



# Effect of algebraic language and problem text wording on problem model accuracy when solving age word problems

Carlos Soneira<sup>1</sup> · José Antonio González-Calero<sup>2</sup> · David Arnau<sup>3</sup>

Accepted: 26 April 2023 / Published online: 19 June 2023  
© The Author(s) 2023

## Abstract

The use of the algebraic method for solving word problems is a challenging topic for secondary school students. Students' difficulties are usually associated with extracting the problem's network of relationships between quantities and with formalizing these relationships into algebraic language in a problem model. Both sources can coexist and interact; thus, it is usually not possible to determine which source of difficulty is more relevant. In addition, there are specific errors, such as the error by multiple referents for the unknown, which are directly linked to the wording of the problem text, and in which the same two sources of error coexist. In this work, we present the results of an experiment conducted with 255 secondary school students assessing the effect of two common difficulties on the accuracy of problem models and on the rate of multiple referents for the unknown. The first difficulty is the use of algebraic language in the construction of the problem model; the second is the use of the same expression to designate different quantities within the problem text. We used a  $2 \times 2$  between-between design, with one factor related to the symbolic language (algebraic or arithmetic) in which the problem model is constructed, and the other factor related to the actual language features of the text problem. Our results indicate that overall, the main source of difficulty for students is the use of algebraic language to formalize a problem model, representing a large effect size.

**Keywords** Word problem · Problem model · Algebraic language · Secondary school

---

✉ Carlos Soneira  
carlos.soneira@udc.es

<sup>1</sup> Department of Pedagogy and Didactics, Faculty of Educational Studies, Universidade da Coruña, Elviña, 15071 A Coruña, Spain

<sup>2</sup> Department of Mathematics, School of Education of Albacete (Edificio Simón Abril), University of Castilla-La Mancha, Plaza de La Universidad, 3, 02071 Albacete, Spain

<sup>3</sup> Department of Didactics of Mathematics, University of Valencia, 46022 Valencia, Spain

## 1 Introduction

Word problem solving is an important topic in mathematics education at all educational levels. To solve these problems, the student must convert the information obtained after a comprehensive reading of the problem text into a set of mathematical relationships. Throughout elementary education, students solve word problems using arithmetic calculations, starting from known quantities to determine the unknown quantity. When they start secondary education, the students are introduced to the algebraic method, which implies assigning letters to unknown quantities and formulating equations. This transition involves significant difficulties (Filloy et al., 2008; Kieran, 2007) due to both the newness of this way of solving and its reliance on algebraic language.<sup>1</sup>

The research community has focused its attention on students' difficulties when solving word problems linked to algebraic language, such as the error termed multiple referents of the unknown (hereinafter multiple referents) by Stacey and MacGregor (1997/1999). This error occurs when students assign the same letter to two different quantities. Another important line of research has studied the influence of students' understanding of the problem text on their abilities to solve problems, an aspect shared with arithmetic solving (Boonen et al., 2013; Fuchs et al., 2015; Walkington et al., 2019). Indeed, the influence of reading comprehension impeding success in solving word problems has been supported by numerous studies over the years, and currently prevails as a relevant subject of study (Boonen et al., 2016; Daroczy et al., 2015). Natural language commonly employs words or phrases whose meanings vary within the same text, making it more difficult to connect different sentences (Fossard et al., 2012). This phenomenon is common in the text of word problems and, specifically, in age problems where the same expression designates different quantities in different sentences (Bloedy-Vinner, 1996; Boero et al., 2008).

Following the terminology of Bloedy-Vinner (1996), we refer to such expressions as expressions with evolving meaning. Consider the example below:

Example 1: Amaya is 9 years older than Andrea. In 3 years' time, Amaya's age will be twice Andrea's age. How old are they?

In example 1, the expressions "Amaya's age" and "Andrea's age" refer to both the current and future ages of the characters in different sentences. In the second sentence, the expressions "Amaya's age" and "Andrea's age" refer to "Amaya's future age" and "Andrea's future age," and the meaning of the written expressions are updated by the expression "In 3 years' time."

