Revista de Psicodidáctica, 2014, 19(1), 93-123 *ISSN: 1136-1034 eISSN: 2254-4372*
www.ehu.es/revista-psicodidactica
DOI: 10.1387/RevPsicodidact.7865

Development of Children's Solutions of Non-Standard Arithmetic Word Problem Solving

Laura Jiménez* and Lieven Verschaffel**

*University of Castilla-La Mancha, **University of Leuven

Abstract

This study investigated the development of children's solutions of four types of non-standard arithmetic word problem, with a view to gather indirect evidence of the following beliefs about word problems that they develop through their immersion in the culture of traditional school mathematics: every word problem is solvable, there is only one numerical correct answer, it is always necessary to do calculations, and all numbers must be used in order to calculate the solution. Children from Grade 1 to 6 solved four word problems that violated these four beliefs. General results revealed, first, that only 37.9% of children's responses were correct. Second, the difficulty was increasing, starting with the *unsolvable* problem (18.3%), and followed by the *multiple solutions* (30.3%), the *given solution* (45.7%) and the *irrelevant data* problem (57.3%). Third, children's correct responses increased from Grade 1 (15.5%) to Grade 6 (56%), but not within the three lower and the three upper grades. The article ends with a discussion of the theoretical, methodological, and educational implications.

Keywords: Beliefs, non-standard problems, problem solving strategies, mathematics education, mathematics culture.

Resumen

En este estudio se ha investigado el desarrollo de las soluciones que dan los niños a cuatro tipos de problemas aritméticos no-estándar. Se ha intentado ofrecer una evidencia indirecta de las creencias que los niños desarrollan sobre los problemas verbales a través de la inmersión en la cultura escolar tradicional de las matemáticas. Dichas creencias son: todo problema tiene solución, solo hay una respuesta numérica correcta, siempre es necesario realizar cálculos y todos los números deben ser usados para hallar la solución. Se pidió a alumnos de 1.º a 6.º de Educación Primaria (E.P.) que resolvieran cuatro problemas verbales contrarios a estas cuatro creencias. Los resultados generales revelaron, en primer lugar, que solo el 37.9% de respuestas de los niños eran correctas. En segundo lugar, el grado de dificultad ante los distintos problemas verbales fue aumentando desde el denominado problema *irresoluble* (18.3%), y seguido por el de *soluciones múltiples* (30.3%), el problema cuya *solución estaba ofrecida en el enunciado* (45.7%), hasta el de *datos irrelevantes* (57.3%). En tercer lugar, se halló que las respuestas correctas de los niños aumentaron desde 1.º de E.P. (15.5%) hasta 6.º de E.P. (56%), pero no entre los tres cursos inferiores ni entre los tres cursos superiores. El artículo finaliza con una discusión de las implicaciones teóricas, metodológicas y educativas.

Palabras clave: Creencias, problemas no-estándar, estrategias de resolución de problemas, educación matemática, cultura de las matemáticas.

Correspondence concerning this article should addressed to Laura Jiménez, Department of Psychology, University of Castilla-La Mancha, Campus Universitario s/n, 16071, Cuenca, Spain. E-mail: laura.jimenez@uclm.es

Introduction

Standard versus non-standard word problems

Word problems constitute an important part of the mathematics curriculum of the elementary school. In word problem solving, children are expected to make proper use of the mathematical concepts and skills that they have acquired in mathematics instruction. This requires not only a good mastery of arithmetic facts, rules, principles and operations —namely, the syntax of mathematics (Ilani & Margolin, 2010) but also to understand the problem statement and to reason how to use the mathematical syntax in relation to the situation described —namely, the semantics of mathematics (Gerofsky, 1996; Verschaffel & De Corte, 1997).

One of the major reasons for including word problems in the mathematics curriculum is that they offer practice for everyday situations in which learners will need what they have learnt in their arithmetic, geometry or algebra lessons at school (i.e., the so-called application function of word problems, see Verschaffel, Greer, & De Corte, 2000).

Most authors consider the competent solution of a word problem as a complex multi-phase *modeling* process the "heart" of which is formed by (1) the construction of an internal model of the problem situation, reflecting an understanding of the elements and relations in the problem situation, and (2) the transformation of this situation model into a mathematical model of the elements and relations that are essential for the solution. These two steps are then followed by (3) working through the mathematical model to derive mathematical result(s); (4) interpreting the outcome of the computational work; (5) evaluating if the interpreted mathematical outcome is computationally correct and reasonable; and (6) communicating the obtained solution (e.g., Verschaffel et al., 2000; Verschaffel & De Corte, 1997).

In the research literature, word problems are differentiated in various ways. An important distinction is between standard and non-standard problems (Verschaffel et al., 2000). Standard word problems can be correctly and unambiguously modeled and solved through the straightforward use of one arithmetic operation with the given numbers, or a combination of two or more operations. In some cases, the correct arithmetic operation(s) can be easily derived from (a) keyword(s) included in the problem statement (e.g., the presence of the word "got" or "received" means that the required operation will be addition, whereas words such as "gave" or "lost" suggest subtraction), but in other cases, the identification of the correct arithmetic solution is less trivial (e. g., when the word problem contains a keyword that leads to the wrong operation, or when it is built around a problem situation with which the solver is not familiar, or when the problem requires multiple calculation steps). However, common to all standard problems is that they nicely fit Gerofsky's (1996; see also Verschaffel et al., 2000) description of a word problem as texts with a typical tripartite structure:

— A "set-up" component, establishing the characters and location of the putative story (however, according to Gerofsky, this component is not essential to the solution of the problem itself).
— An "information" component, which gives the information needed to solve the problem (Gerofsky [1996, p. 37], comments that sometimes extraneous information is added "as a decoy for the unwary").
— A question, that can be solved by the application of (a) mathematical operation(s) on the numerical data provided in the second component.

As argued by Verschaffel, Greer, and De Corte (2007), most of the research on word problems has dealt with standard problems, particularly in the curricular domains of elementary arithmetic but also of geometry and algebra. This research has yielded a great number of results and insights into the influence of various task and subject variables on learners' accuracies, representations, solution strategies and errors, and on the impact of various kinds of instructional materials and approaches on learners' problem solving skills and processes (for an overview see e.g., Verschaffel et al., 2007).

However, for various reasons, researchers have also investigated another type of word problems, called non-standard problems. Contrary to the first category of problems, which fit Gerofsky's (1996) above-mentioned tripartite structural frame, and therefore, by definition, evoke little or no discussion of what has to be considered as the correct answer, non-standard problems share the general feature that they deviate in one or more ways from that tripartite structure and its inherent rules (for examples of studies with non-standard problems see e.g., De Corte & Verschaffel, 1985; Jiménez, 2012; Jiménez & Ramos, 2011; Littlefield & Riesser, 1993; Puchalska & Semadeni, 1987; Reusser & Stebler, 1997; Selter, 1994; Verschaffel, De Corte, & Lasure, 1994). In the present article, we will report a study in which we confronted children from all grades of elementary school to several kinds of non-standard word problems, in the context of an individual interview, with a view to learn more about the development of their (implicit) beliefs about word problems and how these problems should be solved at school. But before presenting the rationale, the research questions, the design and the results of that study, we will briefly review the available research on non-standard arithmetic word problems.

