Title	Comparison of the level of authenticity of arithmetic word problems in Spanish and Singaporean textbooks (Comparación del nivel de autenticidad de los problemas aritméticos verbales de los libros de texto españoles y singapurenses)
Author(s)	Santiago Vicente, Lieven Verschaffel and David Múñez

Comparison of the level of authenticity of arithmetic word problems in Spanish and Singaporean textbooks

Santiago Vicente[1], Lieven Verschaffel[2], & David Múñez[3]

[1]Universidad de Salamanca. Departamento de Psicología Evolutiva y de la Educación. Facultad de Educación. Pso. Canalejas 169. 37008 Salamanca. E-mail: sanvicente@usal.es. Tel. +34 923 194 630 ext. 5735.

[2]Center for Instructional Psychology and Technology. Education and Training Research Group. Katholieke Universiteit Leuven. Van den Heuvel Instituut. Dekenstraat 2. Post Box 3773, B-3000. Leuven, Bélgica. E-mail: Lieven.Verschaffel@ppw.kuleuven.be. Tel. +32 (0)16 326258

[3]Centre for Research on Child Development. National Institute of Education. Nanyang Technological University. 1 Nanyang Walk, Singapore 637616. E-mail: david.munez@nie.edu.sg. Tel. +65 6219 6265

This work was funded by Ministerio de Ciencia, Innovación y Universidades (Spain). Grant no. PGC2018-100758-B-I00.

Abstract

The nature of arithmetic word problems found in math textbooks influences the way students develop the ability to solve them, as teachers use the books in their classes quite frequently. Given that students are better able to reason through and solve authentic problems that are contextualized in situations that sound familiar to them, and that different international assessments have shown the students in Singapore to be more skilled at problem-solving than their Spanish counterparts, this study compares the level of authenticity of the word problems included in the primary school textbooks of the main publishers in each country. Results show that the textbooks in Singapore contain problems with a higher degree of authenticity than the Spanish textbooks across school levels, thus providing students with better opportunities to learn how to solve problems through reasoning.

Comparison of the level of authenticity of arithmetic word problems in Spanish and Singaporean textbooks

Singaporean students have attained outstanding results in the most recent editions of international math competence assessments (618 points in the Trends in International Mathematics and Science Study [TIMSS], 2015 and 569 in the Programme for International Student Assessment [PISA] 2018; see Instituto Nacional de Evaluación Educativa [INEE], 2016, 2018). In contrast, performance of Spanish students, with 505 points on the TIMSS and 481 on the PISA, is average (TIMSS) or lower than average (PISA) among the countries assessed. Given these results, it is worth exploring how math is taught in both countries and whether there are differences in the way that teachers support students' learning.

Textbooks are among the factors that contribute to such differences (Oates, 2014). Specifically, the type of learning they promote (concerning the way word problems are solved). Problem-solving is the cornerstone of the math curriculum, as gleaned from the official curricular documents in the majority of countries and the theoretical frameworks on which the principles of international student performance tests are based. In this sense, according to PISA, math and problem-solving skills should help students recognize the role that math plays in the world, make sound well-grounded judgments, and take decisions as constructive, committed, and reflective citizens (OECD, 2018). The role of textbooks is relevant since teachers in many countries use math books in their classes (Hiebert et al., 2003; Mullis et al., 2008). Specifically, according to the 2007 TIMSS study, 60% of primary school teachers in OECD member countries use textbooks as the foundation of their educational practices (this rises to 74% among Spanish teachers), 34% use them as supplementary material (22% in Spain) and the remaining 6% do not use them (4% in Spain, Mullis et al., 2008). Therefore, it can be assumed that the content of

these books exerts a major influence on the educational practices that teachers implement in their classrooms.

In this sense, and considering that solving arithmetic word problems (hereafter, AWPs) is a key activity in students' learning how to solve any kind of math problem (Verschaffel et al., 2014), we should question whether the AWPs included in primary math textbooks in different countries provide students with similar opportunities to practice problem-solving in a meaningful, well-reasoned way. To answer this question, and considering the results of previous studies that show that students learn to solve problems better when they are contextualized in situations that sound familiar to them (Palm & Nyström, 2009; Vicente & Manchado, 2016), in this study we compare the situations proposed in the AWPs included in primary math textbooks in Singapore and Spain, and the situations that students may encounter in their everyday lives.

Problem-solving and situational understanding

AWPs can be defined as 'verbal descriptions of problems that give rise to one or more questions whose answers can be found by applying mathematical operations using the numerical information found in the problem' (Verschaffel et al., 2014, p. 641). These problems are considered prime tools for promoting the development of students' problem-solving competences (Verschaffel et al., 2014), and more specifically, to develop both the skills and tendency to use math to make sense of everyday situations and solve problems in the real world (Van Dooren et al., 2006). Students can develop these skills by learning how to deal with AWPs through what is called 'mathematical modeling', a problem-solving process based on understanding different aspects of the situation proposed in the problem (for a review of the concept, see Blum, 2015; Lesh, 2012).

