The Properties of Powers: Didactic Contract and Gender Gap

Federica Ferretti[1] · Chiara Giberti[2]

© The Author(s) 2020

Abstract

National and international large-scale assessments of mathematics show that, in most nations, males achieve better results than females and Italy is one of the countries with a greater gap. Many research studies in mathematics education have analysed this issue, using both quantitative and qualitative methods to understand the sources and characteristics of this gap. This study focuses on a specific Grade 10 task that requires algebraic manipulations of powers with the same base. Item-level analysis enables the study of gender differences on specific content, before using the lenses of mathematics education theories to interpret macro-phenomena emergeing from standardized assessment results. The quantitative analysis, carried out using the Rasch statistical model, highlights a gender gap in favour of males in this task and, furthermore, a variance in choices of incorrect options between males and females; the interviews conducted provide a key to understanding this phenomenon in terms of didactic contract.

Keywords Didactic contract · Gender gap · Large-scale assessment

Rationale. This research fits into a socio-constructivist perspective, according to which the student actively builds his/her own knowledge, interacting with the environment and organizing his/her mental constructs; the student does not simply receive information in a passive manner, but constantly re-elaborates each proposal independently. Social interaction plays a fundamental role within our study, framed by the didactic contract construct as defined by Brousseau (1988) and its features and results based on gender. Research (i.e. Lindberg, Hyde, Petersen & Linn, 2010; Primi, Busdraghi,

✉ Federica Ferretti
   federica.ferretti@unibz.it

1 Department of Education, University of Bologna, Bologna, BO, Italy

2 Department of Human and Social Sciences, University of Bergamo, Bergamo, Italy

Tomasetto, Morsanyi & Chiesi, 2014) highlights that the gender gap is often affected by meta-cognitive aspects and that its characterization depends on factors not exclusively within the sphere of cognition.

Specific studies (i.e. Cascella, Giberti, Bolondi, 2020; Leder, 1992; Leder & Forgasz, 2008; Organization for Economic Cooperation and Development [OECD], 2016a) highlight how variables linked to teaching practice, such as curriculum variables and assessment methods, might affect the gender gap in mathematics. One of the aims of this research study is to investigate, quantify and characterize macro-phenomena related to a gender gap which emerges from standardized assessments and may be framed within the context of the didactic contract. This research is carried out in accordance with the concept that standardized assessment results may be useful both for research and teachers, and can provide timely and efficient information to positively affect the learning/teaching process. While the combined analysis of texts and results makes it possible to identify new macro-phenomena (Bolondi, Ferretti, & Giberti, 2018; Ferretti, Giberti, & Lemmo, 2018), at the same time the sizeable amount of data provided by large-scale assessment also allows the quantification of existing phenomena and the definition of further characterisations (Ferretti, & Bolondi, 2019). Specifically, our research is based on data collected by the National Evaluation Institute for the School System (INVALSI). INVALSI is an Italian research institution that carries out annual surveys of all Italian school stagess - in primary school grades 2 and 5, in middle school grade 8, in high school grades 10 and 13 (since SY 2018–2019) – in Mathematics, Italian, and English (since SY 2017–2018). The results of a national sample are released annually and made freely accessible for public use, as are the test items. The framework adopted by INVALSI assessment tests is strictly connected to the National Guidelines, includes aspects of mathematical modelling followed in the PISA research project (Arzarello, Garuti, & Ricci, 2015), and is developed according to the findings of mathematics education research.

In our research study, we analyse data using the Rasch Model, following the example of INVALSI data analysis (INVALSI, 2017). The Rasch Model belongs to Item Response Theory and is a simple logistic model (Barbaranelli & Natali, 2011; Bond & Fox, 2001; Rasch, 1960). It allows us to analyse an entire test by estimating an ability parameter for each student and a difficulty parameter for each item (delta). Specifically, in this paper we consider a macro-phenomenon that emerged from the mathematics grade 10 INVALSI test in the school year 2012 with a mixed method analysis (Hart, Smith, Swars, & Smith, 2009). A quantitative analysis carried out on an INVALSI task featuring operations with powers highlights a gender gap that we assume can be linked to the didactic contract as defined by Brousseau. Subsequent structured interviews allow us to confirm our hypothesis and characterize the phenomenon studied.

Theoretical Framework

Didactic Contract

This research fits into a constructivist perspective in which learning processes cannot be separated from the context in which they are developed, a context which is defined by three main actors: the student, the teacher and the knowledge (Chevallard, 1985). We will

use this perspective to interpret the phenomena highlighted by the quantitative analysis and, in particular, we will investigate the causes of this phenomena using the didactic contract construct defined by Brousseau (Brousseau, 1997; Education Committee of the EMS [EMS-EC], 2012) as the main interpretative key. In line with Brousseau's definition (1980), we consider the didactic contract as the pupil's expectations of teacher behaviour and the teacher's expectations of student behaviour; this construct was introduced to interpret possible reasons for the specific failure in mathematics of some students and it seems very appropriate for interpreting the phenomena we investigated.

The didactic contract is one of the solid findings of mathematics education (EMS-EC, 2012) and it has significant international relevance (Sarrazy, 1995); indeed, this construct is useful for interpreting several classroom situations and, thus, several difficulties encountered by students (Brousseau, 1997).

The socio-cognitive dimension is fundamental for the definition and evolution of this notion (Sarrazy, 1995): the didactic contract regulates classroom activities, influencing the behaviour of the teacher and students, also regarding acquired knowledge. Indeed, the didactic contract has a strong influence on students' learning processes, and the effect of this in mathematics activities cannot be underestimated. The didactic contract imposes rules of behaviour but these rules and norms are often implicit, meaning that they can in fact become obstacles in the teaching/learning processes. Although it is possible that these rules come from didactic practices and may also be useful for students in some situations, difficulties connected to the didactic contract are often due to the application of these rules in a new context where they are not appropriate.

In several studies, the didactic contract was used to interpret phenomena in primary school but some studies have highlighted that it emerges also in higher scholastic grades (D'Amore, Fandiño, Marazzani & Sarrazy, 2010; De Vleeschouwer & Gueudet, 2011) and that it is possible to observe its existence and effects not only through qualitative interviews but also in large-scale assessment processes (Ferretti, & Bolondi, 2019).