Previous research on non-standard word problems

As stated before, most of the research on word problems has focused on standard problems. After all, these are also the problems that occur far more frequently in traditional mathematics textbooks, lessons, and assessments (Reusser & Stebler, 1997; Verschaffel et al., 2000). However, a problem with this research literature is that it sheds only a partial light on the processes and components of skillful word problem solving, and of the skill of mathematical modeling in general. During the past decades, theoretical and empirical work in research fields such as linguistics, pragmatics, anthropology, and ethnomathematics have led to an increased interest in word problems as a peculiar "genre of text" (Gerofsky, 1996, p. 37), "a peculiar cultural device" (Lave, 1992, p. 75), or "a specific game" with its own intent, structure, rules and tactics, that need to be known and respected by all partners involved in that game (De Corte & Verschaffel, 1985, p. 7). Inspired by these theoretical ideas, researchers have started to investigate this phenomenon empirically. A main research method in this empirical research has been to confront learners with problems that deviate from the standard structure of word problems, namely non-standard problems. Hereafter we will briefly review the research literature for four different types of such non-standard problems. Typically, the results of these studies, and more particularly, students' weak performance in solving this type of problems, have been used to demonstrate the problematic nature of the learning outcomes from traditional instruction that only involves standard problems.

Unsolvable word problems. Some researchers have worked with word problems that are unsolvable because one of the "givens" is missing. For example, De Corte and Verschaffel (1985) presented to first graders the following word problem with missing information "Pete had some apples. He gave four apples to Ann. How many apples does Pete have now?", and found that the majority of the children did not realize that it was unsolvable. Puchalska and Semadeni (1987) reported that a significant number of children aged 7 to 12 years concluded that a similar problem was unsolvable, but curiously most of them also believed that a better student or the teacher could solve it for sure. In a recent study, Jiménez and Ramos (2011) also found that 80% of second and third graders failed in detecting that they needed more information to solve the following similar problem with missing information "Mario is playing with marbles in the park. Mario has 17 marbles and his friend Jorge gives him 7. How much marbles does Jorge have now?".

Interestingly, the literature also contains case studies of word problems that are *unsolvable* because the information component and the question are completely unrelated, as in the hilarious classic of the captain's problem "There are 26 sheep and 10 goats in a boat. How old is the captain?", which was erroneously solved by 88% of children at Grades 1-2 and 38% of children at Grades 3-4 (Baruk, 1985). Most typically, pupils answered that the captain was 36 years-old by adding up the number of sheep and gouts (similar results were reported by Reusser, 1988). Contrary to Baruk, Radatz (1983) found that the percentage of children trying to reach some solution tended to increase rather than decrease between Kindergarten (10%) and Grade 4 (60%). Many authors have interpreted this remarkable result, whereby elementary school children computed en masse the age of the captain by adding (or subtracting, multiplying or dividing) goats and sheep, as convincing evidence of their "lack of sense-making" (Baruk, 1985; see also Verschaffel et al., 2000). However, in a few later studies, wherein children were asked to explain their previously given solutions to the captain's problem, it was found that they came up with "magic contexts" – personal elaborations to justify their previously given "absurd" answer, for example: *He bought an animal for each year of his life, so he always knows how old he is* (Selter, 1994).

Word problems with multiple solutions. Other researchers have confronted children with problems involving multiple solutions, due to the presence of one or more ambiguities in the problem statement. Puchalska and Semadeni (1987) asked first and second graders to solve the following problem "Gapcio gave Dolly the following problem: There were sparrows on a tree. I saw 5 sparrows and Dick saw 6 sparrows. How many sparrows were there on the tree? What should Dolly say to Gapcio?".

They also found that many children considered that Dolly should answer: *11 sparrows*. Likewise, Verschaffel et al. (1994) presented upper elementary school pupils a variety of non-standard (or as the authors called them "problematic") word problems, which all involved an ambiguous or disputable relationship between the "set up" of the word problem, on the one hand, and the suggested underlying mathematical structure, at least if one seriously thinks about the realities of the situation described in the problem, such as the following school distance problem "Bruce and Alice go to the same school. Bruce lives at a distance of 17 kilometers from the school and Alice at 8 kilometers. How far do Bruce and Alice live from each other?". The vast majority of the pupils reacted to this problem in the expected way by giving a single, precise numerical response based on a calculation with the numbers given in the problem (e.g., *either* an addition *or* a subtraction with the two given numbers). Actually, the number of pupils who pointed to the existence of more than one solution, was dramatically low, namely 5%.

Given solution problems. Other researches presented children with still another type of non-standard word problems, namely problems that can be solved by giving, as a solution, one of the numbers included in the problem. Likewise, Selter (1994) asked third graders to solve a variant of the captain's problem that included, besides the two additional irrelevant data, a number indicating the captain's age "A shepherd of 27 years old has 19 sheep and 10 goats. How old is the shepherd?". Even explicitly adding the shepherd's age did not help third graders to be successful. Rather than answering that the shepherd was 27 years old, half of the children either added the shepherd's given age to the others two given numbers, or did two other arithmetic operations with the three given numbers. Van Dooren, De Bock, Hessels, Janssens, and Verschaffel (2005) presented children from Grades 2 to 8 different proportional and non–proportional items with the solution included in the text problem, such as the following item: "A group of 5 musicians play a piece of music in 10 min. Another group of 35 musicians will play the same piece of music. How long will it take this group to play it?". Overall, only 30.6% of children realized that the solution was offered, with a general increase from Grade 3 to 8 and stagnation in the middle grades (Grades 5 and 6). Interestingly, the number of proportion-based erroneous answers increased from Grade 2 up to Grade 6. The authors concluded that, probably, the increase of these kinds of errors in fifth and sixth graders was related to the type of word problems recently treated in classroom.

Word problems with extraneous information. A final category of non-standard word problems involves problems that contain extraneous information. In other words, besides the numbers that are needed for the solution, the problem text also contains one or more irrelevant numbers (e.g., Carpenter et al., 1988; Kouba et al., 1988; Littlefield & Riesser, 1993). For instance, Littlefield and Riesser (1993) analyzed the difficulty of adding irrelevant information, presenting more and less mathematically successful fifth graders problems with different levels of similarity between the relevant and irrelevant information, such as the following problem with non-similar irrelevant data: "On the first day of the big race, Anthony ran 12 miles. He ran 9 miles the second day. Shelley sold 15 quarts of lemonade during the race. She gave away 8 quarts of ice water. How far did Anthony run?". For the successful students the percentage of success was around 95%, while for the less successful ones it was around 65%. Moreover it was generally easier for children to detect that the information was superfluous when it was not similar to the relevant data. The authors interpreted these results as another piece of evidence for mathematically unsuccessful children's superficial approach to word problems, involving "doing something with all the given numbers in the problem" with little or no attention to the actual role of these numbers in the problem structure.

In conclusion, despite clear differences between the various types of problems, previous research has provided various pieces of empirical evidence for elementary school children's superficial, routine-based, and/or non-realistic approach to different kinds of non-standard problems, whereby they solve these problems as if they were standard ones. As such, these findings are typically interpreted as evidence for the claim that students in general and weaker students in particular tend to approach and solve word problems with the strong expectation to get a standard one and without an intention to construct a rich situational and mathematical model of it. This expectation is based on (implicit) beliefs that (a) every problem presented in a math class has a solution; (b) there is only one single, precise and numerical correct answer to every word problem; (c) it is necessary to do calculations to solve a problem, and (d) all the given numbers in the statement must be used in the calculation.