A common error when algebraically solving age problems consists of assigning the same letter to two different quantities that are expressed within the same set of words in the text (e.g., the same letter to Amaya's age at present and in the future). This is a particular instance of multiple referents and has attracted interest from the research community (e.g.,

---

<sup>1</sup> In this paper, arithmetic language means the set of symbols made up of the set of figures for numbers, the symbols for arithmetic operations between numbers, brackets, the equals sign, and signs for inequality. Its syntactic rules are those of arithmetic calculations. Algebraic language means the set of symbols which constitutes arithmetic language, together with letters to represent unknown quantities, with the same rules as in arithmetic language and with letters being handled as numbers. Equation means an equality between two expressions in algebraic language, which includes letters—handled just like numbers—to represent unknown quantities involved in relationships between quantities, and which is true only for some values of the unknown(s) (see Drouhard & Teppo, 2004).

Bloody-Vinner, 1996; Boero et al., 2008; Filloy et al., 2008, 2010; Molina et al., 2017; Soneira et al., 2018) because it is common among secondary school students. It has been reported that the occurrence of this error increases when the problem text contains expressions with evolving meaning (Bloody-Vinner, 1996; Soneira et al., 2018). However, these authors claimed that the causes of this error are unclear and, at the very least, there are two possibilities. On one hand, the solver could believe both clauses, “Amaya is 9 years older than Andrea” and “Amaya’s age will be twice Andrea’s age,” refer to the same thing: Andrea’s age. This would be due to a dissociation between the first and second sentence as a consequence of not updating the meaning of the expressions for the characters’ ages through the expression “In 3 years’ time.” Here, the error would be related to the solver’s inability to identify the quantities and their relationships. On the other hand, the solver could correctly recognize both quantities as different, but wrongly preserve the syntax of the natural language (where both quantities are expressed by the same set of words) when translating the words into equations, thus assigning them the same letter. In this case, the error would be linked to a lack of command of the algebraic language. Hence, when analyzing students’ performances, it is usually not possible to specify the relative weight of the two sources of error, as both are involved and can act interconnectedly.

In this context, our study aims to determine whether the source of the different difficulties when algebraically solving age word problems resides in inferring the network of relationships between quantities of the problem, or in the process of formalization into algebraic language. To accomplish this, we compared secondary-school students’ productions when performing tasks that required them to infer the network of relationships of a word problem and formalize it either in equations or by reasoning from hypothetical but concrete numbers. Our experimental design relied on the fact that the network of relationships presented in the text was the same in both tasks, while the language to be used to formalize that network differed (arithmetic or algebraic).

## 2 Theoretical framework

### 2.1 Word problem solving

In order to model the word problem comprehension-solution process, Nathan et al. (1992) refined the work of Reusser (1988) and considered three components: the text base, the situation model, and the problem model. The text base is a propositional network derived solely from the text. The solvers connect the proposed problem to their background knowledge, giving rise to the situation model which describes the problem situation in everyday terms. Based on the latter, the solvers construct a problem model. The problem model is a formalization of the situation model, which includes the whole network of specific relationships between quantities and its representation in terms of computational and symbolic operations (Walkington et al., 2019).

The three components constrain each other in an iterative process, in which reading comprehension plays a salient role (Kintsch, 1998). Each piece of information is processed as soon as it is identified and then integrated with the rest of the text in working memory (Kintsch, 1998). It encompasses progressive integration cycles, where what happens at the end of the sentence has a singular relevance. Thus, this framework supports the idea that the use of expressions with evolving meaning increases the difficulty in connecting different sentences (see example 1).