Solving problems through mathematical modeling is a complex competence (Blum & Niss, 1991; Burkhardt, 1994; Mason, 2001). In this sense, Verschaffel et al. (2000) propose the existence of two ways of solving problems. The first one, called 'genuine modeling', allows students to solve any problem through an understanding of both the situation described (in terms of people, actions, and intentions) and its mathematical structure. Only after having understood both aspects should the student choose the arithmetic operation to solve it, carry it out, and interpret the result. In contrast, the second problem-solving model, called 'superficial modeling', means ignoring the information of the operations to go straight to the result, without checking whether the answer is plausible; therefore, it can hardly be considered an authentic mathematical modeling process.

Only the simplest problems contextualized in situations that require minimal understanding can be solved in this superficial way. Let us consider the following problem: 'Carlos wanted to buy himself a storybook, but since he didn't have enough money, he asked his grandmother for money, and she gave him 23 euros. If Carlos now has 40 euros, how much money did he have at first?' Solving this problem using the genuine model would mean understanding the situation described in terms of the people, their needs and intentions (Carlos wants a storybook but needs more money; his grandmother wants to make him happy), the actions that each person performs following these intentions (Carlos asks for money and his grandmother gives it to him) and the consequences of these actions (Carlos has more money than before). Afterward, mathematical comprehension that is coherent with the situation described is needed; in this case, the initial amount that Carlos had must be less than the 40 euros that he has at the end because his grandmother gave him 23 euros. After this mathematical comprehension, the student has to deduce that subtraction is needed; then, the student has to carry out the

operation and check that the initial amount of money is less than the final amount. In contrast, superficial problem-solving is limited to choosing the information in the problem (23 and 40) and a textual or contextual clue (such as 'his grandmother gives him' as a sign that they have to add) to decide on the operation needed (addition), execute it and give the result as a solution (in this case, a wrong response). This problem has to be solved through genuine modeling for two reasons: first, because the situational context of the problem allows students to generate an appropriate model of the situation, which helps them create a mathematical model of the problem; and secondly, because it is mathematically difficult, as it asks for the initial amount and there is no consistency between the keyword in the problem and the operation that solves it. In contrast, a simple word problem with hardly any situational context (such as: ‘A shepherd had 17 black sheep and 23 white sheep. How many total sheep did he have?’) lends itself to superficial problem-solving.

Different studies have shown that students solve more easily AWPs when statements include information that helps them understand the situation or the mathematics. For example, Palm and Nyström (2009) started with the idea that students are better able to reasoning if problems are posed as 'authentic problems': that is, if they represent situations from real life in such a way that the most important aspects of those situations —such as the event described, the question asked or the information provided in the statements— are simulated to a reasonable degree (Palm, 2008). These authors took the problematic AWPs in the study by Verschaffel et al. (1994) and created authentic versions of these problems (for an in-depth description of the problematic AWPs, see Verschaffel et al., 2020). One example of a problematic item is a non-proportional multiplication situation: 'Louis runs 100 meters in 10 seconds. How long would it take him to run 1,000 meters?’ The authentic version proposed by Palm and Nyström (2009) was

the following: ‘A friend and you are watching a track and field competition on the TV when the fastest man in the world, Maurice Green, wins the 100-meter race in 10.00 seconds. The next race you watch is the 10,000-metre, which Haile Gebrselassie wins in 26 minutes and 5 seconds. How would you answer if your friend asked you how long you thought it would take Maurice Green to run 10,000 meters?' Given that the authentic versions of these problematic items were considerably easier for students, the authors concluded that because more aspects associated with real-life were simulated better than in the real versions, the authentic versions promoted more in-depth understanding and thereby a more effective problem-solving process. Vicente and Manchado (2016) also found that students solved more easily ‘authentic’ versions of problems whose difficulty was not situational, as in the case of the problematic items in the study by Palm and Nyström (2009), but mathematical (similar to the one described above as an example to illustrate the genuine model according to the model by Verschaffel et al. 2000).

Taken as a whole, the results of the studies by Palm and Nyström (2009) and Vicente and Manchado (2016) indicate that presenting AWPs authentically is associated with better performance, regardless of whether the source of the problem’s difficulty is situational or mathematical. These findings fit within the concept of 'situated cognition' (Lave, 1988; see, too, Cobb, 2001; Waite & Pratt, 2015), which claims that thinking and learning are situated; that is, they are produced according to the context in which they take place, even when many of the activities conducted in classrooms tend to imply abstract knowledge lacking specific contexts. Thus, everyday activities could be considered the foundation of student learning to achieve this abstract mathematical knowledge (Lave & Packer, 2008), even if these “everyday situations” may vary from one cultural context to another.

Textbooks

The type of problems that students solve throughout their education and the degree of understanding that is required seem to be crucial in explaining how students learn to solve problems. Given that the role of math textbooks as a prime resource for teaching is a common practice in the majority of countries (Hiebert et al., 2003; Mullis et al., 2008), textbooks may determine the variety of problems that students are faced with throughout primary school. In this sense, we could assume that the curriculum of the majority of schools is not defined so much by the program suggested by different educational institutions as by the textbooks (Apple, 1992; Vincent & Stacey, 2008). Thus, analyzing the textbooks may shed light on what students may be learning when they solve AWPs.