Where do pupils' difficulties in solving non-standard problems come from?

The above interpretation brings us to the question: How do these superficial solution strategies for and beliefs about the solution of school arithmetic word problem develop? The development of students' tactics for and conceptions about word problem solving is assumed to occur implicitly, gradually, and tacitly through being immersed in the culture of the mathematics classroom in which they engage. Putting it another way, students' strategies and beliefs develop from their perceptions and interpretations of the didactical contract (Brousseau, 1997) or the socio-mathematical norms (Yackel & Cobb, 1996) that determine(s), explicitly to some extent, but mainly implicitly, how to behave in a mathematics class, what problems to expect, how to solve them and communicate about them with the teacher, and so on. More specifically, this enculturation seems to be mainly caused by two aspects of current instructional practice, namely (1) the nature of the problems given, and (2) the way in which these problems are conceived and treated by teachers (e.g., Freudental, 1991; Gerofsky, 1996; Orrantia, González, & Vicente, 2005; Reusser & Stebler, 1997; Schoenfeld, 1991; Wyndhamn & Saljo, 1997).

Let's first have a look at the first of these two explanatory factors. In an attempt to summarize the characteristics of traditional word problems that appear in classrooms and textbooks and which lie at the basis of students' inadequate strategies for and beliefs about solving word problems as discussed above, Reusser and Stebler (1997) wrote:

Most word problems used in mathematics instruction are phrased as semantically impoverished, verbal vignettes. Students not only know from their school mathematical experience that all problems are undoubtedly solvable, but also that everything numerical included in a problem is relevant to its solution, and everything that is relevant is included in the problem text. Following this authoring script, many problem statements degenerate to badly disguised equations (p. 323).

If the vast majority of the textbook and test problems have these characteristics, it should not be a surprise that many pupils develop gradually but inevitably superficial strategies for and inaccurate beliefs about word problem solving. In this respect, we point to an analysis of addition problems realized with math textbooks from Grade 1 to 6 produced by three of the most important Editors in Spain – Santillana, SM and Anaya, which showed that, for all grades together, only a total of 3.1, 1.6, and 0.6% of problems, respectively, had missing or superfluous information (Orrantia et al., 2005). Likewise, Schoenfeld (1991) found that in some U.S. textbooks the vast majority of the word problems (in some textbooks up to 90%) could be solved by means of the key-word strategy. Clearly, after frequently experiencing standard problems, pupils can be expected to develop a superficial approach to word problem solving and the accompanying inappropriate beliefs, instead of an authentic and deep mathematical modeling approach combined with more open and productive beliefs (Freudental, 1991; Reusser, 1988; Schoenfeld, 1991; Verschaffel et al., 2000; Wyndhamn & Säljö, 1997).

A second plausible explanatory factor for the development of the observed tactics for and beliefs about word problem solving is the way in which these problems are conceived and actually treated by teachers in the mathematics lessons. Generally speaking, teachers pay little or no attention to the articulation of and reflection on the peculiar genre of word problems during the mathematics lessons (Gerofsky, 1996). Based on an in-depth analysis of the teacher-pupil interactions around word problems in two typical fifth-grade classes, Depaepe, De Corte, and Verschaffel (2010) concluded that spontaneous comments of pupils about the ambiguous or problematic nature of word problems, or suggestions for alternative interpretations or solutions, were seldom picked up and valued in a typical mathematics class. Analogously, Verschaffel, De Corte, and Borghart (1997) asked (future) elementary school teachers, first, to solve a set of word problems that were problematic from a realistic point of view (such as the above-mentioned school distance problem), and, second, to evaluate four alternative answers from (imaginary) pupils to the same set of problems as "absolutely correct answer", "partly correct and partly incorrect answer", or "completely incorrect answer". For each problematic item, the four alternative responses always included were the standard non-realistic answer and the most reasonable realistic answer – as well as other two incorrect answers. Only half of the future teachers' own answers to these problematic items in test 1 were scored as realistic, and, with respect to test 2, their evaluation of the routine-based non-realistic pupil answers was considerably more positive than for the more sensitive answers based on realistic considerations!

In sum, the available evidence suggests that pupils' beliefs about and tactics for word problem solving do not develop as a result of direct teaching, but rather emerge from the nature of the textbook and test problems with which they are confronted and from the permanent interaction between teacher and students around these problems as part of the traditional classroom practice and culture.

Research objectives and questions

The purpose of the present study was threefold. The first objective was to analyze the impact on children's thinking of the four above-mentioned beliefs about arithmetic word problems, by presenting them four non-standard problems that directly contradicted these beliefs: (1) an *unsolvable* problem, (2) a problem with *multiple solutions*, (3) a problem in which the *solution is given*, and (4) a problem containing relevant as well as *irrelevant data*. By comparing pupils' performance on these four types of non-standard problems, we expected to yield information about the strength of the beliefs about math word problems they contradict. No predictions were formulated concerning the relative strength of these various beliefs, except that the problem with extraneous information was anticipated to be the easiest because it violates the problem structure (as described by Gerofsky, 1996) the least.

The second objective was to examine the evolution of children's performance on non-standard problem solving throughout the entire elementary school. Several previous studies yielded evidence for the difficulties experienced by students from different educational levels, but only few involved a developmental perspective (e.g., Baruk, 1985; Radatz, 1983; Reusser, 1988; Van Dooren et al., 2005), and we are not aware of any study that involved the whole elementary school range and that compared at the same time the four types of non-standard problems reviewed above. In the present study we wanted to evaluate whether the different level of children's success improved in upper educational levels. Again, our predictions were not straightforward. On the one hand, it could be argued that, taking in account that children from Grade 1 to 6 undergo an important development in general thinking and reasoning skills (e.g., Inhelder & Piaget, 1958; Montangero, 1996), and in various aspects of mathematical thinking in particular (for an overview see e.g., Nunes, Bryant, Sylva, & Barros, 2009; Schliemann & Carraher, 2002), there would be an increase in performance on the four non-standard problems from

Grade 1 to 6. On the other hand, since some researchers had reported only marginal improvement on such problems with age (Baruk, 1985; Reusser, 1988; Van Dooren et al., 2005), and in some cases, even a decrease in performance with grade (see Radatz, 1983; Van Dooren et al., 2005), one could argue that no strong positive effect of grade was to be expected.

The third objective was to analyze possible changes in the nature of the errors in relation with problem type and school level. In most previous studies, the analysis was focused on the percentage of correct responses. In the present study we aimed at a systematic and fine-grained categorization and analysis of the errors too. By looking more closely at these errors, we hoped to find more evidence of the nature of pupils' beliefs and how they affected their problem solving processes. In this respect, we wondered if pupils' gradual contact with and mastery of the four arithmetic operations would influence the nature of their erroneous solutions. For instance, one of the didactical contract's predictions assumes that pupils who solve word problems superficially tend to apply routinely the last studied or practiced operation (Sowder, 1988). So, we predicted that beginning elementary school children would be more inclined to rely on addition to find a solution for the non-standard problems, whereas in later grades they would misapply all four arithmetic operations (or combinations thereof).

Method

Participants

Participants were 300 primary school pupils, 50 for each group from Grades 1 to 6 (aged between 6 and 12 years) from a Spanish medium socio-cultural level public school (boys = 51%, girls = 49%). They all were randomly selected from the children with a normal level of mathematical performance from two classes of each grade level. Children who received special education teaching assistance, independently of the subject, were excluded.

