

GENDER DIFFERENCES IN RELATION TO PERCEIVED DIFFICULTY OF A MATHEMATICAL TASK

Bianca Nicchiotti* and Camilla Spagnolo*

*Free University of Bozen-Bolzano

The theme of perceived difficulty in mathematics is current, but it has only been considered in the last few years. This study aims to highlight gender differences in relation to perceived difficulty of a mathematical task, and factors influencing it. Italy is one of the countries with the largest gender gap in mathematics, hence there is the interest in analysing perceived difficulty considering it as a key to better understand. We started considering as a benchmark the nationwide quantitative analysis of gender gaps in two INVALSI tasks, characterized by different gender gap levels. Then, we link students' perceived difficulty to these two tasks, analysing qualitatively the differences between boys' and girls' perception. Preliminary findings point out that girls' perception is mainly related to personal consideration.

INTRODUCTION AND THEORETICAL FRAMEWORK

The topic of difficulty in mathematics has been object of research for years and the difficulty of a mathematical task seems to be influenced by a combination of factors, such as text comprehension (Spagnolo et al., 2021), and affective factors (Zan, et al., 2006). Within this frame, the students' perspective is crucial, and in the last few years research in mathematics education started considering the issue of perceived difficulty (PD) (Spagnolo & Saccoletto, 2023). The PD is different from the difficulty, as the latter is usually evaluated in retrospective, considering the ratio between the students who solved the item correctly and the total number of students who faced the item (Mehrens & Lehmann, 1991).

The issue of perception of difficulty has been analysed in the field of cognitive psychology, developing different characterisations and definitions for the concept (e.g. Eccles & Wigfield, 2020). The one we consider the closest to our conception of PD is the “feeling of difficulty” (FOD), defined as a “metacognitive experience that monitors cognitive processing as it takes place” (Efklides & Touroutoglou, 2010, p. 272). However, FOD and PD are strongly related but conceptually different, as the former is based on the experience whereas the latter can be described as a metacognitive judgement built considering a conscious memory of knowledge about oneself and the task. At the moment, in mathematics education research, there is not a clear definition of PD, but some of the factors that appear to influence it have been described and analysed qualitatively (early phases of the studies were presented by the authors during international conferences ICME14, CERME13, CIEAEM74 and MAVI29). They have been distinguished into five different but not mutually exclusive categories, developed basing on students' answers to specific questions about PD. The first category is

Resolution strategy, containing all the references to the procedure, strategy or knowledge needed to solve the task, calculations and reasonings included. The second one is *Capability and experience*, and it contains elements related to students' opinion of their competence as well as their familiarity with the task and similar previous experiences. The references to the time spent to solve the task, possible obstacles encountered and doubts belong to this category too. The third one, *Emotions*, regards the emotional aspects (positive or negative). The fourth one is *Task formulation*, including all the comments about the task in general (structure, text, type, etc.) while the last one, *Personal consideration*, concerns each student's personal opinion about their own success in mathematics or as a student.

In this paper we want to analyse students' PD of a task from a gender perspective. The issue of gender gap in mathematics is well-known and current; in fact national and international surveys state that, in many countries, boys achieve better results than girls in mathematics at all school levels (Giberti, 2019); moreover, Italy is one of the countries with the largest gender gap in favour of boys (OECD, 2019). Some Italian studies confirmed that the cause of gender gap cannot be found in any biological or cognitive difference between boys and girls; instead, a variety of factors can play a role in it, such as metacognitive influences, affective factors and general biases (Giberti et al., 2016). There is the urge to put in place didactical interventions to equally involve boys and girls in mathematics, aiming to a more equitable discipline (Ferrara et al., 2021). To do so, we consider crucial to deepen the description of the phenomenon also considering students' perspective. When solving tasks, boys and girls seem to have different approaches; not only usually more boys than girls solve the tasks correctly but also, among the wrong answers, the two groups appear to prefer different distractors (Giberti et al., 2016). Hence, the aim of the paper is to discuss students' PD of tasks, highlighting whether there are any differences between boys' and girls' perception or not, even in relation to the categories of factors influencing it.

METHODOLOGY

We built a questionnaire composed by two INVALSI tasks, administered in previous years to grade 10 students, each one followed by a set of questions. The selected INVALSI tasks, represented in Figure 1, are both argumentative and dealing with algebra, but they are of a different type, in fact Task 1 is a multiple-choice question while Task 2 is an open-ended one.

Task 1

Antonio states that « $4n-1$ is always a multiple of 3». Is Antonio right? In the following table select the only argumentation that justifies the right answer

Antonio is right...	Antonio is not right...
A. <input type="checkbox"/> because $4n-1=3n$	C. <input type="checkbox"/> because $4n-1$ is always odd
B. <input type="checkbox"/> because if $n=4$ then $4n-1=15$	D. <input type="checkbox"/> because if $n=3$ then $4n-1=11$

Task 2

Marco states that, for every natural number n greater than 0, $n^2 + n + 1$ is a prime number. Is Marco right?