The majority of studies that have analyzed math textbooks have focused on the mathematical aspects of the problems, such as the variety of AWPs proposed according to their semantic-mathematical structure, or the steps that students follow to learn how to solve problems (for a review of these studies, see Vicente et al., 2018). However, less attention has been paid to the situational aspects of the problems. One of the few studies was conducted by Depaepe et al. (2009), who analyzed the level of authenticity (Palm & Burman, 2004) of a sample of word problems in sixth-grade textbooks in Flanders (Belgium). The results showed that even though some of the aspects that were analyzed (event, the existence of information, specificity of information, and realism of information) were appropriately simulated in the majority of problems, other aspects like external tools, guidance and, specially, the purpose of the problem were inadequately simulated. For this reason, the majority of those problems could not be considered authentic situations, such that when solving these problems, students may develop inappropriate beliefs and problem-solving techniques.

However, it is essential to highlight that the authenticity of the problems is considered a matter of all or nothing in the study by Depaepe et al., such that the conclusions these authors reach could be nuanced with an analysis that admitted intermediate categories between authentic and non-authentic problems. In this sense, Vicente and Manchado (2017) took the five main aspects defined by Palm and Burman (2004) —event, question, the existence of information, purpose, and specificity of the information— and, according to whether all, some, or none of them were simulated in a credible, possible but not very credible, or impossible way, they established a categorization of the problems according to their level of authenticity: authentic problems (all the main aspects are well simulated, or all except the purpose), stereotyped problems (some of the main aspects were possible but not very credible) and absurd problems (some of the main aspects were not credible). The results showed that the majority of problems in the mathematics books from two main Spanish publishers were stereotyped and that the problems related to topics in upper levels (fractions, geometry, organization of information, and probability) were very poorly contextualized for different reasons. For instance, while fraction problems showed difficulties with the information, geometry and probability problems showed difficulties with the events.

This study

The goal of this study is to ascertain whether the problems included in primary school math textbooks in Singapore are contextualized in a more authentic way than the problems found in the textbooks in Spain —leading to greater opportunities for Singaporean students to exercise comprehension when solving problems. It is reasonable to expect that at least part of Singaporean students' success in developing their mathematical skills is due to the quality of their textbooks; perhaps this is the reason why many other countries have begun to translate and

adapt the textbooks from Singapore (such as Maths-No Problem in the United Kingdom, the materials from Polygon Education and the 'Piensa Infinito' publishing project by Editorial SM in Spain, along with the math education materials developed by the Ministry of Education in Chile).

In this context, it is feasible that the word problems of the textbooks from the Singaporean publisher are more authentic than those of the textbooks from the two Spanish publishers. This means, that math textbooks in Singapore present more good-fit problems than the textbooks in Spain. Furthermore, it is possible that such differences are also observed across grades, especially at the upper levels.

Method

This study presents a quantitative comparison between all the AWPs included in the textbooks and complementary workbooks from the Singaporean publisher Marshall Cavendish (hereafter, MC) in their 2015 edition, and the problems included in the books of the Spanish publishers Santillana and SM in their 2010 editions (the same books as those analyzed by Vicente & Manchado, 2017). MC textbooks are widely used in primary schools in Singapore (86%; Clark, 2013). In Spain, Santillana (43.16%) and SM (25.76%) account for almost 70% (Vicente et al., 2018).

Sample

We analyzed all of the AWPs with statements describing situations, either real or imaginary, that require the execution of at least one of the four basic arithmetic operations. A total of 6,791 AWPs were analyzed (2,961 from MC, 2,399 from Santillana, and 1,431 from SM).

Instrument: analysis system

We used the system developed by Vicente and Manchado (2017), which was, in turn, based on the aspects described by Palm and Burman (2004) and applied in the study by Depaepe et al. (2009). This system assessed the following aspects of each problem: event, question, information, purpose, and specificity of the information (see Table 1). One of three values was assigned to each aspect: '1', when the aspect was given in the problem in the way the students would tend to find it outside school; '.5', when it is possible but unlikely that the students would find the aspect analyzed as presented in the problem outside school; and '0', when the aspect is presented in such a way that students are highly unlikely to experience it in their daily lives. Therefore, the higher a problem's degree of authenticity, the higher the resulting sum of the scores in all aspects.

Table 1: System of analysis of the level of authenticity of AWPs, taken from Vicente and Manchado (2017).

	Score		
Aspect	1	.5	0
Event	Likely that the student will experience it outside school	Possible but not very likely outside school	Imaginary or fictitious
Question	Reasonable; its answer has practical value in real life	Possible, but of limited interest to students.	Could not be asked in the real world
Existence of information	The information matches information that is accessible in real life	The information may exist in reality but is seldom found	The information does not match information that is accessible in real life
Purpose in the figurative context	Explicit and appropriate to the situation	Implicit but easy to deduce	Implicit and difficult to infer
Specificity of the information	People with names, defined objects and specific places	The situation is not specific, but the objects, role or name of the people are	Neither the subjects nor objects involved are specified

Based on the overall scores assigned to the different aspects in Table 1, each AWP was classified into different types (see Table 2):

- Good-fit problems: equivalent to the authentic problems proposed by Palm and Burman (2004) and Vicente and Manchado (2017). All main aspects are well simulated (or all aspects but the purpose): the event may happen in students' lives outside school, the question of the problem makes sense, the information provided is adequate, there is a purpose for the problem or it can be easily inferred, and the information is specific. In this way, the score of all the aspects is 1 except in the problems where the purpose is not explicitly stated, in which case this aspect is scored as .5, (and the sum across aspects is between 4.5 and 5 points). One example would be: 'Sanjay has 10 USD to buy fruit. He wants to buy three different types of fruit. What will he buy? Help him choose'. (apples 2 USD, bananas 3 USD, apricots 5 USD, grapes 6 USD, oranges 1 USD) (MC, book 1B, p. 132).

