THE PHENOMENON OF THE GENDER GAP AMONG GIFTED STUDENTS: THE SITUATION IN ITALY BASED ON ANALYSIS OF RESULTS IN MATHEMATICS COMPETITIONS

Lorenzo Mazza¹, Alessandro Gambini²

¹Dipartimento di matematica, P.le Aldo Moro 5, Sapienza Università di Roma, Italy, email: lorenzo.mazza@uniroma1.it
²Dipartimento di matematica, P.le Aldo Moro 5, Sapienza Università di Roma, Italy, email: alessandro.gambini@uniroma1.it

Abstract

This paper deals with the gender gap phenomenon as displayed by Italian students participating in mathematical competitions - gifted students. In particular, the results obtained by the entire student population participating in district competitions from 2017 to 2020 were analyzed. The combined analysis of the competition results and of the final questionnaire, allowed us to highlight and deepen possible causes of the gender gap. It has been observed that girls underestimate themselves, sometimes giving up participating in competitions. Particularly during the upper secondary school years, increasingly fewer girls choose to attend mathematical competitions. Analysing the results, although the percentage of correct answers between the two genders is fairly similar, the rate of girls who prefer not to answer to avoid mistakes is higher than that of boys.

Keywords: gender gap, gifted students, high school students, large scale assessment, mathematical ability, mathematical olympiad.

1. Introduction

Analysis of results in the main student competitions and national and international surveys shows that, in many countries (especially in Italy), the gender difference in mathematics performance is particularly marked. Many researchers have established how the gender gap in STEM (Science, Technology, Engineering and Mathematics), both in terms of performance and participation, is correlated to multiple factors of a sociocultural nature, but also to low self-confidence levels in students, to the different perception of their own skill, and to the beliefs and influence of teachers and parents (Andreescu et al., 2008; Bahar, 2021; Ceci & Willimas, 2007, Guido et al., 2008, Stoet et al., 2016). These gaps persists even after they leave school. A recent U.S. study showed that women are more likely to be employed in particular occupations such as social work (60%), while only a small percentage of them are employed in science (26%) and engineering fields (15%; National Science Board, 2018).

The performance gap between male and female students appears particularly pronounced if we limit ourselves to analysing those with high-level mathematical performance (so-called gifted students) who, with good approximation, can be identified in those who successfully take part in various student competitions (Ellison & Swanson, 2010; Makel et al., 2016; Niederle & Vesterlund, 2007, 2010; Wai, Cacchio, et al., 2010).

This paper focuses on an analysis of the gender gap phenomenon limited to the top band of the student population, paying particular attention to the Italian situation without losing sight of the international context. The starting point is an analysis of standardised tests, but considerable space is given to the
study of results obtained by Italian students in district mathematics competitions in the last four years and to the analysis of answers provided by a sample of these students to a questionnaire administered after the district competition of 2020. We also try to provide a key to explain the possible main causes underlying the gender gap among gifted students in mathematics competitions.

1.1 Gender gap in standardised tests

The gender gap represents one of many phenomena studied in the field of mathematics education. As pointed out by Leder and Forgasz (2008), the global introduction over the last twenty years of standardised tests aimed at measuring students' skills and competences, has drawn the attention of the international community to a disparity in performance between males and females. This, in turn, led to analysis of this phenomenon with reference to geographical distribution, the specific education systems of each country and socio-economic and cultural contexts.

The most widespread international measurements in terms of assessment are represented by the OECD PISA and IEA TIMSS tests.

The PISA (Programme for International Student Assessment) tests of the OECD (Organisation for Economic Co-operation and Development) are aimed at assessing the acquisition of key competencies required for a full and active participation of citizens in society. The tests are carried out every three years: the first edition was held in 2000 and the reference sample is composed of a number of 15-year-old schoolchildren from more than 70 countries, including Italy. Although each edition has a specific focus on a certain area, the four aspects analysed are Reading Literacy, Mathematical Literacy and Scientific Literacy, plus another area (e.g. creativity).

The TIMSS (Trends in International Mathematics and Science Study) tests of the IEA (International Association for Evaluation of Educational Achievement) are designed to assess acquisition of content and skills in mathematics and science by students in 4th and 8th grade. The first edition took place in 1995 and was subsequently carried out every four years. There are approximately 60 participating countries.

An additional IEA survey, called TIMSS Advanced, focuses on student learning in mathematics and physics in the last year of secondary school (grade 13), the last edition of which was held in 2015 and involved different types of secondary schools.

At the national level, there may be other standardised tests aimed at verifying the acquisition of certain skills in certain areas, as for example happens with the INVALSI tests in Italy, which are carried out annually in all classes in grades 2, 5, 8, 10 and 13 and regard the Italian language and mathematics. For some school grades there is also an English test. The education system in Italy consists of a first cycle of education that is divided into two distinct school tracks: primary school, lasting 5 years (grades 1 to 5), and lower secondary school for 3 years (grades 6-8, equivalent to middle school), followed by upper secondary school for 5 years (grades 9-13).

Analysis of mathematics results in the different standardised tests mentioned above shows that the gender performance gap phenomenon does not occur equally in all participating countries. On the contrary, it sometimes occurs that in some nations females perform better than males, while at other times the difference in performance is minimal and therefore statistically insignificant (OECD 2012, 2013, 2015, 2016a, 2016b, 2016c, 2016d). Nevertheless, in about half of the countries considered, a significant gap in favour of boys is observed, and this is the case both for those with above-average overall performance and for those with below-average results.
It must be said, however, that the difficulty of reaching some regions to administer the test, or the difficulty that girls have in continuing their studies until the age of 15 (the age at which PISA tests are administered) can significantly influence the results of the measurements (Giberti, 2019).

