

Arithmetic word problems in Primary Education. An analysis of teaching guides

Problemas aritméticos verbales en Educación Primaria. Un análisis de guías didácticas

DOI: 10.4438/1988-592X-RE-2022-396-536

Raúl Tárraga-Mínguez

Julio Tarín-Ibáñez

Universidad de Valencia

Abstract

Introduction: Textbook teaching guides that complement textbooks are key tools that guide teachers when implementing the curriculum in the classroom, especially in terms of assessment, one of the most relevant elements of the curriculum due to the repercussions that it entails. **Methodology:** this study analyzes the current scenario of arithmetic word problems in seventy-eight assessment tests of seventy-two primary mathematics' teaching guides created by six Spanish publishers. We addressed two study objectives: to find out the word problem frequency compared to other routine tasks and to characterize the problems according to semantic structure, level of challenge, and the statement's situational context. **Results:** our findings reveal that these tests contain a low number of arithmetic word problems compared to other types of exercises. In addition, the problems included in these tests show little variability in terms of their semantic structure, with problems mostly belonging to the subcategories of consistent arithmetic word problems (the easiest subcategories), which involve a low level of challenge and lack situational context that increases the understanding of the statements. These results coincide with those obtained by previous research, which has been carried out with curriculum materials published during previous legislative frameworks. This shows that, despite the changes in the educational laws in Spain, publishing houses have barely modified how they handle problem solving. **Conclusions:** the teaching guides do

not constitute adequate tools to evaluate mathematical competence of problem-solving skills in primary school students. In addition, they can contribute to developing superficial and passive resolution strategies.

Keywords: Assessment, teaching guides, mathematics, primary education, problem-solving, textbook.

Resumen

Introducción: Las guías didácticas que complementan a los libros de texto son un material clave que orienta al profesorado sobre algunos aspectos relevantes referentes a la concreción del currículum en el aula, especialmente en la evaluación, uno de los elementos curriculares más relevantes por las repercusiones que conlleva. **Metodología:** En el presente estudio se analiza el tratamiento de los problemas aritméticos verbales en setenta y ocho pruebas de evaluación incluidas en las guías didácticas de matemáticas publicadas por seis editoriales españolas. El análisis se dirige a conocer cuál es la frecuencia y variabilidad de los problemas frente a otras tareas rutinarias y cuál es su caracterización de acuerdo con su estructura semántica, su grado de desafío y el contexto situacional en que aparecen. **Resultados:** Los resultados muestran que estas pruebas contienen una proporción escasa de problemas en relación a los de ejercicios de aplicación mecánica. Asimismo, los problemas incluidos en estas pruebas se caracterizan por presentar una escasa variabilidad en su estructura semántica, por pertenecer mayoritariamente a las subcategorías de problemas aritméticos verbales consistentes (las más sencillas de resolver), por implicar un escaso grado de desafío y por carecer de un contexto situacional enriquecido. Estos resultados son, además, coincidentes con los obtenidos por investigaciones previas, que se han llevado a cabo con materiales curriculares publicados en marcos legislativos anteriores, lo que muestra que, a pesar de los cambios en las leyes orgánicas de educación, las editoriales no han modificado a penas el tratamiento que otorgan a la solución de problemas. **Conclusiones:** Se concluye, por tanto, que estas pruebas editadas en las guías didácticas no constituyen herramientas adecuadas para evaluar la competencia matemática en proceso de la resolución de problemas de los alumnos de Educación Primaria, y que pueden llegar a contribuir al desarrollo de estrategias de resolución superficiales y pasivas.

Palabras clave: Educación primaria, evaluación, guías didácticas, libros de texto, matemáticas, resolución de problemas.

Introduction

The role of textbooks and teaching guides in the classroom.

Textbooks are an inherent part of education as we know it. According to Area (2000), “if we had to choose a representative symbol of education, surely many would be inclined to refer to textbooks” (p.189). In fact, current research has shown that textbooks play a hegemonic role in most educational systems in developed countries (Escudero, 2015; Fuchs and Bock, 2018).

Although it is not the only resource used by teachers, nor is there a homogeneity regarding its use, data is overwhelming. According to the Spanish National Association of Book and Teaching Material Publishers (ANELE), 81.30% of teachers used textbooks as a main resource in 2014. These teachers acknowledged that they used textbooks on a daily basis. Furthermore, 71.90% of parents considered textbooks to be essential in the education of their children, both in educational centers and at home. Its influence is so decisive that for a long time, classic authors such as Apple (1992), or more recently Gimeno (2015), considered the textbook to be the real curriculum embodied in educational practice, or the authentic interpreter of the official curriculum when referring to implementing the different levels of the curriculum.

However, textbooks are not an isolated element in the educational context. The teaching guide that accompanies it is also a decisive instrument in terms of determining which curriculum is actually taught and evaluated in school. In Spain, after the enactment of the General Education Law in 1970, a new publishing modality appears that will replace the previous textbooks that only included answers: teaching guides. From that moment on, a new phase which is still in force begins. It will now be the teaching guide and not the teacher, who will be in charge of interpreting and operationalizing the requirements of the official curriculum: what, how and when to teach; and also what, how and when to evaluate. Thus, like textbooks, teaching guides can be considered windows into school reality or the current curriculum.

The relevance of problem solving in mathematics education.