### 2.1.1 Algebraic method of solution

Given a word problem, the algebraic method of solution (hereinafter algebraic method) consists, essentially, of formalizing the relationships between quantities in a problem model by means of an equation or a system of equations, whose resolution leads to the solution to the problem. For analytical purposes, the process up until the writing of the equation(s) can be broken down into a sequence of steps (Fillooy et al., 2008), which we illustrate by means of example 1:

Step 1. An analytical reading of the problem, by which a set of quantities and the relationships among them is inferred from the text. In example 1, the most usual analytical reading would result in a network made up of four relationships: Amaya's current age equals nine plus Andrea's current age; Amaya's future age equals two times Andrea's future age; Amaya's future age equals three plus Amaya's current age; and Andrea's future age equals three plus Andrea's current age.

Step 2. A letter is assigned to an unknown quantity—or several different letters to several unknown quantities. In example 1, we may assign  $x$  to Andrea's current age.

Step 3. Algebraic expressions are assigned to the remaining unknown quantities. In example 1, Amaya's current age can be expressed as  $x+9$ , Andrea's future age as  $x+3$ , and Amaya's future age both as  $(x+9)+3$  and  $2\cdot(x+3)$ .

Step 4. An equation (or as many as the letters defined in step 2) is written based on equating two algebraic expressions that designate the same quantity (step 3). In example 1, a possible equation is  $2\cdot(x+3)=(x+9)+3$  equating the expressions for Amaya's future age.

Regarding the Nathan et al. (1992) model, in step 1, the solvers construct the text base, the situation model, and form a network of relationships between quantities which can be inferred from the text base and the situation model. Steps 2 to 4 involve formalizing that network by means of equations in a problem model. In addition, it is common for secondary school students to begin writing the equation(s) without having inferred the whole network of relationships. Specifically, students do not fully complete the construction (step 1) before moving on to assign letters to unknown quantities (step 2); they then simultaneously construct the equation(s) (step 4) and assign expressions to the remaining unknown quantities (step 3). For example, let us suppose that students had addressed Example 1 by following a sequence which would lead them to start the equation  $(x+9)+3=$ . This equation would be completed after adding the expression  $2\cdot(x+3)$ , which had not been considered beforehand. Accordingly, steps 3 and 4 have been collapsed into a single phase, and the solver would go back to the problem text to recover information about those relationships not yet integrated (step 1). Another potential source of errors in steps 2, 3, and 4 appears if the students disregard the algebraic language that requires a one-to-one correspondence between quantities and letters and the meaning remains invariable throughout the solving process. In contrast, natural language allows expressions with evolving meaning (Fillooy et al., 2008; Soneira et al., 2018). Another cognitive challenge is that in algebra, letters can represent variables and unknowns. Unknowns stand for specific numbers, as in the case of word problem solving, and variables for variable numbers.

## 2.1.2 Trial-and-error method

An alternative approach to the algebraic method is the trial-and-error method. Essentially, this consists of inferring the whole network of relationships among the problem's quantities and applying these relationships to a concrete—although hypothetical—value to assess whether a contradiction is obtained. Although the solvers do not use letters to construct a problem model, they do not know whether the hypothetical values are solutions. The underlying logic of the tasks used in this study, apart from their utilization of hypothetical values in the formalization, followed the algebraic method to some extent and could be broken down into analogous sequential steps. We illustrate this below by means of example 1:

Step 1. An analytical reading is conducted as in the algebraic method, from which the same four relationships between quantities are inferred: Amaya's current age equals nine plus Andrea's current age; Amaya's future age equals two times Andrea's future age; Amaya's future age equals three plus Amaya's current age; and Andrea's future age equals three plus Andrea's current age. The relationships between quantities to be formalized in a problem model are the same as in the algebraic method. In addition, in this example—and in all the tasks in this study—all the relationships contain two unknown quantities and one known quantity. Thus, it is not possible to find the solution just by applying a sequence of arithmetic operations that begins only from known quantities and ends up determining the quantities that are asked for in the problem text.

Step 2. The solvers, instead of assigning a letter to an unknown quantity, give a concrete value to it, which works as a hypothetical solution. In example 1, for instance, the solvers could assign the value 11 to Andrea's current age.