With reference to the Italian situation and the area of mathematics, when analysing both the results of the 2015 PISA test and those of the TIMSS tests, we can see how Italy is one of the countries where the difference in performance between boys and girls is among the most marked. This is especially evident (in the case of the TIMSS tests) for 8th grade students as compared with 4th grade students (Giberti, 2019; Giberti & Ferretti, 2019). The rather worrying situation is also confirmed by INVALSI measurements carried out at the national level (INVALSI, 2016; INVALSI, 2017; Contini et al., 2017), where the gender gap phenomenon is found significantly in all school levels covered by the test.

1.2 The gender gap phenomenon and different levels of ability

Recent research (Andreescu et al., 2008; Ellison & Swanson, 2010; Makel et al., 2016; Niederle & Vesterlund, 2010; Olszewski-Kubilius & Lee, 2011) has shown that the performance gap between males and females is more pronounced among students with high levels of ability in mathematics, above the 95th percentile. As early as 1980, Benbow and Stanley published an article regarding the SAT-M (Scholastic Assessment Test - Mathematics) test scores of more than 40,000 students up to age 13, which showed that the ratio of males to females (referred to as MFR) at the highest 0.5% range (99.5th percentile and score ≥ 500) was 2 to 1, rising to a ratio of 13 to 1 for the highest 0.01% (score 700-800). Subsequent research (Wai, Cacchio et al., 2010; Olszewski-Kubilius & Lee, 2011; Makel et al., 2016) has shown that, in the years following the research of Benbow and Stanley, especially from 1980 to 2015, this males to females ratio gradually decreased to 2.5 for the top 0.01% range, denoting an attenuation of the gender gap phenomenon, although it is still significant.

The gap between males and females tends to widen not only as a function of ability level, but also with age. (Contini et al., 2017; Hyde & Mertz, 2009, Fryer & Levitt, 2010). Contini et al. (2017) use the results of INVALSI tests administered to the entire Italian student population in grades 2, 5, 8 and 10, in the years between 2010 and 2015. The authors analyse the gender gap at different points of the test scores’ distribution with quantile regression. What emerges from their studies is that the gender gap increases with the age of children and is greater among top performing children. Hyde and Mertz (2009), citing studies published between 1966 and 1974 by the developmental psychologist Eleanor Maccoby and in 2005 confirmed by the research of Elizabeth Spelke in U.S., point out that performance between boys and girls in primary school years is similar. Boys' skills in mathematics increased faster than girls' beginning around 12 or 13 years of age, leading to a significant gap in high school years. Fryer & Levitt (2010) analyse the phenomenon of the gender performance gap in the first six years of school using a particularly large sample of over 20,000 U.S. children. The mathematics test used assesses number recognition, counting, comparing and ordering numbers, solving word problems and interpreting picture graphs. They observed that there is no average difference between boys and girls at the beginning of school career, but that the gap increases over the course of the school years. While in kindergarten, girls occupy 45% of the top 5 percentiles, after 5 years the percentage of females occupying that range drops dramatically to 28%. Some years earlier, Winner (1996) also observed that girls and boys who are particularly gifted in mathematics emerge in equal proportions in the early grades, but in high school the percentage of gifted girls drops to 30% and drops even more among older students.
An analysis of standardised tests (PISA 2009, 2012, 2015) shows, for most participating countries, a particularly pronounced performance gap between boys and girls with high levels of ability. The effect is such that (with reference to PISA 2015) the average gap in OECD countries, equal to 8 points, doubles if we consider only the 10% of students belonging to the high end (better performance). Again, Italy is revealed as one of the countries where this gap is more pronounced, as confirmed by the INVALSI tests, with only 8% of girls reaching the highest level of the PISA scale compared to 13% of boys (Giberti, 2019).

To support this statement, we can cite the response of Leder and Forgasz (2008) to a 2008 paper by Hyde, Lindberg, Linn, Ellis, and Williams that appeared in the journal Science. In the journal, the authors state that they find no differences in performance between boys and girls when analysing data from standardised tests in the United States. Leder and Forgasz point out that the gender performance gap phenomenon is more marked among high levels of ability, while Hyde and colleagues used tests made up of items of medium or low cognitive level and therefore feasible for all, thus explaining the possible reason for the absence of a statistically significant gap.

Of particular interest is the analysis carried out by Niederle and Vesterlund (2010), who interpret the gender gap in mathematical performance among high-achieving students in terms of a different approach to competition between boys and girls. It has been observed that in mixed competitive contexts, i.e., where both girls and boys are called upon to participate, not only do the females achieve less satisfactory results than the males (Gneezy, Niederle & Rustichini, 2003) but they often prefer not to take part (Niederle & Vesterlund, 2007). The same authors (2007) show that the number of students belonging to the lowest quartile who are interested in participating in a mathematical competition is proportionally greater than the number of female students belonging to the first quartile willing to take part in a similar competition. In all-girl (or principally female) situations, girls show less reluctance to compete and a greater belief that they can achieve high positions in the final rankings.

Currently, researchers are leaning toward viewing a test score not as a simple measure of a student's cognitive ability, but as an indicator of both the student's cognitive and non-cognitive abilities. For example, this is recognized by Cunha & Heckman (2007) or Segal (2008), with noncognitive factors (e.g., motivation and sense of responsibility) having a marked influence on test scores. Gneezy and Rustichini (2000) demonstrated the extent to which the motivational aspect can influence performance outcomes, by administering a 20-minute IQ test to a homogeneous sample which revealed lower scores among those who were paid a lower-value “reward”. From this perspective, the response that each individual shows when faced with a competitive context represents a non-cognitive ability that can influence performance. It thus happens that scores obtained in tests carried out in a competitive context (such as mathematics competitions) can reveal a gender gap that would not have emerged significantly in a less competitive context.