In this paper, we will use the didactic contract as a theoretical lens through which we interpret students' answers to a specific task taken from a large-scale assessment test at grade 10. The quantitative analysis of the task allows us to analyse students' answers also as a function of their ability across the whole test, focusing on differences between males and females. The didactic contract is here used to interpret the phenomena highlighted by the quantitative analysis, and the interviews confirm the accuracy of this interpretation. In detail, we analysed the students' approach to solving the proposed task, investigating the presence of those behaviours attributable to contractual rules which are created by the students. These implicit didactic contract rules are not imposed by teachers, but are created by students over time in response to recurrences that have generated standard problem models. These norms are so firmly established that, as already highlighted in the literature (Ferretti, & Bolondi, 2019), they also emerge during standardized assessment situations.

Gender Gap in Mathematics

The strong disparity between mathematical results of males and females (in favour of males), is addressed in the international literature and highlighted by standardized assessment results (Mullis, Martin, Foy, & Hooper, 2016). International mathematics standardized assessment systems, OECD-PISA and IEA-TIMSS, confirm the trend both in primary and secondary schools (OECD, 2016a).

Many studies in recent years (i.e. Leder & Forgasz, 2008; Forgasz et al., 2010; Leder & Forgasz, 2008) have focused on investigating the causes behind the better results of male students in mathematics; this growing interest in the subject can also be linked to widespread evidence emerging in standardized assessment tests. And it is thanks to the data and comparisons of results in different countries that it is possible to investigate the characteristics and causes of the gender gap. From the data of international standardized assessment tests, it emerges that the gap between test results of males and females is not uniform across all the school systems investigated; in most countries, the gap is in favour of males, but there are exceptions where it is in favour of women (OECD, 2016a). This fact supports the theory that causes of a biological and physiological nature, considered in some studies (e.g. Baron-Cohen & Wheelwright, 2004), do not appear to be a predominant factor in the emergence of the gap.

If the causes were predominantly biological in nature, the differences should be more or less the same in all countries, a hypothesis which is contradicted by the results of the OECD-PISA and IEA-TIMSS tests (Hill, Corbett, & Rose, 2010). The non-homogeneity in the distribution of gender gap also supports the numerous studies that uphold the strong impact of social and cultural factors (e.g. Guiso, Monte, Sapienza & Zingales, 2008). For example, several studies have shown that the gender gap in mathematics is closely linked to the emancipation of women in society; in societies where gender equality is reached, this gap tends to disappear (i.e. Guiso, Monte, Sapienza, & Zingales, 2008; González de San Román & De La Rica, 2012).

The performance gap is also strongly influenced by the beliefs and attitudes of teachers, parents and students towards mathematics, which are often dictated also by gender stereotypes (eg Jacobs & Bleeker, 2004; Tomasetto, 2013) and by metacognitive factors closely linked to mathematics (Herbert & Stipek, 2005; Hill et al., 2010; OECD, 2015). A lower level of self-efficacy and self-confidence in mathematics (OECD, 2015) may lead girls to be more afraid of making mistakes, thus preferring well-known resolution strategies already adopted in the classroom rather than trying out new strategies in solving problems (e.g. Bell & Norwood, 2007). Finally, factors closely related to the classroom context such as the curriculum, teaching and evaluation practices may have a different impact on males and females (Leder, 1992; Leder & Forgasz, 2008). Even micro-social factors that arise in the classroom context can lie at the basis of gender differences in mathematics: a greater bond between girls with teaching practices and the teacher can, for example, lead girls to be more influenced by misconceptions and the didactic contract (Bolondi, Cascella, & Giberti, 2017).

Italy displays one of the most marked gender gaps (Mullis et al., 2016; OECD, 2016b; Gilberti, 2019). The clear mathematical gap between males and females in Italy is also confirmed in the various school grades by the results of the national standardized assessment INVALSI surveys.

Research Questions

In this paper we employed the assumption that didactic contract can have a stronger influence on females than on males, focusing particularly on its effect connected to a specific mathematics task.

The links between didactic contract and gender differences are varied and support this hypothesis. Firstly, the relevance of social and cultural factors in determining gender difference in mathematics can be transferred also at a classroom level, where micro social factors (strictly related to the milieu habits, classroom practices and teacher-student relationship) have a different influence on male and female behaviour, both in classroom work and in assessment situations (e.g. Bell & Norwood, 2007; Bolondi et al., 2018). Moreover, the relationship with the teacher and with the discipline itself is another important connection between these two phenomena: the literature shows the importance of metacognitive factors such as math anxiety and math self-efficacy (e.g. Hill et al., 2010) as well as the relevance of beliefs of teachers on students' performances and, also in this case, we observe that these factors disadvantage females more than males in learning mathematics. Then, phenomena such as the didactic contract (emerging within classroom habits and related to implicit and explicit rules set by the teacher) have a stronger influence on females, and may constitute an obstacle to learning processes. This is also observed in this specific task in which the gender gap in favour of boys might be explained by the influence of didactic contract but also to the girls' reluctance to eschew didactic practices and rules established by classroom habits (Bolondi et al., 2018; Giberti, & Ferretti, 2019).

Finally, in recent studies we observed that the didactic contract is particularly influential on students with medium and medium-high ability levels (Ferretti et al., 2018; Giberti, & Ferretti, 2019); this is in line with previous literature and the definition itself of didactic contract i.e. arising from rules and norms of the classroom and teacher habits, influencing mostly those students who follow the teacher's lessons and his/her instructions but are not able to discern correctly when a rule should be applied or not. On the other hand, literature on gender differences highlight that the gap is greater for students with high ability levels; furthermore, the gender gap is almost null at the beginning of schooling but increases over the years (OECD, 2016b), thus confirming the relevance of scholastic factors.

Our assumption, also based on previous research (Giberti, & Ferretti, 2019), is that a lower level of self-efficacy and self-concept in mathematics could lead girls to be more inclined to take "safe refuge" in didactic routines and practices, and be more prone to follow what the teacher says (or does not say, in the case of implicit rules). The expectations which, explicitly or not, arise from the classroom environment constitute this "safe refuge" but can also become a limitation on students' autonomy and an obstacle to learning, as other studies on the didactic contract show (Ferretti, 2020).

In this work, we investigate gender differences in answering a specific task involving an operation with powers, interpreting this phenomena in terms of didactic contract: girls, when having to operate with powers of the same base and high exponents, are more driven by didactic practices to use power properties even though it is not appropriate for the case at hand. In the first part of our analysis, the quantitative tools highlight gender differences (in favour of males) in a specific task belonging to a grade 10 INVALSI math test. In the second part. we will interpret statistical evidence via a qualitative analysis based on interviews. The first research questions are related to the quantitative analysis:

(1) does the quantitative analysis performed via the Rasch Model reveal a gender gap in terms of ability level of the students?

(2) can the gender gap observed regarding the correct answer be due to differences in the selection of wrong options for males and females? In other words: Are males and females attracted by the same incorrect answers?