The importance of problem solving (PS) in the mathematics teaching-learning process is a foundational aspect accepted by the entire community of mathematics educators (Piñero, Castro-Rodríguez, and Castro, 2019). Furthermore, the 126/2014 Royal Decree, which establishes the basic curriculum for primary education, states that: “problem-solving processes constitute one of the main axes of mathematical activity and must be the main source and support of learning throughout the educational stage, given that they constitute the cornerstone of mathematics education” (p.33). The RD organizes the content of the mathematics curriculum into five large blocks: mathematical processes, methods, and attitudes; numbers; measurements; geometry; and statistics and probability, highlighting the importance of the first block, corresponding to PS, which: “has been formulated with the intention of being the backbone of the rest of the blocks” (p.33).

In the same way, different international evaluations (IEA’s TIMSS for primary education or OECD’s PISA for secondary education), consider PS to be a key process for the evaluation of different cognitive levels. In this sense, according to Piñero et al. (2019), these international assessment frameworks use PS as a fundamental indicator when evaluating the mathematical competence of students and the quality of educational systems.

However, the results of these tests show that in Spain, both mathematics, in general, and PS, in particular, are the Achilles heel of a considerable number of students. International reports (TIMSS, 1995, 2011, 2015 and 2019), confirm the need for the Spanish educational system to focus attention on mathematics. In 1995, Spain participated for the first time in this evaluation, assessing levels of 7th and 8th grade students. It ranked 32nd out of 39 participating countries for 7th grade, and 31st out of 41 for 8th grade. With 4th grade students, in the 2011, 2015 and 2019 editions, Spain was below the average of the OECD and EU countries in the four mathematical content domains and in the three cognitive domains, with significant differences in the cognitive domain “reasoning”, which corresponds to PS.

There is no doubt that the reasons behind these results are complex and are due to factors of various kinds. To unravel this issue, research has focused on the one hand, on the indices of the social, economic, and

cultural status of each country (ISEC, according to TIMSS-2019), and on the other hand, on the policies that regulate the different educational systems, especially in terms of teacher professional development and curriculum materials.

Regarding curriculum materials, textbooks take on a special significance. This significance has led researchers to conduct a large number of studies dedicated to analyzing textbooks from multiple perspectives (Fuchs and Bock, 2018; Vojří and Rusek, 2019). In terms of PS, the precursor study in Spain was carried out by Orrantia, González, and Vicente (2005) with mathematics textbooks published in the normative framework of the LOGSE (1990). Based on this pioneering work, other studies have been carried out which aimed to verify the evolution of the scenario described in this initial study (Chamoso, Vicente, Manchado, and Múñez, 2014; Vicente and Manchado, 2017; Vicente, Manchado, and Verschaffel, 2018). At an international level, some of the most recent studies are those of Cai and Jiang, (2017); Tarim (2017); Van Zanten and Van den Heuvel-Panhuizen, (2018); or Yang and Sianturi (2020).

All these studies have made it possible to understand what type of problems are solved by students on a daily basis. However, to the best of our knowledge, these analyses have not been carried out with the teaching guides. Therefore, given the importance of the guides as documents in which students' assessment tests are presented in a pre-prepared way, the main contribution of this study is to analyze the relevance of arithmetic word problems (hereinafter, AWP) in the mathematics assessment tests for primary education, published in the teaching guides of six of the most relevant textbook publishers in Spain.

To achieve this purpose, we addressed two specific objectives: a) to examine the types of items that appear in the assessment tests of the teaching guides by analyzing the difference in frequency between exercises and word problems; and b) to characterize these word problems based on three variables: their semantic structure, their degree of challenge, and the statement's situational context.

We consider that these objectives are of interest, given that it is the first study carried out in Spain that analyzes these variables in the teaching guides of mathematics textbooks. In addition, we believe that the results can help to determine if these materials pay sufficient attention to the evaluation of PS processes in primary education, and if the problems used are appropriate from a pedagogical point of view.

Method

Materials

The study sample included the mathematics teaching guides from six publishing projects: Grupo Santillana (“Know How”); Grupo Anaya (“Learning is growing”); Ediciones S.M (“Savia”); Grupo Vicens Vives (“Active Classroom”); Grupo Edebé (“Talentia”); and Grupo Edelvives (“Superpixépolis”), published between 2014-2015 when the LOMCE came into force (2013).

The analysis focused on the different assessment tests used to evaluate the mathematical learning acquired by students, both at the beginning of a school year and at the end of it, proposed by each publisher.

Taking into account that there were six selected publishers, six grades in primary education, and two assessment tests for each publisher (initial and final assessment), seventy-two assessment tests were analyzed. In addition, Santillana offers complementary “advanced” assessment tests for each grade, so six more tests were added to the total. Therefore, the final number of analyzed tests was seventy-eight.

Word problem variables

To analyze the frequency and variability of each of the items, a coding system was created according to the two following variables:

- a) Item type, distinguishing between exercises and word problems.
- b) Problem characterization according to their semantic structure, degree of challenge, and situational context.

Word problem vs. exercise

A fundamental question for the coding of this variable was the conceptual delimitation of the word problem and the exercise. To do this, we relied on definitions that emphasize the difference between the two concepts: the word problem differs from the exercise in that the solver does not previously have a procedure or algorithm, a solution scheme or

standard procedure that leads with certainty to a solution. Therefore, word problems are conceived as non-routine tasks, as a challenge or a challenging and reflective situation, where there are no ways to reach the solution automatically (Schoenfeld, 1985).