### 1.3 Possible factors behind the gender gap

Regarding possible factors underlying the gender gap, one can find various and different interpretations in the literature, and the debate is quite broad, with references to biological, social, cultural, and psychological factors. It is possible to distinguish between internal factors (i.e., dependent on the individual) and external factors (i.e., dependent on the socio-cultural context in which the individual lives).
With reference to internal factors, the idea that differences of a biological or physiological nature may exist has revealed itself to be limited, and in any case could not have represented the only factor capable of explaining different performance between the two sexes (Contini et al., 2017), especially when also considering the performance gap between boys and girls in different countries in standardised tests (sometimes favouring girls). It should be remembered that towards the end of the 19th century it was believed that women had a smaller brain size than men, thereby explaining a vaunted gender gap which was probably linked, even more than today, to social and cultural factors (Hyde & Mertz, 2009). This position can also be easily refuted in the light of the results of PISA and TIMSS which show that not all nations reveal a male skewed gap.

There are several studies that have analysed the gender gap in terms of performance, participation and affective; we mention some of them knowing that, as previously discussed, there are several factors that can influence this gender gap, and much depends on the jurisdiction and context in which the study is carried out. We are therefore far from being able to say that the research in the literature has universal value.

For instance, some research (Gallagher & De Lisi, 1994; Gallagher et al., 2000) has shown that at the basis of the gender gap, particularly among students who achieve high levels of performance, lies the different approach and different resolution strategies adopted by girls as compared with boys, with the former preferring to apply algorithmic reasoning to more conventional problems, while the latter succeed better in less conventional problems by also taking advantage of reasoning based on intuition. Lawton & Hatcher (2005) note better performance by male students in visuo-spatial skills. A possible explanation could be that girls are more compliant with teachers' expectations while boys are less subject to the didactic 'contract' and are less willing to work according to routines, and therefore perhaps more familiar with problem solving activities. Nevertheless, it is believed that also these reasons are not sufficient alone to explain the gender gap, as certain skills can be acquired by anyone if given appropriate training.

A more delicate issue lies in analysis of the psycho-social factors at the root of motivations, beliefs, and the degree of self-confidence. Following the 2012 PISA survey, which focused on mathematics, in the context questionnaires the OECD analysed mainly three constructs related to this topic, namely: mathematics self-concept (perceived competence in mathematics), mathematics self-efficacy (perceived ability to solve a series of problems in pure and applied mathematics) and mathematics anxiety (feeling of stress and need for support in dealing with mathematics). These are key aspects of learning and performing mathematical tasks; for example, those with less self-confidence will tend to tackle the task at hand with less enthusiasm and inclination.

The questionnaires revealed that female students have less confidence in their mathematics skills and consider themselves inferior to their opposite-sex peers, and this emerges even for those with similar or equivalent mathematics skills or achievement levels (OECD, 2015). Girls possess a level of mathematics self-efficacy quite similar to boys on many particular tasks such as the application of algorithmic procedures, which agrees by Gallagher et al. (2000), but nevertheless at a high level less confident in their strength in mathematics. Finally, as supported by other research (Devine et al. 2012; Primi et al., 2014), girls' level of mathematics anxiety is higher overall than that of boys. In contrast, the mathematical gap disappears when considering female and male students with equal levels in the three aforementioned constructs (OECD, 2015).

The gender gap phenomenon, on the other hand, cannot be attributed to internal factors alone. Social and cultural factors (external factors) must also be considered, such as the level of women's
empowerment, gender-related stereotypes, classroom context habits, and different teaching practices (Fryer & Levitt, 2010; Guiso et al., 2008; Leder & Forgasz, 2008; Tomasetto, 2013).

Fryer and Levitt (2010) document the presence of the gender performance gap across every stratum of U.S. society. Guiso et al. (2008) mainly analyse the situation in Turkey, Sweden and Iceland, with the former having a particularly low level of emancipation and a particularly high performance gap between boys and girls; on the contrary, in the two northern European countries, the level of emancipation is higher while the gender gap is lower (or, as in the case of Iceland, it is in favour of girls). Tomasetto (2013) takes up and analyses a series of studies carried out previously, including Jacobs and Eccles's 1992 study on a sample of around 1,500 American students aged 11-12 and their mothers, which shows that parents' perceptions of competence in relation to their daughters are not the same as those of their sons. Authors point out that it is still widely held among parents that success in mathematics by boys is the result of a natural predisposition, whereas for girls it is the result of effort. In addition, teachers also attribute greater ability in mathematics to males than females (Helwig et al, 2001).

Beginning with data from PISA 2003, Guiso et al. (2008) relate the gender performance gap in mathematics and Italian (the first in favour of males, the second in favour of females) to the GGI - World Economic Forum's Gender Gap Index. This is an index that expresses (for each nation) gender equality based on economic, political, educational and health conditions. In countries with a high level of emancipation of women (and therefore with a high value of the GGI) the gender gap in mathematics tends to disappear. The Italian situation is emblematic; in 2017 it occupied the 82nd position out of 144 countries analysed in terms of gender equality.