At first glance, the postulated interpretation of the task highlights that difficulties of students nationwide might be related to didactic contract. The interviews performed in the second part of our work are used to investigate if, and how, the behaviour of students in this task has been influenced by didactic contract, before answering the following questions:

3 Do interviews confirm that the difficulties encountered by students in this task are due to the didactic contract?
4 Are reasons and type of error linked to a specific difficulty and thus to a specific wrong option, which differs between males and females?

Research Methods

In this research we adopted a mixed method (Johnson & Onwuegbuzie, 2004) approach based on the integration of two different analyses.

The first part of our research is based on the results of standardized assessment and consists in a quantitative analysis. Using the Rasch model and Differential Item Functioning (DIF) analysis on a large national sample of students (more than 40,000 students) we observe macro-phenomena on a specific item of the test, which can be linked to solid findings of mathematics education theories (EMS-EC, 2012). In particular, the quantitative analysis allows us to highlight students' behaviour and students' answers as indicative of their ability on the whole test and, furthermore, it gives us the possibility also to characterize gender differences on the basis of student ability. In this way we can make conjectures concerning the specific difficulties encountered by students in a task and the cognitive processes linked to each possible answer, also considering differences between male and female responses. These conjectures are then investigated via a qualitative analysis based on interviews of students of the same grade, with the aim of understanding the cognitive processes adopted by the students and then the reasons for the emergence of gender gap in this specific task.

The structure of our mixed methodology, in accordance with Johnson and Onwuegbuzie (2004), is thus QUAN→QUAL, with the integration of two different steps of equal importance but with different goals: highlighting the macro-phenomena on a large sample representative of the population, and then interpreting this phenomena by interviewing students.

Quantitative Analysis

The task analysed in this paper belongs to an INVALSI test administered in 2012 at grade 10. The test comprised 45 items covering the entire content domain (*Space and Shapes, Data and Uncertainty, Numbers, Relation and Functions*) and was constructed and analysed by the INVALSI team. We consider the results of the INVALSI sample, which is composed of 41,812 students selected to be

representative of the whole population nationwide. Moreover, using INVALSI sample data guarantees the regularity of the administration process and data entry as it is performed by INVALSI experts. A more detailed description of the construction and representativeness of the INVALSI sample is reported in the technical report (INVALSI, 2012) and based on the work of the INVALSI statistical team. Also, the analysis of the validity and internal consistency of the whole grade 10 INVALSI test in 2012 is reported in the technical report; particularly, the Cronbach's Alpha is very acceptable, equal to 0.89.

Our quantitative analysis is based on the Rasch model (Rasch, 1960) which is a logistic model often used to analyse tests and adopted also by the INVALSI statistical team for the main analysis of results. The Rasch model belongs to Item Response Theory and is useful for analysis both at test and item level as it allows a joint estimate of the difficulty of the items and of the ability of each student by placing them on the same scale from −4 to +4. Student ability is measured across the whole test and is directly comparable with item difficulty: for instance, a student with an ability parameter equal to 0.8 has a 50% probability of answering correctly an item with a difficulty index of 0.8, while the same student has a probability higher than 50% of answering correctly an item with a difficulty parameter lower than 0.8. The Item Characteristic Curve (ICC), output of the Rasch Model, expresses the probability of responding correctly to a specific item depending on the difficulty of the item itself and the ability of the respondent, measured across the entire test. The ICC-plot (see, for example, Fig. 1 and 2) reports students' ability on the x-axes and the probability of choosing the correct answer on the y-axes. Together with the ICC, we can also represent empirical data relative to the same item, reporting for each
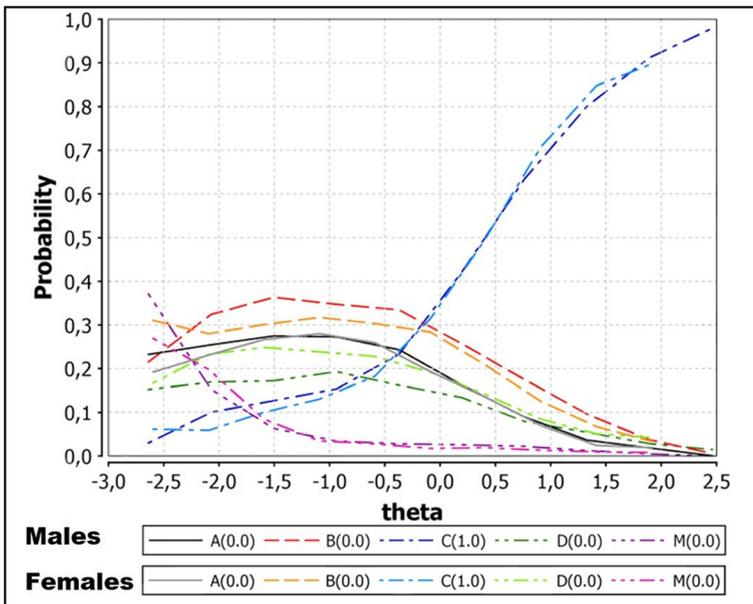

Figure 1   Results of item D21, national sample INVALSI data

decile of the population (grouped on the basis of their ability across the whole test) the percentage of students who choose a specific answer to the task (correct and wrong options). These graphs are named Distractor Plots and are particularly important in analysing the fit of the item with the model: comparison of the ICC and the empirical line of the correct answer gives information regarding the way the theoretical curve of the item represents the empirical data. Moreover, distractor plots are also useful for analysing the trend of incorrect answers as a function of students' ability (observing the empirical lines of the incorrect answers).

In this paper we used JMetrik 4.0 software to perform a Rasch analysis of the INVALSI test administered at grade 10 in 2012. We based our quantitative analysis on this model because it allows us to:

- place all the students of the sample on the same scale according to their ability measured over the entire test
- produce the distractor plot of the task which is studied in this paper
- produce the distractor plot dividing males and females.

The plot results of this procedure allow the comparison of male and female answers as a function of their ability measured over the entire test. The x-axes of the plot report the Rasch parameter for students' ability (measured using the Rasch model for the whole sample, to place male and females on the same ability scale). The sample was then divided by gender and, the two groups were further divided into deciles according to their ability. For each male and female decile, the plot reports the percentage of students choosing each option, which gives us the opportunity to observe the empirical trends of male and female answers with a direct comparison.

In particular, this last point is crucial in our analysis because it allows us to compare answers given by males and females on the basis of their ability and highlight differences not only in choosing the correct answer, but also in choosing incorrect options.