An exercise, on the contrary, is a routine, mechanical and reproductive task that leads directly to the solution through the application of previously learned knowledge. Thus, while the “word problem implies thinking”, “the exercise implies mechanizing” (Alsina, 2006, p.114).

Furthermore, exercises are not contextualized, they are not associated with any specific situational context, while word problems, in addition to having a conceptual or mathematical nature, have a textual and contextual nature, given that the first step to solve any word problem is obviously by reading the problem statement. A widely accepted definition of AWP's assumes that these are verbal descriptions of problematic situations in which one or more questions are posed and the answer must be obtained through reasoning and the application of mathematical operations from the numerical data available in the statement (Verschaffel, Depaepe, and Van Dooren, 2020).

Based on these general criteria and following the coding system used in the pioneering study by Orrantia et al. (2005) with textbooks, items expressed through verbal language were considered word problems. E.g., “The playground for the little children measures 63 steps, and the one for the older children, 97 steps. How many more steps does the playground for older children have?” (Anaya), was considered a word problem. But “ $97 - 63 = ?$ ” was regarded as a mathematical exercise.

However, circumstances expressed through verbal language such as “How much is needed to reach one euro? Data: a 50 cent coin, a 20 cent coin, and a 10 cent coin ”(Vicens Vives), were not considered word problems, because even though it is a verbal description, it does not appear within a situational context.

Word problems are also defined as verbal descriptions at the end of which one or more questions are asked. Therefore, items without explicit questions were not coded as word problems. E.g., “Daniela left home at 8:30 in the morning. The journey to the airport took 30 minutes. Parking and check-in luggage, half an hour. When she finished, she went to the departure lounge and waited 15 minutes before the plane took off. The plane took off at (...)” (Anaya).

Finally, for word problems that asked several questions after presenting the information, each question was coded separately, given that it is the question that determines the semantic structure of the problem, resulting in as many categories as formulated questions. E.g., “The ten books used in second grade cost 235 euros and school supplies cost 97 euros. When buying the books and school supplies, Silvia paid with a €500 bill. How much do books and school supplies cost? How much money did she get back? How much does a family with 3 children pay for books?” (Vicens Vives). In this example, three independently categories are distinguished: combine 1 (first question), change 2 (second question), and multiplication (third question).

Semantic structure

The eighteen categories proposed by Heller and Greeno (1978) were used for the word problem’s semantic structure codification: two subtypes of combine problems, six of change, and six of compare; as well as the six categories of matching problems proposed by Carpenter and Moser (1983).

The consistency hypothesis proposed by Lewis and Mayer (1987), which distinguishes between the language used in problems as being consistent or inconsistent, was also taken into account. Consistent problems, which are easier to solve, have a consistency or coherence between the surface structure of the problem and the algorithm needed to solve it. E.g.: “Juan has 3 marbles. In one game he **wins** 5 marbles. How many marbles does Juan have now?” ($3 + 5 = 8$).

However, in inconsistent problems, this “keyword” indicates the opposite algorithm, so that terms such as “win” are used when subtraction is needed to solve the problem. E.g.: “Juan has some marbles. In one game he **wins** 5 marbles. Now Juan has 8 marbles. How many marbles did he have?” ($8 - 5 = 3$).

For the coding of complex arithmetic word problems, the categorization system proposed by Orrantia et al. (2005) is used. It includes eleven categories, although as the authors point out, the possibility of identifying new categories is contemplated. E.g., category A: “Sergio had 150 euros. On his birthday, his father gave him 35 euros and his mother 46 euros. How much money does Sergio have now?”. This problem presents both

a change and a combination, with the change structure being the main one.

Problems combining various arithmetic operations were also coded. These problems (which are counted in parentheses in Table 3) were coded in the corresponding addition and subtraction category. E.g.: “Lydia made 20 collages and Caesar made half as many as Lydia. Teo made twice as many collages as Lidia. How many collages did they make together?” (Santillana).

Finally, because we were interested in understanding the entire typology of word problems used in the assessment tests, problems that required only multiplication and/or division to answer the questions were also coded, although a further categorization of these problems was not carried out.

Degree of challenge

The second variable that was analyzed was the word problem’s degree of challenge. The expression “degree of challenge” refers to problems that go beyond the selection of data and the execution of the corresponding operation. For analyzing this variable we also used the categorization system from Orrantia et al. (2005), which considers the general categories of information and invention.

- a) Superfluous information (extra data): irrelevant information appears that must be discarded for a correct understanding and resolution of the problem. E.g.: “Ana bought a box of 15 paints. Her friend Marta gives her another box containing 7 pens and 9 paints. How many paints does Laura have now?”
- b) Missing information (minus data): data necessary to find a solution is omitted. E.g.: “Mario has gone to the park to play marbles with his friends. Mario has 17 marbles, and his friend Jorge gives him 7. How many marbles does Jorge have left?”
- c) Total invention: from given elements or other structurally similar or different problems, the student is asked to formulate a totally new problem. E.g.: “Formulate a problem based on these data: children’s tickets cost 8 euros and adult tickets cost 12 euros”.

- d) **Partial invention:** complete the problem with the question or with some data. E.g.: “Marta is 12 years old, her brother Juan is 9 years old, and her cousin Sara is 7 years old.”