Material

The material consisted of a test of six problems: four non-standard problems and two standard distractor problems involving a simple subtraction with the two numbers given in the problem (e.g., "Lucía has 20 euros in his piggy bank and she buys a ring that costs 7 euro. How much money does Lucía have now?"). The non-standard problems were, as stated above, formulated to violate four well-documented beliefs that children have about the word problem solving, namely: (a) an *unsolvable* problem (violating the belief that every word has a numerical solution); (b) a *multiple solutions* problem (contrary to the belief that there is only one single numerical response to a word problem); (c) a problem of which the *solution is given* in the problem statement (opposed to the belief that an arithmetical operation on the given numbers is always required), and (d) a problem with *irrelevant data* (challenging the belief that the solution to a word problem is obtained by performing one or more operations with all the given numbers in the problem). The four non-standard problems are given in Chart 1.

Possible problem sequence effects were prevented by using a Latin Square design leading to a total of four versions of the test with a different presentation order of the four non-standard problems. The two distractor problems were always placed in the third and fifth position.

To avoid difficulties related to other possibly relevant task factors, all problems were formulated as problems with a similar semantic structure, namely a change structure with the unknown variable being the result set. Moreover, all problems were formulated

Chart 1

The Four Non-Standard Problems Used in the Study

Unsolvable	Maria goes to the circus with her friend Ana. Maria has 13 euro and her friend Ana gives her 7 euro to pay for the circus ticket. How much money does Ana have now?
Multiple solutions	Lucía buys a bag with 14 chewing gums of different flavors. Few of these chewing gums have a mint flavor. Since mint is her favorite flavor, she orders 8 mint chewing gums afterwards. How many mint chewing gums does Lucía have now?
Given solution	A shepherd has 17 sheep in his farm. He wants to expand the farm and he buys 8 goats. How many sheep does the shepherd have in the farm now?
Irrelevant data	Lorena buys a box with 12 crayons for her Arts class. Her friend Silvia gives her another box containing 3 pens and 9 crayons. How many crayons does Lorena have now?

Note. Problems were presented in Spanish.

in such a way that an erroneous solution based on superficial modeling would involve an addition with (all) the given numbers in the problem. Furthermore, to prevent that computational difficulties would play a decisive role, all numbers were kept small (i.e., below 20), with the first term between 10 and 19 and the other term(s) below 10.

Procedure

One of the four versions of the set of six problems was individually administered to each child from the six grade levels during the regular school time. Interviews took approximately 20 min. We decided to use this data collection procedure, because it is expected to yield more process-oriented data about pupils' conceptions about (how to approach and solve) word problems than a collective paper-and-pencil test (Verschaffel & De Corte, 1997; Wyndhamn & Säljö, 1997).

Each problem was presented in written form on a separate card, and was read aloud by the interviewer. In case children needed a second reading, they could ask the interviewer to do it or (re)read the problem themselves as many times as they wanted. For each problem, the children were first asked to give their solution and, afterwards, to explain it; or, if they could not answer the problem, to state their doubts about or difficulties with giving an answer. If children did not know how to solve the problem and/or asked for help, the interviewer just re-read the problem and, if they were still unable or unwilling to respond, the interviewer asked them to justify why they could not answer it or why they found it difficult to do so.

Data handing

All children's responses were transcribed during the interviews. Five categories of responses to the four experimental items were differentiated: one category for the correct answer and four categories of incorrect responses. For each problem, children could score either 1 point (if they gave the correct answer) or 0 points (if their answer was categorized in one of the four error categories), resulting in a total score between 4 and 0. The procedure to assign the solutions to one of the five response categories was based on a combination of: (a) the analysis of the child's answer (e.g., the numerical answer to the problem, the child's statement that the problem was unsolvable, or that it had more than one possible solution, etc.); (b) the analysis of the child's explanations of the arguments for the arithmetic operation(s) (s)he performed and/or the non-computation based response given (see the Appendix for a more detailed explanation of the coding scheme). Taking in account the objectives of the study, children's technical computational errors were neglected when categorizing their answers into one of these five categories.

1. Correct answer: The correct (numerical or not) response to the problem followed by an appropriate explanation based on the problem's requests; that is, children needed to make explicit, in function of the problem type, that it was not possible to find a solution because there was some missing numerical information (*unsolvable* problem), that it was possible to offer more than one solution depending on the value of "few" (*multiple solutions* problem), that it was not necessary to do an arithmetic operation since the (numerical) answer to the question was given (*given solution* problem), or that some data were not needed to solve the problem (*irrelevant data* problem).
2. Incorrect addition answer: Responses based on the addition of all the numbers given in the problem statement, namely the two given numbers in the *unsolvable*, *multiple solutions* and *given solution* problems and the three given numbers in the *irrelevant data* problem.
3. Other incorrect operation(s) errors: Responses based on applying one or more *other* arithmetical operations with the given numbers (i.e., a subtraction, multiplication, division or a combination of two or more of the four arithmetical operations).
4. No answer: When children did not give any response at all or responded that they could not answer the problem and were unable to offer any explanation about the reasons for not being able to answer the problem.
5. Other answer: All other solutions that could not be coded in one of the previous categories.

All children's responses were categorized by the first author. All responses from 10% randomly selected children from each of the six grades were independently coded by a second coder, leading to an inter-rater reliability of .96, which is an almost perfect agreement (weighted kappa coefficient, Cohen, 1968).

Results

The results are presented in two sections. First, we present the results of the quantitative analysis of the impact of the independent variables on the number of correct answers, and afterwards we report the results of the qualitative analysis of the pupils' errors.

Quantitative analysis

Table 1 gives the percentages and standard deviations of the correct responses for the four experimental items in the six grade levels. Data were analyzed by a mixed ANOVA 6 (Grade: 1 *vs*. 2 *vs*. 3 *vs*. 4 *vs*. 5 *vs*. 6), as a between-subject factor, x 4 (Problem type: *unsolvable vs. multiple solutions vs. given solution vs. irrelevant data*), as the within-group factor, with repeated measures for the last factor.

The first notable result was children's great difficulties to solve the four non-standard problems, as shown by the overall percentage of correct answers which was only 37.9%. This finding was in line with our first prediction, and with the results of several previous studies, in which an alarmingly high number of incorrect answers in solving different kinds of non-standard problems had been found (see also Baruk, 1985; Jiménez & Ramos, 2011; Littlefield & Riesser, 1993; Radatz, 1983; Reusser, 1988; Selter,

Table 1

Percentage of Correct Answers (and Standard Deviations) in each Grade to the Four Problem Types

	Unsolvable	Multiple solutions	Given solution	Irrelevant data	Mean
Grade 1	2 (.14)	8 (.27)	18 (.39)	34 (.48)	15.5 (.32)
Grade 2	4 (.20)	10 (.30)	26 (.44)	32 (.47)	18 (.35)
Grade 3	18 (.39)	30 (.46)	36 (.49)	48 (.51)	33 (.46)
Grade 4	30 (.46)	36 (.49)	62 (.49)	70 (.46)	49.5 (.48)
Grade 5	22 (.42)	50 (.51)	72 (.45)	78 (.42)	55.5 (.45)
Grade 6	34 (.48)	48 (.51)	60 (.50)	82 (.39)	56 (.47)
Mean	18.3 (.39)	30.3 (.46)	45.7 (.50)	57.3 (.50)	

1994; Van Dooren et al., 2005; Verschaffel et al., 1994). In general, the low level of success for these problems confirms children's misconceptions about word problems and how to handle them in the mathematics classroom. Looking at this overall result from an individual perspective, we found that a high percentage of pupils was unable to adequately consider the specific problems' requests, as evidenced by the fact that 34.7% of children did not correctly solve any of the four non-standard problems, and that only 12% of participants correctly solved all four problems.