Furthermore, using the same software, we also performed a Differential Item Functioning (DIF) analysis in order to compare the performances of the two subgroups (focal group: males; reference group: females) on each item of the test:

> Differential item functioning (DIF) occurs when one group of examinees has a different expected item score than comparable examinees from another group. It indicates that an item is measuring something beyond the intended construct and is contributing to construct irrelevant variance. (Meyer, 2014, p. 69)

INVALSI tests are built to be DIF-free, because the presence of a significant DIF is a threat to the validity of the test and DIF items interfere with the quality of the measured trait.

We performed the Cochran-Mantel-Haenszel (CMH) statistic for testing statistical significance provided by jMetrik, This procedure tests the null hypothesis: item scores are conditionally independent of group membership (in our case, gender). The CMH procedure, might be influenced by the sample size (Meyer, 2014), so the common-odds ratio is used to describe practical significance: no DIF is detected if this index is 1, while values larger/smaller than 1 indicate an item favouring the reference/focal group
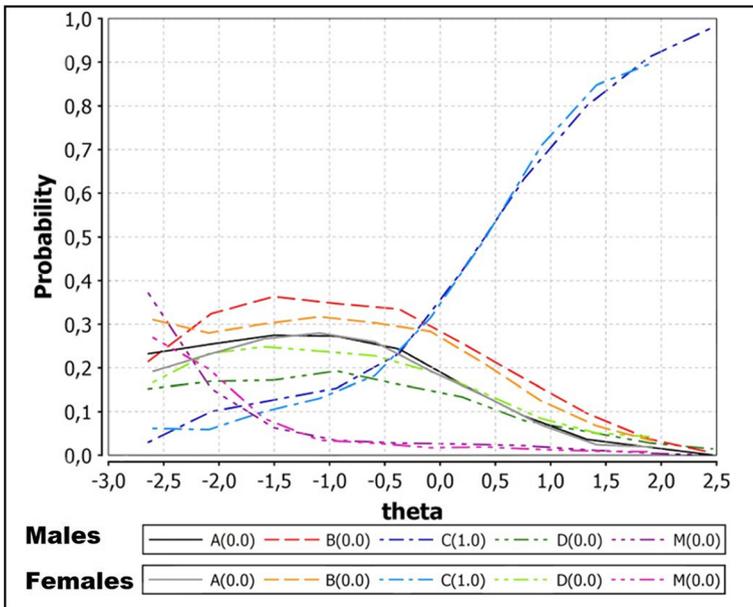
Fig. 2 Distractor Plot by gender

respectively. Based on CMH and common odds ratio results, all the items have been classified, following the criteria suggested by Zwick and Ercikan (1989)[1]: "A" if the DIF is negligible, "B" if the DIF is moderate and "C" if they exhibit a high value of DIF and thus potential problems for the validity of the test.

We performed this analysis for the whole test, considering both the row score and the Rasch parameter (theta) as matching variables; we thus compared male (reference) and female (focal) answers for each item of the test.

Qualitative Analysis

To confirm the quantitative analysis, we interviewed students of the same age as those who had tackled the same INVALSI task years earlier. Already in 1980, Ericsson and Simon (1980) supported the need to move beyond a quantitative approach in order to interpret the situations and so we performed a qualitative analysis from a cognitive and meta-cognitive perspective. We interviewed 18 students attending 6 classes of upper secondary school (with different orientations): 7 Technical Institute students and 11 Liceo (lyceum) students. Teachers were asked to indicate two or three students per class

---

[1] The rules for this classification, in terms of the common odds ratio, are as follows (Meyer, 2014; Zwick & Ercikan, 1989):

- "A" items have (a) a CMH p value greater than 0.05, or (b) the common odds ratio is strictly between 0.65 and 1.53.
- "B" items are neither "A" nor "C" items.
- "C" items have (a) a common odds ratio less than 0.53, and the upper bound of the 95% confidence interval for the common odds ratio is less than 0.65, or (b) a common odds ratio greater than 1.89 and the lower bound of the 95% confidence interval is greater than 1.53.

with an "average" performance compared to the class trend. Task-based structured interviews were conducted (Goldin, 2000), which focused on how the pupils tackled the mathematical activity. The interviews took place individually and were recorded and transcribed with the students' consent - it seemed that the presence of the voice recorder did not affect the conducting of the interview in any way. The structure of the interviews was planned in advance by the three interviewers; the students were given a questionnaire and asked to complete it. At the beginning of the interview, the students were asked to enter a survey ID name and were asked to carry out the tasks (Attachment 1) either aloud or in silence, explaining their strategies later. Most of the students opted for the second option, first carrying out all the tasks and then explaining aloud the reasoning and strategies implemented. Investigating the effects of didactic contract, in line with the theoretical framework (Brousseau, 1988), we tried to create a relaxed environment in all interviews. This made it possible to create conditions suitable for research; in fact, many aspects of the didactic contract are inherent to the difficulty of explicit inner processes and, very often, external factors are decisive. Moreover, as Goldin (2000) recommends, we tried to take precautions with regard to possible unforeseen events and, being willing to temporarily alter the planned structures of the interviews, we identified situations that led to useful observations for the purposes of research. As the author emphasizes, at the time of the interview, attention was focused on the process that the subject undertakes to arrive at the answer rather than the correctness or otherwise of the mathematical task (ibidem); and it is precisely this fact that creates the possibility of investigating important issues in greater depth than allowed by other experimental means. Specifically, in the context of this research, this has been crucial in eliciting feelings and beliefs typical of the cases between students who are explicit or implicit in their strategies.

Results

Quantitative Results

The results of the first part of the quantitative analysis reported in the table below (Fig. 1) derive from the statistical analysis made by INVALSI using the software Acer Conquest, which merges the results of the Rasch Analysis (Item delta and Distractor Plots) with the main index of Classical Test Theory (Discrimination and Weighted MNSQ).

The item shows acceptable psychometric features (Barbaranelli & Natali, 2011; INVALSI, 2012). The percentage of correct answers is 35%, which is low if we consider the fact that this is a multiple-choice item with four options. The difficulty of the item is also confirmed by the item delta which is equal to 0.76. Moreover, we observe a strong difference between males and females in choosing the correct answer: 38% of males answered correctly as opposed to 31% of females. Option B was the most attractive to both males and females and there was no significant gender difference for option A. A large part of the gap observed is due to option D, which was chosen by 14% of males as opposed to 19% of females. Almost all the students answered this question and the percentage of missing answers was only 3%.

The distractor plot allows us to compare the students' choices on the basis of their ability level measured over the entire test (on the x-axes): the dotted lines represent the

empirical trend of each possible answer (percentage of students for each decile who chose a specific option), while the solid line is the ICC output of the Rasch model. The comparison of the empirical line of the correct answer with the ICC curve highlights a little over-discrimination: the model overestimates the probability of choosing the correct answer for students with medium-low ability levels and underestimates it for higher ability levels.