- Stereotyped problems, in which one of the main aspects is not well simulated, describe situations that the student may find in real life with adequate but not specific information, or with specific information but in situations that are not familiar to students. This also includes situations that may be familiar to students but are not common for them in either the event or the specificity of the information, and in which almost any situation with any magnitude of sets and any action on them could fit. In this case, as shown in Table 2, the scores can fluctuate between 2.5 and 4. For example, 'A toy factory made 625 teddy bears in two days. On the first day, it made 277 teddy bears. How many teddy bears did the factory make on the second day?' (MC, workbook 2A, p. 86). In this case, all the aspects were scored 1 except the event (.5, manufacturing teddy bears is an unusual but not wholly unfamiliar activity to students), the specificity of the information (.5) and the purpose (0).

- Poor-fit problems: equivalent to Vicente and Manchado's (2017) absurd problems, in which one or several of the main aspects are poorly simulated, or they propose situations that are

usually vague or absurd (the question of the problem makes little sense). It is obvious that the arithmetic operation is what matters in these problems. In this case, the maximum score is 2. These situations may be absurd for three reasons: (a) event: 'A grouper which is 37 meters under the sea gets close to a cluster of green seaweed which is 18 meters above it. How deep is the seaweed?' (SM, 6th grade, workbook 3, p. 26); (b) question: 'There are 3 men and 2 women in a lift. The three men weigh a total of 245 kg and the 2 women weigh 120 kg less than the men. What is the mean weight of the people in the lift?' (Santillana, 6th grade, p. 35); and (c) existence of information: ‘Mario has 2 C, 5 D and 8 U meters of string. How many total meters does he have?’ (SM, 3rd grade, p. 9).

Table 2: Levels of authenticity according to the score of each problem on each aspect. Following Palm and Burman (2004), when the main aspect scored '0', the analysis was halted and a total score of '0' was assigned.

TYPE	Event	Question	Existence of information	Purpose	Specificity of information	Total
Good-fit	1	1	1	1	1	5
	1	1	1	.5	1	4.5
Stereotyped	1	1	1/.5	.5/0	.5	3.5/4
	1	.5	1/.5	.5/0	1	
	.5	1	1/.5	.5/0	1	
	.5	1	.5	.5/0	.5	2.5/3
Poor-fit	.5	.5	.5	0	.5	2
	0					
	1/.5	0				
	1/.5	1/.5	0			

Results will be presented both aggregated and by grade. To facilitate interpretation, the six primary school grades were grouped into 3 levels: level 1 (first and second grade), level 2 (third and fourth grade), and level 3 (fifth and sixth grade).

Procedure

First, we submitted 20% of the sample to exploratory coding so the specifications of the criteria needed to achieve an acceptable degree of reliability were defined (Table 1 shows a summary of these specifications). Afterward, a different subsample of 20% of the word problems was coded. This subsample was representative of the various grades and publishers. After categorizing the problems, the inter-judge agreement was calculated for each aspect using the Cohen's Kappa coefficient: event, κ =.97; question, κ = .99; existence of information, κ = .98; purpose, κ = .97; specificity of the information, κ = .99, and all these correlation indexes were highly significant ($p < .001$).

The remaining 80% of the problems (including the 20% analyzed in the initial exploratory coding) were then categorized by the first author of this study. The first author checked with a second coder to discuss and clarify the way to proceed with word problems that did not clearly fit the criteria for a given category. To check that the analysis system was reliable in both the Spanish and Singaporean educational settings, 5% of the sample was analyzed by a third coder that was familiar with the Singaporean context. In this case, the inter-judge agreement, calculated by the Cohen's Kappa coefficient, was equal to or higher than κ = .97 for all aspects except for the event, κ = .88, and the correlation indexes were once again highly significant ($p < .001$).

