

# Arithmetic word problems in Castilian-Leonesian and Andalusian textbooks

## Problemas aritméticos verbales en los libros castellano-leoneses y andaluces

<https://doi.org/10.4438/1988-592X-RE-2024-406-641>

### **Santiago Vicente Martín**

<https://orcid.org/0000-0002-2072-8133>

*Universidad de Salamanca*

### **Marta Ramos Baz**

<https://orcid.org/0000-0002-9643-6495>

*Universidad de Salamanca*

### **M<sup>a</sup>. Mercedes Rodríguez Sánchez**

<https://orcid.org/0000-0002-1020-6681>

*Universidad de Salamanca*

### **Beatriz Sánchez-Barbero**

<https://orcid.org/0000-0001-9118-2118>

*Universidad de Salamanca*

### **Rosario Sánchez Fernández**

<https://orcid.org/0000-0002-7984-1894>

*Universidad de Salamanca*

### **Abstract**

The results of mathematics achievement assessments such as the TIMSS, in which problem-solving is the cornerstone, show that the performance of students in Castile and Leon is significantly greater than that of students in Andalusia. These differences seem to be due to some aspects pointed out by the

TIMSS (such as the social and economic index), although they could also be due to certain elements of the teaching-learning process, such as the textbook, since they are used intensively in the classroom. In fact, the books used in countries with more proficient students in mathematics include more problems and more varied and difficult problems than books from other countries. The aim of the present work is to determine whether the primary mathematics textbooks used in Castile and Leon are different from those used in Andalusia in terms of learning verbal arithmetic problem solving. For this purpose, in the first study, we determined which books are most frequently used in each autonomous community, and in the second study, we analysed 3rd grade primary school books from the most widely used textbooks in each autonomous community in relation to the number of tasks aimed at solving problems and the variety and level of semantic-mathematical difficulty of these problems. The results indicated that although the most frequently used books in both communities belong to different publishing projects, the most-used books in Castile and Leon and Andalusia were very similar with respect to the variables analysed, the only difference being a greater number of problems in the Castilian-Leonese books. These results contrast with those obtained in international comparisons of textbooks and point to a greater influence of other variables on differences in student performance in both autonomous communities.

*Keywords:* textbook, primary education, problem solving, arithmetic, comparative analysis, content analysis.

### **Resumen**

Los resultados de evaluaciones del rendimiento en matemáticas como TIMSS, en las que la resolución de problemas es la piedra angular, muestran que el rendimiento los alumnos de Castilla y León es significativamente superior al de los alumnos de Andalucía. Estas diferencias parecen deberse a algunos aspectos señalados por el propio TIMSS (como el Índice Social y Económico), aunque también podrían deberse en algún grado a determinados elementos del proceso de enseñanza-aprendizaje, como el libro de texto, puesto que se usan de manera intensiva en las aulas. De hecho, los libros utilizados en países con alumnos más competentes en matemáticas incluyen más problemas, y problemas más variados y difíciles que los libros de otros países. El objetivo del presente trabajo es determinar si los libros de matemáticas de Primaria que se utilizan en Castilla y León son diferentes a los que se utilizan en Andalucía, en relación con el aprendizaje de la resolución de problemas aritméticos verbales. Para ello, en un primer estudio, se determina qué libros se utilizan con más frecuencia en cada comunidad autónoma, y en un segundo estudio, se analizan libros de 3º de Primaria de los proyectos editoriales más utilizados en cada

comunidad autónoma en relación con la cantidad de tareas destinadas a resolver problemas, y la variedad y nivel de dificultad semántico-matemática de esos problemas. Los resultados indicaron que, aunque los libros más frecuentes en ambas comunidades pertenecen a proyectos editoriales diferentes, los libros más utilizados en Castilla y León y Andalucía eran similares respecto a las variables analizadas, encontrándose como única diferencia una mayor cantidad de problemas en los libros castellano-leoneses. Estos resultados contrastan con los obtenidos en comparaciones internacionales de libros de texto, y apuntan a una mayor influencia de otras variables en las diferencias de rendimiento de los alumnos de ambas comunidades.

*Palabras clave:* libro de texto, enseñanza primaria, resolución de problemas, aritmética, análisis comparativo, análisis de contenido

## Introduction

Problem solving is the cornerstone of the mathematics curriculum of most of the world's education systems (Philpot et al., 2021) and of the theoretical frameworks of international mathematics achievement assessments (Mullis et al., 2020). In this sense, the Trends in International Mathematics and Science Study (TIMSS) theoretical framework defines problem solving as the ability to solve a wide variety of situations, applying the necessary mathematical knowledge to perform reasoning of different types and levels of difficulty (Lindquist et al., 2017).

The results of international assessments show that not all countries and their different regions are equally effective at developing problem-solving skills for their students. To understand the origin of these differences, the TIMSS provides data on several educational and sociocultural variables: family environment, school composition and resources, school climate or school safety and discipline (Mullis et al., 2020). In addition to these variables, other more specific elements of educational practice, such as textbooks, influence student learning (Fagginger et al., 2016; Sievert et al., 2019, 2021).

Considering the above, this paper aims to verify whether different books are used in two Spanish autonomous communities whose students have shown very different levels of mathematical competence. We will examine these differences both in relation to their presence in the classroom and to the number and variety of arithmetic word problems (hereinafter, AWPs) they contain.

## Differences in Problem Solving Performance

One of the objectives of programs such as the TIMSS is to evaluate the success achieved by different educational systems in developing different mathematical skills in students, among which problem solving stands out (Lindquist et al., 2017). The TIMSS establishes different levels of performance based on student scores, which imply different degrees of problem-solving ability. Students at a low level have difficulty solving any problem, students at an intermediate level are able to solve only routine problems, students at a high level solve two-step problems requiring conceptual understanding of integers, and students at an advanced level are able to solve complex multistep problems requiring understanding and reasoning (Mullis et al., 2020).

The TIMSS (2016) results in Spain revealed that, in Castile and Leon (hereafter, C&L), the percentage of 4th grade students with high (33%) and advanced (7%) levels is significantly greater than that in Andalusia (18% and 2%, respectively). Significant differences were also found in all the domains evaluated, both in content (numbers, geometric shapes and measures, and data representations) and cognition (knowing, reasoning and applying).

Differences in performance between students in different countries and regions seem to be due to multiple causes: family involvement in school education, the importance of education for society, the budget allocated to education, or teacher training (Rao et al., 2010). Data from the TIMSS itself and from other reports, such as the PISA, indicate a strong correlation between mathematics achievement and the Social, Economic and Cultural Index of Families (SECI, see Instituto Nacional de Evaluación Educativa, I.N.E.E., 2023). Other studies attributed differences between regions within the same country to other factors (teachers' epistemological beliefs, students' job expectations after schooling, or certain instructional variables, such as the time and type of instruction students receive, see Hippe et al., 2018). In Spain, the differences between Castilian-Leonese and Andalusian students have been attributed to general factors that are difficult to modify, such as greater family SECIs, greater equity of the educational system in C&L, or greater mathematics anxiety among Andalusian students (I.N.E.E., 2023). However, the analysis of certain instructional variables more closely related to educational practice could help us to understand these differences. One of these instructional

variables could be the textbook and the type of activities and problems found in them. In this sense, AWP are particularly relevant.