Furthermore, it is interesting to observe the trend of the incorrect options: they are all constant for low and medium ability levels, while they show a decreasing trend only from medium ability levels upwards. This means that students with low and medium-low ability display the same behaviour when tackling this task and their choices are similar in terms of percentage. Our elaboration of INVALSI data using JMetrik 4.0 software allows us to make a comparison of the distractor plot of males' answers and that of females, and compare them on the same ability scale created with the Rasch model (Fig. 1). It is interesting to note that the differences between males and females observed in terms of percentage are not constant in all ability levels. First, the trend of the correct answer is almost the same for all ability levels; we observe a difference in favour of males only for low ability levels. Indeed, we observe that the difference in the correct answer is almost certainly due to the fact that males reach higher ability levels, and the higher decile of males and females are at a different ability level. This factor also influences the trend of the other options: the male lines reach higher levels, while female deciles are more concentrated on medium values of the ability trait. Our analysis shows also that there is no significant difference in option A, while option B is preferred by males and option D is preferred by females, especially for medium and low ability levels.

Finally, we present the results of DIF analysis (Table 1). INVALSI tests are constructed to be DIF-free because the emergence of a significant DIF value in many items of the test would compromise the validity of the test itself. Table 1 demonstrates that all the items except three are classified as DIF-free (class A). There are only three items demonstrating moderate DIF, two in favour of females (B-) and one in favour of males (B+). Finally, as expected, no item highlights a large value of DIF (class C) - the presence of this kind of item would have affected the validity of the test.

The item used in our analysis (D21) is classified as A and thus presents a negligible DIF value. This is in line with our expectation because the comparison of distractor plots for males and females showed that the difference in percentages of correct answers was almost always due to the fact that males reach higher ability levels, as opposed to a different trend in the empirical curves. The percentage of males and females selecting the correct answer is the same for a given ability level, and thus no DIF was detected. The main results of this research are not based on differences in the correct answer as opposed to differences in choosing the other (incorrect) options.

Interview Analysis

To investigate the features of the phenomena from a qualitative point of view, students were given a questionnaire to complete. This included task D21 from the INVALSI 2011 test, Grade 10. In line with the theoretical framework and in order to create conditions as far removed as possible from the didactic contract, questions were chosen that involve basic skills as it is assumed that high school students solve them without

Table 1 Results of DIF analysis. Matching Variable: theta. Focal Group Code: M. Reference Group Code: F

Item	Chi-square	p value	Valid N	E.S. (95% C.I.)	Class
m1_a	6,50	0,01	41,625	1,11 (1,02, 1,20)	A
m1_b	288,04	0,00	41,754	0,69 (0,66, 0,72)	A
m1_c	266,29	0,00	41,764	0,70 (0,67, 0,73)	A
m2_a	14,76	0,00	41,625	1,11 (1,05, 1,17)	A
m2_b	12,47	0,00	41,111	1,12 (1,05, 1,19)	A
m2_c	1,20	0,27	41,720	0,97 (0,91, 1,03)	A
m2_d	24,11	0,00	41,509	1,13 (1,08, 1,18)	A
m2_e	19,89	0,00	40,998	0,85 (0,80, 0,91)	A
m3	2,35	0,13	41,772	1,04 (0,99, 1,09)	A
m4_a	15,84	0,00	41,338	1,11 (1,05, 1,16)	A
m4_b	114,37	0,00	41,687	1,31 (1,25, 1,38)	A
m4_c	83,21	0,00	41,687	0,81 (0,77, 0,85)	A
m5	81,10	0,00	41,747	0,82 (0,78, 0,85)	A
m6_a	244,07	0,00	41,747	1,56 (1,48, 1,65)	B-
m6_b	12,57	0,00	41,711	1,08 (1,04, 1,13)	A
m6_c	8,62	0,00	41,395	1,10 (1,03, 1,17)	A
m7_a	14,60	0,00	41,748	0,92 (0,88, 0,96)	A
m7_b	16,97	0,00	41,688	1,10 (1,05, 1,16)	A
m8	18,61	0,00	41,560	1,11 (1,06, 1,16)	A
m9_a	56,95	0,00	41,747	1,16 (1,12, 1,21)	A
m9_b	130,69	0,00	41,754	0,78 (0,75, 0,81)	A
m9_c	0,31	0,58	41,593	1,01 (0,97, 1,06)	A
m9_d	1,34	0,25	41,772	1,02 (0,98, 1,07)	A
m10_a	88,13	0,00	41,738	1,23 (1,18, 1,28)	A
m10_b	10,76	0,00	41,638	0,93 (0,89, 0,97)	A
m11_a	77,61	0,00	41,359	1,32 (1,24, 1,41)	A
m11_b	69,08	0,00	41,574	1,25 (1,19, 1,32)	A
m11_c	234,05	0,00	41,239	1,54 (1,46, 1,63)	B-
m12	35,73	0,00	41,550	1,22 (1,14, 1,30)	A
m13	57,74	0,00	41,748	1,19 (1,14, 1,25)	A
m14_a	56,23	0,00	41,761	0,83 (0,79, 0,87)	A
m14_b	20,66	0,00	41,737	1,12 (1,07, 1,18)	A
m14_c	14,97	0,00	41,550	0,90 (0,85, 0,95)	A
m15	185,99	0,00	41,737	0,73 (0,70, 0,77)	A
m16	0,66	0,42	41,728	0,98 (0,93, 1,03)	A
m17	34,58	0,00	41,504	1,20 (1,13, 1,28)	A
m18	12,54	0,00	41,198	1,09 (1,04, 1,14)	A
m19	71,40	0,00	41,782	0,82 (0,78, 0,86)	A
m20	0,81	0,37	41,296	0,98 (0,93, 1,03)	A
m21	**0,20**	**0,66**	**41,553**	**0,99 (0,94, 1,04)**	**A**
m22	20,82	0,00	41,771	0,91 (0,87, 0,95)	A
m23_a	32,91	0,00	41,633	1,15 (1,10, 1,20)	A

Table 1 (continued)