Data analysis

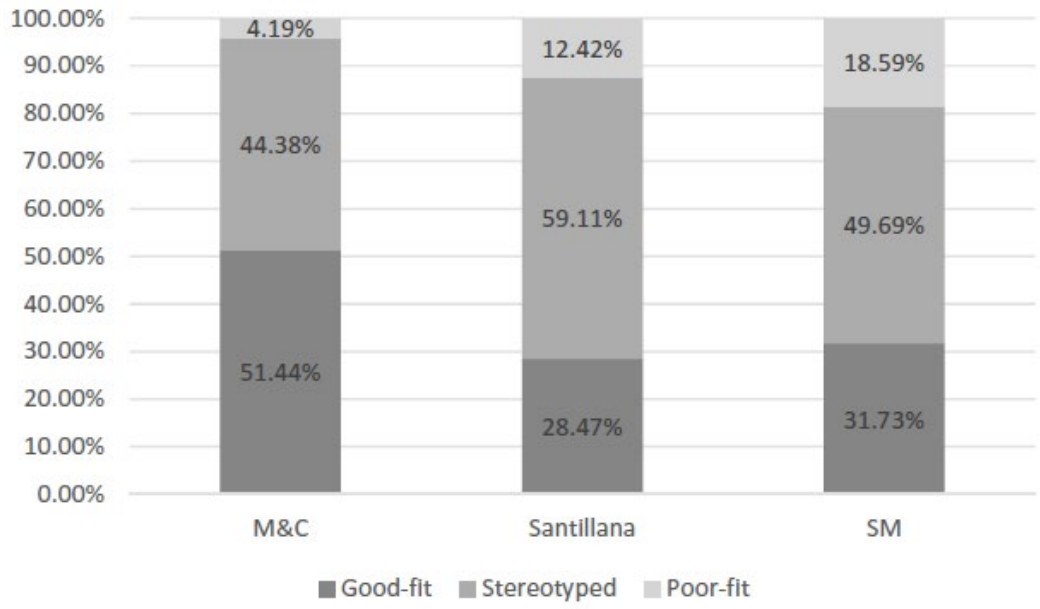
Non-parametric statistics were used to check the significance of the differences found. The chi-squared statistic was used to analyze the global differences found among the three publishers in the level of authenticity of the problems (calculated with planned tables). To compare the specific differences between any two publishers, the tests were conducted with a level of significance of .05 via comparisons of pairs of column proportions between the variables

on the planned tables. To simplify the explanation of the results, only the differences that achieve the level of statistical significance (.05) are described in the results section.

Results

Significant differences were found among the three publishers in the proportion of word problems categorized at each of the three levels of authenticity, $\chi^2(4, n = 6{,}791) = 477.87$, $p < .001$. The MC textbooks included more good-fit problems than those from Santillana and SM, while the books from Santillana and SM included more stereotyped and poor-fit problems than the textbooks in Singapore. Likewise, while Santillana included more stereotyped problems than SM, SM included more poor-fit problems than Santillana (see Figure 1).

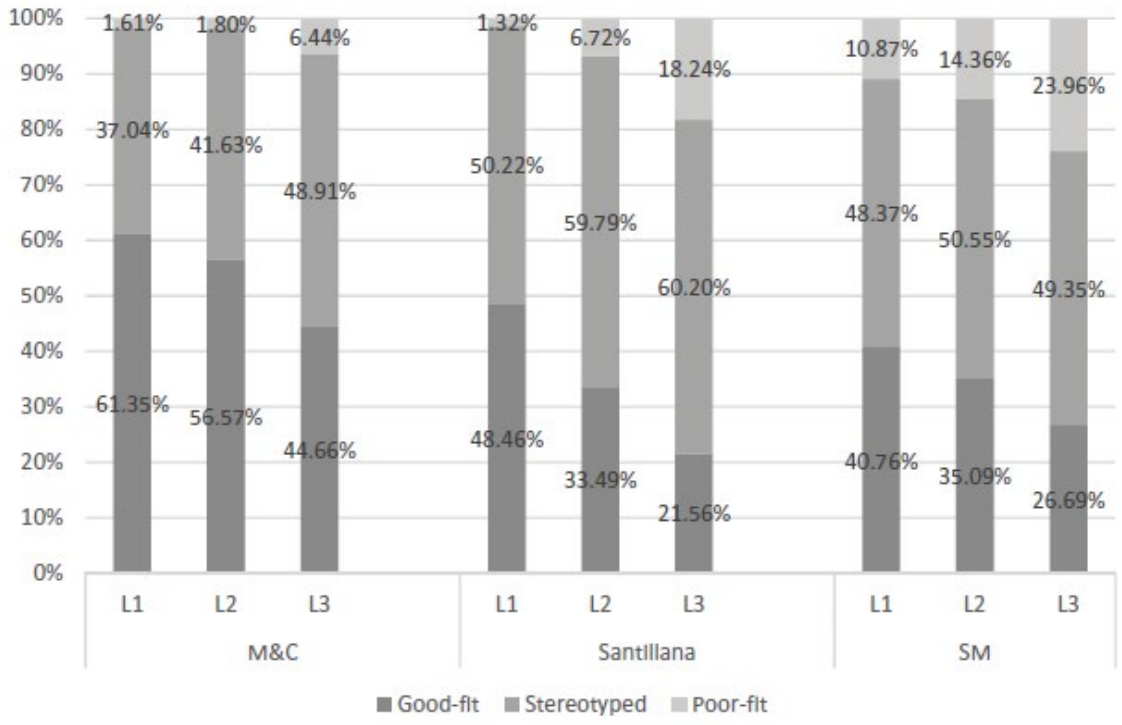
Figure 1: Proportions of problems by level of authenticity and publisher.



These results were also observed at the three educational levels that were specified, and they were accentuated at the upper level (see Figure 2). At Level 1 (grades 1 and 2), textbooks in Singapore contained a higher proportion of good-fit problems than those of the Spanish

publishers, $\chi^2(4, n = 1{,}094) = 65.89$, $p < .001$, (61.3% compared to 48.5% in Santillana and 40.8% in SM). The textbooks in Spain contained more stereotyped problems than those in Singapore (50.2% in Santillana and 48.4% in SM compared to 37% in MC). Furthermore, SM included substantially more poor-fit problems (10.9%) than MC (1.6%) and Santillana (1.3%).

Figure 2: Proportions of problems by level of authenticity, publisher, and grade.