## What are Arithmetic Word Problems?

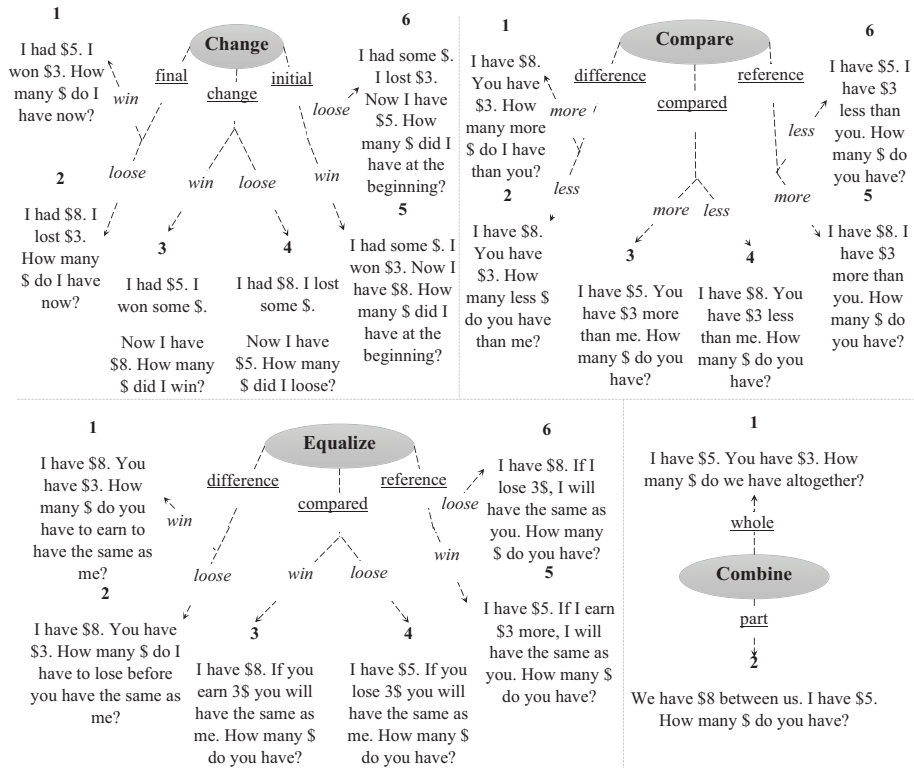
AWPs are verbal descriptions of situations in which one or more questions are posed whose answers can be obtained by applying arithmetic operations to the numerical data of the problem (Verschaffel et al., 2020). AWP can involve different levels of difficulty for students. One of the most influential aspects of the level of difficulty of an AWP is its semantic-mathematical structure (Carpenter and Moser, 1984; Greer, 1992; Vergnaud, 1991). For example, AWP of additive structure (those that can be solved by adding or subtracting) can be categorized as change, compare, combine, and equalize problems (Carpenter and Moser, 1984; Heller and Greeno, 1978). The subcategories described in Figure I can be established based on the unknown set and the relationships (additive or subtractive) between the sets.

Some of these problems can be solved in a routine manner. For example, the Change 2 problem in Figure I can be solved using the “keyword strategy” (Hegarty et al., 1995), taking the problem data and the expression “lost” as the key to subtract, without understanding its mathematical structure. However, difficult problems can only be solved by understanding the relationships between quantities (Verschaffel et al., 2000). For example, to solve the Compare 5 problem in Figure I, it is necessary to understand that, if I have more money than you, you will have less than me, and therefore subtract, even though the expression “more than” suggests that you must add.

Similarly, four types have been distinguished in multiplicative structure AWP (Greer, 1992; Vergnaud, 1991): ratio (or equal groups), multiplicative comparison (or scalars), Cartesian product and rectangular matrix. Depending on the operation required to solve the AWP and the unknown set, the subcategories shown in Figure II can be established.

Given that 34 semantic-mathematical structures have been described for additive or multiplicative structure AWP and that a lack of practice with some types of problems seems to hinder their learning (Siegler and Oppenato, 2021), students need to face as wide a variety of problems as possible to learn to solve them. In fact, the TIMSS includes problems

FIGURE I. Types of additive semantic-mathematical structures



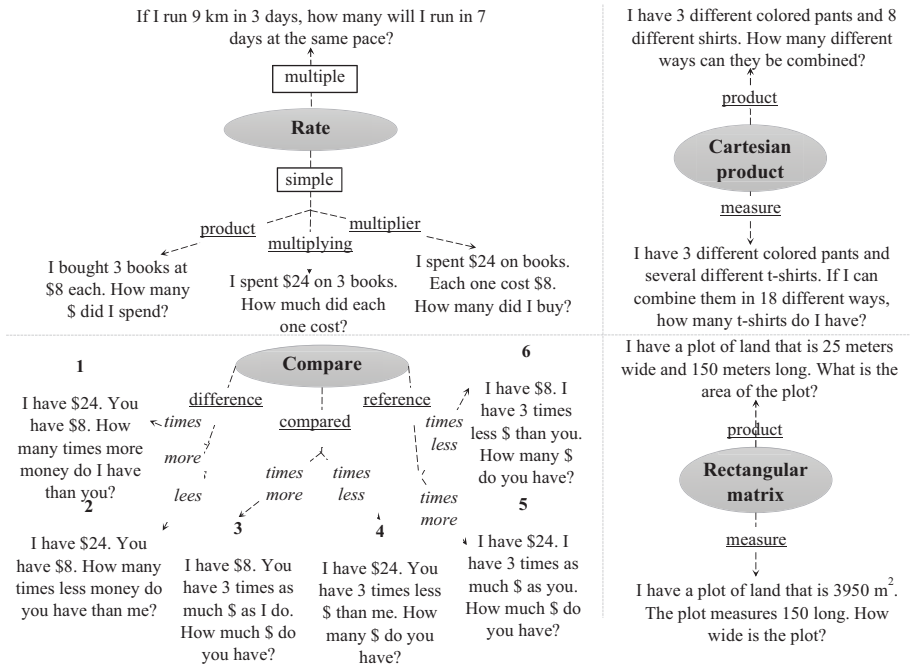
Source: Vicente et al., (2022). Compiled by the authors.

of different structures and levels of difficulty in its assessments (I.N.E.E., 2016). An analysis of textbooks can provide relevant information about the variety of problems that students face in mathematics classes.

## Textbooks and Problem Solving

The textbook is considered a very influential teaching resource in schools for three reasons. First, it is used by the vast majority of teachers (94% of teachers in countries belonging to the Organization for Economic

FIGURE II. Types of multiplicative semantic-mathematical structures



Source: Vicente et al. (2022). Compiled by the authors.