Second, the ANOVA showed a significant main effect for Problem type, $F(3,882) = 73.18$, $p < .01$, $\eta_p^2 = .199$. More specifically, the order of difficulty was increasing starting with the *unsolvable* (18.3%), and followed by the *multiple solutions* problem (30.3%), the *given solution* problem (45.7%) and, finally, the *irrelevant data* problem as the easiest one (57.3%) (see Table 1). Post hoc analyses showed that all these increases in the overall difficulty level of the four problems were significant (Bonferroni's test, $p < .01$). Assuming that the above-mentioned differences for each problem type are a reflection of the strength of the misbelief they aim to address, this second finding shows, as expected, that certain beliefs are more established in the children's thinking than others.

Third, a significant effect of Grade was found, $F(5,294) = 17.57$, $p < .01$, $\eta_p^2 = .230$. Post hoc analyses showed significant differences between each of the three lower grades and each of the three upper grades (Bonferroni's test, $p < .01$), but no significant differences within the three lower (15.5 *vs.* 18 *vs.* 33%, respectively, for Grades 1, 2 and 3) or within the three upper grades (49.5 *vs.* 55.5 *vs.* 56%, respectively, for Grades 4, 5 and 6) (see Table 1). So, while we found an increase of the level of success throughout the elementary school, the generally low level of performance in the upper grades and the lack of any significant progression during these grades were quite striking. As shown in Table 1, in Grades 4, 5 and 6 still at most 50% of children's responses could be considered as correct.

Finally, the interaction between Problem type and Grade was also significant, $F(15,882) = 1.71$, $p < .05$, $\eta_p^2 = .028$. As shown in Figure 1, the developmental pattern of children's performance was very similar for each of the four non-standard problems because pupils' performance always increased with grade, except in a few cases wherein the level stagnated (i.e., Grades 5 and 6 in the *multiple solutions* problem) or even decreased (Grade 5 in the *unsolvable* prob-

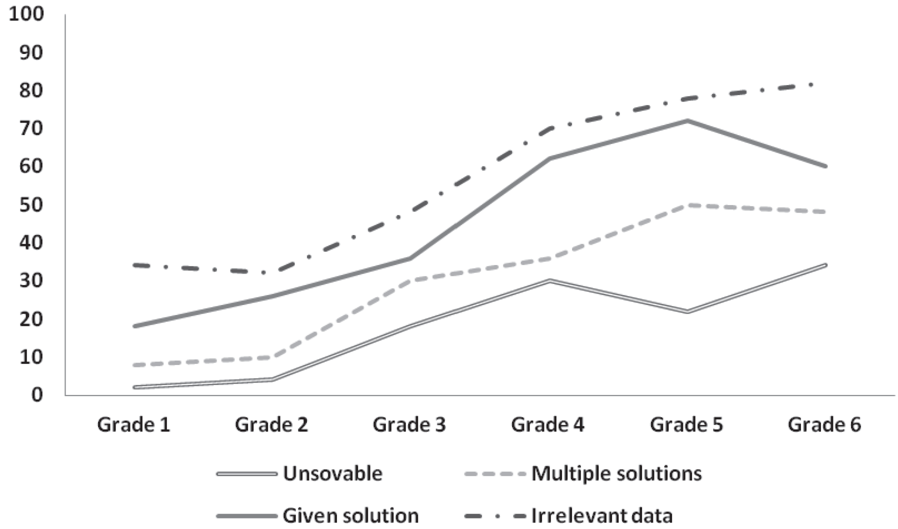

Figure 1. Percentages of level of success for each problem type per grade.

lem and Grade 6 in the *given solution* problem) (see also Table 1). As a result of this consistency in the developmental pattern for each type of problem, the above-mentioned overall order of difficulty was reflected in each grade (see Figure 1).

Mean differences for *unsolvable* problems between Grades 4 and 5 (M = .30 *vs*. M = .22, respectively) and for *given solution* problems between Grades 5 and 6 (M = .72 *vs*. M = .60, respectively) were calculated by independent unpaired Student's t tests. Results showed significant mean differences only in the latter case ($p < .05$).

Qualitative error analysis

In this section we first comment on the overall distribution of the errors over the different error categories. Then we report the results of the qualitative error analysis for the four types of problems, and, finally, for the six grade levels. Taking in account that the overall percentage of errors for every Problem type and Grade was different (see Table 1), we have computed the percentages of incorrect answers in relation to the total number of errors being made.

The overall analysis of errors showed that, in every Problem type, most of them were addition incorrect answers (64.5%), followed by other incorrect operation(s) errors (21.1%), then other answers (11.5%), and finally no answers (2.8%). As expected, the addition appearance of the four non–standard problems caused that most errors were produced by a superficial analysis of the problem, resulting in adding up all the numbers offered in the statement while neglecting the real problem's requests. A rather high percentage of errors came from the erroneous application of one of the three other incorrect operations, namely subtraction, multiplication or division, with all the given numbers, or a combination thereof. The category of errors that was ranked third, namely other answers, involved a variety of numerical errors that will be discussed later on. Finally, only 2.8% of totality of the incorrect answers occurred because children did not offer any solution (see Table 2). This percentage of no answers is remarkably low, most probably, because the combination of the nature of our non-standard problems strongly evoked an "add all numbers" solution strategy. But also, the fact that children were not collectively but individually tested may have led to such a low percentage of no-answer errors.

Table 2

Percentage of Incorrect Answers, in Relation to the Total Percentage of Errors, over the Problem Type

	Unsolvable	Multiple solutions	Given solution	Irrelevant data	Mean
Addition incorrect answer	39.6	60.2	79.2	82.7	65.4
Other incorrect operation(s) errors	37.1	21.1	15.2	11.0	21.1
No answer	2.1	4.7	1.9	0.0	2.8
Other answer	21.2	14.0	3.7	6.3	11.3

Note. Total percentage of errors for each problem was *unsolvable* = 81.7%, *multiple solutions* = 69.7%, *given solution* = 54.3%, and *irrelevant data* = 42.7%.

Furthermore, we found a different pattern of distribution of children's incorrect answer categories for the four non-standard problems (see also Table 2). The frequency of errors based on the addition of all given numbers increased with decreasing problem difficulty, from only 39.6% for the most difficult problem, namely the *unsolvable* problem, to 82.7% for the easiest problem, namely the *irrelevant data* problem. To the contrary, the other incorrect operation(s) errors and other answers tended to augment with increasing problem difficulty.