Situational context

The last of the characteristics we analyzed was the situational context of the problems. Standard word problems are those that are devoid of any kind of background information. These are very short problems in terms of the information they provide: only premises with data and questions. According to Staub and Reusser (1995), all the necessary information to solve these problem is present in the statement and all the information in the statement is necessary to answer the question. However, these problems can be enriched by including background information to help understand the problem statement. To characterize this variable, the study by Orrantia et al. (2005), which establishes a series of categories based on the Reusser (1990) model was used: description, intention, action, cause, and time. E.g., intentional information referring to the protagonist’s needs, purposes, goals, aims, or motives: “Ivan wants to buy some swimming goggles...”, (Santillana); causal information: “A farmer collected 450 kilos of grapes. He removed 63 kilos because they were damaged...” (Santillana). Furthermore, the possible combinations of the previous categories were coded: e.g., action + intention: “This week we collected money to help children in a country where there had been a flood...” (Anaya).

Content analysis procedure and reliability

To ensure that the item codification process was reliable, an inter-rater reliability analysis was carried out.

Regarding the distinction between word problems and exercises, the second author of the study codified all the items included in the teaching guides. Subsequently and independently, the first author encoded 100 items randomly selected from the set of items in the unit of analysis. Additionally, and in order to ensure the reliability of the process, four researchers with a PhD in Education or Educational Psychology carried out the coding of a total of 40 items also randomly selected from among the items that made up the unit of analysis.

In terms of the analysis of the problem's semantic structure, degree of challenge, and situational context, again the second author carried out the coding of all the items. In this case, the first author independently coded 120 word problems, and then five researchers with a PhD in Education or Educational Psychology coded 10 word problems according to the semantic structure and 5 word problems according to their degree of challenge and situational context.

Finally, Cohen's Kappa index was calculated with the SPSS 27 statistical package (see Table 1), to determine the degree of agreement between the different codifications. This index takes into account, not only the degree of agreement between coders, but also the degree of agreement that can be attributed to chance, thus providing a more reliable indicator than just the percentage of agreement.

TABLE I. Value and interpretation of Cohen's Kappa index for inter-rater reliability analysis

Aspect subject to inter-rater reliability	Number of coders and items evaluated	Overall agreement %	Cohen's κ	I.C (95%)	Consistency (Landis and Koch,1977)
Exercises vs. Word problems	Two coders, 100 items	95.83%	.95	.90-.99	Almost perfect
	Five coders, 40 items	95%	.90	.77-1.0	Almost perfect
Semantic structure	Two coders, 120 problems	88.33%	.83	(.76-.90)	Almost perfect
	Five coders, 10 problems.	68%	.67	(.76-.90)	Substantial
Degree of challenge	Two coders, 120 problems	95.83%	.94	(.90-.99)	Almost perfect
	Five coders, 5 problems	100%	1	-	Perfect
Situational context	Two coders, 120 problems	90.83%	.89	(.93-.95)	Almost perfect
	Five coders, 5 problems	84%	.82	(.60-.1)	Almost perfect

Source: Own elaboration

Results

Item distribution in the assessment tests

First, the results corresponding to the distribution of the analyzed items from the assessment tests are presented, distinguishing between exercises and word problems. As can be seen in Table 2, the total distribution of the items (1904) is very uneven. The assessment tests consisted mainly of routine tasks, that is, exercises (82.70%) and to a lesser extent word problems (17.30%). Table 2 also reveals that the six publishers present a similar scenario in terms of the low proportion of items dedicated to evaluating problems.

TABLE 2. Total results of item distribution in the assessment tests, distinguishing between exercises and word problems.

PUBLISHER	ITEMS	EXERCISES	WORD PROBLEMS
SANTILLANA	421	315 (74.8%)	106 (25.2%)
ANAYA	159	139 (87.4%)	20 (12.6%)
S.M.	110	88 (80%)	22 (20%)
VICENSIVIVES	384	311 (81%)	73 (19%)
EDEBÉ	203	169 (83.3%)	34 (16.7%)
EDELVIVES	627	554 (88.4%)	73 (11.6%)
TOTAL	1904	1576 (82.8%)	328 (17.2%)

Source: Own elaboration

AWP characterization according to semantic structure and degree of challenge

The analysis of all 328 word problems (see Table 3) shows that there were 163 simple problems (49.70% of the total), 42 complex problems (12.80%), 115 multiplication and/or division problems (35.10%), and 8 that implied some degree of additional challenge (2.44%).

The first relevant result reveals the low variability in terms of the word problem's different semantic categories and subcategories. Most are concentrated in the simplest subcategories: combine 1 and change 2 (between them they concentrate more than 40% of the simple problems). The rest of the categories presented a practically marginal frequency. According to the consistency hypothesis, there is a tendency to overrepresent problems that are easier to solve. Thus, of the 163 simple problems presented by publishers, 143 (87.73%) were consistent problems (easier to solve), while only 20 were inconsistent problems (12.27%).

The analysis of complex AWP's offered a similar scenario, characterized by low subcategory variability. Of the eleven categories of complex word problems proposed by Orrantia et al. (2005), only six were considered, although most were concentrated in a single category: "A".

Finally, both the multiplication and/or division word problems, as well as the mixed problems, which combine addition, subtraction, multiplication, and division (in parentheses in Table 3), begin to be included by publishers in second grade assessment tests, at which time the multiplication algorithm is introduced into the official curriculum.