Step 3. The rest of the unknown quantities are computed. This step is feasible thanks to the replacement of an unknown quantity by a number in the second step. Thus, Amaya's current age would be 20 ( $11 + 9$ ), Andrea's future age, 14 ( $11 + 3$ ), and Amaya's future age, 23 ( $20 + 3$ ) and 28 ( $2 \cdot 14$ ).

Step 4. The solvers, instead of constructing an algebraic equation, compare the two numbers calculated for the same quantity, using the equals sign with the algebraic meaning of equivalence. In example 1, those correspond to Amaya's future age (23 and 28).

Since the compared values are different, the solvers would repeat the actions from step 2 to step 4, modifying the number assigned to Andrea's current age. Completing steps 2 to 4 also implies representing the network of relationships between quantities in a problem model in arithmetic language inferred in step 1.

A key point is that, unlike in the algebraic method, secondary school students should deviate to a lesser extent from the sequential order of steps when they use a trial-and-error strategy. This is because in this case starting step 4 without completing step 3 is unnatural. Indeed, this means that when performing pure arithmetic operations, the student maintains all the operations indicated without calculating the result (i.e., maintaining  $(11 + 9) + 3$  instead of performing the sequence  $11 + 9 = 20$ ,  $20 + 3 = 23$ ). However, solvers may also identify relationships in the text whenever they are needed to conduct computations instead of inferring the complete network beforehand.

This issue is important because following the sequential order of steps in a specific manner may affect the cognitive demands of the task. In particular, strict compliance

with the sequence of steps within the algebraic method would reduce working memory load. In fact, in step 4, memory can be partially unloaded in the external representations constructed in step 3 (Koedinger & Nathan, 2004). This allows the solver to focus solely on identifying which mathematical relationship (only one per equation) should be used to construct an equation. However, as mentioned above, secondary school students usually begin step 4 before having completed the previous steps, and manage several mathematical relationships, as well as their representation, concurrently. This fact increases working memory load.

In addition, conducting the four steps within the trial-and-error method would imply a lower cognitive demand than within the algebraic method. Our analysis above shows that within the trial-and-error approach, there is a greater likelihood of precisely complying with the strict sequential order of the four steps. Therefore, the solver simply takes into account one relationship at each step and then removes it from working memory.

### 2.1.3 The role of natural language

When solving problems, the language features of word problems have also been reported to impact the level of perceived difficulty (e.g., Clinton & van den Broek, 2012; Walkington et al., 2019). In particular, the use of expressions with evolving meaning—one and the same expression designate different quantities within the same text—raises issues. As a universal rule, within arithmetic and algebraic languages, each symbol must have a unique referent. For instance, in example 1, if the letter  $x$  represents Amaya's age at present, then  $x$  cannot be used to represent Amaya's age in ten years' time; either a different letter or  $x + 10$  must be used. Yet, according to Kintsch (1998) because each sentence is processed in a different integration cycle, the readers may not take into account that "Amaya's age" in the second sentence had already been processed but with a different meaning in the first sentence. As a consequence, they would use the same letter to represent it, which is an error by multiple referents.

On the contrary, let us consider the following text framing the same problem model as in example 1, but in which expressions with evolving meaning are avoided by adding an explicit hint to avoid the error by multiple referents in each sentence (indicated in italics):

Example 2: *At present*, Amaya is 9 years older than Andrea. In 3 years' time, Amaya's *future* age will be twice Andrea's *future* age. How old are they *at present*?

The information in the second sentence is self-sufficient when attaching referents to verbal expressions and the solvers do not need to update the meaning of any expression. In example 2, the addition of explicit information would make it less likely to believe that the relationship "twice" refers to the current ages and to commit the consequent error by multiple referents.

### 3 Research purpose

Within the Nathan et al. (1992) model, solving a problem algebraically requires inferring the network of relationships between quantities expressed in the problem text and subsequently formalizing the network into a problem model in algebraic language. Both aspects could be a source of difficulties and errors (Soneira et al., 2018). In practice, knowing the significance of each source could improve teaching sequence design by focusing on either the processing of natural language to infer the network of relationships or on the formalization into algebraic language.

