-class variance within schools; σ^2 is the individual level variance.

Using INVALSI data (described below) for children enrolled in second grade in Piedmont, we have estimated $\rho = 0.067$, $R_2^2 = 0.50$, $R_1^2 = 0.40$, thus $MDES(\overline{Y}_t - \overline{Y}_c) = 0.196$.[4]

5. Key data sources

5.1. INVALSI national test

Since 2010, the National Evaluation System of the Ministry of Education (INVALSI) has administered Math and Italian language tests on all the children attending grades 2, 5, 6, 8 and 10 (more than half a million students in each grade sit this test each year).

In addition to test scores, INVALSI data includes information about parental characteristics and family background, collected from a students' survey and from school board records. In selected years, INVALSI provides a synthetic indicator of economic and socio-cultural status (ESCS) similar to that the one available in PISA. The ESCS index is calculated by taking into consideration parental educational background, employment and occupation, and home possessions.

These data are used in the project in four distinct ways:

i) The national test for mathematics at the end of grade 2 for years 2013-2017 have been used to explore the gender gap in mathematics in different dimensions and areas, as defined by INVALSI, and to identify the area in which the laboratories will be developed (Numeracy).

ii) The national test for mathematics at the end of grade 2 for year 2017 has been used for the calculation of the MDE (see section 4).

[4] The same parameters indicated by Bloom (2008) are $\rho = 0.10$, $R_2^2 = 0.60$, and so our estimates are not unlikely.

iii) The data of children attending grade 2 in 2018 will be linked to each child in the trial (see section 4a) and will be used to recover relevant covariates to be included in model (2).

iv) 2018 INVALSI information for grade 2 will be used to position the children taking part to the RCT within the general population in Piedmont and Torino in terms of general achievement and socio-economic background.

5.2. Ad-hoc tests on numeracy skills

The experts in mathematical education will design a pre- and post-test for math numeracy skills to be conducted in treated and control classes, before and after the intervention. The structure of the national INVALSI test is such that for each domain (Numeracy, Space and figures, Data and previsions) there are only few questions. Therefore, given that the intervention is specifically designed to enhance competencies in the domain of Numeracy, we need to build an *ad hoc* tool to measure these competencies before and after the intervention. Pre-test and post-test will be substantially different but will follow the same conceptual framework. The pre-test will be delivered 1-2 weeks before the beginning of the treatment in each class by the project staff with the presence of the teacher. The post-test will be delivered approximately 4 weeks after the last laboratory session in each class.

The post-test will be used as the main outcome variable to assess the effectiveness of the intervention, whereas the pre-test, being likely highly correlated to the post-test, will be included as a control variable to improve precision of the estimates.

6. Balance test

In order to assess whether baseline covariates are balanced across the treatment and control groups, we will first compare descriptive statistics between the two groups. Then, we will estimate a regression with treatment status as the dependent variable and individual characteristics as independent variables, and conduct a F-test of the joint significance of the variables.

Archive

The pre-analysis plan is archived before we start the intervention. We archive it at the registry for randomized controlled trials in economics held by the American Economic Association.

References

Boaler, J. (2002a). The development of disciplinary relationships: Knowledge, practice and identity in mathematics classrooms. *For the Learning of Mathematics* 22(1), p. 42-47.

Bloom, H.S. (2008). Chapter 9. The core analytics of randomized experiments for social research, in Alasuutari, P., Bickman, L. & Brannen, J. (eds.) *The SAGE Handbook of Social Research Methods*. London: SAGE Publications Ltd doi: 10.4135/9781446212165

Contini D., Di Tommaso M.L., Mendolia S. (2017), The gender gap in mathematics achievement: Evidence from Italian data. *Economics of Education Review*, 58, pp 32-42. doi 10.1016/j.econedurev.2017.03.001

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103–127. doi:10.1037/a0018053

Gutierrez, A., and Boero, P. (2006). *Handbook of Research on the Psychology of Mathematics Education Past, Present and Future*. Rotterdam: Sense publ., pp. 305-428

Ichino, A. and Tabellini, G. (2014). Freeing the Italian school system, *Labour Economics*, 30, 113–128.

Lubiensky, S., Robinson, J., Crane, C., Ganley, C. (2013) Girls' and Boys' Mathematics Achievement, Affect, and Experiences: Findings from ECLS-K. *Journal for Research in Mathematics Education*, 44, 634-645.

Lucifora, C.and Tonello M. (2015), Cheating and social interactions. Evidence from a randomized experiment in a national evaluation program, *Journal of Economic Behavior & Organization*, 115, 45-66.

Mullis, I. V. S., Martin, M. O., Foy, P., and Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Retrieved from Boston College, TIMSS & PIRLS International Study Center.

OECD (2016), *PISA 2015 Results (Volume I): Excellence and Equity in Education*, Paris: OECD Publishing.

Zohar, A., and Sela, D. (2003). Her physics, his physics: gender issues in Israeli advanced placement physics classes. *International Journal of Science Education*, 25(2), 245-268.

Appendix: Sampling procedure plan

If the number of schools applying to participate is between 26 and 30 we will randomly select 25 schools. If the number of schools is larger than 30, these schools will be stratified according to size and district. The scope is to have a heterogeneous sample of 25 schools and therefore we will select randomly within each stratum in order to ensure an appropriate representativeness of each stratum.

If schools have enrolled to the program with two classes, one will be randomized in the treatment and the other in the control group. If schools have enrolled with more than two classes we will randomly select the 2 classes that will take part to the study.

If the number of schools is lower than 25, we will send a reminder to the principals. If the number is still less than 25 we will use the following procedure. Assume we have only k<25 schools. We will first randomize for each school one treated and one control class (step 1). If all schools participate with 2 classes we will end up with a smaller sample of k classes in the treated and k classes in the control. If some schools participate with more than 2 classes, we will gather together the classes not selected in step 1) and randomly select 25-k additional classes for the treatment and 25-k for the control group.