



Gender-Based Performance in Mathematical Facts and Calculations in Two Elementary School Samples From Chile and Spain: An Exploratory Study

Violeta Pina^{1*}, Diana Martella^{2*}, Salvador Chacón-Moscoso^{3,4}, Mahia Saracostti^{5,6} and Javier Fenollar-Cortés⁷

¹ Departamento de Psicología Evolutiva y de la Educación, Facultad de Educación, Economía y Tecnología de Ceuta, Universidad de Granada, Ceuta, Spain, ² Instituto de Estudios Sociales y Humanísticos, Facultad de Ciencias Sociales y Humanidades, Universidad Autónoma de Chile, Región Metropolitana, Chile, ³ Departamento de Psicología Experimental, Universidad de Sevilla, Sevilla, Spain, ⁴ Departamento de Psicología, Universidad Autónoma de Chile, Santiago, Chile, ⁵ Núcleo Científico y Tecnológico en Ciencias Sociales y Humanidades, Universidad de la Frontera, Temuco, Chile, ⁶ Escuela de Trabajo Social, Facultad de Ciencias Sociales, Universidad de Valparaíso, Valparaíso, Chile, ⁷ Departamento de Psicología, Universidad Loyola Andalucía, Sevilla, Spain

OPEN ACCESS

Edited by:

Pietro Cipresso,
University of Turin, Italy

Reviewed by:

Isabel Ramirez,
National University of Distance
Education (UNED), Spain
Begoña Delgado,
National University of Distance
Education (UNED), Spain

*Correspondence:

Violeta Pina
violetapina@ugr.es
Diana Martella
diana.martella@uautonoma.cl

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 30 April 2021

Accepted: 08 July 2021

Published: 17 August 2021

Citation:

Pina V, Martella D,
Chacón-Moscoso S, Saracostti M and
Fenollar-Cortés J (2021)
Gender-Based Performance in
Mathematical Facts and Calculations
in Two Elementary School Samples
From Chile and Spain: An Exploratory
Study. *Front. Psychol.* 12:703580.
doi: 10.3389/fpsyg.2021.703580

Gender differences in mathematical performance are not conclusive according to the scientific literature, although such differences are supported by international studies such as the Trends in International Mathematics and Science Study (TIMSS). According to TIMSS 2019, fourth-grade male students outperformed female students in Spanish-speaking countries, among others. This work approaches the study on gender difference by examining the basic calculation skills needed to handle more complex problems. Two international samples of second and third graders from Chile and Spain were selected for this exploratory study. Tests on basic mathematical knowledge (symbolic and non-symbolic magnitude comparisons, fluency, and calculation) were administered. The tests did not show significant difference or size effect between genders for mean performance, variance in the distribution of performance, or percentiles. As noted in the existing literature on this topic and reiterated by these findings, great care should be exercised when reporting on possible gender differences in mathematical performance, as these can contribute to low self-concept among female students.

Keywords: gender differences, primary education, mathematical fluency, calculation, children

INTRODUCTION

As technology grows by leaps and bounds, new fields of study and analysis tools are also expanding: Big Data, artificial intelligence, modeling engineering, software architecture, etc. This has increased the demand for professionals, both men and women, with the solid knowledge of mathematics needed for such roles (Belloum et al., 2019). However, women continue to be underrepresented in this kind of professional roles but overrepresented in lower paying jobs (Adams et al., 2019). Determining the origin of these differences is challenging as cognitive, social, and cultural factors over the course of one's life may all contribute, making it difficult to connect mathematical abilities during childhood to gender differences on the job market. However, if this gender gap were

intrinsic, it could appear during childhood, making it essential to determine whether this is the case and, thus, intervene to avoid future underrepresentation of women.

The reduced presence of women in mathematical fields is nothing new. As indicated by Kane and Mertz (2012, p. 10), differences between males and females have been found in the mathematics participation rate, mean and high-end performance, and variance in the distribution of performance, making gender difference the subject of extensive debate. Havelock Ellis (1894) was the first to intend to empirically justify the differences between men and women with his variability hypothesis. According to this hypothesis, men could be expected to show greater variability than women in physical and intellectual abilities. This would explain the predominance of men in higher academic and scientific positions, but also their prevalence in crime. This hypothesis coincided with the expanded presence of women in higher education and their demands for recognition (Shields, 1982) and continued to be defended, with minor modifications, until the 21st century, when the focus turned to extreme scores and percentiles. In particular, some studies found more females in the lower percentiles and more males in the higher percentiles for different math assignments (Barbaresi et al., 2005; Reigosa-Crespo et al., 2012; Anaya et al., 2017).

Nevertheless, in a wide study carried out by Baker and Jones (1993) with 77,602 students from 19 countries, mathematical performance between genders was found to vary by country and the type of education students received; additionally, the differences were noted to be decreasing over time. The authors also largely debunked the hypothesis of variability when they pointed out that gender differences in performance decrease when women have more equitable access to higher education and qualified work. Indeed, several studies (for instance, Kane and Mertz, 2012) have moved away from the deterministic hypothesis of variability, providing empirical evidence for a hypothesis of sociocultural factors.

Coinciding with the argument that sociocultural factors have the greatest bearing on mathematical performance, differences between genders should gradually disappear as society becomes more progressive and inclusive. Lindberg et al. (2010) support this hypothesis in a broad meta-analysis. Based on 242 studies published between 1999 and 2007 (a total sample of 1,286,350 individuals both males and females), the authors concluded that mathematical performance is similar for both genders, noting that any gender difference can be attributed to unrelated factors. The authors conclude that it is crucial to disseminate these results to counteract stereotypes about [alleged] female math inferiority (Lindberg et al., 2010, p. 1134). More recent studies have replicated these results (for instance, Voyer and Voyer, 2014; Scheiber et al., 2015; Kersey and Cantlon, 2018), and others have limited the differences to a minority of tasks (Hutchison et al., 2019).