We now turn to a more fine-grained analysis of these error data, by looking at the distribution of the three main error categories and the precise nature of these errors in each of the four Problem types. As reported before, the *unsolvable* problem, violating the idea that "every problem has a solution", was the most difficult one. Remarkably, the distribution of the three main errors categories was most equal in this problem type. That is, 39.6% of all errors were produced by adding up the two given numbers, 37.1% came from the application of one or more other incorrect arithmetic operations, and 21.2% were considered as other answers. A closer look at the errors in the latter category revealed three different types of other answers. First, 15.9 of these 21.2% other answers were produced because children noticed that a numerical element was missing and modified the problem to make it solvable, by adding themselves that missing number (e.g., *Ana has now 0 euro, because she had before 7 euro and she gave to María 7, so she has no more money left. She was supposed to have 7 euro, because, if not, it would have been written in the problem statement*). Despite such responses can be considered as nice examples of children's "sense-making" attempts (Schoenfeld, 1991; Selter, 1994; Verschaffel et al., 2000), their solutions were considered as incorrect, in the context of the present study, because they elaborated the problem in a "random" way to allow the computation of a solution, instead of working within the (problematic) task constraints. Second, 2.5% of other answers were solutions whereby children also did respond with a precise numerical answer, showing again the need to come up with a numerical solution anyhow, but without providing a clear rationale for how they arrived at that number (e.g., *the solution is 6 because this is the answer to the problem*). Third, 2.8% of other answers showed a misunderstanding about which was the missing data (e.g., *I do not know the answer, because the prize of the tickets is not included in the problem statement*).

In order of decreasing difficulty, children found fewer —but still considerable— difficulties in solving the *multiple solutions* problem, which contradicted the belief that for every word problem "there is only a precise numerical correct answer". For this problem, most errors were again categorized as addition incorrect answers (60.2%), but there were also quite some other incorrect operation(s) errors (21.1%) and other answers (14%). Interestingly, it was again possible to distinguish different types of responses in this last category. The first one involved again various kinds of self-made adaptations of the problem story to transform it into a standard problem solvable by just one single numerical answer (10.5 of 14% of other answers). For example, some pupils provided themselves additional semantic information with a view to get rid of the imprecise numerical element "a few" (e.g., *I think she already ate the few mint chewing gums she had initially, so we should not take in account these 'few' and so she now has only 8 mint chewing gums*); in other occasions they converted the imprecise information element "a few" in a precise number (e.g., *a few can be 1, only 1, so now she has 9 chewing gums*). The second type of erroneous other answers (the remaining 3.5 of the 14% of other answers) were solutions whereby children offered an invented precise numerical answers, without any clear explanation, showing the need to offer a solution for the problem even when they did not know how to solve it (e.g., *5 because I have guessed*).

The *given solution* problem, explicitly going against the misbelief that "it is always necessary to do an arithmetic operation to solve a problem", was easier for the children. In this problem, most errors were expected addition incorrect answers (79.2%) and, to a much lesser extent, other incorrect operation(s) errors (15.8%). Only 3.7% of total errors solutions were categorized as other answers: 2.4% of them were also produced by the need to offer a numerical solution different from a given number but resulting from some arithmetical operation, without understanding or explaining the procedure (e.g., *105 sheep because he shepherd expands the farm*), and in 1.3 of these 3.7% of other answers because they incorrectly considered the problem as an unsolvable problem rather than as a given solution problem, realizing that they could not add the data (e.g., *it cannot be solved because the shepherd has bought goats and the problem is asking for the sheep, so I cannot do anything with the goats*).

Finally, the easiest problem was the *irrelevant data* one, in which pupils had to do an arithmetic operation, namely an addition, but the challenge was to detect that one of the data included in the statement – the second one, was irrelevant and should therefore not be added to the two other ones. In this problem, most errors came, as expected, from adding all three given numbers (82.7%), in line with pupils' belief that "all the numbers including in a problem must be used in the calculation". Only 11% of responses came from the incorrect application of one or more arithmetic operations different from addition. To a less extent, 6.3% of the errors were other answers: 5.6% of them because children added up the two first numbers, the relevant and the irrelevant one, ignoring the other relevant number (e.g., *15 because Lorena had 12 crayons and she received 3 more*) and 0.7 of the 6.3% of other answers because children offered a solution without providing a clear or acceptable explanation for how they arrived at their numerical answer (e.g., *100 crayons because she received a gift*).

Table 3

Percentage of Incorrect Answers, in Relation to the Total Number of Errors, over the Grade

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Mean
Addition incorrect answer	75.1	73.3	58.2	53.5	43.8	45.5	58.2
Other incorrect operation(s) errors	9.5	17.7	32.8	25.7	40.5	26.1	25.4
No answer	3.0	1.2	1.5	0.0	3.4	6.8	2.7
Other answer	12.4	7.9	7.5	20.8	12.3	21.6	13.8

Note. Total percentage of errors for each grade was *Grade 1* = 84.5%, *Grade 2* = 82%, *Grade 3* = 69.7%, *Grade 4* = 50.5%, *Grade 5* = 44.5%, and *Grade 6* = 44%.

The analysis of errors in function of Grade, as shown in Table 3, revealed some age-related differences in the nature of the errors being made, according with the didactical contract's predictions. As shown in this table, the frequency of the dominant error category, namely adding up all numbers, tended to decrease with Grade, except for the last two grades. More specifically, the percentages were, from Grade 1 to Grade 6, respectively, 75.1%, 73.3%, 58.2%, 53.5%, 43.8%, and 45.5%. To the contrary, the percentages of errors due to the incorrect application of other arithmetic operations, such as subtraction, multiplication, division and different combinations of them, tended to be higher in the upper grades than in the lower grades, although that pattern was somewhat less clear.

It is not so surprising that the nature of the wrong-operation errors change with grade. While first graders in Spain are typically confronted almost exclusively with the operation of addition, in Grade 2, they are systematically and intensively confronted with subtraction, while in Grade 3 the two other operations get a lot of instructional attention, and afterwards all four operations get more or less equal attention. So, the decrease in the percentage of incorrect addition errors and the accompanying increase in the percentage of other wrong operation errors during the first years of elementary school seem to be a direct reflection of the kind(s) of operation(s) with which the pupils had been confronted most frequently or most recently in their mathematics classes, as reported by Sowder (1988; see also Verschaffel et al., 2000).

Discussion

Previous studies have found an alarmingly high level of incorrect answers when elementary school children have to solve different kinds of non-standard word problems. As a possible explanation, it has been claimed that children's failure come from mathematics-related beliefs that determine how they approach word problems in the mathematics classroom. These beliefs are claimed to be the individual-psychological pendant of the socio-mathematical classroom norms (Yackel & Cobb, 1996) or, as Brousseau (1997) would call it, the *didactical contract*. For example, some studies in which some children have been interviewed about their previous incorrect responses to different non-standard word problems, have evidenced that some of these errors, result from the belief(s) that: (1) every word problem is solvable, (2) there is only one numerical and precise correct answer, (3) it is necessary to do calculations to solve a word problem, and (4) all numbers that are part of the word problem must be used in order to calculate the solution. (Caldwell, 1995; Hildalgo, 1997; Reusser & Stebler, 1997; for a review see Verschaffel et al., 2000).

The present study systematically investigated the development of the four above-mentioned beliefs about word problems. Groups of children from Grade 1 to 6 were asked, in the context of an individual interview, to solve four problem types of word problems that violated these four beliefs. The study has confirmed, first, children's great difficulties in solving non-standard problems that contradicted these four specific beliefs.