Item	Chi-square	p value	Valid N	E.S. (95% C.I.)	Class
m23_b	205,22	0,00	41,525	0,66 (0,62, 0,70)	A
m24	69,07	0,00	41,121	0,80 (0,75, 0,84)	A
m25	337,74	0,00	41,728	0,62 (0,59, 0,65)	B+
m26_a	235,56	0,00	41,666	1,42 (1,36, 1,48)	A
m26_b	132,75	0,00	41,529	1,29 (1,24, 1,35)	A
m26_c	41,00	0,00	41,687	1,15 (1,10, 1,20)	A
m27	43,86	0,00	41,560	0,86 (0,83, 0,90)	A
m28	0,56	0,46	41,771	1,02 (0,97, 1,06)	A
m29_a	0,07	0,79	41,669	1,01 (0,96, 1,05)	A
m29_b	0,27	0,60	41,772	1,01 (0,97, 1,05)	A
m29_c	4,16	0,04	41,772	0,96 (0,92, 1,00)	A
m30	36,90	0,00	41,728	0,86 (0,82, 0,90)	A

difficulty, thus putting them at ease. Students who correctly answered task 3 (the INVALSI task, object of the investigation) were told about the general failure nation-wide among their peers and then asked to suggest reasons for the phenomenon. Meanwhile, with the students who provided the wrong answer, we tried to explore in depth the reasons for the difficulties encountered. Below are some significant extracts from interviews, in which some features of the phenomena under investigation are highlighted. Out of the 18 students interviewed, regarding task 3, 8 students provided the correct answer and 10 gave the wrong answer. The first two tasks of the question-naire were actually simple - in fact, all the students interviewed provided the correct answer. Although it is not possible to generalize on such a small sample, we observed that the performances are in line with the national sample while another relevant fact is that all the females interviewed who made a mistake chose option D. Only the extracts deemed most significant from the point of view of the research hypotheses are shown below (ST = Student, F = female, M = male).

Two students (ST_F_01 and ST_F_02) supplied the wrong answer only to task 3 of the questionnaire, namely to task 21 of Grade 10 INVALSI (giving the correct answer to the three other questionnaire tasks); they are both female and both chose option D. ST_01 first performed all the tasks in silence and autonomously, then the researcher asked her to explain aloud the procedure followed for each task. The following section refers to the INVALSI task.

- R: To solve this task, what reasoning did you follow?
- ST_F_01: To be honest, since I do not consider myself smart, I do not feel very smart ... I think I made a mistake because ...well, first of all, we cannot apply the properties of the powers because there is no multiplication, so we cannot add the powers, so we cannot have 75.

ST_F_02 spoke aloud while solving tasks. The following sentences show her reasoning as explained during the INVALSI task solution.

- ST_F_02: this! ... no, wait ...

    … Then surely I exclude option A
- R: Why do you exclude option A?
- ST_F_02: because it resolves only with the exponents ... but here ... it should be multiplication, it is D ... because it should be a property!

ST_M_01 chose answer B for the INVALSI task and explained his reasoning aloud while tackling the tasks. The following sentences refer to the resolution of the INVALSI task.

- ST_M_01: then ... I have 37 so I try to calculate the same exponent ... no, well ... the same basis. So, certainly not this, and certainly not this

    [ST_M_01indicates options A and C]
- R: So why do you exclude options A and C?
- ST_M_01: Because I do not think we can do 2a, and then after a(a + 1) I do not think we can do that either ...
- R: So you are thinking in reference to the bases?
- ST_M_01: yes, because since I have the same bases ... but it is not a multiplication ... so ... the, this, no [ST_03 indicates option D]
- R: Why not? That answer is very popular around the country…
- ST_M_01: Because ... because there is a "sign of addition" and not the "multiplication sign" ... [ST_03 indicates the text of the task]

Interesting observations also arose when the interviewer asked students about the possible reasons for students' difficulty nationwide. Indeed, some of them referred explicitly to what they usually do during mathematics lessons and to the way they are used to working with powers. This is the case of ST_M_07 who, with regard to the correct answer and the algebraic manipulation, affirmed:

- ST_M_07: We are used to seeing and always using the property, so you do not see it. It does not even cross your mind to look for it.

Another interesting interview comes from student ST_F_05 who only fails to answer INVALSI task 21 correctly.

- R: What "comes to mind" when looking at this question?
- ST_F_05: Because, theoretically, there are theories ... like for the powers, only there was no "sign of multiplication" in the middle here?
- ST_F_05: or option B ...
- R: Why option B?
- ST_F_05: I added the two exponents of the numbers, but I do not remember correctly whether in order to add up the exponents I should have a "sign of multiplication" here in the middle ... So maybe it's D ... either B or D …

Finally, explicit reference to didactic practice is also made by this student:

ST_M_03: Because, in my opinion, many exercises are done on powers, and so it becomes stored in the brain as a form of reasoning. It is normal, perhaps, from our point of view to perform an addition or multiplication.

Discussion

The quantitative analysis highlights that this task creates great difficulties for students of grade 10, even though power properties (and, more generally, manipulation of algebraic expression) are extensively addressed in the first years of high school. Only 35% of students supplied the correct answer to this task and interviews confirm their difficulties in answering, even when they recognized the topic and referred to other exercises on powers.

Gender differences in choosing the correct answer are evident and in favour of males. The gap on the correct answer is of 7% but the distractor analysis reveals a particularity of this gap: if we consider males and females with the same ability, the percentage of correct answers is the same, while the gap is due to the fact that there are fewer girls reaching highest ability levels than boys. This is also confirmed by the absence of DIF in this item and it is in line with literature on gender differences: the gap is more evident if we consider top performers - in this case, we can state that the gap is due to the lower number of girls who reach the higher levels.

Furthermore, we can observe that there are differences between males and females of all ability levels in the way they choose an incorrect answer. Distractor D ($a^{37 \cdot 38}$) is more attractive for females of low and medium ability levels, while distractor B (a^{75}) is more attractive for males. Those two options consist in a power with the same base of the two added in the task and, as exponent, an operation between the two exponents. As we expected, these two options are chosen also by students of medium ability levels who try to solve the task by locating some kind of rule for which the base must be the same and requires them to operate with exponents. This is explicit for example in the ST_F_01 interview in which she expressed the necessity to operate only with exponents and then chose D, affirming "it must be a property!". This behaviour is attributable to attitudes concerning the didactic contract, according to which many students (when faced with the text of a problem) activate a sort of *selective reading*, based on the identification of numerical data and on some rules that suggest the right operation to 'combine' the numbers in the text. As we can read in D'Amore et al. (2010), these norms are often caused by implicit norms generated by the didactic contract.

The necessity of finding some norms and rules to solve the task is also expressed by other students. For instance, the first student (ST_F_02) knew how to multiply powers with the same bases and understood that she "cannot apply the properties of the powers because we do not have multiplication; therefore, we cannot add the powers, so we cannot have 75" but then she went on to choose option D.