Given the empirical evidence reported in the literature, males and females would be expected to perform similarly on international studies such as the TIMSS (Instituto Nacional de Evaluación Educativa [INEE], 2016, 2020). Unfortunately, this is not the case. According to the TIMSS, the gender gap in

mathematical performance is present as early as fourth grade and, according to the last edition (2019), has even increased in countries such as Spain. Gender differences at the fourth grade level can also be noted, although to a lesser extent, in other Spanish-speaking countries that participated in TIMSS 2019. Chile, for example, went from having no significant gender differences in mathematics in the 2015 report to having significant ones in the 2019 report (Instituto Nacional de Evaluación Educativa [INEE], 2016, 2020). In fact, samples from Spanish-speaking countries show a large gender gap in international reports, but few studies include them in their analysis (e.g., the meta-analysis by Voyer and Voyer, 2014).

This exploratory study considers two Spanish-speaking samples from Chile and Spain, both of which are underrepresented in the literature. First, it is important to analyze potential gender differences for these countries. Second, by looking into the samples from these two countries, common patterns may be found that could allow an early intervention to be designed.

Both Spain and Chile have similar math curricula at the elementary school level. In both Chile (MINEDUC, 2018) and Spain (e.g., Decreto n.º, 198/2014), students are doing single-digit addition and subtraction at the end of the first grade (at around age seven). In second grade, they learn multiplication and begin adding and subtracting beyond single digits. In third grade, they begin using multiplication and division to solve math problems. Although both samples are socially and culturally distinct, the objective of this study was to investigate whether gender differences exist and, if so, whether the pattern is the same for the two Spanish-speaking countries. Since both Chile and Spain have similar math curricula, if this was the case, the gender gap could be considered intrinsic.

International evaluations and even math tests administered at schools often do not consider the skills and basic knowledge for acquiring mathematical abilities, and are instead focused on complex calculation or problem solving (Nosworthy et al., 2013). Hence, when gender differences are observed in such tests, it is hard to determine whether such differences are also present for more basic math skills. A logical hypothesis is that the basic abilities needed for complex tasks would also show a gender gap. In recent years, different studies have explored what basic numerical skills necessary for mathematical achievement are responsible for gender differences (for instance, Kersey and Cantlon, 2018; Hutchison et al., 2019).

In this regard, the literature has widely reported on the basic skills required for mathematical achievement. On one hand, non-symbolic magnitude comparison (e.g., the ability to look at two groups of objects and determine which is greater in number), is considered a stepping stone to learning numbers (see Landerl, 2019). For non-symbolic comparison, visuospatial perception, not the ability to count, is required (Kersey and Cantlon, 2018). Once it is acquired, boys and girls can relate quantities with symbolic numbers (for instance, relating the number four with four objects) and learn the meaning of the Arabic numeral. On the other hand, symbolic magnitude comparison (the ability to choose which of two numbers is larger) requires participants to efficiently access the analog magnitude representations that

correspond to the Arabic numerals in order to determine which number is greater (Landerl, 2019). Two tasks are widely used to evaluate these abilities: the symbolic magnitude comparison task and the non-symbolic magnitude comparison task. A recent meta-analysis has confirmed that both tasks correlate with subsequent mathematical performance, although the correlation is higher for the symbolic comparison task (Schneider et al., 2016).

Once children understand the meaning of the Arabic numerals, they start to count. The first operations are simple single-digit addition and subtraction with a sum of <10 . Depending on one's knowledge of arithmetic facts, two strategies are mainly used. In the beginning, the child will have to count to solve the operation; and through repetition, the answers become lodged in their memory (De Smedt et al., 2019). The automation of these operations has been considered critical to developing robust complex calculations (Royer et al., 1999).

Calculation and mathematical problems are considered more complex areas of mathematical performance. Mathematical problems also require other abilities such as reading comprehension (Abedi and Lord, 2001; Donlan et al., 2007). Calculation usually includes the four basic mathematical operations (addition, subtraction, multiplication, and division) and generally cannot be solved from memory, since it usually involves double digits or more. Children can use a wide variety of strategies for approaching these operations, such as decomposition or sequencing (Hickendorff et al., 2019). Calculation also requires executive functions, such as working memory. The strategy children use depends not only on their individual skills but also on the instruction they have received. Gender differences have been found (e.g., Diamantopoulou et al., 2012) for these complex tasks. Winkelmann et al. (2008) proposed that disparities in different studies on gender differences can be attributed to what is considered "basic ability" in a particular study and on the cognitive resources required for each task. The breakdown of tasks by gender differences allows for the design of learning and support strategies in specific areas to gradually reduce the gender gap.

Gender differences tend to be more evident in upper elementary. Some studies do present evidence of a gender gap for lower elementary (Jordan et al., 2006; Scheiber et al., 2015), but others do not (Lachance and Mazzocco, 2006). Third grade seems to be a critical year for the detection of gender differences in mathematical performance, although the studies at this level are scarce. As an example, Germany presents gender differences starting in third grade (Winkelmann et al., 2008). In Spain, the TIMSS (2015 and 2019) reports a large gender gap in fourth grade, thus suggesting a possible gender gap in third grade as well. Another international study, TERCE 2015 (UNESCO, 2016), shows that the number of Latin American countries where males have an advantage over females increases between third and sixth grade. Therefore, second and third grade are considered crucial. The aim of this study is not to determine the age at which gender differences, if they exist, appear, but to observe whether these differences are detected in lower elementary in Spanish-speaking countries.