The scarce presence of problems that include an additional degree of challenge is also noteworthy: only 8 out of the 78 problems from the assessment tests required students to partially invent a problem (simpler task). The category of total invention is not contemplated, nor the information category: problems with superfluous or omitted information.

TABLE 3. Total results of word problem frequency and variability in the assessment tests for each grade

Problem category/grade		1°	2°	3°	4°	5°	6°	TOTAL
SIMPLE PROBLEMS	CA1	1	1	1	1	1	0	5
	CA2	1	8(3)	7(4)	4(8)	3(3)	2(7)	50
	CB1	8	10(1)	12(6)	3(17)	2(18)	4(4)	85
	CB2	1	0	0	2(1)	0	(5)	9
	CPI	2	(2)	4	1	0	2	11
	CP2	0	(2)	0	0	0	0	2
	CP3	1	0	0	0	0	0	1
CONSISTENCY HYPOTHESIS	CONSISTENT	11	25	30	33	27	17	143
	INCONSISTENT	3	2	4	4	0	7	20
TOTAL SIMPLE		14	19(8)	24(10)	11(26)	6(21)	8(16)	163 (49.70%)
COMPLEX PROBLEMS	A	0	0	3	9	6	8	26
	B	0	0	0	0	2	2	4
	C	0	0	0	0	1	0	1
	D	0	1	1	0	1	1	4
	E	0	1	0	2	0	1	4
	F	0	0	2	0	1	0	3
TOTAL COMPLEX		0	2	6	11	11	12	42 (12.80%)
MULTIPLICATION AND/OR DIVISION PROBLEMS		0	6	23	25	30	31	115 (35.06%)
PROBLEMS WITH ADDITIONAL DEGREE OF CHALLENGE		1	4	2	1	0	0	8 (2.44%)
TOTAL		15 (4.57%)	39 (11.89%)	65 (19.82%)	74 (22.56%)	68 (20.73%)	67 (20.43%)	328 (100%)

CA = Change; CB = Combine; CP = Compare. In parentheses, problems with addition and/or subtraction + multiplication and/or division.

Source: Own elaboration.

After the general analysis of the results, we proceeded to compare the role that the problems play in the assessment tests of the six publishers. The results (see Table 4) show that there are three publishers that include a significant number of word problems (Santillana, Edelvives, and Vicens Vives), while three publishers (Edebé, S.M., and Anaya) include a significantly lower number of word problems. However, a detailed analysis shows that how publishers use word problems in assessment

tests is similar, given that the proportion of consistent problems is in all cases greater than that of inconsistent. Furthermore, the distribution of complex problems in textbooks from all publishers is very low in terms of frequency and variability.

TABLE 4. Total results of word problem frequency and variability according to each publisher

	SANTILLANA	EDELVIVES	VICENS VIVES	EDEBÉ	S.M	ANAYA
TOTAL SIMPLE	24(32) (17.0%)	28(16) (13.4%)	13(13) (8.0%)	6(7) (4.0%)	8(4) (3.0%)	9(2) (3.0%)
CONSISTENT	52 (15.8%)	40 (12.10%)	22 (6.7%)	12 (3.6%)	11 (3.3%)	9 (2.7%)
INCONSISTENT	4 (1.2%)	4 (1.20%)	4 (1.2%)	1 (0.3%)	1 (0.3%)	2 (0.6%)
TOTAL COMPLEX	26 (8.0%)	5 (1.5%)	3 (1%)	5 (1.5%)		4 (1.2%)
MULTIPLICATION AND/OR DIVISION	24 (7.3%)	19 (5.7%)	44 (13.4%)	15 (4.5%)	9 (2.7%)	4 (1.2%)
CHALLENGE		5		1	1	1
TOTAL	106 (32.3%)	73 (22.2%)	73 (22.2%)	34 (10.3%)	22 (6.7%)	20 (6.1%)

Source: Own elaboration

AWP characterization according to situational context

Of the 328 problems, only 48 (14.6%) contained situational information of some kind (see Table 5). The analysis by grade showed that the teaching guides begin to enrich the background of the word problems from third grade on. In first and second grade, this support for mathematical and contextual understanding of word problems is non-existent, precisely at a moment when they are most necessary, given that students begin to solve problems in a formal way.

The most frequent categories were those referring to the actions of the protagonists, which in theoretical terms are the least relevant for the understanding and creation of the episodic situation model proposed by Reusser (1990). The rest of the categories appeared in a very low proportion. Given that the number of word problems is so low, and because there are hardly any differences between publishers, the data for this variable are presented as a whole, without specifying the distribution of problems by publishers.

TABLE 5. Results of word problem frequency and variability taking into account the situational context

Situational information/grade	1°	2°	3°	4°	5°	6°	TOTAL
Action	1		2	7	7	2	19
Description			3	2	2		7
Time							0
Cause						1	1
Intention			3	1	3	1	8
All							0
Action + description			2	2	3		7
Action + time							0
Action + cause			1		1		2
Action + intention		1		1	1	1	4
TOTAL	1	1	11	13	17	5	48

Source: Own elaboration.

Discussion

In order to examine to what extent the assessment tests of the teaching guides constitute effective instruments to evaluate the mathematical competence of schoolchildren, our study aimed to analyze (a) how publishers use word problems compared to exercises and (b) the characterization of these AWP according to their semantic structure, degree of challenge, and the situational context depicted in the problems.