Cooperation and Development, Mullis et al., 2008). In Spain, 81.3% of teachers use it daily (ANELE, 2014); other studies report greater use in autonomous communities such as Madrid (97.14%, Fernández et al., 2013) or the Basque Country (96%, Mullis et al., 2008). Second, they translate the official curriculum into a sequence of concrete actions that teachers and students can follow, acting as a mediator between the official curriculum and the implemented curriculum (Valverde et al., 2002) and largely determining what is taught in the classroom (Oates, 2014). Third, because they influence student learning, in relation to mathematics, students are more proficient in the content to which books devote more space (Schmidt et al., 2001), more exercise and more problems (Törnroos, 2005), and they learn arithmetic principles better and use commonly-appearing problem-solving strategies more frequently (Fagginger et al.,

2016; Sievert et al., 2019, 2021). In contrast, students are less proficient at solving certain fraction and decimal problems that, being mathematically simple, rarely appear in books (Siegler and Oppenzato, 2021).

Some specific textbook variables can influence student performance, such as certain socioeconomic factors (e.g., the type of school), support (digital or paper) or the design of key aspects of the book, such as the distribution of tasks (Behnke, 2018). In relation to problem solving, some studies have shown that books from Eastern countries such as Japan, China or Singapore (whose students have demonstrated a high level of mathematical competence) contain a more diversified and balanced distribution of additive and multiplicative AWP than books from countries such as the United States (Schoenfeld, 1991; Stigler et al., 1986; Xin, 2007) and Spain (Orrantia et al., 2005; Tárraga et al., 2021; Vicente et al., 2018), whose students are less proficient.

In summary, mathematics achievement assessment programs such as the TIMSS show the differences in mathematics achievement and problem solving that exist among students from different geographic and cultural contexts. These differences could be due, in part, to the use of textbooks, since teachers use them often, and students learn better the content that appears most frequently. Therefore, to learn to solve APWs of different levels of difficulty, students need books that include a large number and variety of problems. In fact, books from some countries whose students have shown a greater ability to solve problems include more varied and more difficult AWP than books from other countries whose students seem to be less proficient. Thus, it would be interesting to analyse the existing differences in the most used books in two Spanish autonomous communities, in relation to the most used books, respectively, and to the presence and variety of semantic-mathematical structures of the AWP included in those books. This is a particularly valuable study because the students in these communities (C&L and Andalusia) have shown very different levels of mathematical competence.

## The present work

The aim of this study is to answer the question of whether there are differences in the AWP included in the most commonly used textbooks in the primary school classrooms of C&L and Andalusia by means of



two different studies. In the first study, we will check whether the textbooks most used in 3rd-grade primary education classrooms<sup>1</sup> in C&L are different from those most used in Andalusia. In the second study, we will check whether there are differences in the AWP of the different textbooks according to three aspects: a) percentage of activities aimed at solving AWP versus other mathematical activities; b) variety of semantic-mathematical structures; and c) percentage of problems according to their level of semantic-mathematical difficulty.

## Design and methodology

### Sample

Study 1 included mathematics textbooks used in the 3rd grade of primary schools in public and subsidized schools in C&L (851) and Andalusia (2460) during the 2020-2021 academic year. Books from 709 centres in C&L (83.31% of the total) and 2425 in Andalusia (98.57%) were identified. For Study 2, we selected the books that, being the most used in each autonomous community, were used in at least 80% of the centres in each community.

### Procedure

In Study 1, the books in use in C&L were determined by consulting the web pages of all the primary schools in the community. This information is available on the website of the Regional Department of Education. Those schools that did not publish the list of books on their websites were contacted by e-mail. In the case of Andalusia, the books were published on the website of the Regional Department of Education. The percentage of the total number of centres in which each book was used

---

<sup>1</sup> A single grade's books were analysed so that the sample would be comprehensive, considering the number of variables and books analysed. The 3rd grade of primary school was chosen because at this level, both additive and multiplicative structures appear, and problems with fractions, decimals, or transformations of units of measurement, which involve additional levels of difficulty, are rare.

in each autonomous community was then calculated. To preserve the privacy of the publishers, the name of each publisher was coded with a combination of two numbers, the first referring to the publisher and the second to the publishing project. Subsequently, a second coding was made of the most frequently used books in each autonomous community, representing the order of their percentage of use. Correlative letters were assigned for C&L, and numbers were assigned for Andalusia. The books used in both communities were coded with a combination of letters and numbers. This second coding was used for Study 2.

In relation to Study 2, we first analysed all the activities in the books, understanding “activity” as a task or set of tasks grouped under the same heading, in which the student had to answer one or more questions that required calculations or the application of mathematical knowledge. Each activity was categorized as a) AWP resolution or b) other activities (hereafter OAs). AWP resolution activities were categorized as those that included one or several AWP. AWP resolution activities were considered to be those tasks in which a) a verbal description of real or imaginary situations was made and a mathematical question was posed that was answered by applying at least one of the four basic arithmetic operations; and b) they could be classified in one of the structures described in Figures I and II. The remaining activities were considered OAs, including other types of verbal problems (e.g., statistical, geometric) not classifiable in any of the semantic-mathematical structures of our study. We identified 8,105 activities, of which 2,057 were AWP-solving activities, including 3,834 AWP.

Second, the variety of semantic-mathematical structures was analysed. Different classifications were used to analyse the additive and multiplicative structures; the former were classified as problems of change, compare, combine, and equalize (Carpenter and Moser, 1984; Heller and Greeno, 1978), establishing 20 different subcategories according to the unknown set and the additive or subtractive relationships established between the quantities of the problem (see Figure I). Multiplicative structure AWP resolution activities were classified as ratio (or equal groups), compare (or scalar), Cartesian product and rectangular matrix problems (Greer, 1992; Vergnaud, 1991), establishing 14 different subcategories (see Figure II).

Finally, each structure was categorized with respect to the level of semantic-mathematical difficulty, following the model of Riley and Greeno’s resolution strategies (1988) for additive structures and Greer

(1992) and Vergnaud's (1991) approaches for multiplicative structures. The additive structures of change 1 and 2, compare 1 and 2, equalize 1 and 2, and combine 1 were considered easy, as were the multiplicative structures of simple ratio, since they are the closest to those of additive structure (they can be solved as repeated additions or subtractions). Additive structures of change 3 and 4, compare 3 and 4, equalize 5 and 6, and combine 2 (see Riley and Greeno, 1988, for combination 2 structures); and multiplicative structures of multiple ratios and those of consistent multiplicative comparison—that is, in which the problem terms 'times more' or 'times less' coincided with the operation needed to solve it (a multiplication and a division, respectively, see Xin, 2007)—were considered of medium difficulty. Finally, the additive structures of changes 5 and 6, compare 5 and 6, and equalize of 3 and 4, and the multiplicative structures of inconsistent comparisons, rectangular matrices, and Cartesian products were considered difficult.

## Data coding and analysis

The authors jointly coded 10 didactic units from different books to establish the categorization criteria. Then, two of the authors analysed another 10 didactic units, after which interjudge agreement was calculated using Cohen's kappa coefficient, obtaining highly significant correlation indices ( $\kappa = .98$  for the categorization of AWP and OA resolution activities;  $\kappa = .96$  for the semantic-mathematical structures of the AWPs). The first two authors analysed the structures of the AWPs of the rest of the sample, and the last three authors categorized the activities as AWPs or OAs.