This study delves into these differences to determine if they are present in the first years of school for basic math skills. After controlling for reading proficiency, mathematical performance by Spanish-speaking second and third graders in Chile and Spain will be compared for non-symbolic comparison, symbolic comparison, fluency, and calculation. As part of this study, any differences in mean performance, variance in the distribution of performance, and percentiles will be assessed. According to the literature, differences are not expected in basic mathematical tasks such as symbolic and non-symbolic comparison but could appear for tasks that require mathematical fluency and calculation. We hypothesize that a greater number of males will be in the higher percentiles and a greater number of females will be in the lower percentiles.

METHODOLOGY

Participants

The participants were second and third graders at four schools in the O'Higgins region (Chile) and seven schools in Murcia (Spain). These children were organized into two separate samples, one for each country, with a mean age of 7.75 for the second graders ($M = 7.68$, $SD = 0.31$ for the Spanish children, and $M = 7.88$, $SD = 0.48$ for the Chilean children) and 8.77 for the third graders ($M = 8.7$, $SD = 0.33$ for the Spanish children, and $M = 8.89$, $SD = 0.39$ for the Chilean children). The sample was formed by 201 Spanish school children (107 second graders and 94 third graders; 55.2% male and 44.8% female) and 184 Chilean school children (90 second graders and 94 third graders; 48.9% male and 51.1% female) and was collected through incidental sampling. The following exclusion criteria were established: vision difficulties, diagnosis of autism spectrum disorder or cognitive impairment, insufficient knowledge of the Spanish language, and, in the case of the third graders, significant difficulties reading. **Table 1** details the number of males and females by grade and country. Possible differences regarding age (Mann-Whitney U) and gender proportion by grade and country (χ^2) were examined.

This study was carried out following the recommendations of the Chilean Commission for Scientific and Technological Investigation (CONICYT in its Spanish acronym). The protocol was approved by the Ethics Committee of Universidad de la Frontera (Act 066-2017, on Sheet 036-17). The study complied with the standards of the ethical committee of Universidad Autónoma de Chile and with the agreement between the Department of Education and Universities of the autonomous community of Region of Murcia and Universidad de Murcia. Informed consent was requested to participate in this study.

Instruments

The following paper-and-pencil tests were administered to measure mathematical performance and reading fluency.

Numerical magnitude comparisons (Nosworthy et al., 2013). This task includes two different parts:

TABLE 1 | Distribution in the Spanish language sample of second and third graders: scoring on symbolic test by gender.

<i>M(DT)</i>	Spanish sample					Chilean sample				
	Male	Female	<i>U</i>	<i>p</i>	<i>r</i> ^a	Male	Female	<i>U</i>	<i>p</i>	<i>r</i>
Second grade										
N	107					90				
Gender <i>n</i> (%)	56(52.3)	51(47.7)				49(54.4)	41(45.6)			
Magnitude comparison										
Symbolic	29.3(9.3)	33.0(6.8)	1055.0	0.020	-0.26	29.9(10.1)	30.9(10.6)	961.0	0.727	
Num. errors	0.1(0.4)	0.1(0.9)	1572.2	0.060		0.6(1.3)	0.7(1.2)	931.5	0.850	
Non-symbolic	29.1(7.6)	31.6(5.7)	1103.5	0.042	-0.23	28.8(13.8)	33.9(12.0)	796.5	0.092	
Num. errors	0.3(1.0)	0.6(0.1)	1419.5	0.932		1.4(2.7)	2.2(2.9)	826.0	0.110	
Third grade										
N	94					94				
Gender <i>n</i> (%)	55(58.5)	39(41.5)				55(58.5)	39(41.5)			
Magnitude comparison										
Symbolic	37.7(7.0)	39.8(6.2)	837.0	0.069		39.3(10.2)	37.4(11.2)	1158.0	0.513	
Num. errors	0.2(0.6)	0.1(0.4)	1126.5	0.443		0.6(1.1)	0.6(1.1)	1051.5	0.847	
Non-symbolic	32.3(7.3)	31.7(4.8)	1196.0	0.344		37.0(11.3)	35.5(11.4)	1193.5	0.355	
Num. errors	0.5(1.0)	0.3(0.6)	1135.5	0.412		1.7(2.7)	1.1(1.6)	1133.5	0.618	
PreDisCal										
Sentences	13.6(14.7)	14.7(2.9)	905.5	0.199		13.2(6.9)	12.8(6.2)	1090.0	0.896	
Num. errors	2.2(1.9)	1.6(1.2)	1237.0	0.193		6.5(9.3)	3.2(2.4)	1252.5	0.163	
Mathematical fluency	17.7(6.8)	17.1(4.5)	1079.5	0.960		12.9(5.5)	10.7(5.1)	1369.5	0.023	0.27
Num. errors	1.5(2.5)	1.1(1.7)	1127.0	0.652		1.8(2.6)	1.3(1.4)	1050.5	0.864	
Calculation	11.6(4.2)	11.0(3.4)	1143.0	0.589		9.0(4.3)	9.0(4.3)	1126.0	0.683	
Num. errors	0.2(0.6)	0.3(0.6)	1202.0	0.313		4.1(8.0)	1.9(8.0)	1294.5	0.083	

Scores are shown in terms of mean and standard deviation (not mean scores) to facilitate understanding.

^aEffect size corresponding to the biserial correlation between ranges. Only values corresponding to significant differences between the groups are provided.

- Symbolic task (56-digit pairs), where the participants were asked to compare numerical pairs (numbers one through nine) and indicate which number was higher.
- Non-symbolic task (56 dot arrays), where dot arrays of varying quantities are present and the participants were asked to indicate which side had more dots.