Second, it has revealed substantially different levels of difficulty among the non-standard problems depending on which belief the problem contradicted. Specifically, the *unsolvable* problem was the most difficult one, followed, in decreasing order of difficulty, by the *multiple solutions*, the *given solution*, and the *irrelevant data* problem. These differences indicate that some beliefs are more established in the children's thinking than others. The finding that the *irrelevant data* problem elicited most correct responses suggests that the belief that all given numbers should be used in a word problem, is the least entrenched one. In this respect, we remind that Gerofsky's description of a (standard) word problem (1996) does not exclude the possibility of irrelevant data (although it considers it as highly atypical). The fact that the *given solution* problem was found to be more difficult, seems to indicate that children believe more strongly that a proper solution of a word problem requires at least one arithmetic operation (with the given numbers). The finding that the *multiple solution* problem was still more difficult, can be interpreted as evidence that it is even harder for children to overrule the belief that every word problem has (only) one numerical answer. Finally, given that the *unsolvable* problem was the most difficult of all four non-standard problems, we may conclude that the belief that every word problem is altogether solvable is extremely strong. In this respect, we remind that Puchalska and Semadeni (1987) observed that many children, who (rightly) described a problem as unsolvable, also thought that a better student or the teacher could solve it for sure. The present study did not investigate the origins of the differential strength of these four beliefs, but our literature review (see: Introduction) suggests that they lie in the nature of the textbook problems given to the pupils and the way in which these problems are conceived and treated by teachers. Although there is already some evidence for this claim (e.g., Freudental, 1991; Gerofsky, 1996; Reusser & Stebler, 1997; Schoenfeld, 1991; Wyndhamn & Saljo, 1997), more research is needed.

Third, the percentages of correct responses tended to increase with grade, but we only found a significant difference between the first and last three grades of elementary school, and still far from flawless performance in the upper grades, as previous researches with elementary and secondary school children and, even, student-teachers (e.g., Baruk, 1985; Reusser, 1988; Van Dooren et al., 2005; Verschaffel et al., 1994). Moreover, this general finding was observed for each of the four problem types. In comparison to the amount of formal knowledge that pupils are accumulating in the course of six years of elementary mathematics education, their relatively low performance on these four types of non-standard word problems, even at the end of elementary school, is quite alarming, suggesting that what they have learnt during the word problem solving lessons is essentially to routinely solve standard word problems but not to react properly and reflect thoughtfully upon problems that deviate from that standard.

Finally, the qualitative error analysis showed that the most frequent type(s) of errors are those resulting from a process whereby the non-standard problem was —mostly unnoticed but sometimes consciously— treated as or transformed into a standard one.

From a methodological perspective, we acknowledge that the present study provided only *indirect* behavioral evidence for the role of beliefs about word problems on children's actual solutions of these problems. Therefore, it would be interesting to complement the systematic analysis of children's correct and incorrect responses to non-standard problems that violate certain beliefs, with an analysis of their reactions to questionnaires and/or interviews in which they are more directly questioned about these beliefs. The confrontation of these different kinds of data could yield deeper insight into the nature of the relationship between children's approaches to and beliefs about word problems, as well as about children's level of consciousness of these beliefs.

Anticipating on the forthcoming empirical evidence of the needed further ascertaining studies, from an educational perspective, our study adds to the claims made by several authors (Staub & Reusser, 1995; Verschaffel et al., 2000) to seriously alter the dominant practice and culture of word problem solving lessons, by changing both the nature of the problems given to the pupils and the way these problems are treated by the teacher. With respect to the nature of the problems, many authors have pleaded for more purposeful variety in the problems that are given to the pupils. More specially, besides the standard problems that are cleanly and unquestionably modeled by applications of (one or combinations of) the four basic arithmetic operations with the numbers given in the problem, children should at regular times also be confronted with non-standard problems including (a) problems with superfluous, misleading, or missing data, targeting the belief that all data in the problem statement, and no others, should be fed into the calculations, (b) problems where alternative situation models and solutions are possible, countering the belief that there is a single correct interpretation and solution for every problem, (c) problems calling for an appropriate mix of forms of answers, such as estimations instead of precise calculations, answers in the form of intervals, and explanations why a given problem is unsolvable, but also (d) tasks requiring pupils to pose or reformulate word problems themselves instead of solving them. Such variation in the problem set would break up the beliefs among children about the rules involved in the didactical contract. With respect to the way the problems are treated by the teacher, researchers have pleaded for alternative ways of working and communicating with pupils around these problems. This could involve explicit discussions with children about to the genre of word problems, their function, and their implicit expectations and rules, as well as their relation with solving modeling problems in the real world out of school (Gerofsky, 1996; Staub & Reusser, 1995; Verschaffel et al., 2000). In this respect, we refer to studies by Chapman (2009) and Depaepe, De Corte, and Verschaffel (2009), pleading for a better balance between a "paradigmatic" and a "narrative" approach to word problem solving in the classroom discussions. We acknowledge that the above recommendations constitute a major program for a thorough overhaul of word problems and their use in the mathematics lessons. However, the research literature contains examples of studies showing significant positive effects on elementary school children's approaches to and beliefs about word problems of experimental intervention programs that deliberately and systematically aim at changing the nature of the problems and their instruction (see e.g., Mason & Scrivani, 2004; Verschaffel et al., 1999; Verschaffel & De Corte, 1997).

References

Baruk, S. (1985). *L'âge du capitaine. De l'erreur en mathematiques*. Paris: Seuil.

Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Dordrecht, The Netherlands: Kluwer academic publishers.

Caldwell, L. (1995). *Contextual considerations in the solution of children's multiplication and division word problems*. Unpublished undergraduate thesis, Queen's University, Belfast, Northern Ireland.

Carpenter, T. P., Lindquist, M. M., Brown, C. A., Kouba, V. L., Silver, E. A., & Swafford, J. O. (1988). Results of the four NAEP Assessment of Mathematics: Trends and conclusions. *Arithmetic Teacher, 36*, 38-41.

Chapman, O. (2009). Teachers' conceptions and use of mathematical contextual problems in Canada. In L. Verschaffel, B. Greer, W. Van Dooren, & S. Mukhopadyay (Eds.), *Words and worlds. Modelling verbal descriptions of situations* (pp. 227-244). Rotterdam, The Netherland: Sense Publishers.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*, 213-220. doi: 10.1037/h0026256

De Corte, E., & Verschaffel, L. (1985). Beginning first graders' initial representation of arithmetic word problems. *Journal of Mathematical Behavior, 4*, 3-21.

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. Mukhopadyay (Eds.), *Words and worlds. Modelling verbal descriptions of situations* (pp. 245-263). Rotterdam, The Netherland: Sense Publishers.

Depaepe, F., De Corte, E., & Verschaffel, L. (2010). Teachers' approaches towards word problem solving: Elaborating or restricting the problem context. *Teaching and Teacher Education, 26*, 152-160. doi: 10.1016/j.tate.2009.03.016

Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht, The Netherlands: Kluwer academic publishers.

Gerofsky, S. (1996). A linguistic and narrative view of word problems in mathematics education. *For the Learning of Mathematics, 16*(2) 36-45.

Hidalgo, M. C. (1997). *L'activation des connaissances à propos du monde réel dans la résolution de problèmes verbaux en arithmétique*. Unpublished doctoral dissertation, Université Laval, Québec, Canada.

Ilany, B., & Margolin, B. (2010). Language and mathematics: Bridging between natural language and mathematical language in solving problems in mathematics. *Creative Education, 1*, 138-148. doi: 10.4236/ce.2010.13022

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London, UK: Routledge and Kegan Paul.