For Study 1, the percentage of use of each publisher in each autonomous community was calculated. For Study 2, two measures were calculated. First, for each variable, a weighted measure of the frequencies observed in the books was calculated according to the percentage of use in each autonomous community using the following formula:

$$M_{CA} = (F_1 \times P_1) + (F_2 \times P_2) + \dots (F_n \times P_n)$$

where  $M_{AC}$  is the weighted measure of books in each autonomous community,  $F$  is the absolute frequency of each book in that variable and  $P$  is the percentage of use of that book in that autonomous community.

Although this measure implies a slight overestimation of the frequencies observed in Andalusian books (the sum of the percentages of use is slightly greater in Andalusia than in C&L), this does not affect the validity of the results.

The second measure of Study 2 was the frequency observed for each variable in each book. This measure allowed comparisons to be made between all the books. To facilitate the interpretation of the results, the percentages corresponding to each observed frequency were calculated.

To test the statistical significance of the differences found, the nonparametric chi-square test was used in all the analyses to analyse the overall differences through planned tables and Fisher's exact test when the chi-square test was not appropriate. To verify the effect of these differences, we used Cramer's V statistic, which, according to Cohen (1988), indicates whether the effect of the differences is small (.1), medium (.3) or large (.5). To compare differences between individual books, z tests with a significance level of .05 were used, using pairwise comparisons of column proportions from the planned tables.

## Hypotheses

According to the theoretical framework presented in Study 1, if books influence the differences in performance in solving AWP between students in C&L and Andalusia, the books most used in 3rd grade mathematics classes in C&L will be different from those most used in Andalusia (Hypothesis 1).

In Study 2, considering the results of previous studies on the influence of books on mathematics learning and the differences in the TIMSS scores of Castilian-Leonese and Andalusian students, it would be expected that the books most used in C&L, compared to those most used in Andalusia, would present several of the following differences:

- a) A greater percentage of AWP resolution activities (Hypothesis 2a).
- b) A greater variety of semantic-mathematical structures (Hypothesis 2b).
- c) A greater percentage of AWP of structures of medium and high semantic-mathematical difficulty that were both additive (Hypothesis 2c1) and multiplicative (Hypothesis 2c2).

## Results

### Study 1

In C&L, 46 books from 16 publishers were used, whereas in Andalusia, 15 books from eight publishers were used. The most frequently used books in C&L were different from those most used in Andalusia, confirming Hypothesis 1. No significant association was found between the number of books used in both autonomous communities ( $p < 0.001$ ), and the difference was large (.81). In C&L, the most commonly used books were 1-1 (30.97%), 1-2 (12.78%), 2-1 (13.35%), and 3-1 (14.48%). However, in Andalusia, the most frequently used books were 1-2 (26.18%), 1-3 (9.6%), 2-2 (31.92%), and 3-2 (11.91%), and 4-2 (8.08%). Similarly, the percentage of use in C&L of 1-1, 2-1 and 3-1 was greater than that in Andalusia; in contrast, in Andalusia, books 2 and 3 from publishers 1 and 2 and books 3-2, 4-2 and 6-1 were used more frequently (see Table I).

Considering the results of Study 1, the five books most used in each autonomous community were selected for Study 2. These books were coded following the procedure described above (see Table I). It should be noted that two books frequently used in Andalusia (4-2 and 1-3) were also used in C&L even though they were not among the five most used publishers in that community, so they were also included in Study 2. Since these two books were the 10th and 7th most-used publishers in C&L, respectively, they were coded with the 10th and 7th letters (“J” and “G”).

### Study 2

#### Percentages of AWP resolution activities

The weighted frequencies of activities dedicated to solving AWP in C&L books were 262 and 693 for OAs (27.4% and 72.6%, respectively). In the Andalusian books, a weighted frequency of 207 AWP resolution tasks and 654 OAs was observed (24% and 76%, respectively). Although the distributions of the tasks in the books of both communities differed—the activities dedicated to solving AWP represented a greater percentage of the C&L books—this difference did not reach statistical significance.

TABLE I. Frequencies and percentages of use of the most used books in 3rd grade of primary schools in the C&L and Andalusia schools

EDITORIAL	BOOK	C&L		ANDALUSIA		Code Study 2
		N	%	N	%	
1	1-1	218	30.97*	0	0	A
	1-2	90	12.78	635	26.19*	D2
	1-3	11	1.56	233	9.61*	J4
	1-4	7	1	0	0	
	<b>Total</b>	326	46.31	868	35.80	
2	2-1	94	13.35*	0	0	C
	2-2	31	4.40	774	31.92*	E1
	2-3	17	2.41	122	5.03*	
	<b>Other</b>	4	0.56	0	0	
	<b>Total</b>	146	20.72	896	36.95	
3	3-1	102	14.49*	0	0	B
	3-2	34	4.83	289	11.92*	F3
	<b>Other</b>	6	0.85	0	0	
	<b>Total</b>	142	20.17	289	11.92	
4	4-1	20	2.84	0	0	
	4-2	18	2.56	196	8.08*	G5
	<b>Other</b>	6	0.85	0	0	
	<b>Total</b>	44	6.25	196	8.08	
5	<b>Unspecified</b>	1	0.14	55	2.27*	
	<b>Other</b>	6	0.86	0	0	
	<b>Total</b>	7	1	55	2.27	
6	6-1	1	0.14	118	4.87*	
	6-2	7	1	0	0	
	<b>Total</b>	8	1.13	118	4.87	
7	7-1	8	1.14	0	0	
<b>OTHER PUBLISHERS</b>		17	1.72*	3	0.11	
<b>NO BOOK</b>		11	1.56*	0	0	
<b>TOTAL</b>		709	100	2425	100	

Source: Compiled by the authors.

Note: "Other publishers" includes books from publishers with absolute frequencies lower than 5 or percentages lower than 0.7 of the total in each community. Asterisks indicate statistically significant differences.

However, the distribution of tasks dedicated to solving AWP and OAs differed among the different books analysed,  $\chi^2(7, n = 8105) = 42.74$ ,  $p < .01$ , although the effect was small (.07). Books A and B (the most used in C&L) devoted a greater percentage of activities to solving AWP; in fact, this difference was significant with respect to book E1, the most-used in Andalusia (see Table II). Similarly, in the two most-used books in C&L, we found higher frequencies than expected, while in the most-used book in Andalusia, we found a lower frequency than expected. Other common books in Andalusia, such as J4 and G5, also included fewer AWP resolution activities than A did and had lower frequencies than expected. These results confirm Hypothesis 2a. Finally, the difference in the total number of tasks between the different books, especially between books A and F3 and J4, is striking.