Significant differences were also found at Level 2, $\chi^2(4, n = 2{,}151) = 154.39$, $p < .001$. Textbooks in Singapore included a higher proportion of good-fit problems (56.6%) than those of Santillana (33.5%) and SM (35.1%). Conversely, textbooks from Santillana and SM included a higher proportion of stereotyped problems (59.8% and 50.5% respectively) and poor-fit problems (6.7% and 14.4%) than MC (41.6% stereotyped and 1.8% poor-fit). Whereas the books from Santillana included a higher proportion of stereotyped problems than those from SM, SM included a higher proportion of poor-fit problems than Santillana.

Significant differences were also found among the three publishers at Level 3, $\chi^2(4, n = 3,545) = 271.76$, $p < .001$. Once again, the textbooks from Singaporean publisher included a higher proportion of good-fit problems (44.7%) than those from Santillana (21.6%) and SM (26.7%). The Spanish publishers included a higher proportion of poor-fit problems in their textbooks (18.2% and 24% in Santillana and SM) than MC (6.4%). Furthermore, SM included a higher proportion of poor-fit problems than Santillana and MC, while Santillana included a higher proportion of stereotyped problems (60.2%) than MC (48.9%) and SM (49.4%).

Taken as a whole, these results confirmed the second prediction: differences in the level of authenticity of word problems that were presented in the textbooks were evident across grade levels and, specially, at the higher level.

Discussion

Textbooks are key elements that influence how students develop their word problem solving competencies, as teachers all over the world use textbooks frequently in their classes (Hiebert et al., 2003; Mullis et al., 2008). In this sense, some studies have found differences between the textbooks of high-performing countries like Singapore and those in countries where students perform well below the level of Singaporean students. Those studies have shown, for instance, that textbooks in Singapore provide more appropriate instruction for students to learn how to solve problems through reasoning (for example, by providing a greater variety of word problems in terms of their mathematical structure, see Vicente et al., 2018) than those in Spain or the United States.

The results of this study expand the description of such differences in a similar sense. Singaporean students may have more opportunities to learn how to solve problems through

genuine modeling ~~(Verschaffel et al., 2000)~~, as the level of authenticity of the word problems ~~(Palm & Burman, 2004)~~ included in their textbooks is higher than that of textbooks in Spain. In contrast, Spanish textbooks pose a higher proportion of stereotyped problems that do not encourage students to reasoning in the way they should outside of school. Furthermore, math textbooks in Spain present more poor-fit problems that can inhibit any type of reasoning related to an understanding of the situation from the standpoint of the real world. In this sense, math textbooks in Spain are similar to those analyzed by Depaepe et al. (2009) in Flanders, and the same conclusion can be reached in both studies: the majority of the word problems in the textbooks are not well simulated.

Therefore, many of the word problems included in the math textbooks in Spain may be inhibiting the use of reasoning and an understanding of the situations described in the statements of the problems. This means, encouraging students to learn how to solve word problems using superficial modeling. To illustrate this point, we can analyze the following example: 'Julian filled his fish tank by pouring in 9 one-liter pitchers, 18 half-liter pitchers, and 24 quarter-liter pitchers. How many liters fit in Julian's fish tank?' (Santillana, 3rd grade, p. 207). To genuinely solve this poor-fit problem, one would have to ask why Julian does not always use the same pitcher to fill his fish tank such that to ascertain the capacity of the fish tank he only has to count the number of pitchers he pours into the tank. This word problem does not seem to help students become the reflective, critical, and constructive citizens that PISA (OECD, 2018) considers desirable. Instead, it leads them to become mere operators of arithmetic procedures. Similarly, stereotyped problems do not promote the use of situational understanding in the problem-solving process, as the situations they describe are unspecific and relatively distant from students'

experiences, such that students probably perceive these word problems as mere exercises aimed at automating calculation processes.

Furthermore, the results by grade show that the decrease in the level of authenticity of the word problems presented in textbooks corresponding to the upper years (described by Vicente & Manchado, 2017) is also observed in the textbooks in Singapore, albeit to a lesser extent. This is probably because it is more difficult to properly contextualize the more abstract mathematical content that is introduced at the upper levels. Such math topics —e.g., geometry, angles, powers— are less common in students' lives than other topics that are studied at lower grades —e.g., arithmetic operations with natural numbers— which are easier to contextualize more authentically. Notwithstanding, the textbooks in Singapore provide evidence that it is possible to lower the presence of poor-fit problems concerning more abstract math topics (if certain aspects are carefully designed). For instance, whereas textbooks in Singapore only presented the most frequently used units of measurement in word problems that involve measurement topics, textbooks in Spain included units of measurements that are never used in real life (e.g., decametres).