Choose one of the two answers and complete the sentence.

- Marco is right because
-
- Marco is not right because
-

Figure 1: Task 1 and Task 2 (original texts from www.gestinv.it, translation provided by the authors).

In particular, Task 1 asks the students to select the argumentation supporting the right answer, whereas Task 2 requires the students to construct and provide themselves an argumentation for a given statement.

The tasks selected were quantitatively analysed and used as benchmarks. From the national data, 40.3% of Italian students answer correctly to Task 1 and 17.8% respectively to Task 2, meaning that they have a different difficulty in the traditional sense of the term. Furthermore, Task 1 and 2 were afflicted by gender gap in different proportion. For each one, we calculated the value of the gender gap index GGI_k (Spagnolo & Nicchiotti, 2023), defined as follows

$$GGI_k = \begin{cases} \frac{M_k - F_k}{F_k}, & M_k > F_k \\ \frac{M_k - F_k}{M_k}, & M_k < F_k \end{cases},$$

where M_k is the ratio between the number of the correct answers to the item given by boys and the total number of answers to the item given by boys, while F_k is the equivalent for girls. The value of the index is positive when boys overperform girls, negative in the opposite case and equal to 0 when there is no gender gap. Referring to the threshold values provided, Task 1 resulted to be balanced ($GGI_k = 1.4\%$) while Task 2 has a moderate gender gap in favour of boys ($GGI_k = 16.5\%$).

In the questionnaire, students were asked to solve the tasks; each task was followed by the request for students to rate it according to their PD, on a scale from 1 to 10 (being 1 “very easy” and 10 “very difficult”) and from the request to motivate their rating. The last part of the questionnaire asked students which task was the more difficult in their opinion and why; these last questions gave us more elements to analyse their PD in general. After these preliminary analyses, we carried out a qualitative study involving 7 classes from two Italian high schools (5 grade 9 and 2 grade 10 classes) for a total of 148 students, of which 61 boys and 87 girls. The two schools were a technical scientific high school (Istituto tecnico in Italian) and an educational humanistic high school (Liceo delle Scienze Umane in Italian); the classes from the former had a predominance of boys, while from the latter had predominantly girls. The questionnaire was administered through Google Forms and students answered it during class hours using school computers. The answers were then collected and analysed qualitatively, carrying out a text analysis on students’ answers referring to the categories of PD previously discussed. In addition to that, mean values of the students’ ratings of difficulty were calculated as well as the GGI_k referring to the results of the students of the sample. All the above was done considering the gender perspective and highlighting possible similarities and differences.

RESULTS AND DISCUSSION

Firstly, quantitative analysis allowed us to select Task 1 and 2 (Figure 1), since they are meaningful from a gender gap perspective. Then, we analysed qualitatively the answers of 148 students to the two tasks (determining whether they had solved them correctly or not) and we calculated the value of the gender gap index for each task referring to the results of our sample. In contrast to the values obtained for the national data, in this case we observed for both tasks the presence of a severe gender gap in favour of males. In fact, the value of the index for Task 1 (T1) is the same as for Task 2 (T2), namely $GGI_{T1} = GGI_{T2} = 41\%$; $GGI_{T1} = GGI_{T2}$ by coincidence, but the presence of a gender gap in favour of males clearly emerges for the two tasks, considering our sample. We consider important to treat this as a characteristic of our sample when deepening the analysis of PD and the categories of factors; to be statistically significant the study should be implemented quantitatively.

Analysis of PD for Task 1

Students regarded Task 1 as moderately difficult, with the girls seeing it on average more difficult than the boys. The average difficulty assigned by the girls is, in fact, equal to 5.5 while for the boys it is 4. Therefore, this result seems consistent with the presence of a gender gap highlighted by the value of GGI_{T1} : not only girls performed more poorly on this task, but they also found it more difficult than the boys did.

The textual analysis of the reasons of the rating allowed us to outline a more precise picture of the issue. We identified and classified 171 references, meaning that some answers presented elements recalling more than one category. 39 of them were not considered in the analysis as they did not give any information, being answers like “It was difficult” or left blank. The remaining ones (59 from boys and 73 from girls) were classified into the five categories, as summarized in Table 1.

Category	Number of references (boys)	Number of references (girls)
Resolution strategy	28	35
Capabilities and experience	24	15
Emotions	2	1
Text formulation	5	10
Personal consideration	0	12

Table 1: Distribution of the references among the categories for Task 1.