### Variety of semantic-mathematical structures

The weighted frequency of structures in the Castilian-Leonese and Andalusian books was very similar (17.44 and 19.18, respectively); in fact, the difference in the distribution of these structures in the books of one and the other autonomous community was not significant. Significant differences were not found in the distribution of structures between

TABLE II. Frequencies and percentages of AWP and OA resolution activities by publisher

BOOK	AWP			OAs			N TOTAL
	N	St. Res.	%	N	St. Res.	%	
<b>A</b>	383	7.85	29.6 <sup>E1, J4, G5</sup>	911	5.64	70.4	1294
<b>B</b>	289	1.99	27.9 <sup>E1</sup>	748	-0.29	72.1	1037
<b>C</b>	269	0.74	24	850	3.42	76	1119
<b>D2</b>	307	3.11	28.3 <sup>E1, J4</sup>	776	0.73	71.7	1083
<b>E1</b>	216	-2.56	20.7	829	2.65	79.3 <sup>A, B, D2</sup>	1045
<b>F3</b>	195	-3.87	26	556	-7.27	74	751
<b>J4</b>	150	-6.68	21.6	546	-7.64	78.4 <sup>A, D2</sup>	696
<b>G5</b>	248	-0.57	23	832	2.76	77 <sup>A</sup>	1080

Source: Compiled by the authors.

Note: St. Res. = standardized residual.

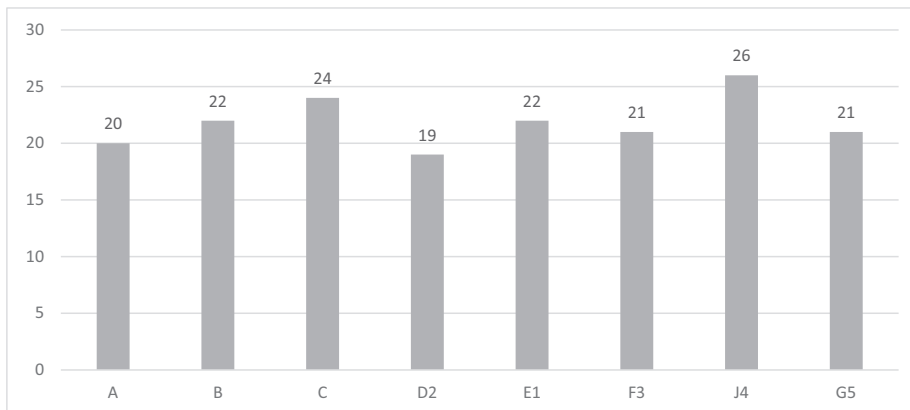
books or in the comparisons made between pairs of books (see Figure III), so the results do not confirm Hypothesis 2b.

### Semantic-mathematical level of difficulty. Additive structures.

The distributions of the weighted frequencies of the three levels of difficulty were very similar in the books of both communities, and no significant differences were found. The C&L books included more easy and medium difficulty structures, although the percentages they represented with respect to the total were practically identical (see Figure IV).

A comparison between the different books showed that the vast majority of the additive structures were easy, while difficult structures were very rare, except in book J4. The distribution of difficulty levels across books was significantly different between books ( $p < .001$ ), although the size of this difference was small (.16). There were no significant differences between the most frequent publishers in C&L and those in Andalusia (see Table III). The only significant differences found were the following: a) books A, D2 and G5 included more easy problems than B and J4, although only in A and D2 were frequencies

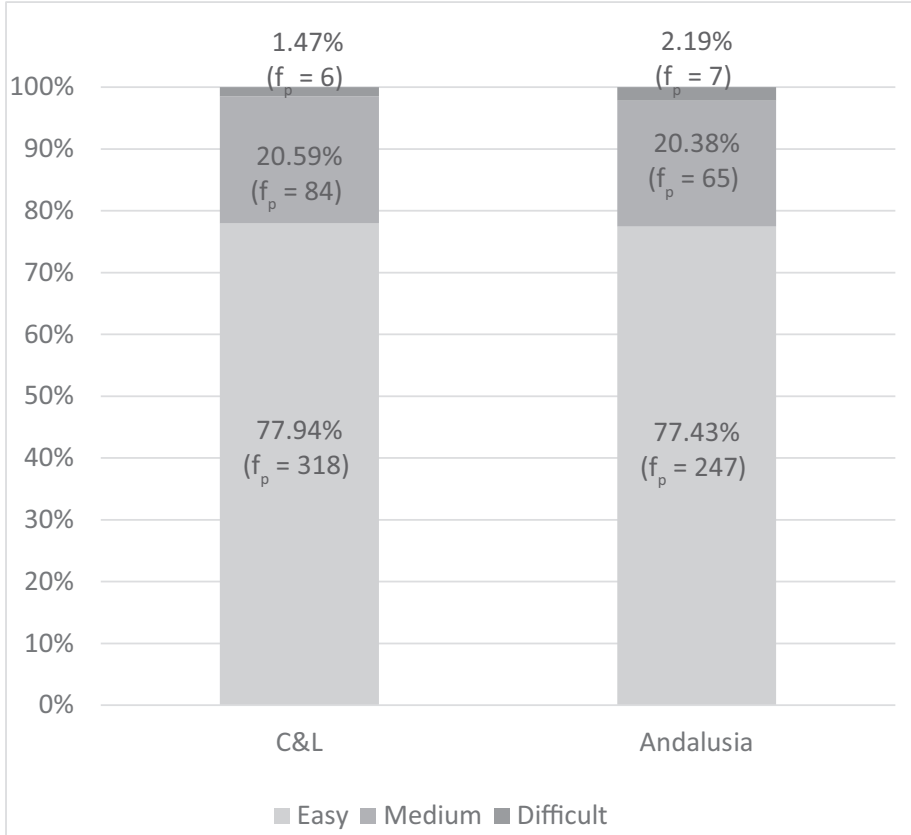
FIGURE III. Number of semantic-mathematical structures included in each book



Source: own elaboration.



FIGURE IV. Frequencies and percentages of additive structures by level of semantic-mathematical difficulty, weighted by autonomous community



Source: Compiled by the authors.

higher than expected; b) book B included more structures of medium difficulty than did A (although the standardized residual of A was higher) and G5, where a frequency well below the expected was found (as was F3). Finally, regarding difficult structures, book J4 included more difficult problems than did the other books, with a frequency well above the expected frequency, while book C included more difficult problems than did A. These results do not confirm Hypothesis 2c1.

TABLE III. Frequencies and percentages of additive structures of each level of semantic-mathematical difficulty by book

BOOK	EASY			MEDIUM			DIFFICULT			N TOTAL
	N	St. Res.	%	N	St. Res.	%	N	St. Res.	%	
A	584	18.13	79.9 <sup>B,J4</sup>	141	7.77	19.3	6	-1.48	0.8	731
B	227	-3.19	68.8	92	2.08	27.9 <sup>A,G5</sup>	11	0.03	3.3	330
C	164	-6.95	77.4	40	-3.96	18.9	8	-0.88	3.8 <sup>A</sup>	212
D2	451	10.19	78.8 <sup>B,J4</sup>	117	4.98	20.5	4	-2.09	0.7	572
E1	226	-3.25	76.6	65	-1.06	22	4	-2.09	1.4	295
F3	132	-8.86	77.6	34	-4.66	20	4	-2.09	2.4	170
J4	248	-1.93	69.3	67	-0.82	18.7	43	9.72	12.0 <sup>A,B,C,D2, E1,F3,G5</sup>	358
G5	211	-4.14	82.7 <sup>B,J4</sup>	37	-4.31	14.5	7	-1.18	2.7	255