Interviews thus confirm that students who tackle this task are led not to reason on the meaning of this operation, for instance by working on orders of magnitude, but set out immediately to find a rule to apply. This behaviour caused students to exclude options A and C because they expect the answer to be a power with the same base: for instance, ST_M_01 explicitly declares that he excluded option A and C because the base is different.As predicted, for all the students interviewed, the task is not new in terms of content and almost all of them refer to theories studied and other exercises already carried out with powers. The

need to find some rule to solve this task is inherent with didactical practices, as stated even by students who have answered the task correctly and then who have been asked to interpret the widespread nationwide difficulties. Students ST_M_03 and ST_M_04 answered the task in question correctly when trying to identify possible causes of error, they hypothesise that other students have drawn heavily on the identification of rules and properties. They said that, most probably, students who made mistakes have been trained to solve exercises with powers using properties without reasoning on the meaning of the power itself; such students know, on the basis of previous experience, that if the bases are the same they have to operate with exponents and this led them to exclude the correct option. Therefore, we confirm that the difficulties encountered by students in this task are not connected to a lack of knowledge, but rather to the incorrect use of knowledge related to power properties and to students' conviction that they need to apply procedures and strategies already used during the maths lessons. This behaviour is, therefore, connected to an influence of didactic contract both according to analysis of the processes adopted by those students who make mistakes, and the ideas of students considering possible causes of errors. Students often apply rules and properties when they operate with powers, and then assume they always have to do so, even when they understand that the properties cannot be applied in that particular case. Moreover, we note that the fact that choosing option D considerably increases the order of magnitude does not actually hinder students' choice. The causes of this behaviour can be traced to a clause of the didactic contract, that of *formal proxy* as defined by D'Amore:

> "The student reads the text, then decides on the operation to carry out and the numbers with which he/she has to work. At that point, indeed, the clause of the formal proxy is triggered. It is no longer for the student to reason and check; he/she no longer considers what follows as his/her personal responsibility […]. The involvement of the student is finished; now it's up to the algorithm or, better still, to the machine, to work for him. The student's next task will be that of transcribing the result, whatever it is, and it does not matter what it means within the problematic context he/she started with." (D'Amore, 2008, p. 16).

This behaviour is evident of course also in students of medium and high ability levels, who are those most influenced by the didactic contract and misconceptions, as already observed in other studies (Giberti, 2018).

Finally, the interviews also confirmed the strong influence of metacognitive factors, especially for females, in answering this task: female student ST_F_01 expressed all her insecurity and low self-confidence regarding mathematics before explaining her reasoning and this, of course, led her to try to solve the task using previously-adopted strategies instead of focusing on the meaning of the sum of the two powers, even though she demonstrated correct knowledge of the properties of powers in the case of multiplication.

Conclusions

The results of this research fit within the international panorama of research on the gender gap, which is increasingly directed at detecting underlying causes of different performances by males and females. The analysis of gender gap on specific tasks can lead to the

identification of factors of a didactic nature that have a different impact on males and females, and which can be interpreted throughout solid findings of mathematics education theories.

From a quantitative point of view, the analysis allows us to answer our research questions. The analysis performed using the Rasch Model highlights characterization of gender gap in terms of students' ability level. Moreover, it shows not only that there are fewer girls reaching highest ability levels than males but, among the students who give the wrong answers, different choices are made by girls and boys. Males and females with different abilities are attracted by different incorrect options; nonetheless, the wrong options most attributable to didactic contract are chosen by students, both male and female, of "average" mathematical ability. This trend, in line with the literature (Bolondi et al., 2018), is confirmed also by the interviews; the interviews conducted allowed us to frame the phenomena that emerged from the comments by students who chose certain wrong options through the theoretical framework of the didactic contract. Indeed, the qualitative analysis highlights how the students' difficulties in tackling this task are not directly connected to a lack of knowledge – instead, the causes of the errors are attributable to the didactic contract. The students interviewed who chose those options often referred to classroom habits, their relationship with the teacher and the habits of the milieu (Brousseau, 1988). The interviews reveal behaviour specifically related to didactic contract and, although to a small extent, differences emerge in terms of gender. As is clearly stated in the literature (D'Amore et al., 2010; Sarrazy, 1995), although the intrinsic link with mathematics differentiates the didactic contract from the social contract, social factors still have an impact, exactly as with the gender gap. This paper contributes to the field of knowledge on the relationship between the didactic contract and the gender gap, a topic of increasing international interest in the field of mathematics education research (Bolondi et al., 2018). By combining quantitative and qualitative analysis, it can be concluded that the gender gap that manifests itself when answering these items is influenced by the didactic contract. The small number of students involved has made it possible to distinguish and characterize male and female behaviour; a research study with a larger sample of students to interview, focusing on the specific connection between didactic contract and gender gap, is currently underway.

We have no conflicts of interest to disclose.

References

Arzarello, F., Garuti, R., & Ricci, R. (2015). The impact of PISA studies on the Italian National Assessment System. In K. Stacey & T. Ross (Eds.), *Assessing mathematical literacy. The PISA experience* (pp. 249–260). Switzerland: Springer International Publishing.

Barbaranelli, C., & Natali, E. (2011). *I test psicologici: teorie e modelli psicometrici [Psychological tests: psychometrical theories and models]*. Rome, Italy: Carocci.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*(2), 163–175.

Bell, K. N., & Norwood, K. (2007). Gender equity intersects with mathematics and technology: Problem-solving education for changing times. In D. Sadker & E. S. Silber (Eds.), *Gender in the classroom* (pp. 225–258). Mahweh, NJ: LEA.

Bolondi, G., Cascella, C., & Giberti, C. (2017). Highlights on gender gap from Italian standardized assessment in mathematics. In J. Novotnà & H. Moravà (Eds.), Diversity in Mathematics Education. *Proceedings of the International Symposium Elementary Maths Teaching, SEMT* (Vol. 17, pp. 81-90). Prague, The Czech Republic: Universita Karlova Press.

Bolondi, G., Ferretti, F., & Giberti, C. (2018). Didactic contract as a key to interpreting gender differences in maths. *Journal of Educational, Cultural and Psychological Studies (ECPS Journal), 2018*(18), 415–435. https://doi.org/10.7358/ecps-2018-018-bolo.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Brousseau, G. (1980). Les échecs électifs dans l'enseignement des mathé-matiques à l'école élémentaire [Elective failures in mathematics education in elementary school]. *Revue de laringologie otologie rinologie, 3*, 107–131.

Brousseau, G. (1988). Le contrat didactique: le milieu [The didactic contract: The milieu]. *Recherches en Didactique des Mathématiques, 9*(3), 309–336.

Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Dortrecht, The Netherlands: Kluwer Academic Publishers.