For both boys and girls, the broadest category resulted *Resolution strategy* with many references to the reasoning for boys and to examples for girls. However, for boys, this category seems to have almost the same importance as the category of *Capabilities and experience*, which for girls too is the second broadest but not with a comparable importance. Hence, most students focused on the resolution of the task as the crucial

element to determine the reasons of their PD, considering primarily the objective aspects and leaving aside the more subjective side. This might be an explanation for the almost total absence of any reference to *Emotions*: students might not consider this factor when rating the difficulty of a task because they are used to leave out the subjectivity when dealing with mathematics and they might think about the emotional side only if directly asked about it. The category of *Text formulation*, for this task, appears to play a role in the PD for both boys and girls but it is not the most significant factor. Almost all the references belonging to this category mentioned the text while only a few ones referred to the question format.

The most evident difference between boys' and girls' answers is represented by the category of *Personal consideration*. As already described, it contains the answers mentioning students' consideration of themselves under different points of view: them as students, their relationship with mathematics, their reflection about their progresses in their course of study and others. For Task 1, elements recalling these aspects were mentioned only by girls. None of the boys considered this kind of factors important when determining their PD; girls instead considered them even more important than the text formulation. They reflected about their self-perception ("I do not have logic"; "I do not understand these things and I cannot do them even though I practice a lot"), their perception as mathematics students ("I am not good at maths"; "Probably I am not very good at solving these tasks") and their preparation ("I should practice more"; "I should study better").

Analysis of PD for Task 2

Like Task 1, Task 2 was considered of a medium difficulty but in general slightly more difficult than the former. The average rating given by boys is equal to 4.1 while the one from the girls is 5.7. In this case, the difference between the average ratings is as evident as before and it confirms the findings described for Task 1: girls tend, on average, to perceive tasks as more difficult than the boys do. Despite Task 2 resulting very difficult, as national data show, the task was perceived as only moderately difficult. However, we observed that in this case, students' perception was not aligned with their actual results, in fact the percentages of right answers were even lower than the national ones (11% of the students answered correctly, 9% of the girls and 13% of the boys). Students seem to lack awareness of their mistakes and they are not always coherent rating the task and explaining the reasons, almost as if they are worried about giving a too high rating to the PD of the task.

Analysing students' answers and explanations we identified 177 references, 39 of which were not meaningful. The remaining 138 (59 from boys' answers and 79 from girls' ones) were classified into the five discussed categories, as reported into Table 2.

Category	Number of references (boys)	Number of references (girls)
Resolution strategy	27	35

Capabilities and experience	28	22
Emotions	0	0
Text formulation	4	13
Personal consideration	0	9

Table 2: Distribution of the references among the categories for Task 2.

The results for Task 2 are similar to the ones obtained for Task 1, making the findings stronger. For girls, once again, the most represented category is *Resolution strategy* with references to examples needed to solve the task, reasonings that can be put in place and calculations. In particular, this last element was one of the biggest concerns for girls, as they considered difficult dealing with the powers. For boys too this category is one of the most important ones, as almost all the references are divided evenly between *Resolution strategy* and *Capabilities and experience*. In the former, boys did not refer to calculations as much as girls did; they instead focused on the examples and the reasoning. The reasoning in particular has both a positive and a negative connotation for boys, because some of them stated that “more reasoning was needed to solve the task” while others wrote that “the task required not a lot of reasoning”, being this last one the prevalent thought. Regarding the category of *Capabilities and experience*, it is well represented both for boys and girls, being the most important one for the former. Boys mentioned very often in their answers their confidence in the answer provided and the absence of obstacles encountered, as a reason for considering the item easy or very easy. From the answers belonging to this category, we can observe the absence of awareness of mistakes mentioned before, especially from boys. The girls’ answers from this category contain many references to obstacles but considered only in the negative sense. In other words, girls considered the item more difficult because they dealt with many different obstacles solving it, and they seem more aware of this aspect. The other factor considered by them is the previous experience with similar tasks, because girls are quite honest about the fact that they consider a task more difficult if they never solved something similar. Once again, we could not find any reference belonging to the category *Emotions*. For Task 2 as well as for Task 1, the category *Text formulation* results not to be the most important one. In this case, boys’ and girls’ answers are similar, considering confusing above all the presence of “many letters in the text” of which they do not know the values. Many students affirmed that the task would have been easier if in the text there had been some “example of numbers to substitute to n”.

Finally, the answers belonging to the category of *Personal consideration* are again the element that mainly differentiates boys’ and girls’ answers. In fact, among boys’ answers, also for Task 2 there were not references to this category, while girls mentioned elements related to it. In this case, the category is slightly smaller than in Task 1, but the previously described characteristics are preserved. Girls seem more conscious about their mistakes and difficulties, sometimes even making general and

hard remarks about themselves, which do not seem to descend directly from the task itself (“I cannot do it, I have basic gaps”, “I am ignorant”). An interesting aspect to evaluate is the fact that the girls whose answers can be classified into this category for Task 2, referred to this category also in their answers regarding Task 1. Moreover, in most cases, they considered the two tasks equally difficult and they attributed very high ratings to them. This evidence might suggest that personal consideration and PD are intertwined for girls and a negative personal consideration could be both a cause and an effect of a high PD of a task.