Hyde & Mertz (2009) analysed the composition of the top 30 teams in the overall ranking among those who took part in the IMOs (from 1989 to 2008) and observed how the percentage of girls is statistically correlated to the value of the GGI of the respective countries considered. A similar study was conducted independently by Ellison & Swanson (2010), but they found no statistically significant correlation. The same authors attribute the difference in results to the fact that their research covered a larger number of countries participating in the IMOs (91 nations) and for a more restricted period (2006-2008). With reference once again to the Italian context, the low value of the GGI is in alignment with the fact that in the 41 competitions in which the country took part from 1967 to 2020, there were only 5 girls in total, although one of these (Maria Colombo) took part in three editions.

Ellison and Swanson (2008) also analysed the performance of approximately 120,000 students who participated in the 2007 AMC, noting that the male:female ratio for those scoring higher than 100 (top 6%) is 4, and that ratio rises to 10 when analysing the first percentile.

Some recent research seems to show that the gender gap phenomenon has narrowed over the last few decades, a fact also related to a greater number of women employed in STEM disciplines, the availability of advanced courses for girls, a greater emphasis towards the pursuit of gender equality and a general attention towards this issue (Freeman, 2004; Hyde & Mertz, 2009; Makel et al., 2016).

1.4 The present study

In Italy, there are several student competitions in mathematics. The most widespread is represented by the Giochi di Archimede (Archimedes’ Games), an individual competition organised by the UMI (Italian Mathematical Union) and held generally in November within each institute adhering to the project (as many as 1440 institutes throughout Italy in school year 2017/18). The primary objective of this event is the dissemination of a different way of seeing mathematics by setting more challenging
problems than those proposed during normal school lessons. The number of participants, as has been observed for other mathematical competitions held in other countries (for example, for the AMC - American Mathematics Competition; see Bahar, 2021), has been steadily decreasing, from about 260,000 students in the 2010/11 school year to about 200,000 students in the 2017/18 school year. Participants appear to be evenly split by geographic area and age. The ratio of boys to girls (about 4 to 3) also shows that such an event attracts both male and female students. The number of questions that make up the competition varies, ranging from a minimum of 16 items for students in the first two years of high school (grades 9 and 10) to 20 items for students in the next three years (grades 11, 12, and 13). Since the youngest students (grade 9) could risk being left out of the Level II competition, a "repechage" competition reserved for them (in which about 2,000 students throughout Italy participate) is staged in early February.

The next step is the district competition at the end of February, which is attended by the best performers in the Archimedes Games and the Grade 9 student competition in each district. About 10-12 thousand students take part in this competition every year. The competition, the same for all participants, proposes 12 multiple-choice items, 2 numerical response items and 3 open-ended demonstrative questions. The best 300 students then compete in the national final, which is held every year in early May in Cesenatico, near Rimini. The duration of the competition is 4 and a half hours and participants must solve 6 open-ended demonstration exercises. In this final stage, the difference between girls and boys is very marked, with girls accounting for around 10% of the total number of participants.

Six students (selected partly on the basis of the national competition results and partly on performance during some internships that take place during the year) are sent to the IMO (International Mathematics Olympiads) which are held each year, usually in July, in a different host country.

Figure 1

Following the district competition in February 2020, a questionnaire was administered with the purpose of investigating the gender gap among gifted students, with reference to their states of mind and perception of their own skills during student competitions.

The questionnaire with 8 questions was submitted only to students in the district of Rome on the day of the test and was aimed at determining:

- The level of difficulty perceived, during the competition, in the different areas that characterize the mathematics of the olympiad (algebra, combinatorics, geometry, number theory);
- Whether or not they felt capable in each of the above areas;
Whether or not they had seen exercises at school similar to those in the competitions;
School performance;
Possible placement in the ranking;
The level of anxiety experienced during tests in class and during the mathematical competitions.

2. Methodology and data analysis

To study the phenomenon of the gender gap among gifted students, we first analysed the sample of Italian students, divided by gender, who competed in district competitions from 2017 to 2020 and the results they achieved. Subsequently, and with reference to the restricted sample of Roman students participating in the February 2020 competition, the answers to a questionnaire were examined.

A particularly useful index in the analysis is represented by the ratio between the number of males and number of females (MFR, which we will indicate with γ). It can be deduced that if γ=1, the number of male and female students is the same, if γ>1 there is a predominance of males and if γ<1 there is a greater number of females in the sample considered.

The analysis of the participants, divided by gender, showed that the value of the MFR, on the total number of participants (TMFR) is about 3 for each of the years considered, which means that 75% of the competitors are male and only 25% female. This ratio, just under 3 in the first two years of secondary school (grades 9 and 10), tends to increase with the age of the participants, as shown in the table below.

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFR I year</td>
<td>2.98</td>
<td>2.48</td>
<td>2.91</td>
<td>2.70</td>
</tr>
<tr>
<td>MFR II year</td>
<td>3.04</td>
<td>2.77</td>
<td>2.87</td>
<td>2.51</td>
</tr>
<tr>
<td>MFR III year</td>
<td>3.15</td>
<td>2.80</td>
<td>2.83</td>
<td>3.17</td>
</tr>
<tr>
<td>MFR IV year</td>
<td>3.69</td>
<td>3.36</td>
<td>3.12</td>
<td>3.44</td>
</tr>
<tr>
<td>MFR V year</td>
<td>3.99</td>
<td>4.06</td>
<td>4.31</td>
<td>3.64</td>
</tr>
<tr>
<td>TMFR (total)</td>
<td>3.32</td>
<td>3.08</td>
<td>3.12</td>
<td>2.99</td>
</tr>
<tr>
<td>No. of students</td>
<td>10,544</td>
<td>9,688</td>
<td>9,905</td>
<td>9,995</td>
</tr>
</tbody>
</table>

Table 1: MFR divided per year

Such an analysis supports the observations of Ellison & Swanson (2010) and Bahar (2021) regarding AMCs.