Cascella, C., Giberti, C., & Bolondi, G. (2020). An analysis of Differential Item Functioning on INVALSI tests, designed to explore gender gap in mathematical tasks. *Studies in Educational Evaluation, 64,* 100819.

Chevallard Y. (1985). La transposition didactique. Du savoir savant au savoir enseigné. Grenoble: La Pensée Sauvage.

D'Amore, B. (2008). Epistemology, didactics of mathematics and teaching practices. *Mediterranean Journal of Research in Mathematics Education., 7*(1).

D'Amore, B., Fandiño, M. I., Marazzani, I., & Sarrazy, B. (2010). *Didattica della matematica alcuni effetti del "contratto" [Mathematics education some effects of the "contract"]*. Bologna, Italy: Archetipolibri.

De Vleeschouwer, M., & Gueudet, G. (2011). Secondary - tertiary transition and evolutions of didactic contract: The example of duality in linear algebra. In M. Pytlak, E. Swoboda, & T. Rowland (Eds.), *Proceedings of CERME 7* (pp. 1359–1368). Poland: Univ. Of Rzeszów.

Elia, I, Gagatsis, A, Panaoura, A, Zachariades, T, & Zoulinaki, F. (2009). Geometric and algebraic approaches in the concept of limit and the impact of the didactic contract. *International Journal of Science and Mathematics Education, 7*(4), 765–790.

EMS-EC (Education Committee of the EMS). (2012). What are the reciprocal expectations between teacher and students? Solid Findings in Mathematics Education on Didactical Contract. *Newsletter of the European Mathematical Society, 84,* 53–55.

Ericsson, A., & Simon, H. (1980). Verbal reports as data. *Psychological Review, 8*(215), 251.

Ferretti, F. (2020). Valutazioni standardizzate di matematica: Evidenze di difficoltà dalla scuola primaria alla scuola secondaria di secondo grado [Standardized mathematics assessments: Evidence of difficulty from primary to upper secondary school]. Didattica della matematica. *Dalla ricerca alle pratiche d'aula, 7,* 36–49. https://doi.org/10.33683/ddm.20.7.2

Ferretti, F., & Bolondi, G. (2019). This cannot be the result! The didactic phenomenon "the Age of the Earth". *International Journal of Mathematical Education in Science and Technology*.

Ferretti, F., Giberti, C., & Lemmo, A. (2018). The Didactic Contract to interpret some statistical evidence in mathematics standardized assessment tests. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(7), 2895–2906.

Forgasz, H. J., Becker, J. R., Lee, K., & Steinthorsdottir, O. (Eds.). (2010). *International perspectives on gender and mathematics education*. Charlotte, NC: Information Age Publish.

Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics, 2*(2), 210–240.

Giberti, C. (2019). Gender differences in mathematics: from the international literature to the Italian situation. *Didattica della matematica. Dalle ricerche alle pratiche d'aula d'aula, 2019*(5), 9–26.

Giberti, C. (2018). Differenze di genere e misconcezioni nell'operare con le percentuali: Evidenze dalle prove INVALSI [Gender differences and misconceptions in working with percentages: Evidence from INVALSI tests]. *CADMO, 2018*(2), 97–114. https://doi.org/10.3280/CAD2018-002007

Giberti, C, & Ferretti, F. (2019). Ci deve essere qualche proprietà! Proprietà delle potenze: Contratto didattico e gender gap ["There must be some property!" Properties of powers: didactic contract and gender gap]. *L'insegnamento della Matematica e delle Scienze Integrate*.

Goldin, G. A. (2000). A scientific perspective on structured, task-based interviews in mathematics education research. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 517-545). Mahwah, NJ: Lawrence Erlbaum Associates.

González de San Román, A., & De La Rica, S. (2012). Gender gaps in PISA test scores: The impact of social norms and the mother's transmission of role attitudes. *Estudios de Economía Aplicada, 34*(1).

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *SCIENCE-NEW YORK THEN WASHINGTON, 320*(5880), 1164–1165.

Hart, L. C., Smith, S. Z., Swars, S. L., & Smith, M. E. (2009). An examination of research methods in mathematics education (1995-2005). *Journal of Mixed Methods Research, 3*(1), 26–41.

Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology, 26*(3), 276–295.

Hill, C., Corbett, C., & Rose, A. (2010). *Why so few? Women in Science, technology, engineering, and mathematics*. Washington, DC: American Association of University Women Retrieved from: https://www.aauw.org/aauw_check/pdf_download/show_pdf.php?file=why-so-few-research.

INVALSI. (2012). Rilevazione nazionale sugli apprendimenti 2011–12 [National survey on learning 2011-12.]. Rapporto tecnico. Retrieved from: https://www.invalsi.it/snv2012/documenti/Rapporti/Rapporto_tecnico_SNV2012.pdf

Jacobs, J. E., & Bleeker, M. M. (2004). Girls' and boys' developing interests in math and science: Do parents matter?. New directions for child and adolescent development, 2004(106), 5–21.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14–26.

Leder, G., & Forgasz, H. (2008). Mathematics education: New perspectives on gender. *ZDM – The International Journal on Mathematics Education, 40*(4), 513–518.

Leder, G. C. (1992). Mathematics and gender: Changing perspectives. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 597–622). New York, NY: Macmillan.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*(6), 1123–1135.

Meyer, J. P. (2014). *Applied measurement with jMetrik*. New York, NY: Routledge.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Boston, MA: TIMSS & PIRLS International Study Center.

OECD. (2015). The ABC of Gender Equality in Education: aptitude, behaviour, confdence. Paris, FR: OECD Publishing.

Organization for Economic Cooperation and Development (OECD). (2016a). *PISA 2015 – Results (volume I): Excellence and equity in education*. Paris, France: OECD Publishing.

Organization for Economic Cooperation and Development (OECD). (2016b). *ITALY – Country note – Results from PISA 2015*. Paris, France: OECD Publishing.

Primi, C., Busdraghi, C., Tomasetto, C., Morsanyi, K., & Chiesi, F. (2014). Measuring math anxiety in Italian college and high school students: Validity, reliability and gender invariance of the abbreviated math anxiety scale (AMAS). *Learning and Individual Differences, 34*, 51–56.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Denmarks Paedagogiske Institut.

Sarrazy, B. (1995). Le contrat didactique [The didactic contract]. *Revue Française de Pédagogie, 112,* 85–118.

Tomasetto, C. (2013). Matematica per i maschi, italiano per le femmine: Stereotipi de genere e atteggiamenti verso lematerie scolastiche tra ggenitori e figli, IN-MIND ITALIA, 5, 19–24.

Zwick, R., & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of educational measurement, 26*(1), 55–66.