Jiménez, L. (2012). La aplicación del conocimiento contextualizado en la resolución de problemas matemáticos: un estudio sobre las dificultades de los niños en la resolución de problemas no rutinarios. *Cultura y Educación, 24*, 351-362. doi: 10.1174/113564012802845640

Jiménez, L., & Ramos, F. J. (2011). El impacto negativo del contrato didáctico en la resolución realista de problemas. Un estudio con alumnos de 2.º y 3.º de Educación Primaria. *Electronic Journal of Research in Educational Psychology, 9*, 1155-1182.

Kouba, V. L., Brown, C. A., Carpenter, T. P., Lindquist, M. M., Silver, E. A., & Swafford, J. O. (1988). Results of the fourth NAEP Assessment of Mathematics: Number, operations, and word problems. *Arithmetic Teacher, 35*, 14-19.

Lave, J. (1992). Word problems: A microcosm of theories of learning. In P. Light, & G. Butterworth (Eds.), *Context and cognition: Ways of learning and knowing* (pp. 74-92). New York: Harvester Wheatsheaf.

Littlefield, J., & Rieser J. J. (1993). Semantic features of similarity and children's strategies for identification of relevant information in mathematical story problems. *Cognition and Instruction, 11*, 133-188. doi: 10.1207/s1532690xci1102_2

Mason, L., & Scrivani, L. (2004). Enhancing students' mathematical beliefs: An intervention study. *Learning and Instruction, 14*, 153-176. doi: 10.1016/j.learninstruc.2004.01.002

Montagero, J. (1996). *Understanding changes on time: The development of diachronic thinking in 7-12 year old children*. London, UK: Taylor & Francis.

Nunes, T., Bryant, P., Sylva, K., & Barros, R. (2009). *Development of math capacities and confidence in Primary school*. Oxford, UK: University of Oxford.

Orrantia, J., González, L. B., & Vicente, S. (2005). Un análisis de los problemas aritméticos en los libros de texto de Educación Primaria. *Infancia y Aprendizaje, 28*, 429-451. doi: 10.1174/021037005774518929

Puchalska, E., & Semadeni, Z. (1987). Children's reactions to verbal arithmetical problems with missing, surplus or contradictory data. *For the Learning of Mathematics, 7*(3), 9-16.

Raddatz, H. (1983). Untersuchungen zum Lösen eingekleideter Aufgaben. [Research on solving word problems]. *Zeitschrift für Mathematik-Didaktik, 4*, 205-217.

Reusser, K. (1988). Problem solving beyond the logic of things: Contextual effects on understanding and solving word problems. *Instructional*

Science, 17, 309-338. doi: 10.1007/BF00056219

Reusser, K., & Stebler, R. (1997). Every word problem has a solution: The suspension of reality and sense-making in the culture of school mathematics. *Learning and Instruction, 7*, 309-328. doi: 10.1016/S0959-4752(97)00014-5

Schliemann, A. D., & Carraher, D. W. (2002). The evolution of mathematical understanding: Everyday versus idealized reasoning. *Developmental Review, 22*, 242-266. doi: 10.1006/drev.2002.0547

Schoenfeld, A. H. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce of formal and informal mathematics. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 311-343). Hillsdale, NJ: Lawrence Erlbaum Associates.

Selter, C. (1994). How old is the captain? *Strategies, 5*, 34-37.

Sowder, L. (1988). Children's solution to story problems. *Journal of Mathematical Behavior, 7*, 227-238.

Staub, F. C., & Reusser. K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In C. A. Weaver, S. Marines, & C. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 285-305). Hillsdale, NJ: Lawrence Erlbaum Associates.

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2005). Not everything is proportional: Effects of age and problem type on propensities for overgeneralization. *Cognition and Instruction, 23*, 57-86. doi: 10.1207/s1532690xci2301_3

Verschaffel, L., & De Corte, E. (1997). Teaching realistic mathematical modeling in the elementary school: A teaching experiment with fifth graders. *Journal for Research in Mathematics Education, 28*, 577-601. doi: 10.2307/749692

Verschaffel, L., De Corte, E., & Borghart, L. (1997). Pre-service teachers' conceptions and beliefs about the role of real world knowledge in mathematical modelling of school word problems. *Learning and Instruction, 4,* 339-359. doi: 10.1016/S0959-4752(97)00008-X

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

Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Design and evaluation of a learning environment for mathematical modeling and problem solving in upper elementary school children. *Mathematical Thinking and Learning, 1*, 195-229. doi: 10.1207/s15327833mtl0103_2

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, The Netherland: Swets & Zeitlinger.

Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 557-628). Reston, VA: NCTM.

Wyndhamn, J., & Säljö, R. (1997). Word problems and mathematical reasoning: A study of children's mastery of reference and meaning in textual realities. *Learning and Instruction, 7,* 361-382. doi: 10.1016/S0959-4752(97)00009-1

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*, 458-477.

Appendix

Overview of the Criteria for Coding Pupils' Answers to the Four Non-Standard Problems

		Unsolvable	Multiple solutions	Given solution	Irrelevant data
Correct answer	Answer	"I do not / cannot know how much money Ana has left."	—More than one numerical solution (e.g., "9 or 10") or an estimation (e.g., "approximately 10")	The given number "17"	The result of adding up the two relevant data ("21")
	Additional comments*	"I do not know the money that Ana has before"	"Because I do not know how many 'a few' are" or "Because 'a few' can be any (small) number"	"Because it is stated in the problem that this is the number of sheep is 17" or "Because the number of sheep has not changed"	"Because pens are not crayons" or "Because the question does not ask for the pens"
Addition incorrect answer	Answer based on the addition of the given number	13 + 7 = 20	14 + 8 = 22	17 + 8 = 25	12 + 3 + 9 = 24
Other incorrect operation(s) errors	Answer based on other arithmetical operation	13 – 7 = 6	14 – 8 = 6	17 – 8 = 9	(12 + 9) – 3 = 18
	or different combinations between them	(13 + 7) – 7 = 13	(14 - 8) × 8 = 48	(17 × 8) – 8 = 128	(12 x 9) / 3 = 36
No answer	Answer	No answer, or "I do not know"	No answer, "I do not know"	No answer, "I do not know"	No answer, "I do not know"
	Additional comments*	—"I do not understand the problem" or "I do not know what to do with this problem"	—"I do not understand the problem" or "I do not know what to do with this problem"	—"I do not understand the problem" or "I do not know what to do with this problem"	-"I do not understand the problem" or "I do not know what to do with this problem"
Other answer	Answer	Answers that cannot be coded in any of the previous categories	Answers that cannot be coded in any of the previous categories	Answers that cannot be coded in any of the previous categories	Answers that cannot be coded in any of the previous categories

Note. *Additional comments were needed to code correct answer and other answer categories.

Laura Jiménez is Assistant lecturer of Psychology at the University of Castilla-La Mancha, Cuenca, Spain. Her research interests focus on elementary school students' beliefs about and approaches towards mathematical problems, in relation to the features of the educational system. She has already published several articles in this field.

Lieven Verschaffel is full Professor of Educational Sciences at the University of Leuven, Belgium. His major research interests are teaching and learning of problem-solving and learning skills and mathematics education. He has already published numerous articles, books and book chapters about these topics.

Received date: 23-03-2013 Review date: 14-05-2013 Accepted date: 04-09-2013