Results concerning the number of word problems vs. exercises are totally irregular. Most of the tests designed by the six publishers present a similar scenario, characterized by a high number of exercises compared to a very limited number of items dedicated to problem solving, which according to the international framework established by TIMSS (2019) would encourage and promote reasoning skills.

On the other hand, the analysis of the word problem's semantic structure, degree of challenge, and situational context offers a discouraging picture. In terms of semantic structure, the most relevant findings show a low variability of AWP types and subtypes, given that, of the twenty categories of simple problems, the six publishers only include a total of

six subcategories in their assessment tests. In addition, we observed a low frequency of inconsistent problems (more difficult to solve), compared to consistent ones, which can be solved by using superficial strategies. This scenario coincides both with national studies that have analyzed textbooks in Spain (Chamoso et al., 2014; Orrantia et al., 2005; Vicente et al., 2018), as well as international studies that have also studied this variable (Despina and Harikleia, 2014; Tarim, 2017). Thus, it is common to find the same word problem categories in textbooks and in the teaching guides from this study: on the one hand, combine 1, change 1 and 2, or compare 2 and 3 (of a consistent nature) AWP; on the other hand, word problems that are categorized as combine 2 and compare 1, the inconsistent problems that are easier to solve from a structural point of view. The rest of the word problems have a minimal or even residual presence. Likewise, the presence of complex AWP does not compensate for the lack of complexity of the simple ones, given that the results, coinciding with the study by Orrantia et al. (2005), show that the majority of complex problems are concentrated in category “A” where a consistent change structure is combined with another equally consistent combine structure. Therefore, the most numerous problems used by publishers to assess students’ mathematical competence are the easiest to solve.

However, the semantic structure is not the only variable that causes word problems in the teaching guides to be the easiest to solve. The “challenging” problems, that is, those non-routine problems in which the application of an arithmetic operation does not lead to solving the problem, are practically nil. These results are similar to previous studies, which have either focused on the information variable (Orrantia et al., 2005; Wijaya et al., 2015), or on the invention variable (Cai and Jiang, 2017; Orrantia et al., 2005). As a consequence, students infer that solving a problem is doing something with all the numbers present in the statement, given that it will always contain the necessary data for answering the question. In this way, reasoning skills, as a tool to obtain additional information (problems with less data) or to select only the necessary information (problems with more data), are not promoted. In addition, teaching guides do not contemplate the invention of problems as an essential task for assessing mathematical competence.

Regarding the situational context, these assessment tests present word problems in highly standardized or stereotyped contexts (very precise premises with data and questions), with little or even no relevant

background information that could help students solve them. In fact, of the small proportion of problems enriched with background information, the most numerous categories (action and description) are precisely the least relevant to generate the episodic situation model (Reusser, 1990), and those that would be the most relevant when connected to the mathematical model of the problem are the least numerous (character's intentions, goals, and purposes) (Orrantia, Tarín, and Vicente, 2011). The results of the most recent international studies that have analyzed this variable (Brehmer, Ryve, and Van Steenbrugge, 2016; Wijaya et al., 2015) have also shown that textbooks include problems in "purely mathematical" contexts. For these authors, it is necessary to include situationally enriched contexts that arouse the interest of students and that help them integrate mathematical information with non-mathematical information, an aspect that would improve the teaching-learning process of PS.

To summarize, we consider that the PS proposals in the assessment tests contribute to the students developing superficial and passive resolution strategies, which demand little cognitive effort. This approach also favors the development of inaccurate beliefs about what it really means to solve a problem, given that this meaning depends as much on the type of tasks carried out in the classroom, as on the evaluation methods.

Limitations and future lines of research

This study has some limitations that must be taken into account. The first of these is not having offered a more exhaustive analysis of the multiplication and/or division problems, problems that, although they have been coded as such, have not been analyzed in detailed according to the subcategories presented in the study of Chamoso et al. (2014).

We should also point out that in the present study, only the teaching guides of six publishers have been analyzed. Although these publishers cover a good part of the textbook publishing industry in Spain, there are other minority proposals on the market and other PS projects that have not been considered in this study for reasons of space. Therefore, although our review is comprehensive, it cannot be considered fully exhaustive.

Finally, as a future line of research, we consider it necessary to update the AWP textbook scenario. As we have pointed out, based on the pioneering study by Orrantia et al. (2005), carried out with textbooks published during the LOGSE (1990), various studies have been developed with the purpose of updating this issue. However, these reviews have been carried out with textbooks published during the legislative framework of the LOE (2006) and LOMCE (2013). The promulgation of the new education law (LOMLOE, 2020), and its fortieth additional provision in which it is stated that the educational authorities will provide textbooks free of charge, would allow us to expand this analysis and check whether the legislative changes and, therefore, changes in school textbooks are effective when it comes to addressing the learning process of PS or, on the contrary, as research has shown so far, publishers remain oblivious to successive educational reforms, teaching and evaluating in the same way.

Conclusions

The analysis carried out allows us to conclude that the assessment tests included in the teaching guides are characterized by a reduced number of word problems, a low variability of the different categories and subcategories, a high frequency of consistent problems, a very limited proportion of challenging problems, and a standardization of problem statements.