Given these results, two conclusions can be drawn. First, the results suggest that the math textbooks in Singapore proactively support students' mathematical competence, as textbooks include hardly any poor-fit problems (especially at the upper levels). This may encourage students to gain an understanding to better solve word problems (Verschaffel et al., 2000). In this sense, specific guidelines can be deduced by analyzing the problems in those textbooks so the level of authenticity of the word problems in the textbooks in Spain (and other countries) may be aligned accordingly. For instance, stereotyped problems —the most frequent type of word problem in math textbooks in Spain— should be turned into good-fit problems by i) specifying

the situations described in terms of objects, places, and people (for example, by using everyday situations like shopping, proper names for people, references to familiar places and people such as the classroom, the teacher, the family, etc.), ii) positing explicit purposes, and iii) asking thorough questions. Furthermore, word problems that involve units of measurement that students will only find in mathematics classes (decameters, hectograms, etc.) should be avoided, as well as those that mix different units in situations in which a single unit would be used in real life. Similarly, we should question whether poor-fit problems that cannot be contextualized, at least like stereotyped problems, should be left out of math textbooks. In this sense, it is true that all teaching-learning situations require a certain degree of decontextualization (Gravemeijer, 1997; Sethole, 2005) and that this decontextualization increases as the contents to be learned are more complex; nonetheless, it is essential to consider to what extent transferring this decontextualization to the statements of certain types of problems, such as units of measurement, geometry, and fractions, is benefitting students' learning and word problem solving competencies.

The second conclusion of the current study is that all of these suggestions, which are meant for publishers of math textbooks, are also valid for teachers. Math teachers can actively enrich and modify the word problems of the textbooks that they use in their classes. In other words, we should not forget that teachers interpret AWPs and provide students with guidelines according to their knowledge and professional beliefs.

Limitations and prospects

This study has a series of limitations. First, the sample of textbooks that was analyzed could be expanded (the third main publisher in Spain, Anaya, was not analyzed). Likewise, to increase the scope of our findings, future studies should analyze textbooks from other countries

whose students have average or low performance on evaluations like TIMSS or PISA, as well as those from countries with students performing at the level of Singaporean students (a small step in this direction is described at the end of this section). Secondly, the analysis conducted does not include certain elements of the textbooks, such as suggestions for teachers included in teachers' guides, so future studies should consider these elements. Thirdly, the categorization used in this study does not consider certain aspects that have been used in other studies aimed at analyzing the level of authenticity of word problems (e.g., use of language, guidance provided, presence of external tools). Thus, future studies should consider these aspects as part of the analytical approach. Furthermore, other components of the problem statement, such as the syntax (see Poulsen & Gravgaard, 2016), can influence comprehension of the problem. Thus, future studies, should approach the issue of word problems in textbooks attending to the syntax of the statement. Given that the authenticity of word problems in math textbooks depends on sociocultural contexts, more evidence is needed regarding the sensitivity of the system of analysis that is used in this study.

Finally, it is worth mentioning that textbooks are not sole determinants of students' learning processes concerning solving AWPs; for instance, students in Flanders (Belgium) performed significantly higher than Spanish students on the TIMSS (2015), even though the math textbooks in both countries seem very similar (at least in terms of the level of authenticity of word problems). The influence of textbooks is mediated by many factors, including teachers. Therefore, analyses of textbooks like that presented in this study should be complemented with systematic and detailed analyses of how teachers use them. In this sense, teacher training, along with other cultural and societal aspects (i.e., families' involvement in school education, the importance of education in society, and the value of meritocracy) may help us to understand the

success of the educational system in Singapore and why developing students' mathematical competence goes beyond the design of math textbooks (Rao et al., 2010).

References

Apple, M. (1992). The text and cultural politics. *Educational Researcher, 21*(7), 4–11. https://doi. org/10.3102/0013189X021007004

Blum, W. (2015). *Quality teaching of mathematical modelling: What do we know, what can we do?* In S. J. Cho (Ed.) Proceedings of the 12th International Congress on Mathematical Education: Intellectual and attitudinal challenges (pp. 73–96). Springer. https://doi.org/10. 1007/978-3-319-12688-3_9

Blum, W., & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects. State, trends and issues in mathematics instruction. *Educational Studies in Mathematics, 22*(1), 37–68. https://doi.org/10. 1007/BF00302716

Burkhardt, H. (1994). *Mathematical applications in school curriculum*. In T. Husen, & T. N. Postlethwaite (Eds.), The International Encyclopedia of Education (pp. 3631–3634). Pergamon Press.

Clark, A. (2013). *Singapore math: A visual approach to word problems*. Houghton Mifflin Harcourt. http://www.hmhco.com/~/media/sites/home/education/global/pdf/whitepapers/mathematics/elementary/math-in-focus/mif_model_drawing_lr.pdf?la=en

Cobb, P. (2001). *Situated cognition: Origins*. In N. J. Smelser, & P. B. Baltes (Eds.), International encyclopedia of the social & behavioral sciences (pp. 14126–14129). Elsevier. https://doi.org/10. 1016/B0-08-043076-7/01601-6

Depaepe, F., De Corte, E., & Verschaffel, L. (2009). *Analysis of the realistic nature of word problems in upper elementary mathematics education in Flanders*. In L. Verschaffel, B.

Greer, W. Van Dooren, & S. Mukhopadhyay (Eds.), Words and worlds: Modeling verbal descriptions of situations (pp. 245–263). Sense Publishers.

Gravemeijer, K. (1997). Solving word problems: A case of modelling? *Learning and Instruction, 7* (4), 389–397. https://doi.org/10.1016/S0959-4752(97)00011-X

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., Chui, A. M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). *Teaching mathematics in seven countries. Results from the TIMSS 1999 Video Study*. National Center for Education Statistics (NCES).