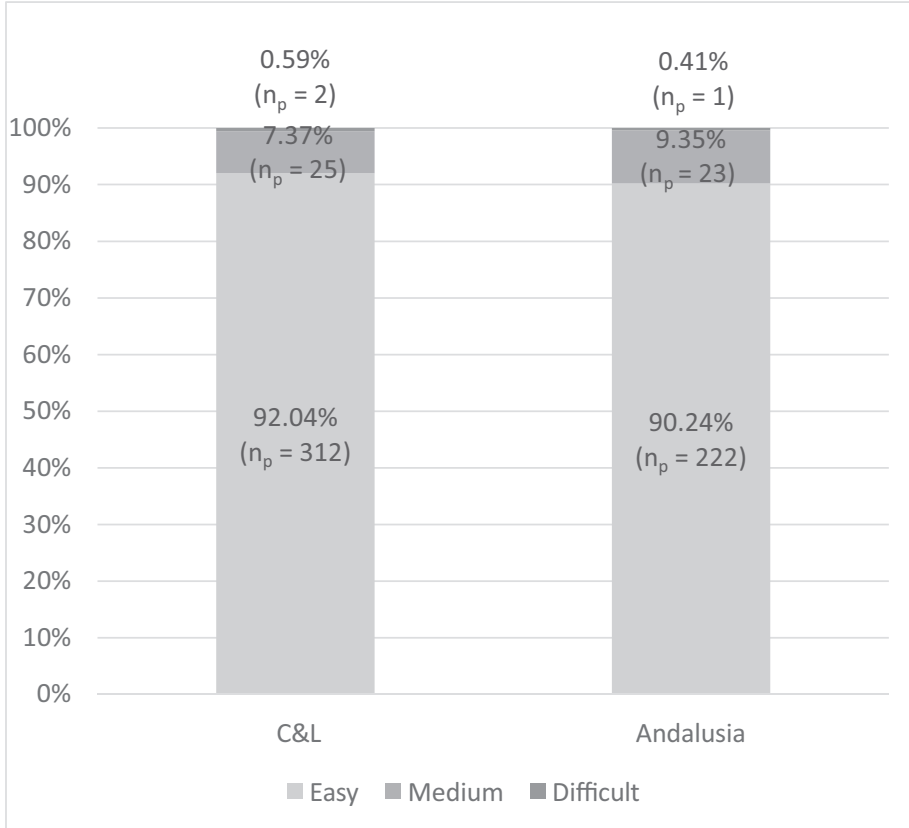
Source: Compiled by the authors.  
 Note St.Res. = standardized residual.

### Semantic-mathematical level of difficulty. Multiplicative structures

Once again, the weighted frequencies of the three levels of difficulty were distributed very similarly in the books of both communities, with no significant differences between the two distributions. As shown in Figure V, in the C&L books, a higher weighted frequency of the three levels of difficulty was observed than in the Andalusian books, although they represented practically the same percentages with respect to the total.

Regarding the comparison between the different books, all showed a vast majority of easy structures and very few difficult structures. The distribution of difficulty levels was different among the books analysed ( $p < .001$ ), but the size of this difference was small (.14). No significant differences were found between the most frequent books in C&L and those in Andalusia (see Table IV). The following differences were found: first, books A and D2, both with frequencies well above what was expected, included more easy structures than did books B, E1, F3 and J4. In addition, books C and G5 included more easy structures than did books B and J4, the latter appearing more frequently than

FIGURE V. Frequencies and percentages of multiplicative structures by level of semantic-mathematical difficulty, weighted by autonomous community



Source: Compiled by the authors.

what was expected. Second, J4 included more structures of medium difficulty than anything other than B and E1, although its difference with respect to the expected value was smaller than that of B. In addition, B and E1 also included more structures of medium difficulty than did A, C, D2 and G5, the latter being the ones that fell farthest below the expected value. Taken together, these results do not confirm Hypothesis 2c2.

TABLE IV. Frequencies and percentages of multiplicative structures of each level of semantic-mathematical difficulty by book.

BOOK	EASY			MEDIUM			DIFFICULT			N TOTAL
	N	St. Res.	%	N	St. Res.	%	N	St. Res.	%	
A	522	15.47	94.2 <sup>B.E1.F3.J4</sup>	32	1.15	5.8	0	-1.55	0	554
B	267	-0.09	84.8	43	3.31	13.7 <sup>A.C.D2.G5</sup>	5	1.68	1.6	315
C	248	-1.25	93.6 <sup>B.J4</sup>	13	-2.56	4.9	4	1.03	1.5	265
D2	440	10.47	93.8 <sup>B.E1.F3.J4</sup>	29	0.57	6.2	0	-1.55	0	469
E1	186	-5.03	86.5	29	0.57	13.5 <sup>A.C.D2.G5</sup>	0	-1.55	0	215
F3	169	-6.07	86.2	23	-0.61	11.7	4	1.03	2	196
J4	113	-9.49	76.4	31	0.96	20.9 <sup>A.C.D2.G5</sup>	4	1.03	2.7	148
G5	203	-4.00	94.9 <sup>B.J4</sup>	9	-3.35	4.2	2	-0.26	0.9	214

Source: Compiled by the authors.  
 Note St.Res. = standardized residual.

## Discussion

According to the TIMSS results (I.N.E.E., 2016), students from C&L obtain better scores on mathematics tests than do Andalusian students; therefore, according to the theoretical frameworks of the TIMSS itself, they are more skilled in solving AWP of different levels of difficulty (Mullis et al., 2020). This difference in performance, which has been attributed to general factors such as differences in SECIs or a more equitable education system (I.N.E.E., 2023), could also be due to the use of textbooks. Books are the most used didactic resource used by teachers to teach problem solving to their students in the classroom (Hiebert et al., 2003; Mullis et al., 2008). Moreover, the variety and level of difficulty of the problems that children face from books influence their ability to solve them (Siegler and Oppenzato, 2021), with the semantic-mathematical structure of these problems being one of the variables that most influences their level of difficulty (Carpenter and Moser, 1984; Greer, 1992; Vergnaud, 1991). On the other hand, some studies have shown that books from countries with more proficient students include problems of varied semantic-mathematical structures (Schoenfeld, 1991; Stigler et al., 1986; Xin, 2007), unlike books from other countries, such as Spain (Orrantia et al. 2005; Tárraga et al., 2021; Vicente et al., 2018). This being so, one

might wonder whether part of the difference in the achievement level of students in C&L and Andalusia could be attributed to the textbooks used in both places. If there was such an influence, one would expect that the textbooks most frequently used in C&L would be different from those most frequently used in Andalusia and, in that case, that the textbooks most frequently used in C&L would contain (a) a greater percentage of AWP resolution activities, (b) a greater variety of semantic-mathematical structures, and (c) a greater percentage of AWP of medium or high semantic-mathematical difficulty, in accordance with the structures and levels of difficulty described in the literature (Carpenter and Moser, 1984; Greer, 1992; Riley and Greeno, 1988; Vergnaud, 1991). The greater the abovementioned differences are, the greater the influence of book design on the differences in the performance of Castilian-Leonese and Andalusian students.

