WHEN GENDER MATTERS: A STUDY OF GENDER DIFFERENCES IN MATHEMATICS

Ferrara, F., Ferrari, G., Robutti, O., Contini, D. & Di Tommaso, M.L.

Università degli Studi di Torino, Italy

This paper addresses gender differences in mathematics at the early grades of primary school, based on a research study conducted in Italy, in the region with the largest gender gap in mathematics in the National panorama. Borrowing from the literature around gender and its different conceptualizations, we focus attention on the possible relationship between the gap and the cognitive demand, task and formulation of mathematical test questions. Restricting the analysis to the content area of numbers, the one with the largest gap, we will highlight some of the variables that seem to affect the gender gap, arguing for a more equitable mathematical practice.

INTRODUCTION

In this paper we want to contribute to the current discussion on gender differences in mathematics. Differences in mathematical performances in favour of boys exist and are considered as having implications on the fact that females are substantially under-represented in STEM university subjects and in highly innovative and technological careers (Miyake et al., 2010). We refer to the difference in mathematical performance between males and females as the gender gap in mathematics (GGM). Research has shown that the GGM is a matter of concern for policies that address equity both at school and in the labour market (Di Tommaso et al., 2018), especially at a time of social crisis, like the current one in regard to the pandemic. On the other hand, patterns of gendered inequity provide a sobering counterpoint to claims of an equitable mathematical experience, thus troubling and disrupting given gender performances within contexts and conditions does matter more than ever (Walshaw et al., 2017). As Walshaw and colleagues underline, other constructs of social difference such as class, race, ethnicity also become significant, as do histories of mathematical access, success, production, underachievement or exclusion. Speaking of GGM is therefore important in relation to a wider perspective of binaries between diversity and equity.

The latest international assessments of mathematics (like PIRLS and PISA) show Italy as one of the countries with the largest GGM. This emerges from the primary through upper secondary school test scores. In particular, Italy possesses the largest gap among the 57 countries taking part in TIMSS grade 4 evaluation (Mullis et al., 2016), and is in the second position in the case of 15-

2021. In Inprasitha, M., Changsri, N., Boonsena (Eds.). *Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 2, pp. 255-263. Khon Kaen, Thailand: PME.

year-old students (OECD, 2016). These results are further problematized looking at data from the National Institute for the Evaluation of the Education System (INVALSI) in Italy, according to which a GGM is observable since grade 2 and becomes more prevalent during secondary school. The primary purpose of this paper is to address issues concerning the GGM in grade 2 in Italy, starting from the results of the assessment of mathematics of years 2013 to 2017. We are particularly interested in studying variables that might affect the GGM in this context, and in designing classroom-based interventions to reduce it in mathematics. To this aim, the research team is interdisciplinary and involves mathematics educators and social economic researchers. In the next section, we frame the research study into the literature that we see as relevant to highlight and discuss differences between male and female performances in mathematics.

THEORETICAL HIGHLIGHTS

Much international literature shows unique achievement trends of males and females in mathematics and reading across a number of countries (e.g., Robinson & Lubienski, 2011; Ajello *et al*., 2018). Math gaps favouring males were found to increase between kindergarten and third grade (Rathbun, West, & Germino-Hauskin, 2004). Also, the GGM is particularly pronounced among high-performing than among low-performing students and widens as children grow older even if it does not widen during lower secondary school (grade 4 through 8; Contini *et al*., 2017). In the broader literature, developments in gender research endeavour to think differently about the GGM, with understandings of gender ranging from biological or cultural and environmental factors to family and teacher beliefs and biases, to girls' low self-confidence and self-efficacy in terms of mathematical ability and performance within gendered identity-work (Else-Quest *et al*., 2010; Lubienski *et al*., 2013). The role of stereotypes and other socio-cultural forces is well established (see Aronson & Steele, 2005 for a detailed review). Some available research studied gender differences in mathematics in relation to performance and highlighted that they seem to be related to the cognitive processes that are investigated by the question and linked to the type of question. For example, Bolger and Kellaghan (1990) discovered that while boys outperform girls in multiple-choice questions, girls outperform boys on open-ended questions. Other studies indicated strong association between aspects of reading and of mathematics tests (Marks, 2008; Caponera *et al*., 2016). Robinson and Lubienski (2011) further claimed that given that gender patterns in math performance tend to run counter to those in reading, examinations of both subjects together provide a more complete picture of girls' and boys' learning. Ajello *et al*. (2018) claim that the reading burden of mathematics questions is associated with student performance in mathematics, independently of mathematical ability. Due to the fact that girls are better performers than boys when facing reading tests, they

seem to be advantaged in mathematics questions with a high reading demand, independent of their level of reading literacy. Questions with a low reading demand are instead more in favour of boys. According to Ajello and colleagues (2018), question difficulty and task can also be related to such differences, therefore further research should investigate the type of cognitive process involved in answering the task, for example whether a computation or problem solving. Other research stresses that variations on question formulation affect differently male and female performances and that this might be concerned with different strategies used by the two populations (e.g., Bolondi *et al*., 2018).