CONCLUDING REMARKS

The study presented in this paper gives a first insight into the PD and the factors characterizing it, comparing boys’ and girls’ perspectives and analysing them on the basis of the issue of gender gap in mathematics. The analysis of students’ ratings and answers allowed us to state that there seem to be some differences between boys’ and girls’ PD. Boys tend to evaluate mathematical tasks as easier than girls do, even regardless of their actual performance solving them.

The categorization of the references into the five sets allowed us firstly to confirm their usefulness to describe the factors influencing students’ PD. Moreover, we observed that the important categories to determine PD seem to be almost the same for boys and girls. The main difference is represented by the connotation they give to the elements: the same aspect, in fact, is ambivalent for boys, being both positive and negative, and usually negative for girls. Namely, when motivating their PD, girls tend to be very severe highlighting all the aspects that make a task more (or less) difficult, but they almost never consider the factors making it easy. This implies that the same element such as the text of the task, might result in a facilitating element for boys and a distractor for girls. However, the biggest difference between boys and girls regarding the categories is the complete absence for boys only of any reference to the *Personal consideration*. It is one of the categories involving subjectivity, which instead appears to be discarded by boys when elaborating about their PD. On the other hand, girls always make reference to it and although it is not the most important category for them either, it is always considered. Element pertaining to it are expressed especially by the girls that had difficulties or were not able to solve the tasks. They appear to have a very low personal consideration and make hard judgement about themselves not only as students but as people in general. This led us to hypothesize the existence of a cause-effect link between PD and personal consideration for girls that works both ways. More studies are needed to confirm this, but it is possible that girls’ low personal consideration makes them perceive mathematical tasks as more difficult, and it seems also reasonable that perceiving mathematical tasks as very difficult leads them to build a bad self-opinion resulting in low personal consideration.

The validity of these findings could be strengthened considering expanding the study working with a bigger sample and more tasks, of different types and regarding other topics than algebra. This is important with the prospect of understanding the reasons

behind the existence of a gender gap in mathematics and, above all, to put in place concrete actions to fill the gap and make mathematics fairer.

References

- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61. Art 101859. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Efklides, A., & Touroutoglou, A. (2010). Cognitive interruption as an object of metacognitive monitoring: Feeling of difficulty and surprise. In: A. E. & P. Misailidi (Ed.), *Trends and prospects in metacognition research* (pp. 171–208). New York, NY: Springer. https://doi.org/10.1007/978-1-4419-6546-2_6
- Ferrara, F., Ferrari, G., Robutti, O., Contini, D. & Di Tommaso, M.L. (2021). When gender matters: A study of gender differences in mathematics. In Inprasitha, M., Changsri, N., & Boonsena, N. (Eds), *Proc. 44th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 2, pp. 255-263). PME.
- Giberti, C., Zivelonghi, A. & Bolondi, G. (2016). Gender differences and didactic contract: Analysis of two INVALSI tasks on power properties. In Csíkos, C., Rausch, A., & Szitányi, J. (Eds), *Proc. 40th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 2, pp. 275-282). PME.
- Giberti, C. (2019). Differenze di genere in matematica: dagli studi internazionali alla situazione italiana. *Didattica della matematica. Dalla ricerca alle pratiche d'aula*, 44-69.
- Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology*. Fort Worth: Holt, Rinehart & Winston.
- OECD. (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. Paris: OECD Publishing. <https://doi.org/10.1787/b5fd1b8f-en>
- Spagnolo, C., Capone, R., & Gambini, A. (2021). Where do students focus their attention on solving mathematical tasks? An eye tracker explorative study. In Inprasitha, M., Changsri, N., & Boonsena, N. (Eds), *Proc. 44th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 4, pp. 84-92). PME.
- Spagnolo, C. & Nicchiotti, B. (2023). Interpreting gender gap issues in standardized tests: definition and application of a theoretical tool. *Frontiers in Education*, 8. <https://doi.org/10.3389/educ.2023.1303041>
- Spagnolo, C., & Saccoletto, S. (2023). How students view the difficulty of mathematical tasks: factors that influence their perceptions. In P. Drijvers, C. Csapodi, H. Palmér, K. Gosztonyi, & E. Kónya (Eds.), *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)* (pp. 1498–1506). Alfréd Rényi Institute of Mathematics and ERME.
- Zan, R., Brown, L., Evans, J., & Hannula, M. (2006). Affect in mathematics education: an introduction. *Educational studies in mathematics*, 63(2), 113–121. <https://doi.org/10.1007/s10649-006-9028-2>