Instituto Nacional de Evaluación Educativa–INEE. (2016). TIMSS 2015. *Estudio internacional de tendencias en Matemáticas y Ciencias. Informe español: Resultados y contexto*. Ministerio de Educación, Cultura y Deporte.

Instituto Nacional de Evaluación Educativa–INEE. (2018). Informe PISA 2018. Programa para la Evaluación Internacional de los Estudiantes. Ministerio de Educación y Formación Profesional.

Lave, J. (1988). Cognition in Practice: Mind, mathematics, and culture in everyday life. Cambridge University Press. https://doi.org/10.1017/CBO9780511609268

Lave, J., & Packer, M. (2008). Towards a socio ontology of learning. In K. Nielsen (Ed.), A qualitative stance: Essays in honor of Steinar Kvale (pp. 17–46). Aarhus University Press.

Lesh, R. (2012). WISDOMe monograph: Quantitative reasoning and mathematical modeling: A driver for STEM integrated education and teaching in context. In R. Mayes, & L. Hatfield

(Eds.), Research on models & modeling and implications for common core state curriculum standards (Vol. 2, pp. 197–203). University of Wyoming.

Mason, J. (2001). *Modelling modelling: Where is the centre of gravity of-for-when teaching modelling?* In J. F. Matos, W. Blum, K. Houston, & S. P. Carreira (Eds.), Modelling and mathematics education. ICTMA9: Applications in science and technology (pp. 39–61). Horwood. https://doi.org/10.1533/9780857099655.1.39

Mullis, I., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grade*. TIMSS & PIRLS International Study Center, Boston College. http://pirls.bc.edu/timss2007/mathreport.html

O.E.C.D. (2018). *PISA for Development Assessment and Analytical framework: Reading, Mathematics and Science*. OECD Publishing. https://doi.org/10.1787/9789264305274-en

Oates, T. (2014). *Why Textbooks Count: A policy paper*. Cambridge Assessment.

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 67*, 37–58. https://doi.org/10.1007/s10649-007-9083-3

Palm, T., & Burman, L. (2004). Reality in mathematics assessment: An analysis of task-reality concordance in Finnish and Swedish national assessments. *Nordic Studies in Mathematics Education, 9*(3), 1–33. http://umu.diva-portal.org/smash/record.jsf?pid=diva2% 3A147409&dswid=9031

Palm, T., & Nyström, P. (2009). Gender aspects of sense making in word problem solving. *Journal of Mathematical Modelling and Applications, 1*(1), 59–76. https://proxy.furb.br/ojs/ index.php/modelling/article/view/1623/1079

Poulsen, M., & Gravgaard, A. K. D. (2016). Who did what to whom? The relationship between syntactic aspects of sentence comprehension and text comprehension. *Scientific Studies of Reading, 20*(4), 325–338. https://doi.org/10.1080/10888438.2016.1180695

Rao, N., Ng, S. S. N., & Pearson, E. (2010). *Preschool pedagogy: A fusion of traditional Chinese beliefs and contemporary notions of appropriate practice*. In C. Chan, & 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

Sethole, G. (2005). *From the everyday, through the inauthentic, to mathematics: Reflection on the process of teaching from contexts*. In H. L. Chick, & J. L. Vincent (Eds.) Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education (pp. 169–175). PME.

Van Dooren, W., Verschaffel, L., Greer, B., & De Bock, D. (2006). *Modelling for life: Developing adaptative expertise in Mathematical modelling from early age*. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), Instructional psychology: Past, present and future trends (pp. 91–109). Elsevier.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modelling of school arithmetic word problems. *Learning and Instruction, 4(*4), 273–294. https://doi.org/10.1016/0959-4752(94)90002-7

Verschaffel, L., Depaepe, F., & Van Dooren, W. (2014). *Word problems in mathematics education.* In S. Lerman (Ed.), Encyclopedia of mathematics education (pp. 641–645). Springer.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Swets & Zeitlinger Publishers. https://doi.org/10.1023/A:1004190927303

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM Mathematics Education, 52*(1), 1–16. https://doi.org/10. 1007/s11858-020-01130-4

Vicente, S., & Manchado, E. (2016). Resolución de problemas aritméticos verbales. ¿Se resuelven mejor si se presentan como problemas auténticos? *Infancia y Aprendizaje, 39*(2), 349–379. https://doi.org/10.1080/02103702.2016.1138717

Vicente, S., & 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. http://revistaseug.ugr.es/index.php/pna/article/view/6242

Vicente, S., Manchado, E., & Verschaffel, L. (2018). Resolución de problemas aritméticos verbales. Un análisis de los libros de texto españoles. *Cultura y Educación, 30*(1), 71–104. https://doi.org/10.1080/11356405.2017.1421606

Vincent, J., & Stacey, K. (2008). Do mathematics textbooks cultivate shallow teaching? Applying the TIMSS video study criteria to Australian eighth-grade mathematics textbooks. *Mathematics Education Research Journal, 20*(1), 82–107. https://doi.org/10.1007/BF03217470

Waite, S., & Pratt, N. (2015). *Situated learning (Learning In Situ)*. In J. D. Wright (Ed.), International encyclopedia of the social & behavioral sciences (2nd ed., pp. 5–12). Elsevier. https://doi.org/0.1016/B978-0-08-097086-8.92042-7