Thus, in the context of age word problems, in this work, we conducted an experiment designed to answer the following research questions about tasks that require the construction of a problem model based on a problem text:

(RQ1) Are there differences in the accuracy of the problem model depending on whether:

- a. the task implies using the algebraic language or the arithmetic language—with hypothetical values—to formalize the relationships between quantities?
- b. the problem text uses expressions with evolving meaning?

(RQ2) Are there differences in the rate of errors by multiple referents depending on whether:

- a. the task implies relying on the algebraic language or on the arithmetic language—with hypothetical values—to construct a problem model?
- b. the problem text uses expressions with evolving meaning?

Note that RQ2 is a particularization of RQ1 and that, as discussed previously, multiple referents are related to the use of expressions with evolving meaning.

We will answer the above questions by testing two hypotheses—H01 and H02—in the context of word age problems, followed by applying a modus tollens argument and interpreting the effects of the factors (see next section):

(H01) There are no differences in the rate of problem model accuracy depending on any of the task variants described in RQ1.

(H02) There are no differences in the rate of errors by multiple referents depending on any of the task variants described in RQ2.

### 4 Method

We used a between-between, 2-way design with two crossing factors: task and text explicitness. The task factor had 2 levels: the equations task and the concrete value task. The text explicitness factor had 2 levels: with an explicit hint to a typical source of error (hereinafter more explicit text) and without an explicit hint to a typical source of error (hereinafter less explicit text).

## 4.1 Participants

The sample consisted of 255 students, aged 15 to 16 years old, in their fourth year of secondary school (aimed towards higher scientific studies). The students were from nine classes in four different Spanish public high schools placed in middle class neighborhoods of the same city. All the students had been exposed to the same contents, evaluation criteria, and learning standards, stipulated by the Spanish region's educational laws. In these high schools, each class had only one mathematics teacher and was made up of a mix of students of different academic abilities. Following the educational regulations for mathematics, individual and group work were applied in the class setting. Students had been solving problems algebraically for two school years and solving problems by means of just arithmetic calculations for at least nine years. According to their teachers, the average proficiency in mathematics of these nine classes was intermediate. In each high school, each class was randomly assigned to a different level of the task factor.

## 4.2 Instruments and procedure

We used a specific questionnaire composed of six tasks for each level of each factor. Each task contained a text made up of two parts: (1) a declarative section describing a network of relationships among quantities in an age problem setting and (2) a question about the relationships described in the declarative section. For the sake of exhaustiveness, we included six different networks of relationships. The declarative sections were exactly the same in each level of the task factor; thus, both texts only differed in the question section (see Tables 1 and 2).

**Table 1** Problems proposed to less explicit text groups

Task factor	
Equation(s) task	Concrete value task
Melisa's age is three times the age of her daughter Marta. In 12 years' time, Melisa's age will be only twice Marta's age. Find the age of each person	Melisa's age is three times the age of her daughter Marta. In 12 years' time, Melisa's age will be only twice Marta's age. Is it possible for Marta's age to be 6?
Anibal is 15, his sister is 12 and his mother is 40. How many years need to pass so that the children's ages, taken together, equal their mother's age?	Anibal is 15, his sister is 12 and his mother is 40. Is it possible that 23 years have to pass so that the children's ages, taken together, equal their mother's age?
Amaya is 9 years older than Andrea. In 3 years' time, Amaya's age will be twice Andrea's age. How old are they?	Amaya is 9 years older than Andrea. In 3 years' time, Amaya's age will be twice Andrea's age. Is it possible for Andrea's age to be 4?
One sister is 3 years older than the other sister, and their father is 43. In 7 years' time their father's age will be twice the sum of the sisters' ages. How old is each sister?	One sister is 3 years older than the other sister, and their father is 43. In 7 years' time their father's age will be twice the sum of the sisters' ages. Is it possible for the younger sister's age to be 6?
A father is twice the age of his son. Seventeen years ago, the father was three times the age of his son. How old is each person?	A father is twice the age of his son. Seventeen years ago, the father was three times the age of his son. Is it possible for the son's age to be 19?
Eight years ago, Ana's age was four times Maria's age. In 12 years' time Ana's age will only be twice Maria's age. Find the age of each person	Eight years ago, Ana's age was four times Maria's age. In 12 years' time Ana's age will only be twice Maria's age. Is it possible for Maria's age to be 10?