Each test had a time limit of 1 min, and the dependent variable was the number of correct answers.

PreDisCal (Pina et al., 2020). This is a set of three tests that measure reading fluency, mathematical fluency, and calculation, in that order. Although the PreDisCal scale was developed in Spain, a cross-validation process was carried out in several Latin American countries. Some of the items were modified because of interpretation issues (related to sociocultural factors) in Latin American countries. The test included the following tasks:

- Sentences. This task consists of 47 sentences that assess reading fluency. Each sentence has a missing word, for instance: “The strawberries are ...” followed by five answer options (one is correct). The incorrect alternatives are close grammatically or semantically. All the sentences are easy to understand. The test had a time limit of 3 min. The dependent variable is the number of correct answers. It is a validated test with adequate test-retest reliability [$r(169) = 0.8, p < 0.001$].

- Mathematical fluency. The arithmetic facts are tested through 63 single-digit addition and subtraction operations. All sums are <10 . The maximum time is 1 min in this test, and the dependent variable is the number of correct answers. The reliability of the test is adequate [$r(169) = 0.85, p < 0.001$].
- Calculation. This test comprises 45 items of increasing difficulty and assesses complex calculation. Participants have to determine what number or symbol, or if both are missing in a comparison of two operations (e.g., “ $3 + _ = 5 + 1$ ”). The test had a time limit of 3 min, and the dependent variable was the number of correct answers. The test-retest reliability of the validated scale was acceptable [$r(169) = 0.75, p < 0.001$].

Procedure

The tasks were administered in February for Spain and in September of the same year for Chile in order to reflect the same moment of the academic calendar in both countries. Trained evaluators collectively administered the test during the school day. Given that PreDisCal required a certain degree of reading proficiency and complex calculation ability, it was administered only to the third graders, while the magnitudes comparison tests were administered to both the second and third grade samples. In the case of the magnitude comparisons, the symbolic test was administered first, followed by the non-symbolic test.

The PreDisCal was administered in keeping with the order established for that test, e.g., sentences, mathematical fluency, and calculation.

Data Analysis

First, the comparability of the groups was verified by assessing age and gender differences by grade and country (χ^2), and mean performance in the sentences test between countries (Student *t*). Second, basic parametric assumptions for data analysis were verified (normality using the Kolmogorov–Smirnov test, homoscedasticity using the Levene test). Non-parametric tests were carried out using the Mann–Whitney *U* and statistical biserial correlations ($r = 0.1$ as low; $r = 0.3$ as medium, and $r = 0.5$ as high effect; Cohen, 1988); and χ^2 to compare gender and percentile performance distribution in third grade.

RESULTS

Both samples showed a similar gender distribution of participants ($\chi^2 = 0.09$, $p = 0.768$ for second grade; $\chi^2 = 0$, $p = 1$ for second grade). There was no significant evidence between countries in relation to the mean performance in the sentences test ($t = 1.35$; $p = 0.178$). However, significant differences were detected between countries regarding the mean age of students ($t = -5.04$, $p < 0.001$, $d = -0.52$ for the third grade; $t = -4.18$, $p < 0.001$, $d = -0.86$ for the second grade), with moderate to high effect sizes (Cohen, 1988). Nevertheless, these differences in age are not relevant to this study, since the analysis between countries is carried out comparing gender, a variable for which there are no significant differences in the samples of both countries.

The distribution of data was examined for each dependent variable according to country and course. Except for the mathematical fluency and calculation tests in the Chilean third grade sample, the remaining variables followed a non-normal distribution ($K-S < 0.05$); thus, non-parametric analyses were applied.

The Levene test showed significant difference in variance between genders in the symbolic test for the Spanish sample of second graders ($F = 3.97$, $p = 0.049$). The scores of boys showed higher variability ($\sigma^2 = 86.5$) with skewness of 0.18, ($SE = 0.32$) and kurtosis of -0.19 ($SE = 0.63$), compared with those of the girls ($\sigma^2 = 46.3$) with skewness of 0.1 ($SE = 0.33$), and kurtosis of 0.27 ($SE = 0.66$), as shown in **Figure 1**. Differences in variance were not found for the other tests for either sample.

Spearman's correlation was carried out between tests for each country (**Table 2**). The correlation between scores on the symbolic and non-symbolic tests was crucially positive and high in the second grade sample. As for the third graders, mathematical fluency positively correlated to calculation and with sentences; these last two subscales correlated positively. The symbolic test also correlated positively with the non-symbolic one ($r_s = 0.46$, $p < 0.001$).

Mean Performance by Gender and Grade

If differentiated by gender, the Mann–Whitney tests indicated that the girls scored significantly higher (See **Table 2**) in the symbolic test of the Spanish second graders ($U = 1,055$, $p = 0.02$,

$r = -0.26$), but this was not the case for the Chilean sample ($p = 0.727$). However, the effect size of the significant differences ranged from low to moderate. Regarding the non-symbolic test, the results were similar, with girls outperforming the boys in the Spanish sample ($U = 1,103$, $p = 0.42$, $r = -0.23$), but not in the Chilean sample ($p = 0.092$). In this case, the effect size was lower than in the symbolic test. The differences detected in the sample disappear in the third grade. In terms of the number of mistakes made on both tests, no significant differences were found between genders in either country.

In the third-grade sample, significant differences were only found between genders for the mathematical fluency test in the Chilean sample ($U = 1,369$, $p = 0.023$, $r = 0.27$). Here, the girls had lower scores ($Md = 9$) than the boys ($Md = 13$). In terms of the number of mistakes made on both tests, no significant differences were found between genders in either country at the third grade level.