With reference to the total number of students, it should be specified that the actual number of those who competed was on average about 10% higher than stated in the table. The value indicated is the result of a selection of the data, uploaded by the local managers on the portal of the Unione Matematica Italiana (the organiser of the competition), eliminating strings that presented inconsistencies, errors or lacks (e.g., lacking gender or score) and thus not usable. Despite this, the
numbers we have in our possession are particularly large and therefore the population under investigation provides us with statistically significant information.

The calculation of the MFR can also be made as a function of a specific percentile \( \theta \), by means of the relation \( \gamma(\theta) = \frac{m(\theta)}{f(\theta)} \), where \( m(\theta) \) and \( f(\theta) \) indicate the number of males and the number of females respectively in the distribution considered (Bahar, 2021). The table below shows the values of \( \gamma \) for the 99th and 95th percentiles.

**Table 2**

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1%</td>
<td>12.50</td>
<td>24.00</td>
<td>18.20</td>
<td>15.00</td>
</tr>
<tr>
<td>Top 5%</td>
<td>6.03</td>
<td>12.16</td>
<td>8.86</td>
<td>7.69</td>
</tr>
</tbody>
</table>

Table 2: male to female ratios for the first and fifth percentiles

The participation gap in the ratio between the two sexes becomes even more evident if one considers the upper end of the ranking, where the number of males is clearly higher than the number of females, leading to particularly emblematic situations such as the 2018 competition, where of the first 100 classified (top 1%) 96 were males and only 4 females. In particular, it was not possible to calculate \( \gamma(99.9) \), i.e. the ratio of males to females for the 0.01% of the top band (which, for samples of around 10,000 students, is equivalent to analysing the gender distribution of the top 10 students in the ranking) as all students of this band, for all 4 years considered, were male.

The following table shows the average scores obtained by Italian students in the four competitions in question, distinguishing between the average score of females only, males only, and the entire student population participating in the competitions.

**Table 3**

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average score (females)</td>
<td>24.8</td>
<td>22.2</td>
<td>22.3</td>
<td>24.7</td>
</tr>
<tr>
<td>SD score (females)</td>
<td>11.4</td>
<td>11.0</td>
<td>11.3</td>
<td>11.6</td>
</tr>
<tr>
<td>Average score (males)</td>
<td>27.6</td>
<td>28.0</td>
<td>27.0</td>
<td>28.6</td>
</tr>
<tr>
<td>SD score (males)</td>
<td>13.8</td>
<td>14.0</td>
<td>14.3</td>
<td>14.4</td>
</tr>
<tr>
<td>Average total score</td>
<td>27.0</td>
<td>26.6</td>
<td>25.9</td>
<td>27.6</td>
</tr>
<tr>
<td>SD global</td>
<td>13.3</td>
<td>13.6</td>
<td>13.8</td>
<td>13.9</td>
</tr>
</tbody>
</table>

Table 3: male and female average scores and standard deviation

An initial analysis of the above values would seem to suggest greater aptitude on the part of males for solving questions in mathematical competitions. The analysis of the mean square deviation suggests that girls tend to be much more likely than boys to be close to their mean score. The latter, on the other hand, have a larger mean square deviation, indicating greater volatility and difference in performance from student to student. The following table would seem to confirm this, as well as offering interesting food for thought.
Table 4:

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct answers females ex. 1-12</td>
<td>27%</td>
<td>25%</td>
<td>23%</td>
<td>26%</td>
</tr>
<tr>
<td>Correct answers males ex. 1-12</td>
<td>30%</td>
<td>32%</td>
<td>28%</td>
<td>31%</td>
</tr>
<tr>
<td>Wrong answers females ex. 1-12</td>
<td>44%</td>
<td>45%</td>
<td>45%</td>
<td>38%</td>
</tr>
<tr>
<td>Wrong answers males ex. 1-12</td>
<td>43%</td>
<td>40%</td>
<td>43%</td>
<td>37%</td>
</tr>
<tr>
<td>Missing answers females ex. 1-12</td>
<td>29%</td>
<td>30%</td>
<td>31%</td>
<td>35%</td>
</tr>
<tr>
<td>Missing answers males ex. 1-12</td>
<td>27%</td>
<td>28%</td>
<td>29%</td>
<td>32%</td>
</tr>
<tr>
<td>Correct answers males ex. 13-14</td>
<td>20%</td>
<td>11%</td>
<td>21%</td>
<td>04%</td>
</tr>
<tr>
<td>Correct answers females ex. 13-14</td>
<td>16%</td>
<td>04%</td>
<td>17%</td>
<td>02%</td>
</tr>
<tr>
<td>Wrong answers females ex. 13-14</td>
<td>13%</td>
<td>35%</td>
<td>39%</td>
<td>24%</td>
</tr>
<tr>
<td>Wrong answers males ex. 13-14</td>
<td>14%</td>
<td>37%</td>
<td>42%</td>
<td>29%</td>
</tr>
<tr>
<td>Missing answers females ex. 13-14</td>
<td>71%</td>
<td>61%</td>
<td>44%</td>
<td>74%</td>
</tr>
<tr>
<td>Missing answers males ex. 13-14</td>
<td>65%</td>
<td>52%</td>
<td>37%</td>
<td>67%</td>
</tr>
<tr>
<td>Missing answers females ex. 15-17</td>
<td>88%</td>
<td>83%</td>
<td>87%</td>
<td>82%</td>
</tr>
<tr>
<td>Missing answers males ex. 15-17</td>
<td>84%</td>
<td>75%</td>
<td>79%</td>
<td>76%</td>
</tr>
</tbody>
</table>