**Table 2** Problems proposed to more explicit text groups

Task factor	
Equation(s) task	Concrete value task
Melisa's <i>current</i> age is three times <i>the current age</i> of her daughter Marta. In 12 years' time, Melisa's <i>future</i> age will be only twice Marta's <i>future</i> age. Find the <i>current</i> age of each person	Melisa's <i>current</i> age is three times <i>the current age</i> of her daughter Marta. In 12 years' time, Melisa's <i>future</i> age will be only twice Marta's <i>future</i> age. Is it possible for Marta's current age to be 6?
<i>At present</i> , Anibal is 15, his sister is 12 and his mother is 40. How many years need to pass so that the children's <i>future</i> ages, taken together, equal their mother's <i>future</i> age?	<i>At present</i> , Anibal is 15, his sister is 12 and his mother is 40. Is it possible that 23 years have to pass so that the children's future ages, taken together, equal their mother's future age?
<i>At present</i> , Amaya is 9 years older than Andrea. In 3 years' time, Amaya's <i>future</i> age will be twice Andrea's <i>future</i> age. How old are they <i>at present</i> ?	<i>At present</i> , Amaya is 9 years older than Andrea. In 3 years' time, Amaya's <i>future</i> age will be twice Andrea's <i>future</i> age. Is it possible for Andrea's current age to be 6?
<i>At present</i> , one sister is 3 years older than the other sister, and their father is 43. In 7 years' time their father's <i>future</i> age will be twice the sum of the sisters' <i>future</i> ages. How old is each sister <i>at present</i> ?	<i>At present</i> , one sister is 3 years older than the other sister, and their father is 43. In 7 years' time their father's <i>future</i> age will be twice the sum of the sisters' <i>future</i> ages. Is it possible for the younger sister's current age to be 6?
<i>At present</i> , a father is twice the age of his son. Seventeen years ago, the father's <i>past</i> age was three times the <i>past</i> age of his son. How old is each one <i>at present</i> ?	<i>At present</i> , a father is twice the age of his son. Seventeen years ago, the father's <i>past</i> age was three times the <i>past</i> age of his son. Is it possible for the son's current age to be 19?
Eight years ago, Ana's <i>past</i> age was four times Maria's <i>past</i> age. In 12 years' time Ana's <i>future</i> age will be only twice Maria's <i>future</i> age. Find the <i>current</i> age of each person	Eight years ago, Ana's <i>past</i> age was four times Maria's <i>past</i> age. In 12 years' time Ana's <i>future</i> age will only be twice Maria's <i>future</i> age. Is it possible for Maria's current age to be 10?

In the equation(s) task, the question asked for the value of an unknown quantity. The instructions were to write an equation or a system of equations whose solution provides the answer to the question, but the students were not asked to solve the equation(s). In the concrete value task, the question asked if a given concrete—although hypothetical—value fulfilled the conditions of the declarative section. The instructions were to answer the question without using algebra, record all the arithmetic operations (even those mentally conducted) required to be sure of the correctness of the answer, and to also write a textual answer to the question. This implied constructing a problem model in arithmetic language without knowing whether the given value is a solution until the end of the process. This would correspond to steps 1 to 4 when using the trial-and-error strategy. Moreover, it was not possible to provide the correct answer without having represented a correct problem model.