Ratio of Boys to Girls at Various Percentiles

Four groups for each sample were formed according to the scores they obtained and their corresponding percentiles. Group 1 scored in the 1–25th percentile; group 2 in the 26–50th percentile; group 3 in the 51–75th percentile; and group 4 in the 76th percentile or higher. χ^2 and its significance were calculated to examine significant differences in the ratio of males and females in the different percentile groups. As no significant differences were found in terms of performance of boys and girls in the third-grade tests (except for mathematical fluency, which had a low effect size), the groups were formed by students from both countries. In the case of the second graders, the calculations varied by country.

For the third group, there were no significant gender differences in terms of the ratio of boys to girls, the different percentile groups, or for any of the tests (see **Table 3**). However, significant differences were found in the symbolic test for the Spanish sample of second graders ($\chi^2 = 9.65$, $p = 0.022$). The percentages were different for the first group (scorings between the 1–25th percentile; $\chi^2 = 6.54$, $p = 0.011$), with a larger quantity of boys (30.4%) than girls (9.8%). Significant differences were not found in the Chilean sample of second graders.

DISCUSSION

This study set out to compare the basic mathematical abilities of boys and girls, which prove essential for the acquisition of complex math skills. The sample was composed of boys and girls from two Spanish-speaking yet culturally different countries in the second and third grade, ages at which the gender gap starts to appear. The results allow us to conclude, in keeping with the scientific literature on the topic, that in terms of basic mathematical knowledge, boys and girls show no significant differences in mathematical performance. Previous investigations, including various meta-analyses (Lachance and Mazzocco, 2006; Hyde et al., 2008; Lindberg et al., 2010; Voyer and Voyer, 2014; Scheiber et al., 2015), have noted as much. Similarly, Hutchison et al. (2019) found no differences in numerical and magnitude comparison, numerical ordering,

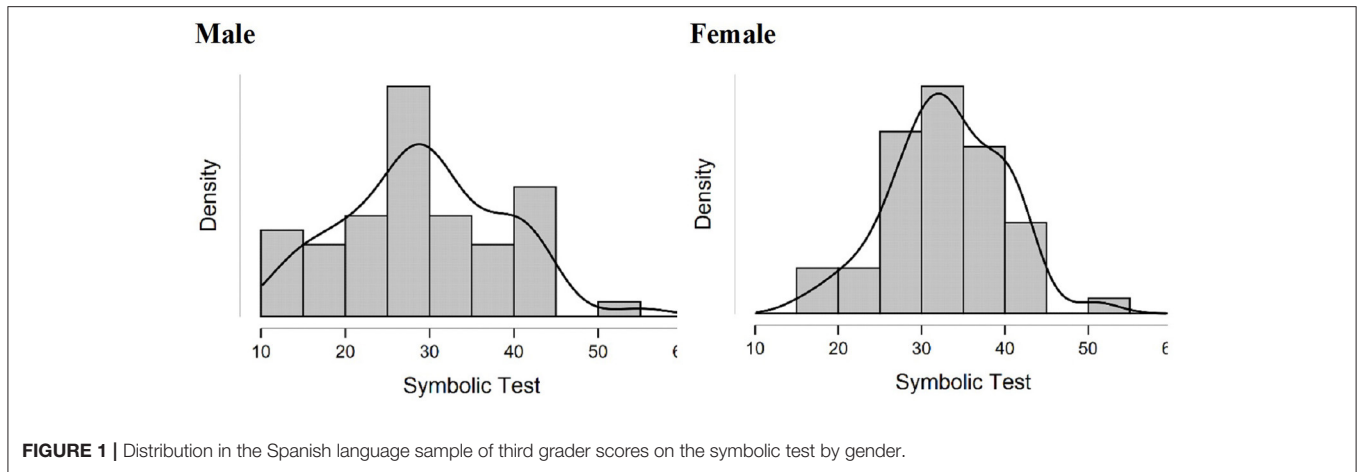


FIGURE 1 | Distribution in the Spanish language sample of third grader scores on the symbolic test by gender.

TABLE 2 | Spearman’s correlations between the PreDisCal and magnitude comparison (third grade tests).

	PreDisCal			Magnitude comparison
	Sentences	Mathematical fluency	Calculation	Symbolic
PreDisCal				
Mathematical fluency	0.194			
Calculation	0.253*	0.451***		
Comparison of magnitude				
Symbolic	0.054	0.389***	0.065	
Non-symbolic	0.025	0.271**	0.314**	0.318**

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

magnitude estimation, multiplication, and division, although the population analyzed for their study hailed from the Netherlands, meaning that results could be different in countries with a higher gender gap. The results of this study, however, reveal no significant differences in most of the tasks considered for two Spanish-speaking samples from Chile and Spain, despite the high gender gap detected in international tests for the two countries (such as TIMSS 2019).

The main differences are found in the symbolic test (comparison of two numbers to determine which one is higher) in the Spanish sample of second graders. In particular, the girls performed better and the boys showed a greater variability. Additionally, females had a higher mean performance, and a greater ratio of males fell below the 25th percentile. However, the effect size is small in second grade, and no such results were found in the third grade sample. For the non-symbolic comparison test, the girls performed better once again, although this difference disappeared in third grade. The direction of these gender differences goes against the hypothesis of higher performance in males. In line with these results, previous studies have shown gender differences in favor of females in basic abilities (for instance, Halpern and Wright, 1996).

The only difference in favor of males was found in the Chilean sample in average performance for mathematical fluency in the third grade. Differences in addition and subtraction have been found previously, although there has not always been evidence of

these differences (Hutchison et al., 2019). In this case, the effect size was small and no differences were found for calculation, meaning that caution is critical when interpreting the results. In the case of the Chilean third graders, test abilities were recently acquired, and the differences would likely disappear with practice. Importantly, mathematical fluency is decisive for the robust development of complex calculation and is an ability that can be practiced until it is learned (Royer et al., 1999).