The results of Study 1 showed that different books are used in C&L and Andalusia. However, of the three measures considered in Study 2 to analyse the AWP of these books, only the percentage of activities dedicated to solving AWP points to the fact that the books most used in C&L are more appropriate than those in Andalusia for students to learn to solve problems. If, in addition to the percentage, we consider that the number of activities dedicated to solving AWP in the books most used in C&L is much greater than that in Andalusia (more than double, in some cases), it could be concluded that this more intense practice in problem solving could be a determining factor. These results show that it might not be necessary to have a high percentage of problems with medium or high difficulty (the most commonly used books in C&L did not include a significantly greater variety of semantic-mathematical structures or more problems of medium or high semantic-mathematical difficulty) but rather many opportunities to practice easy problems and, in some cases, more complex AWP. These results qualify those obtained in previous studies comparing book problems from different countries (Orrantia et al., 2005; Schoenfeld, 1991; Stigler et al., 1986; Tárraga et al., 2021; Xin, 2007) and align in part with those of Vicente et al. (2022), which compared Spanish books with those from Singapore. These authors found that the Singaporean books contained a greater proportion of AWP-solving activities than did the Spanish books and more difficult problems, although the effect of the latter difference was small. In the Singaporean books, as in those analysed in our study, the vast majority of the problems were easy. The differences between the books from both

countries seem to lie in the reasoning aids provided to students, such as the use of graphical representations (Kaur, 2015; Vicente, 2022), more complete and reasoning-based solving models (Vicente et al., 2020) or a higher level of problem authenticity (Vicente et al., 2021).

Similarly, it is reasonable to think that the influence of books on student performance is mediated by the ways in which teachers solve problems in class. There is evidence that only teachers with more knowledge about teaching problem solving promote greater reasoning when solving any type of problem. In contrast, teachers with less knowledge tend to avoid the most difficult problems (Ramos et al., 2024). Although some of the most widely used books in Andalusia contain problems of medium difficulty, many may not implement them in the classroom because they consider them inappropriate for their students.

## Conclusions

The results of the present study allow us to draw three conclusions. First, although the textbooks most used in C&L are not the same as those used in Andalusia, these textbooks are, in essence, very similar. Of the three measures analysed, only the percentage of AWP resolution activities seems to indicate a clear difference between the most commonly used textbooks in C&L and Andalusia. The absence of differences in the variety of structures and levels of semantic-mathematical difficulty, both in additive and multiplicative problems, leads us to think that the influence of the books on the differences in the performance of students in C&L and Andalusia seems to be limited to a greater amount of practice with the AWP in the books from Castile and Leon.

A second conclusion is that even though the most commonly used books in C&L contain AWPs with a very small number of structures and low semantic-mathematical difficulty, they seem to be sufficient material for students to develop problem-solving skills above the national average. Beyond the number of AWPs they contain, this may be due in part to characteristics of the books not described in this study and to the ways in which teachers use them in the classroom.

A third conclusion is that some books could be more suitable than others for teaching certain topics related to AWPs since, although no major differences were found between the most commonly used books

in C&L and Andalusia, there are differences between specific books. For example, while A or D2 provide a large bank of problems and problems with a higher level of procedural complexity, J4 presents both a greater variety of structures and more difficult problems at the semantic-mathematical level (also B, for additive structure problems).

## Limitations and foresight

The results of the present study should be interpreted with caution due to the existence of several limitations. First, the sample of books analysed is limited since only 3rd-grade books from two autonomous communities were included. Therefore, it would be advisable to carry out additional studies to extend the sample of books analysed to all grades of primary school, to other autonomous communities and even to other countries. In fact, it would be convenient to replicate this study by comparing books from different regions of other countries. Second, of all the sources of complexity of the problems, only semantic-mathematical difficulty has been considered, leaving aside other variables of the problems, such as their linguistic characteristics, their stereotyped or nonstereotyped character, or the location of the unknown. These issues should be analysed in the future, together with other aspects of AWP, such as resolution aids or their level of authenticity. Third, the AWP included in the books have been analysed, but not their role within the didactic unit. It would be interesting to know what role the small percentage of problems of medium or high semantic- mathematical difficulty play in the didactic units (e.g., as a challenge at the beginning of the didactic unit, as a practice at the end of the unit). Finally, it would be very interesting to compare the ways in which teachers from both autonomous communities use books in class (e.g., their teaching methodology or the way they assess learning).

## Bibliographic references

- ANELE. (2014). *La edición de libros de texto en España*. ANELE.
- Behnke, Y. (2018). Textbook Effects and Efficacy. In E. Fuchs & A. Bock, (Eds), *The Palgrave Handbook of Textbook Studies*. Palgrave Macmillan. [https://doi.org/10.1057/978-1-137-53142-1\\_28](https://doi.org/10.1057/978-1-137-53142-1_28)

- Carpenter, T.P., & Moser, J.M. (1984). The acquisition of addition and subtraction concepts. In R. Lesh y M. Landau (Eds.), *The acquisition of mathematical concepts & processes*, (pp. 7–44). Academic Press.
- Cohen, J. (1988). *Statistical power and analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Inc. <https://doi.org/10.1002/bs.3830330104>
- Fagginger, M., Hickendorff, M., van Putten, C. Beguin, A., & Heiser, W. (2016). Multilevel latent class analysis for large-scale educational assessment data. Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education*, 29(2), 144–159. <https://doi.org/10.1080/08957347.2016.1138959>
- Fernández, P., Caballero, P., & Fernández, J. (2013). ¿Yerra el niño o yerra el libro de Matemáticas? *Números. Revista de Didáctica de las Matemáticas*, 83, 131-148.
- Greer, B. (1992). Multiplication and division as models of situations. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*, (pp. 276–295). Macmillan.
- Hegarty, M., Mayer, R.E., & Monk, C.A. (1995). Comprehension of arithmetic word problems: a comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, 87, 18–32. <https://doi.org/10.1037/0022-0663.87.1.18>
- Heller, J., & Greeno, J. (1978). *Semantic processing in arithmetic word problem solving*. Comunicación presentada en la Midwestern Psychological Association Convention, Chicago.
- Hiebert, J., Gallimore, R., Givvin, K.B., Hollingsworth, H., Jacobs, J., Chui, A.M. et al. (2003). *Teaching mathematics in seven countries. Results from the TIMSS 1999 Video Study*. National Center for Education Statistics (NCES).
- Hippe, R., Jakubowski, M., & Araújo, L. (2018). *Regional inequalities in PISA: the case of Italy and Spain*. Publications Office of the European Union. <https://doi.org/doi:10.2760/495702>
- Instituto Nacional de Evaluación Educativa (2016). *TIMSS 2015. Estudio internacional de tendencias en Matemáticas y Ciencias. (Informe español: resultados y contexto)*. Ministerio de Educación, Cultura y Deporte. [https://sede.educacion.gob.es/publiventa/descarga.action?f\\_codigo\\_agc=18230](https://sede.educacion.gob.es/publiventa/descarga.action?f_codigo_agc=18230)