Borrowing from these considerations, we shift attention to studying the possible relationships between the type of task, formulation and cognitive demand in mathematical questions and the existence of a GGM, as we have defined it above. In this way, the paper wants: (a) to contribute to current discussions on mathematical gender differences at primary school, in a double manner: by confirming findings from the literature, and by expanding these focusing on variables strictly related to the questions; and (b) to examine the local context of Piedmont, which shows to be the Italian region with the largest GGM in grade 2, supported by territorial funding for dedicated research. In the next section, we introduce context and method of the study.

CONTEXT AND METHOD

As mentioned above, in our research we take the GGM as the difference between average male and female scores in their mathematical performance. Our original data source is given by the scores of the National grade 2 assessment tests of mathematics over the period 2013 to 2017. In order to avoid possible bias related to cheating, the estimation sample was reduced to including only those classes that were supervised by external inspectors during the tests. In addition, the sample was further restricted to a sub-sample including only the classes in Piedmont, where we work with an active network of policy makers and schools.

The assessment test of mathematics delivered each year by INVALSI approximately contains 25 to 28 questions, each of which can be composed by more than one item, like in the case of True or False multiple complex choice questions. The scores to which we associate the GGM take into account all the items of the grade 2 assessment of mathematics in the period mentioned above, for a total of 6.732 observations. The items are associated to a content area, a dimension (the main cognitive process implied by the item) and a question intent (the item purpose). According to the Mathematics Assessment Framework of INVALSI (INVALSI, 2018), which follows the National Guidelines for the curriculum, three are the possible content areas for grade 2: *Numbers*, *Data and previsions*, *Space and figures*, and three the cognitive dimensions: *Knowing*, *Problem solving* and *Arguing*. The question intent is

concerned with typical forms of mathematical thinking, like text comprehension, calculation, use of different representations or measurement tools, reasoning, data research, and problem solving.

Table 1 offers the results of the initial descriptive statistics of our sample by content area, with average score and GGM, and the percentage of items for each area. The score provided for each student is measured as the percentage of correct answers over the total items. The results show that the average score is lowest in the case of Numbers for both males and females, but also contains the majority of items. On average, the total gap is 0.028 (2.8 percentage points, or p.p.): while females answer correctly to 53.9% of the items, for males we get 56.7%. Additionally, the area of Numbers has the largest GGM (3.7 p.p.), moving us to centre our investigation on this particular area. The number of items belonging to Numbers between 2013 and 2017 is 82 (the number of observations in the table; each observation was assessed on about 1340 subjects). Focus was on these items to better understand which of their characteristics could partake of the GGM revealed by the statistics. The analysis was centred on the study of constant differences in the GGM concerned with item characteristics over the entire period rather than on the trend over time. Therefore, we adopted a mixed method, both qualitative and quantitative. The qualitative part borrows from the literature we refer to and regards an initial search for variables that constitute each item formulation and structure, beyond those variables that are considered already by the assessment framework. The second part of the analysis involves descriptive statistics of all these variables. This allows us to study the relationships between the GGM and the type of task, formulation and cognitive demand of the mathematical items.

Variable	**Overall**	**Males**	**Females**	**GGM (M-F)**	**% items**
Average score	0.554	0.567	0.539	**0.028*****	100
Content area					
Numbers	0.517	0.535	0.498	**0.037*****	56.9
Data and previsions	0.614	0.620	0.608	0.012*	16.0
Space and figures	0.613	0.618	0.608	0.010**	27.1
N. observations	*6,732*	*3,387*	*3,345*		

*p-value < 0,10; **p-value < 0,05; ***p-value < 0,01

Table 1: GGM: Average score (% of correct answers) by content area

QUALITATIVE ANALYSIS AND VARIABLE IDENTIFICATION

As anticipated above, we identified the variables that characterise item formulation and structure through a qualitative analysis of all the selected items. This process brought forth the following as relevant variables:

A. Cognitive dimension: Arguing, Knowing, Problem solving.

B. Question intent: Calculation, Text comprehension, Reasoning, Different representations, Data research, Problem solving, Measurement tools.
C. Type of item: Open-constructed response, Multiple choice.
D. Item formulation: Situation, No Situation, Objective, No objective.
E. Kind of figure: No figure, Drawing, Figure in context, Representation.