Prior studies have indicated that differences in mathematical performance rely on the kind of task students are asked to complete. These studies have found that females perform better in arithmetic and calculus, while males perform better in mathematical problem solving (Byrnes and Takahira, 1993). However, there have also been conflicting results in the research. For instance, Royer et al. (1999) identified that males perform better in math-fact retrieval than females. In another study, Winkelmann et al. (2008) showed that females had poorer basic skills, while males had a slight advantage in mathematical problems, such as equations containing a missing number. The authors explained that the results depended on basic elementary skills and on how math problems are defined, i.e., on the cognitive resource that each task demanded.

One hypothesis is that gender differences can be found in complex areas of mathematics that demand more cognitive resources. In this sense, for example, the literature suggests a relationship between spatial abilities and gender differences

TABLE 3 | Percentage and comparison of third grade boys and girls by groups corresponding to the 25th, 50th, 75th, and 100th percentile.

	Group 1 (P25)	Group 2 (P50)	Group 3 (P75)	Group 4 (P100)	Differences between groups	
					χ^2	<i>P</i>
PreDisCal						
Sentences						
Male	30,0	20,9	29,1	20,0	3.80	0.284
Female	21,8	29,5	23,1	25,6		
<i>N</i>	61	44	39	44		
Calculation						
Male	32,7	20,0	23,6	23,6	2.39	0.495
Female	32,1	28,2	16,7	23,1		
<i>N</i>	61	44	39	44		
Mathematical fluency						
Male	24,5	29,1	22,7	23,6	1.60	0.659
Female	32,1	29,5	19,2	19,2		
<i>N</i>	52	55	40	41		
Magnitude comparison						
Symbolic						
Male	27,3	32,7	17,3	22,7	7.10	0.069
Female	24,4	21,8	33,3	20,5		
<i>N</i>	61	44	39	44		
Non-symbolic						
Male	27,3	18,2	30,9	23,6	5.65	0.130
Female	26,9	32,1	25,6	15,4		
<i>N</i>	61	44	39	44		

Groups are organized by percentiles: 1–25th percentile (Group 1), 26–50th percentile (Group 2), 51–75th percentile (Group 3), and 76–100th percentile (Group 4).

in mathematical performance. However, a recent study did not confirm these results and instead concluded that gender differences may depend on the test and the strategy used to solve each item (Ramírez-Uclés and Ramírez-Uclés, 2020). The results support this hypothesis. In particular, the calculation is the task that involves the highest cognitive demand, but no gender differences were found. Another possibility is that gender differences in mathematical performance emerge later in more complex mathematical tasks or are influenced by cultural or social stereotypes, although the results show that at least lower elementary school children do not exhibit such differences.

Another hypothesis for gender differences in mathematical performance is related to the way in which the answer options are presented on tests. For instance, in a study with PISA 2000 data, Lafontaine and Monseur (2009) found that in all the countries analyzed, open-ended (as opposed to multiple choice) answers for reading tasks favored females. For mathematics, Routitsky and Turner supported these conclusions using PISA 2003 data but pointed out that this difference disappeared as the complexity of items increased (Routitsky and Turner, 2003). Future investigations should assess how the questions and type of answer options influence the gender gap.

Previous studies have revealed that the learning routine shows no differences between genders, although boys and girls do express different attitudes toward mathematics

(Barbero-García et al., 2007). This could explain why differences can be found in higher grades despite being absent when basic math skills are evaluated. Importantly, gender differences may well be related to the wording of tasks or instructions, or social behaviors that influence how parents, teachers, and even female students view their own mathematical ability, not by any intrinsic math difficulties (see Lindberg et al., 2010).

This study presents several limitations. All the tasks considered had a time limit, which can influence the results. Processing speed is essential for school performance (Dodonova and Dodonov, 2012), and it is important for the diagnosis of learning disabilities, since children with a diagnosis are identical to those with mean performance if we provide them sufficient time (Jordan and Montani, 1997). The second constraint of this study is that no information was obtained on the socioeconomic level of the children. However, both samples attended public schools, and none were disadvantaged. Lastly, this study is exploratory, and despite having a large sample, it cannot be deemed representative of the total population of both countries.

Gender differences in mathematical performance have been less studied in Spanish-speaking countries. This study was laid out as an initial approach to researching this population. Based on existing literature and the results, gender differences vary by task, strategies used to solve problems, cohort, age, and instructions provided. For future research, the sample should be expanded geographically for higher representativeness

in relation to social, educational, and cultural differences. Furthermore, it would be interesting to expand the number of courses evaluated to observe the pattern between genders and to include more complex tasks; this should allow specific mathematical areas to be identified in which gender differences can easily be detected (Chacón-Moscoso et al., 2019). These new developments would potentially reveal differences related to the country, factorial invariance, differential item functioning (DIF), and the influence of open-ended (as opposed to multiple choice) questions (Chacón-Moscoso et al., 2016).

Practical Value, Implications for Educational Intervention

It should be noted that the tests applied in this study could be used for an early diagnosis of gender difference, and with respect to children with low mathematical performance.