The low quantity of word problems undermines the important role that PS should have in both teaching and assessing mathematical competence in primary education. A role which has been described as serving as the “backbone” of the rest of the mathematical contents (RD, 126/2014). Given the relevance of PS in the mathematics curriculum, this content should be reflected in its assessment. However, as we have found, PS is not a priority in the teaching guide’s assessment tests of the analyzed publishers.

On the other hand, there is a relationship between the most frequent problems and the degree of difficulty. In this way, the most numerous problems are precisely the simplest to solve, that is, problems whose resolution does not require advanced conceptual knowledge or the application of sophisticated resolution strategies. In addition, the scarce variability (only seven subcategories of the twenty possible appear),

represents an obstacle to the advancement of students. As Lester (2013) points out, students will improve as problem solvers “only if they are given opportunities to solve a variety of types of tasks” (p.272).

In addition, problems that imply a certain level of challenge, or problematic situations formulated beyond what is considered a stereotyped situational context, are very limited. In this regard, no problem with superfluous or omitted information is included. All the problems presented contain what Wijaya et al., (2015) call “coincident information”, that is, sufficient and necessary data for their resolution. The approach to this type of problem is key to developing the ability to solve problems, given that it helps students consider the context as a relevant element when addressing the solution. Otherwise, students end up adopting mechanical solving strategies in which only the statement data must be selected and operated with (Salado, Chowdhury & Norton, 2019). In addition, problem invention tasks are underrepresented, despite the fact that research has highlighted this type of task as essential for the development of mathematical competence (Cai, Hwang, Jiang & Silber, 2015).

Acknowledgment

The authors would like to thank the Department of Teaching and Educational Organization of the University of Valencia for providing the financial resources to perform the translation of this article.

References

- Alsina, A. (2006). ¿Para qué sirven los problemas en la clase de matemáticas? *UNO, Revista de didáctica de las matemáticas*, 43, 113-118. Recuperado de: <https://dugi-doc.udg.edu/handle/10256/10636>
- ANELE (2014). *La Edición de Libros de Texto en España. Octubre de 2014. Asociación Nacional de Editores de Libros y material de Enseñanza*. Recuperado de: <https://anele@anele.org>

- Apple, M.W. (1992). The text and cultural politics. *Educational Researcher*, 21(7), 4-11. doi: 10.3102/0013189X021007004
- Area, M. (2000). Los materiales curriculares en los procesos de diseminación y desarrollo del currículum. En J.M. Escudero (Edit.), *Diseño, desarrollo e innovación del currículum* (pp. 189-204). Madrid: Síntesis.
- Brehmer, D., Ryve, A. y Van Steenbrugge, H. (2016). Problem solving in Swedish mathematics textbooks for upper secondary school. *Scandinavian Journal of educational research*, 60 (6), 577-593. doi: 10.1080/00313831.2015.1066427
- Cai, J., Hwang, S., Jiang, C. y Silber, S. (2015). Problem posing research in mathematics: some answered and unanswered questions. En F. M. Singer, N. Ellerton y J. Cai (Eds.), *Mathematical problem posing: From research to effective practice* (pp. 3-34). New York, NY: Springer.
- Cai, J. y Jiang, C. (2017). An analysis of problem-posing tasks in Chinese and US elementary mathematics textbooks. *International Journal of Science and Mathematics Education*, 15(8), 1521-1540. doi: 10.1007/s10763-016-9758-2
- Carpenter, T. y Moser, J. (1983). The acquisition of addition and subtraction concepts. En R. Lesh y M. Landau (Eds.), *Acquisition of mathematics: Concepts and processes* (pp.7-44). NY: Academic Press. doi: 10.2307/748348
- Chamoso, J.M., Vicente, S., Manchado, E. y Múñez, D. (2014). Los problemas de matemáticas escolares de primaria, ¿son solo problemas para el aula? *Cuadernos de Investigación y Formación en Educación Matemática*, 12, 261-279. Recuperado de: <https://revistas.ucr.ac.cr/index.php/cifem/article/view/18924/19038>
- Despina, D. y Harikleia, L. (2014). Addition and Subtraction Word Problems in Greek Grade A and Grade B Mathematics Textbooks: distribution and Children's Understanding. *International Journal for Mathematics Teaching and Learning*, 8, 340-356. Recuperado de: <https://www.cimt.org.uk/journal/desli.pdf>
- Escudero, J.M. (2015). *Prologue. Digital Textbooks: What's New?* (pp.4-6) Santiago de Compostela: Servizo de Publicacións da USC/IARTEM.
- Fuchs, E. y Bock, A. (2018). *The Palgrave Handbook of Textbook Studies*. New York: Hanbooks. doi: 10.1057/978-1-137-53142-1
- Gimeno, J. (2015). El currículum como estudio del contenido de la enseñanza. En J. Gimeno, M.A. Santos, J. Torres, P. Jackson y A.