In both levels, the tasks involved constructing a problem model based on the network of relationships between quantities expressed in the declarative section of the text, which was exactly the same in both levels. Thus, if the students were not able to infer these relationships, there would be little difference between both levels of the task factor. In addition, the task factor assesses the weight of the difficulties in constructing a problem model when the task implies representing the unknown quantities with letters (algebraic language) instead of hypothetical values (arithmetic language).

Concerning the text explicitness factor, in the less explicit text level, the texts were problems taken from textbooks corresponding to two previous school years (13–14 years old) and contained expressions with evolving meanings (Table 1). In the more explicit text level, each text expressed exactly the same network of relationships between quantities

as a corresponding text in the less explicit text level, though these still contained all the necessary words to make explicit that the characters' ages were updated in each sentence. Hence, the problem model could be constructed without connecting the meaning of the same expressions in different sentences (Table 2).

Each subgroup dealt with just one text explicitness level and task level. In each secondary school, the students' classes were randomly assigned to one of the experimental groups beforehand and each group completed the corresponding task in a different classroom. In the end, there were 63 students in the more explicit text and equation(s) task subgroup, 54 in the less explicit text and equation(s) task subgroup, 52 in the less explicit text and concrete value task subgroup, and 86 in the more explicit text and concrete value task subgroup.

The questionnaires were administered by one of the authors, who explained the corresponding task and illustrated it with an example. Once all the participants asserted that they understood the task, each task was shown on a screen at the front of the classroom for 3 min, before being replaced by the next task.

### 4.3 Codification and statistical analyses

We defined two variables for each problem and student:

- $RelRatio_{problem}$  (relationship ratio): this is defined as the number of correctly represented relationships divided by the total number of relationships in the problem.
- $MRP_{problem}$  (multiple referent presence): this receives a value of 1 if there is at least one multiple referents, and 0 otherwise.

We defined the variable  $RelRatio$  as the arithmetic mean of  $RelRatio_{problem}$  of the six problems for each participant. Similarly, we defined  $MRP$  from  $MRP_{problem}$ . The  $RelRatio$  variable assesses the accuracy of the problem model (RQ1), and  $MRP$  measures the incidence of multiple referents (RQ2). In particular, H01 and H02 can be reformulated by saying that, apart from variability due to randomness, there is no effect on  $RelRatio$  or on  $MRP$ , respectively, for either the task or text explicitness factors. A task was considered unanswered if (i) it was blank, (ii) it showed just some letters or numbers without showing relationships, or (iii) it included only text fragments. Regarding  $RelRatio_{problem}$ , the unanswered problems were codified as 0. Regarding  $MRP_{problem}$ , we computed the arithmetic means taking into account only the answered problems, because an unanswered task was neither correct nor was there any empirical evidence of an error by multiple referents.

To code the students' production, we employed tree diagrams representing the generated equations or the arithmetic calculation (Figs. 1 and 2, respectively). We started from the