Table 4: percentage of correct, incorrect and missing answers, for males and females

Looking only at the first 12 exercises (multiple-choice), the percentage of wrong answers for females is comparable (except in the case of 2018) to that of males. However, the percentage of correct answers varies, with females being about 5 percentage points below males. Consequently, the percentage of answers left blank also varies, with girls tending on average to skip more questions than their male counterparts. If we then analyse the answers given to exercises 13 and 14 (with numerical answers, therefore open but without any arbitrary assessment by the markers) we find once again a gap between the correct answers of females and those of males, favouring the latter, but it is interesting to note that in all four years the percentage of wrong answers by males is higher (sometimes even markedly, as in 2020) than that of females, and the percentage of answers left blank between girls and boys is strongly marked. In detail, males respond better and more frequently than females, trying to avoid leaving the exercises blank, unlike girls who, if they are not sure, prefer not to provide any answer.

Finally, also in the last three open-ended exercises (questions 15-17), the percentage of girls scoring zero points is on average higher than that of boys. It should be noted that zero points are generally awarded to those who leave the exercise blank, since the correction grids produced by the competition organisers provide for 1-2 points to be awarded to those who carry out even simple and basic (provided they are correct) considerations on the exercise. This agrees with Gallagher et al. (2000), who in their studies have shown that, when carrying out some SAT-M tests which are carried out according to different times and modalities (conventional or unconventional questions, open or closed answers, 10' or 50' time), girls tend to leave more exercises blank than boys.

As previously reported, a questionnaire was administered to the 486 students in the district of Rome who participated in the February 2020 competition. This is a sample of the entire national population (about 5% of all those who took part in the February 2020 district competition). We know that it
cannot be representative of the overall Italian situation because it involves students from the Rome district only.

This questionnaire was printed on a single sheet of paper stapled to the rest of the test so that the name and surname of the respondent could be known. The decision not to have anonymous results was necessary to link each participant's answers with their results in the competition, so that correlations could be sought. The data was then entered into a spreadsheet and processed using statistical software. The table below shows the breakdown of the sample according to class and gender. The last row shows the value of $\gamma$ for each year of the course.

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>24</td>
<td>36</td>
<td>14</td>
<td>10</td>
<td>21</td>
<td>105</td>
</tr>
<tr>
<td>Males</td>
<td>72</td>
<td>101</td>
<td>60</td>
<td>77</td>
<td>71</td>
<td>381</td>
</tr>
<tr>
<td>MFR</td>
<td>3.00</td>
<td>2.81</td>
<td>4.29</td>
<td>7.70</td>
<td>3.38</td>
<td>3.63</td>
</tr>
</tbody>
</table>

Table 5: numbers of students divided by gender and schoolyear

The proportion of girls participating was 21.6% compared to 78.4% of boys ($\gamma=3.63$, slightly higher than the national value). A total of 541 students were invited to participate in the competition, of whom 55 did not show up, 21 girls (38%) and 34 boys (62%). These values show that the percentage of girls who did not participate in the competition is much higher in proportion to their attendance. Niederle and Vesterlund (2007) also observed that many girls opt out of competitions even though they have the opportunity. The difficulty in contacting absent students meant we were not able to identify the real reasons for their withdrawal; however, it was observed that the division by type of school and class in the group of 55 absent students did not show significant differences compared to the total sample of students summoned. In fact, the absentees were evenly distributed by type of school and school grade, thus ruling out the possibility that the students who were less interested in the event were older or younger students or students from non-scientific high schools.

Careful analysis of the answers in the questionnaires allowed us to identify a possible explanation for the gender gap phenomenon.

First, it should be noted that the average level of difficulty attributed to the questions in the February 2020 test is fairly comparable between boys and girls, with the former estimating the average difficulty (in tenths) at 7.17 and the latter 7.30. It is also interesting to note that, when asked to indicate how good they consider themselves to be in the individual areas of the mathematical olympiad, the boys gave themselves an overall score of 6.56, compared with 6.42 for the girls, which suggests that the latter also believe they have the skills and abilities to solve the competition exercises. It is worth remembering that the questionnaire was administered at the end of the competition, thus after having seen and reasoned about the questions, and not before the competition, where many students could have underestimated or overestimated their abilities, especially in the case of first-time participants.

The difference between the actual and estimated ranking is worth mentioning. It should be remembered that, during administration of the final test, the participants were not asked to assign themselves a possible precise position, but only a range within which they thought they could fall (see question 6 of the questionnaire in Appendix). During the processing and transcription of the data, the range 1-50 was given a value of 1, the range 51-100 a value of 2, and so on, up to the range 451-
500 which was given a value of 10. From the analysis of the final ranking, it emerges that the average of actual placings of the males is around the 225th position, while that of the females is around the 275th position. The same actual placings, reported in values from 1 to 10 depending on the range, correspond to an average of 5 for boys (range 201-250) and 6 for girls (range 251-300).

As far as the estimated value in the ranking is concerned, the average of all the values associated with the different ranges for boys is 5.97 (range 251-300, i.e., estimated to be around the 295th position, 70 places of difference from the real value) while this average for girls is 7.49 (range 301-350 which corresponds approximately to the 375th position, thus 100 positions away from the real value). Also, in this case the result obtained agrees with Gallagher et al. (2000) who, from the analysis of the answers to the information questionnaire administered to participants in their experiments, showed that girls tend to underestimate their abilities in mathematics and science much more than boys and to assign themselves a possible placement in the lowest band of abilities (below average).