- Marrero (Eds.), *Ensayos sobre el currículum: teoría y práctica* (pp.29-62). Madrid: Morata.
- Heller J.I. y Greeno, J.G. (1978). *Semantic processing in arithmetic word problem solving*. Paper presentado en Midwestern Psychological Association Convention. Chicago.
- Landis, J.R. y Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159- 174.
- Lester, F.K. (2013). Thoughts about research on mathematical problem solving instruction. *The Mathematics Enthusiast*, 10 (1), 245–278. Recuperado de: <https://scholarworks.umt.edu/tme/vol10/iss1/12/>
- Ley Orgánica 14/1970, de 4 de agosto, General de Educación y Financiamiento de la Reforma Educativa. B.O.E núm. 187, de 6 de agosto de 1970.
- Ley Orgánica 1/1990 de Ordenación General del Sistema Educativo, de 3 de octubre. B.O.E núm. 238, de 4 de octubre de 1990.
- Ley Orgánica 2/2006, de 3 de mayo de Educación. B.O.E núm. 106, de 4 de mayo de 2006.
- Ley Orgánica 8/2013, de 9 de diciembre, para la Mejora de la Calidad Educativa. B.O.E núm. 295, de 10 de diciembre de 2013.
- Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo de Educación. B.O.E núm. 340, de 29 de diciembre de 2020.
- Lewis, A.B. y Mayer, R. E. (1987) Student´s miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology*, 79(4), 363-371. doi: 10.1037/0022-0663.79.4.363
- Orrantia, J., González, B. y Vicente, S. (2005). Un análisis de los problemas aritméticos en los libros de texto de Educación Primaria. *Infancia y aprendizaje*, 28(4), 429-451. doi: 10.1174/021037005774518929
- Orrantia, J., Tarín, J. y Vicente, S. (2011). El uso de la información situacional en la resolución de problemas aritméticos. *Infancia y Aprendizaje*, 34 (1), 81-94. doi: 10.1174/021037011794390094
- Piñeiro, J. L., Castro-Rodríguez, E. y Castro, E. (2019). Componentes de conocimiento del profesor para la enseñanza de la resolución de problemas en educación primaria. *PNA* 13(2), 104-129. Recuperado de: <https://revistaseug.ugr.es/index.php/pna/article/view/v13i2.7876>
- Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la Educación Primaria. B.O.E núm. 52, de 1 de marzo de 2014.

- Reusser, K. (1990). From text to situation to equation: cognitive simulation of understanding and solving mathematical word problems. En H. Mandl, E. De Corte, N. Bennett y H.F. Friedrich (Eds.), *Learning and Instruction* (pp.477-498). Oxford: Pergamon.
- Salado, A., Chowdhury, A. H. y Norton, A. (2019). Systems thinking and mathematical problem solving. *School Science and Mathematics*, 119(1), 49-58. doi: 10.1111/ssm.12312
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. San Diego (CA): Academic Press.
- Staub, F. y Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. En C.A. Weaver, S. Mannes y C.R. Fletcher (Eds.), *Discourse Comprehension: Essays in honor of Walter Kintsch*, (pp.285-305). Hillsdale, NJ: Lawrence Erlbaum.
- Tarim, K. (2017). Problem Solving Levels of Elementary School Students on Mathematical Word Problems and The Distribution of These Problems in Textbooks. Çukurova University. *Faculty of Education Journal*, 46(2), 639-648. doi: 10.14812/cuefd.306025
- TIMSS. Estudio Internacional de Tendencias en Matemáticas y Ciencias. Marcos e Informes de Evaluación de los años 1995, 2011, 2015 y 2019. Madrid: Instituto Nacional de Evaluación Educativa-INEE.
- Van Zanten, M. y Van den Heuvel-Panhuizen, M. (2018). Opportunity to learn problem solving in Dutch primary school mathematics textbooks. *ZDM: The International Journal on Mathematics Education*, 50(7), 827-838. doi: 10.1007/s11858-018-0973-x
- Verschaffel, L., Depaepe, F. y Van Dooren, W. (2020). Word problems in mathematics education. En S. Lerman (ed.): *Encyclopedia of mathematics education* (pp. 908-911). Springer Nature. doi: 10.1007/978-3-030-15789-0
- Vicente, S. y Manchado, E. (2017). Dominios de contenido y autenticidad: un análisis de los problemas aritméticos verbales incluidos en los libros de texto españoles. *PNA*, 11(4), 253-279. Recuperado de: <https://revistaseug.ugr.es/index.php/pna/article/view/6242>
- Vicente, S., Manchado, E., y Verschaffel, L. (2018). Resolución de problemas aritméticos verbales. Un análisis de los libros de texto españoles. *Cultura y Educación*, 30(1), 87-104. doi: 10.1080/11356405.2017.1421606

- Vojíš, K. y Rusek, M. (2019). Science education textbook research trends: a systematic literature review. *International Journal of Science Education*, 41(11), 1496-1516. doi: 10.1080/09500693.2019.1613584
- Wijaya, A., Van den Heuvel-Panhuizen, M. y Doorman, M. (2015). Opportunity-to-learn context-based tasks provided by mathematics textbooks. *Educational Studies in Mathematics*, 89(1), 41-65. doi: 0.1007/s10649-015-9595-1
- Yang, D. C. y Sianturi, I. A. J. (2020). Analysis of algebraic problems intended for elementary graders in Finland, Indonesia, Malaysia, Singapore, and Taiwan. *Educational Studies*, 1-23. doi: 10.1080/03055698.2020.1740977

Contact address: Raúl Tárraga-Mínguez. Universidad de Valencia. Facultad de Filosofía y Ciencias de la Educación. Dpto. de Didáctica y Organización Escolar. Avda. Blasco Ibáñez, 30, CP: 46010, Valencia. E-mail: raul.taraga@uv.es