- Instituto Nacional de Evaluación Educativa (2023). *PISA 2022. Programa para la evaluación Internacional de los estudiantes. Informe español*. Ministerio de Educación, Formación Profesional y Deportes. [https://www.libreria.educacion.gob.es/ebook/184935/free\\_download/](https://www.libreria.educacion.gob.es/ebook/184935/free_download/)
- Kaur, B. (2015). The model method: A tool for representing and visualizing relationships. In X. Sun, B. Kaur & J. Novotna (Eds.), *Conference proceedings of ICMI Study 23: Primary mathematics study on whole numbers*, (pp. 448-455). [http://www.umac.mo/fed/ICMI23/doc/Proceedings\\_ICMI\\_STUDY\\_23\\_final.pdf](http://www.umac.mo/fed/ICMI23/doc/Proceedings_ICMI_STUDY_23_final.pdf)
- Lindquist, M., Philpot, R., Mullis, I., & Cotter, K. E. (2017). TIMSS 2019 Mathematics Framework. In I.V.S. Mullis & M.O. Martin (Eds), *TIMSS 2019 Assessment Frameworks*. Retrieved from: <http://timssandpirls.bc.edu/timss2019/frameworks/>
- Mullis, I., Martin, M., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grade*. TIMSS and PIRLS International Study Center, Boston College. <http://pirls.bc.edu/timss2007/mathreport.html>
- Mullis, I., Martin, M., Foy, P., Kelly, D., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. TIMSS and PIRLS International Study Center, Boston College. <https://timssandpirls.bc.edu/timss2019/international-results/>
- Oates, T. (2014). *Why textbooks count*. Cambridge assessments. <http://www.cambridgeassessment.org.uk/Images/181744-why-textbooks-count-tim-oates.pdf>
- Orrantia, J., González, L. B., & Vicente, S. (2005). Analysing arithmetic word problems in Primary Education textbooks. *Journal for the Study of Education and Development*, 28(4), 429–451. <https://doi.org/10.1174/021037005774518929>
- Philpot, R., Lindquist, M., Mullis, I. V. S., & Aldrich, Ch. E. A. (2021). TIMSS 2023 Mathematics Framework. In I.V.S. Mullis, M.O. Martin & von Davier, M. (Eds), *TIMSS 2023 Assessment frameworks*. TIMSS and PIRLS International Study Center.
- Ramos, M., Vicente, S., Rosales, J., & Chamoso, J. (2024). Influence of teachers' pedagogical knowledge on their classroom practice when solving arithmetic word problems with their students. An exploratory study. *Journal for the Study of Education and Development*, 47(2), 321-345. <https://doi.org/10.1177/021037022412534>.

- Rao, N., Ng, S. S. N., & Pearson, E. (2010). Preschool pedagogy: A fusion of traditional Chinese beliefs & contemporary notions of appropriate practice. In C. Chan, y N. Rao (Eds.), *Revisiting the Chinese learner. CERC studies in comparative education*, (pp. 255–279). Springer. [https://doi.org/10.1007/978-90-481-3840-1\\_9](https://doi.org/10.1007/978-90-481-3840-1_9)
- Riley, M., & Greeno, J. (1988). Developmental analysis of understanding language about quantities of solving problems. *Cognition and Instruction*, 5, 49–101. [https://doi.org/10.1207/s1532690xci0501\\_2](https://doi.org/10.1207/s1532690xci0501_2)
- Schmidt, W., McKnight, C., Houang, R., Wang, H., Wiley, D., Cogan, L., et al. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. Bass.
- Schoenfeld, A. H. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce of formal and informal mathematics. In J.F. Voss, D.N. Perkins & J.W. Segal (Eds.), *Informal reasoning and education*, (pp. 311–343). Lawrence Erlbaum Associates.
- Siegler, R., & Oppenzato, C. (2021). Missing Input: How Imbalanced Distributions of Textbook Problems Affect Mathematics Learning. *Child Development Perspectives*, 15(2), 76–82. <https://doi.org/10.1111/cdep.12402>
- Sievert, H., van den Ham, A. K., & Heinze, A. (2021). Are first graders' arithmetic skills related to the quality of mathematics textbooks? A study on students' use of arithmetic principles. *Learning and Instruction*, 71(101401), 1–14. <https://doi.org/10.1016/j.learninstruc.2020.101401>
- Sievert, H., van den Ham, A. K., Niedermeyer, I., & Heinze, A. (2019). Effects of mathematics textbooks on the development of primary school children's adaptive expertise in arithmetic. *Learning and Individual Differences*, 74(101716), 1–13.
- Stigler, J., Fuson, K., Ham, M., & Kim, M. (1986). An analysis of addition and subtraction word problems in American and Soviet elementary mathematics textbooks. *Cognition and Instruction*, 3, 153–171.
- Tárraga, R., Tarín, J., & Lacruz, I. (2021). Analysis of Word Problems in Primary Education Mathematics Textbooks in Spain. *Mathematics*, 9(17), 2123. <https://doi.org/10.3390/math9172123>
- Törnroos, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation*, 31(4), 315–327.

- Valverde, G., Bianchi, L. J., Wolfe, R., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Kluwer Academic Publishers.
- Vergnaud, G. (1991). *El niño, las matemáticas y la realidad*. Trillas.
- Verschaffel, L., Depaepe, F., & Van Dooren, W. (2020). Word problems in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education*, (pp. 908–911). Springer.
- Verschaffel, L., Greer, B. y De Corte, E. (2000). *Making sense of word problems*. Swets y Zeitlinger Publishers. <https://doi.org/10.1023/A:1004190927303>
- Vicente, S., Machado, E., & Verschaffel, L. (2018). Solving arithmetic word problems. An analysis of Spanish textbooks. *Culture and Educación*, 30, 71-104. <https://doi.org/10.1080/11356405.2017.1421606>
- Vicente, S., Sánchez, R., & Verschaffel, L. (2020). Word problem solving approaches in mathematics textbooks: a comparison between Singapore and Spain. *European Journal of Psychology of Education*, 35, 567–587. <https://doi.org/10.1007/s10212-019-00447-3>
- Vicente, S., Verschaffel, L., & Muñoz, D. (2021). Comparison of the level of authenticity of arithmetic word problems in Spanish and Singaporean textbooks. *Cultura and Education*, 33(1), 106-133. <https://doi.org/10.1080/11356405.2020.1859738>
- Vicente, S., Verschaffel, L., Sánchez, R., & Muñoz, D. (2022). Arithmetic word problem solving. Analysis of Singaporean and Spanish textbooks. *Educational Studies in Mathematics*, 111, 375–397. <https://doi.org/10.1007/s10649-022-10169-x>
- Xin, Y.P. (2007). Word problem solving tasks in textbooks and their relation to student performance. *The Journal of Educational Research*, 6, 347–359. <https://doi.org/10.3200/JOER.100.6.347-360>

**Contact address:** Santiago Vicente Martín. Universidad de Salamanca. Facultad de Educación. Dpto. Psicología Evolutiva y de la Educación. Paseo Canalejas 169. 37008, Salamanca. E-mail: sanvicente@usal.es