The following table shows some percentage values concerning the overestimation or underestimation of boys and, in the next table, of girls. In order to understand the meaning of the items shown in the table, the wording "overestimated by 1" (resp. by 2, by ≥3, or underestimated by 1, by 2, by ≤3) indicates that a student has assumed to be in the range X (with 1≤X≤10), but in reality his actual position is in the range X+1 (resp. in the range X+2, X+k with k≥3 or in the range X-1, X-2, X+k with k≤-3 and k∈Z).

<table>
<thead>
<tr>
<th>Correctly estimated males (k=0)</th>
<th>15.52 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overestimated males (k=1)</td>
<td>8.33 %</td>
</tr>
<tr>
<td>Overestimated males (k=2)</td>
<td>6.32 %</td>
</tr>
<tr>
<td>Overestimated males (k≥3)</td>
<td>10.34 %</td>
</tr>
<tr>
<td>Total overestimated males (k&gt;0)</td>
<td>25.00 %</td>
</tr>
<tr>
<td>Underestimated males (k=−1)</td>
<td>20.40 %</td>
</tr>
<tr>
<td>Underestimated males (k=−2)</td>
<td>16.09 %</td>
</tr>
<tr>
<td>Underestimated males (k≤−3)</td>
<td>22.99 %</td>
</tr>
<tr>
<td>Total underestimated males (k&lt;0)</td>
<td>59.48 %</td>
</tr>
</tbody>
</table>

Table 6: percentage of correctly-judged, overestimated and underestimated males

<table>
<thead>
<tr>
<th>Correctly estimated females (k=0)</th>
<th>10.64 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overestimated females (k=1)</td>
<td>5.32 %</td>
</tr>
<tr>
<td>Overestimated females (k=2)</td>
<td>5.32 %</td>
</tr>
<tr>
<td>Overestimated females (k≥3)</td>
<td>11.70 %</td>
</tr>
<tr>
<td>Total overestimated females (k&gt;0)</td>
<td>22.34 %</td>
</tr>
<tr>
<td>Underestimated females (k=−1)</td>
<td>15.96 %</td>
</tr>
<tr>
<td>Underestimated females (k=−2)</td>
<td>13.83 %</td>
</tr>
<tr>
<td>Underestimated females (k≤−3)</td>
<td>37.23 %</td>
</tr>
<tr>
<td>Total underestimated females (k&lt;0)</td>
<td>67.02 %</td>
</tr>
</tbody>
</table>

Table 7: percentage of correctly-judged, overestimated and underestimated females
What emerges from analysis of the previous tables is that the number of girls and boys who overestimate themselves is globally the same (22.34% vs. 25.00%), and the percentage of girls and boys who underestimate themselves are, respectively, 67.02% and 59.48%, with a slight predominance of girls underestimating themselves, as compared with boys. However, the most macroscopic data, and probably the most significant, is that the range within which girls and boys assign themselves varies significantly. While, for example, only about one student out of five greatly underestimates himself or herself (i.e., by a difference in the range between the actual position and the estimated position, in absolute value, greater than or equal to 3), this value among female students alone reaches a percentage of 37.23%, i.e., more than one girl out of three. In other words, not only do girls tend to underestimate themselves more than boys, but when they do underestimate themselves, they do so much more notably than their male counterparts.

Again referring to the above-mentioned questionnaire, the same girls, when asked to describe with one or more words the state of mind they had during the competition, tended to use expressions such as "frustrated", "discouraged", "insecure", "stressed", "anxious" more frequently than their male mates. The final part of the questionnaire focused specifically on the theme of anxiety. Questions 7 and 8 aimed at determining the level of anxiety possessed during class tests and mathematical competitions. Despite not being a more complete battery of questions, the information that can be deduced from these is nevertheless indicative (Gogol et al., 2014; Núñez-Peña et al., 2014). It emerged that females possess a higher level of anxiety than boys, especially during classroom tests (4.28 compared with an average of 3.66 for boys).

3. Discussion and conclusion

Though to a lesser extent than a few years ago and varying from country to country (Hyde & Mertz, 2009), the gender gap phenomenon can also be found among gifted students and within student competitions, and is particularly marked when analysing the highest levels (top 1% or 5%). Moreover, the tendency of girls to avoid participating in a mathematics competition becomes increasingly evident as they get older. These observations, carried out on data related to Italian district competitions of the last 4 years, agree with other international studies previously carried out in other countries (Andreeescu et al. 2008; Bahar, 2021; Benbow & Stanley, 1980; Ellison & Swanson, 2010; Makel et al., 2016; Niederle & Vesterlund, 2010; Olszewski-Kubilius & Lee, 2011). These studies do not cover the international panorama regarding the participation gap between girls and boys in competition; jurisdictions where participation is the same cannot be excluded. From an analysis of the questionnaire administered to a sample of Italian students (see appendix), it was observed that girls tend not only (and not so much) to underestimate their own skills, but also to underestimate the possibility of achieving good positions in the competition, sometimes even giving up competing. Thus, the results, presented in Tables 6 and 7 above, seem to suggest that the percentage of girls who underestimate themselves by many places (even more than 150 places out of a population of about 500 students) is much higher than that of boys. Further research could be carried out not only to confirm this trend, but also by using different criteria (e.g. by asking students to estimate also the total score they assume they have obtained).