While the first three classes of variables (A, B, C) refer to INVALSI framing of the items, the other two classes (D, E) were added to account for: the presence or absence of a situation which provides the context of the task, or of an objective which gives the aim of the task (D); the absence or presence of a figure and the eventual kind of figure (E). We distinguish figures according to three kinds: drawings, figures in context, and representations (Fig. 1 shows three examples from specific questions of the 2017 assessment test). A drawing simply contains a number of objects to which the task refers (asking for example to count them, Fig. 1a). A figure in context implies an understanding of the sense to attribute to objects in specific contexts (like in the case of money, Fig. 1b). A representation requires a step forward to infer the relationships between objects (like when lengths of different objects need to be compared, Fig. 1c).

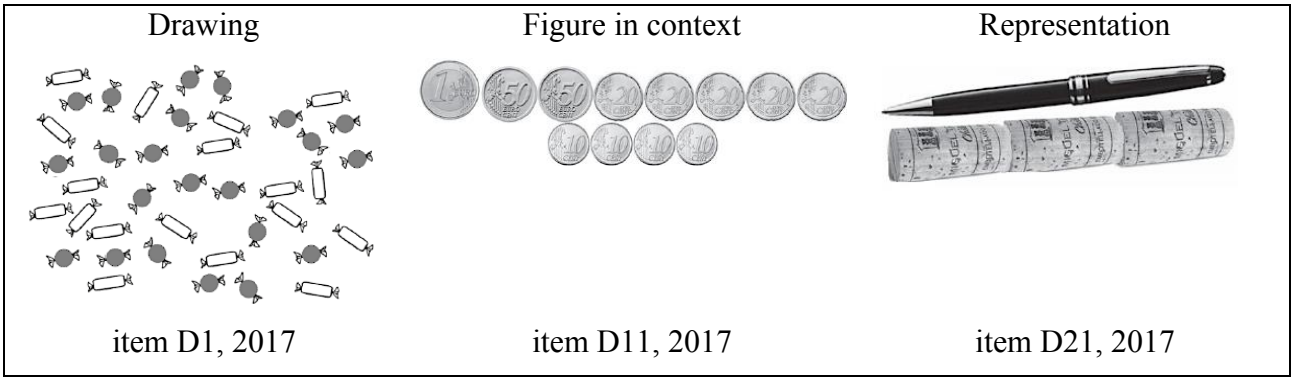

Figure 1: Examples of different kinds of figures

After this identification process, we selected the particular variables for each of the 82 items of our sample and created a table, in which each row refers to a specific item D*n* while the column cells are targeted with value 0 or 1 depending on whether the corresponding variable is absent or present in that item.

QUANTITATIVE ANALYSIS AND RESEARCH FINDINGS

The attribution of values 0 and 1 to item variables was used to develop the new statistics for our quantitative analysis through simple linear regression, which allowed us to get some descriptive measure of the influence of particular variables on the presence of the GGM. In so doing, we focused on the difference across single items obtaining some information from which to begin:

the mean percentage of correct answers across items is 52.5%, while the gender gap across items is 0.039; there is large variability embedded in this gap, with the minimum -0.10 (in favour of females) and the maximum 0.23 (in favour of males). This relevantly suggested that, as a matter of fact, the nature of the items (briefly, their formulation and structure) actually affects the gap, although without saying in which terms. Investigating the variables above exactly allows us to see how and to which extent this occurs. Tables 2 to 4 below help to better explain this. In all the tables standard errors are in parentheses and the number of asterisks defines how significant the gap is (the lower the *p*-value the more significant the gap is). In particular, Tables 2 and 3 are concerned with the influence of the variables from the INVALSI framework, that is, cognitive dimension and question intent. Table 4 instead refers to the additional variables we identified.

Cognitive dimension	**Item GGM**
Arguing	0.018 (0.015)
Knowing	**0.036***** (0.008)
Problem solving	**0.052***** (0.010)
Obs.	*82*
R^2 *adj.*	*0.348*

**p*-value < 0,10; ***p*-value < 0,05; ****p*-value < 0,01

Table 2: Item GGM: influence of Cognitive dimension

Question intent	**Item GGM**
Calculation	0.027** (0.011)
Text comprehension	0.021 (0.018)
Reasoning	0.012 (0.027)
Different representations	**0.048***** (0.013)
Data research	0.067** (0.031)
Problem solving	**0.050***** (0.011)
Measurement tools	0.038 (0.038)
Obs.	*82*
R^2 *adj.*	*0.348*

**p*-value < 0.10, ***p*-value < 0.05, ****p*-value < 0.01

Table 3: Item GGM: influence of Question intent

	Item GGM				
Open constructed-response	**0.028*** **	(0.008)			
Multiple-choice	**0.053*** **	(0.009)			
No situation		**0.061*** **	(0.016)		
Situation		**0.035*** **	(0.006)		
No objective			**0.039*** **	(0.006)	
Objective			0.036**	(0.018)	
No figure				0.033**	(0.011)
Drawing				0.010	(0.013)
Figure in context				**0.064*** **	(0.013)
Representation				**0.045*** **	(0.010)
Obs.	*82*	*82*	*82*	*82*	
R^2 *adj.*	*0.364*	*0.345*	*0.327*	*0.386*	