CONCLUSIONS

In conclusion, results emphasizing that males outperform females in math do not hold true in the key mathematical areas analyzed with the exception of mathematical fluency in the Chilean sample, where the effect size was small. Therefore, the acquisition of complex abilities should be the same for males and females. At a scientific level and in the news media, care should be exercised when reporting on gender differences because it can influence own opinion of the girls of their mathematical ability. To the best knowledge of the authors, this study is the first to examine gender differences in basic math skills in Spanish-speaking populations that reveals considerable gaps in international studies on the topic.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Abedi, J., and Lord, C. (2001). The language factor in mathematics tests. *Appl. Measure. Edu.* 14, 219–234. doi: 10.1207/S15324818AME1403_2
- Adams, R. B., Barber, B. M., and Odean, T. (2019). *The Math Gender Gap and Women's Career Outcomes*. Available online at SSRN: <https://ssrn.com/abstract=2933241> (accessed December 17, 2019).
- Anaya, L., Stafford, F. P., and Zamorro, G. (2017). Gender Gaps in Math Performance, Perceived Mathematical Ability and College STEM Education: The Role of Parental Occupation. EDRE Working Paper No. 2017-21. Available online at SSRN: <https://ssrn.com/abstract=3068971> (accessed November 10, 2017).
- Baker, D. P., and Jones, D. P. (1993). Creating gender equality: cross-national gender stratification and mathematical performance. *Sociol. Edu.* 66, 91–103. doi: 10.2307/2112795
- Barbarelli, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., and Jacobs, S. J. (2005). Math learning disorder: incidence in a population-based birth cohort, 1976 – 82, Rochester, Minn. *Ambul. Pediatr.* 5, 281–289B. doi: 10.1367/A04-209R.1
- Barbero-García, I., Holgado-Tello, F. P., Vila-Abad, E., and Chacón-Moscoso, S. (2007). Actitudes, hábitos de estudio y rendimiento en Matemáticas: diferencias

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Chilean Commission for Scientific and Technological Investigation (CONICYT, in its Spanish acronym). The protocol was approved by the Ethics Committee of Universidad de la Frontera (Act 066-2017, on Sheet 036-17). The study was in accordance with the standards of the ethical committee of Universidad Autónoma de Chile and with the agreement between the Region of Murcia's Comunidad Autónoma through the Department of Education and Universities, and Universidad de Murcia. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

VP, JF-C, and DM conceptualized and designed the study. MS acquired the data. VP, JF-C, and SC-M analyzed and interpreted the data. VP, JF-C, and DM wrote the manuscript. JF-C prepared the figures and tables. SC-M discussed results and reviewed writing. All the authors read and approved the final version of the manuscript.

FUNDING

This study has been supported by the Chilean national projects Fondecyt Regular 2018 ID1181472, Anid/Fondecyt Regular ID1210172 and Fondecyt Regular 2019 (ref. number 1190945) and by a fellowship for residencies abroad of young researchers from Mare Nostrum (2017). Also, this research has been supported by the ERDF Andalusia Operational Program 2014–2020, government of Andalusia, Spain (ref. US-1263096); the VI Own Research and Transfer Plan (VIPPITUS), Universidad de Sevilla, Spain (ref. VIPP PRECOMPETI 2020/1333).

- por género. *Psicothema* 19, 413–421. Available online at: <http://www.psicothema.com/pdf/3379.pdf>
- Belloum, A. S. Z., Koulouzis, S., Wiktorski, T., and Manieri, A. (2019). Bringing the demand and the offer in data science. *Concurrency Comput. Pract. Exp.* 31:E5200. doi: 10.1002/cpe.5200
- Byrnes, J. P., and Takahira, S. (1993). Explaining gender differences on SAT-math items. *Dev. Psychol.* 29, 805–812. doi: 10.1037/0012-1649.29.5.805
- Chacón-Moscoso, S., Anguera M. T., Sanduvete-Chaves, S., Losada-López, J. L., Lozano-Lozano, J. A., and Portell, M. (2019). Methodological quality checklist for studies based on observational methodology (MQCOM). *Psicothema* 31, 458–464. doi: 10.7334/psicothema2019.116
- Chacón-Moscoso, S., Sanduvete-Chaves, S., and Sánchez-Martín, M. (2016). The development of a checklist to enhance methodological quality in intervention programs. *Front. Psychol.* 7:1811. doi: 10.3389/fpsyg.2016.01811
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Lawrence Erlbaum Associates.
- De Smedt, B., Peters, L., and Ghesquière, P. (2019). “Neurobiological origins of mathematical learning disabilities or dyscalculia: a review of brain imaging data,” in *International Handbook of Mathematical Learning Difficulties*, eds A. Fritz, V. G. Haase, & P. Räsänen (Cham, Switzerland: Springer), 367–384. doi: 10.1007/978-3-319-97148-3_23