These results also seem to suggest that the gender performance gap is not based on differences in actual skills or abilities, but rather on a different perception of one's own abilities and a different
emotional state. The gap, perceived by these girls to be bigger than it is, leads them to a higher level of insecurity and therefore to a preference for "not taking risks" by leaving questions blank. The findings revealed about exercises 13 and 14 are particularly indicative (see Table 4): boys answer these questions better and more frequently but also make more mistakes than the girls, as if the former were more inclined to give an answer in all cases or at least to try, while the latter maintain a more cautious profile and prefer to leave blank spaces when they do not know how to proceed or are not sure. The last three exercises of the same test (district competitions) confirm the trend just mentioned, i.e. the fact that girls leave more exercises blank than boys. Further research could be carried out to better understand whether the girls' decision is an "a priori" renunciation (without even reading the text of the exercise but possibly concentrating on a smaller number of questions) or whether some of those who do not answer have nevertheless tried to read and think about it. It is also interesting to see whether, when confronted with tests with a different type of scoring (e.g., negative values for wrong answers and zero scores for answers left blank), the boys take a more cautious attitude, thus reducing the differences found between the two genders.

A possible interpretation of the performance gap could therefore be related to the expectations and the beliefs envisaged by the society and therefore the gap factor could be linked to the didactic contract in the sense of Brousseau. This would explain the tendency of the girls to leave answers blank when unsure.

In any case, a simple analysis of the results of the competitions can give general indications but cannot explain the causes of the gap. Another possible cause could be the different emotional state in which girls and boys approach the study of mathematics (and, in the specific case of this research, the performance of competitions), with girls experiencing more performance and mathematical anxiety than boys. As we saw in the data analysis in the previous section, the fact that girls may have a higher level of mathematical anxiety than boys emerges from different studies (Lengfelder & Heller, 2002; Callahan et al., 1996, ), which show that girls attribute great importance to the opinions of adults, by whom they are influenced and conditioned in their choices, unlike boys, who attribute the merit of their abilities to intrinsic characteristics (personal effort and commitment, their own curiosity). Specifically, Lengfelder & Heller's study looked at a group of 235 German females aged between 19 and 42 years (mean age 27.8 years) who took part in mathematics competitions in their upper secondary school years. Stoet et al. (2016), using PISA data on a sample of 15-year-old students from 68 countries (years 2003 and 2012), observed that in more developed countries with greater gender equality, the difference in the level of mathematical anxiety between boys and girls is particularly marked; furthermore, families believe that sons are more likely to succeed in mathematics than daughters. This is true despite the fact that many more women hold STEM jobs in these countries, as compared to countries with less gender equality. Such an observation, which flies against all forecasts in many respects, gives some insight into why girls are more anxious and eager not to disappoint adults when performing in mathematical competitions.

In light of the previous considerations, work is being carried out which is aimed at exploring more deeply the issue of mathematical anxiety among gifted students. In particular, starting from the administration of a new questionnaire, we want to understand if, and to what extent, anxiety during competitions influences students' performance and if there are significant gender differences.

Appendix
Questionnaire administered to Roman students at the end of the competition on 20 February 2020

1. Indicate, on a scale of 1 (very easy) to 10 (very difficult), the average difficulty of the problems in today's mathematics competition:

   Algebra problems (e.g., no. 9,12,14) - Geometry problems (e.g., no. 4,7,11,17)
   Number theory problems (e.g., no. 1,2,3,6,8,15) - Combinatorial problems (e.g., no. 5,10,13,16)

2. Apart from today's test, indicate on a scale of 1 (not at all) to 10 (very much) how good you think you are at these subjects:

   Algebra - Geometry - Number theory - Combinatorial/Probability

3. Indicate with a cross, which moment you found most difficult in solving the problems of today's mathematics competition.

<table>
<thead>
<tr>
<th>FORMULATING HYPOTHESES AND WORKING STRATEGIES</th>
<th>USING YOUR KNOWLEDGE TO SOLVE THE QUESTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra problems</td>
<td></td>
</tr>
<tr>
<td>Geometry problems</td>
<td></td>
</tr>
<tr>
<td>Number theory problems</td>
<td></td>
</tr>
<tr>
<td>Combinatorial/probability problems</td>
<td></td>
</tr>
</tbody>
</table>

4. How many times have you encountered in your school experience (not in connection with mathematics competitions), problems similar to those faced today?

<table>
<thead>
<tr>
<th>Never</th>
<th>Almost never</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra problems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geometry problems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number theory problems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combinatorial/probability problems</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. What was your mathematics final mark at the end of this year's term/quarter (if you had two separate written and oral grades, please give the average of the two)?

6. Knowing that there are about 500 students competing in this competition, which position segment do you think you will be in??

   1-50  51-100  101-150  151-200  201-250  251-300  301-350  351-400  401-450  451-500

7. Give a score from 1 (not at all) to 10 (very much) for the degree of anxiety you feel when doing a mathematics assignment in class.

   Write one or more words that describe how you feel during a mathematics assignment in class

8. Give a score from 1 (not at all) to 10 (very much) for the degree of anxiety you experience during mathematics competitions.

   Write one or more words that describe how you feel during mathematics competitions.

Acknowledgements
We wish to thank Marta Desimoni, Valentina Vaccaro and Chiara Giberti for their useful insight and suggestions which improve our paper.

References


**Biographies**

Lorenzo Mazza, Professor of mathematics and physics since 2006, he started working on mathematical competitions in 2009, coordinating the running of them for the Rome district. He has participated in several local and national stages, also as a speaker. He actively collaborates with the three Roman state universities in the management of team competitions. Since 2019, he has undertaken a PhD in mathematics education, whose focus is on the issue of the gender gap and the difficulties gifted students have in carrying out geometry questions.

Alessandro Gambini, PhD in Mathematics with a thesis on Analytical Number Theory, is Associate Professor of Mathematics Education at Sapienza University of Rome. He has been working for years in the field of Mathematics education working on large-scale assessment and non-Euclidean geometry.