*p-value < 0,10; **p-value < 0,05; ***p-value < 0,01

Table 4: Item GGM: influence of Item type and formulation

From Table 2 we see that problem solving is the cognitive dimension that affects the GGM the most. Table 3 shows that the use of different representations and problem solving are the two most problematic aspects implicated by the items concerning the GGM. Regarding item type and formulation variables (Table 4), our results confirm (in the local context) the findings of the literature according to which males perform better than females in answering multiple-choice questions, showing a gap of 53%. On the contrary, open constructed-response items are more favourable to females (in fact, the gap is 28%). Further, the absence or the presence of a situation does not seem to affect the GGM in any particular manner (both contribute to it to an almost equal extent), while the presence of an objective seems to act in the direction of reducing the gap with respect to its absence (two asterisks instead of three). The bearing of a drawing is marginal as regards that of a representation or (even more) of a figure in context, while the absence of figures affects the GGM on average. These findings move us to make didactical considerations. For example, more work with representations seems to be needed within the mathematics classroom, both in terms of the treatment of different representations and in relation to their meanings, with the aim to reduce the

documented GGM. Similarly, attention should be devoted to contextualising mathematical activity, like in the case that we use figures requiring knowledge of the context to be understood. The dimension of problem solving is another delicate one that calls for didactical intervention.

CONCLUSIVE REMARKS

Our study wants to contribute to existing discussions about gendered disciplines by shifting emphasis from available gender research to material, concrete experiences of gendered performances in mathematics. Borrowing from the existing literature and the findings from these performances, we suggest that lines of didactical intervention are needed to deeply engage both females and males in mathematical doings. This is particularly relevant in a time of social crisis like that of the pandemic, which also showed to intensify differences. A rethinking of educational practice is needed towards a more equitable mathematics, one that disrupts boundaries to overcome gendered identity discourses within the classroom, for example by de-centring consensus about practice mainly based on calculation and procedural knowledge and shifting attention to problem solving. Focusing on the local context of Piedmont, the Italian region with the largest GGM already at grade 2, we offered reflections about variables that seem to affect the presence of the gap and that we see as relevant to any discourse of mathematics teaching and learning. Future research is necessary to widen the horizon on possible interventions and efficiently inform policy making in these directions.

References

Bolger, N. & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement, 27*(2), 165–174.

Bolondi, G., Branchetti, L. & Giberti, C. (2018). A quantitative methodology for analyzing the impact of the formulation of a mathematical item on students learning assessment. *Studies in Educational Evaluation, 58*, 37–50.

Caponera, E., Sestito, P. & Russo, P.M. (2016). The influence of reading literacy on mathematics and science achievement. *The Journal of Educational Research, 109*(2), 197–204.

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.

Di Tommaso, M.L., Maccagnan, A. & Mendolia, S. (2018). The gender gap in attitudes and test scores: A new construct of the mathematical capability. *IZA Discussion Paper 11843*.

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), 101–127.

INVALSI (2018). *Quadro di Riferimento delle prove INVALSI di Matematica*, retrieved from: https://invalsi-areaprove.cineca.it/docs/file/QdR_MATEMATICA.pdf

Lubienski, 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. & 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.

Marks, G.N. (2008). Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. *Oxford Review of Education, 34*(1), 89–109.

Miyake, A., Kost-Smith, L.E., Finkelstein, N.D., Pollock, S.J., Cohen, G.L. & Ito, T.A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science, 330*(6008), 1234–1237.

Mullis, I.V.S., Martin, M.O., Foy, P. & Hooper, M. (2016). TIMSS 2015 international results in mathematics. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College; International Association for the Evaluation of Educational Achievement (IEA).

OECD. (2016). *PISA 2015 results (volume I): excellence and equity in education*. Paris: OECD Publishing.

Rathbun, A.H., West, J. & Germino-Hausken, E. (2004). *From kindergarten through third grade: Children's beginning school experiences* (NCES 2004-007). Washington, DC: National Center for Education Statistics.

Robinson, J.P. & Lubiensky, S.A. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school examining direct cognitive assessments and teacher ratings. *American Educational Resources Journal, 48*, 2268–2302.

Walshaw, M., Chronaki, A., Leyva, L., Stinson, D. W., Nolan, K., & Mendick, H. (2017). Beyond the box: Rethinking gender in mathematics education research. In A. Chronaki (Ed.), *Proceedings of the 9th International Mathematics Education and Society Conference* (MES9, Vol. 1, 184–188). Volos, Greece: MES9.