- Decreto n.º (198/2014). BORM de 5 de septiembre, por el que se establece el currículo de la Educación Primaria en la Comunidad Autónoma de la Región de Murcia (Región de Murcia: Boletín Oficial de la Región de Murcia (BORM), n.º 206, 2014).
- Diamantopoulou, S., Pina, V., Valero-García, A., González-Salinas, C., and Fuentes, L. J. (2012). Validation of the Spanish version of the woodcock-johnson mathematics achievement tests for children aged 6 to 13. *J. Psychoedu. Assess.* 30, 466–477. doi: 10.1177/0734282912437531
- Dodonova, Y. A., and Dodonov, Y. S. (2012). Processing speed and intelligence as predictors of school achievement: Mediation or unique contribution? *Intelligence* 40, 163–171. doi: 10.1016/j.intell.2012.01.003
- Donlan, C., Cowan, R., Newton, E. J., and Lloyd, D. (2007). The role of language in mathematical development: Evidence from children with specific language impairment. *Cognition* 103, 23–33. doi: 10.1016/j.cognition.2006.02.007
- Halpern, D. F., and Wright, T. M. (1996). A process-oriented model of cognitive sex differences. *Learn. Individ. Dif.* 8, 3–24. doi: 10.1016/S1041-6080(96)90003-5
- Havelock Ellis (1894). *Man and Woman: A Study of Human Secondary Sexual Characters*. New York, NY: Scribner's.
- Hickendorff, M., Torbeyns, J., and Verschaffel, L. (2019). “Multi-digit addition, subtraction, multiplication, and division strategies,” in *International Handbook of Mathematical Learning Difficulties*, eds A. Fritz, V. Haase & P. Räsänen (Cham, Switzerland: Springer), 543–560. doi: 10.1007/978-3-319-97148-3_32
- Hutchison, J., Lyons, I., and Ansari, D. (2019). More similar than different: gender differences in basic numeracy are the exception, not the rule. *Child Dev.* 90, 6600–79000. doi: 10.1111/cdev.13044
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender similarities characterize math performance. *Science* 321, 494–495. doi: 10.1126/science.1160364
- Instituto Nacional de Evaluación Educativa [INEE] (2016). TIMSS 2015. Estudio internacional de tendencias en Ciencias y Matemáticas. IEA. Informe español: Resultados y contexto. Madrid: Ministerio de Educación, Cultura y Deporte.
- Instituto Nacional de Evaluación Educativa [INEE] (2020). TIMSS 2019. Estudio internacional de tendencias en Ciencias y Matemáticas. IEA. Informe español: Resultados y contexto. Madrid: Ministerio de Educación, Cultura y Deporte.
- Jordan, N. C., Kaplan, D., Olah, L., and Locuniak, M. N. (2006). Number sense growth in kindergarten: a longitudinal investigation of children at risk for mathematics difficulties. *Child Dev.* 77, 153–175. doi: 10.1111/j.1467-8624.2006.00862.x
- Jordan, N. C., and Montani, T. O. (1997). Cognitive arithmetic and problem solving: a comparison of children with specific and general mathematics difficulties. *J. Learn. Disabil.* 30, 624–634. doi: 10.1177/002221949703000606
- Kane, J. M., and Mertz, J. E. (2012). Debunking myths about gender and mathematics performance. *Not. AMS* 59, 10–21. doi: 10.1090/noti790
- Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., and Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. *NPJ Sci. Learn.* 3, 1–10. doi: 10.1038/s41539-018-0028-7
- Lachance, J. A., and Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learn. Individ. Dif.* 16, 195–216. doi: 10.1016/j.lindif.2005.12.001
- Lafontaine, D., and Monseur, C. (2009). Gender gap in comparative studies of reading comprehension: to what extent do the test characteristics make a difference? *Eur. Edu. Res. J.* 8, 69–79. doi: 10.2304/eeerj.2009.8.1.69
- Landerl, K. (2019). “Neurocognitive Perspective on Numerical Development,” in *International Handbook of Mathematical Learning Difficulties*, eds A. Fritz, V. G. Haase, and P. Räsänen (Cham, Switzerland: Springer), 9–25. doi: 10.1007/978-3-319-97148-3_2
- Lindberg, S. M., Hyde, J. S., and Petersen, J. L., and Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychol. Bull.* 136, 1123–1135. doi: 10.1037/a0021276
- MINEDUC (2018). *Bases Curriculares Primero a Sexto Básico*. Santiago: Ministerio de Educación.
- Nosworthy, N., Bugden, S., Archibald, L. A., Evans, B., and Ansari, D. (2013). A two-minute paper and pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS ONE* 8:e67918. doi: 10.1371/journal.pone.0067918
- Pina, V., Hernández-Pérez, E., Rabadán-Rubio, J. A., Hernández-Pallarés, L., and Fenollar-Cortés, J. (2020). *PreDisCal: Prueba de cribado para la detección de dificultades lectoras o matemáticas de segundo a sexto de primaria*. Madrid: TEA Ediciones.
- Ramírez-Uclés, I. M., and Ramírez-Uclés, R. (2020). Gender differences in visuospatial abilities and complex mathematical problem solving. *Front. Psychol.* 11:191. doi: 10.3389/fpsyg.2020.00191
- Reigosa-Crespo, V., Valdes-Sosa, M., Butterworth, B., Estevez, N., Rodríguez, M., Santos, E., et al. (2012). Basic numerical capacities and prevalence of developmental dyscalculia: the Havana survey. *Dev. Psychol.* 48, 123–135. doi: 10.1037/a0025356
- Routitsky, A., and Turner, R. (2003). “Item format types and their influences on cross-national comparisons of student performance,” in *Paper Presented at the Annual Meeting of the American Educational Research Association* (Chicago).
- Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. G., and Marchant, H. G. (1999). Math fact retrieval as the cognitive mechanism underlying gender differences in math achievement test performance. *Contemp. Edu. Psychol.* 24, 181–266. doi: 10.1006/ceps.1999.1004
- Scheiber, C., Reynolds, M. R., Hajovsky, D. B., and Kaufman, A. S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychol. Schools* 52, 335–348. doi: 10.1002/pits.21827
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., and DeSmedt (2016). Association of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372
- Shields, S. A. (1982). The variability hypothesis: the history of a biological model of sex differences in intelligence. *Signs. J. Women Cult. Soc.* 7, 769–797. doi: 10.1086/493921
- UNESCO, Gelber et al. (2016). *Inequidad de género en los logros de aprendizaje en educación primaria? ¿Qué nos puede decir TERCE?* Santiago: Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación.
- Voyer, D., and Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychol. Bull.* 140, 1174–1204. doi: 10.1037/a0036620
- Winkelmann, H., van den Heuvel-Panhuizen, M., and Robitzsch, A. (2008). Gender differences in the mathematics achievements of German primary school students: results from a German large-scale study. *ZDM* 40, 601–616. doi: 10.1007/s11858-008-0124-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Pina, Martella, Chacón-MoscOSO, Saracostti and Fenollar-Cortés. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.