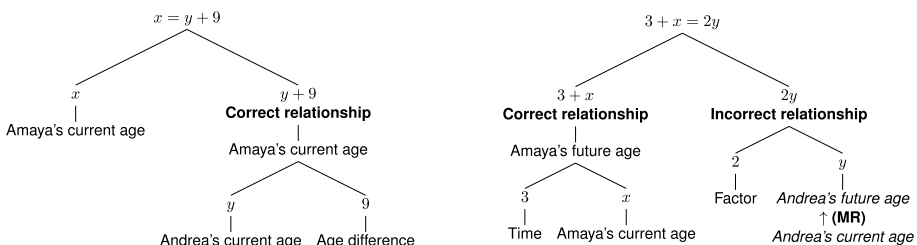


Fig. 1 Tree-diagram corresponding to the answer in Fig. 3

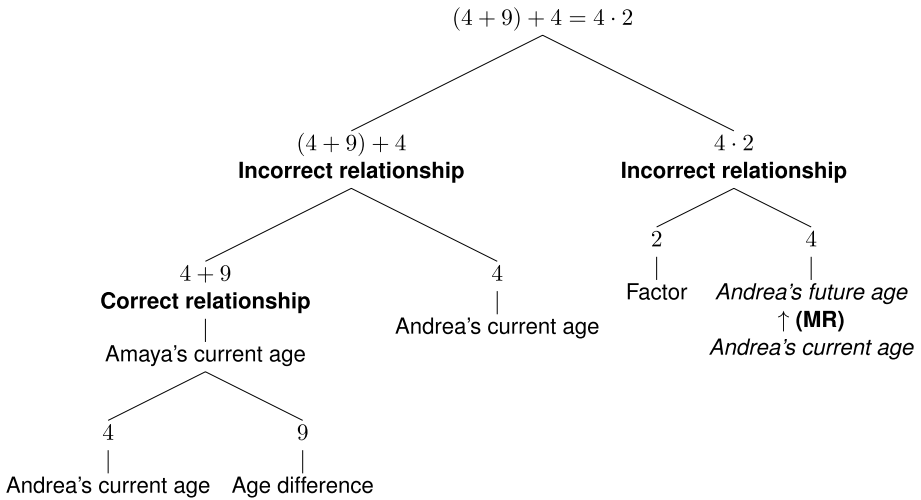


Fig. 2 Tree-diagram corresponding to the answer in Fig. 4

lower level, placing there either known quantities and letters assigned to a quantity (equation task) or the concrete value to be assessed (concrete value task). Then, we ascended in order to make sense of the output.

Regarding the equation(s) task (Fig. 1), in the link nodes, we found algebraic expressions representing relationships between quantities coming from lower-level nodes. The highest nodes—those where equations were expressed—involved assigning two different algebraic expressions to the same quantity. Therefore, these nodes did not correspond to any relationships between quantities. Starting from a set of leaf nodes, we checked the validity of algebraic expressions which appeared in each link node derived from that set. If the expression represented in a link node was correct, the process continued checking the higher-level link nodes. Otherwise, we stopped the process in that subtree and considered that the expressions in the higher-order nodes derived from the incorrect one were also incorrect.

Regarding the answers to the concrete value task (Fig. 2), the procedure and analysis were analogous, though in this case the order in which the computations were conducted facilitated structuring the mathematical operations hierarchically. This procedure resulted in complete agreement between three coders. Below, we illustrate the codification procedure with examples from experimental subjects corresponding to the less explicit text level of the tasks (Figs. 3 and 4). Each example includes the tree diagram we used (Figs. 1 and 2, respectively).

In Fig. 3—codified by Fig. 1—the first equation ( $x = y + 9$ ) correctly represents the additive relationship between the current ages of the protagonists (highest node on the

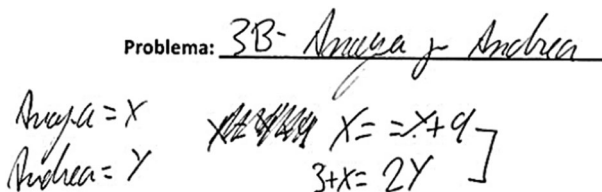


Fig. 3 Student's answer to the less explicit text and equation(s) task in Table 1

$$\begin{aligned} \text{Edad Amaya} &\rightarrow 4 + 9 = 13 \\ \text{Edad Andrea} &\rightarrow \cancel{4} \\ 13 + 4 &= 4 \cdot 2 \\ &\text{NO} \end{aligned}$$

Note. Edad Amaya means Amaya's age; Edad Andrea means Andrea's age

**Fig. 4** Student's answer to the less explicit text and concrete value task in Table 1

left-hand side tree of Fig. 1). The second equation correctly identifies the additive relationship between Amaya's present and future ages ( $3+x$ ) (highest node on the right-hand side tree of Fig. 1). However, we interpreted that the student used the letter  $y$  to represent both Andrea's current and future ages (lowest nodes of both trees in Fig. 1). Therefore, we assigned a value of 1 to the  $MRP_{\text{problem}}$  variable. As a consequence of the previous student's action, the relationship between Andrea's current and future ages was not represented, which could be done by means of  $y+3$ . Moreover, according to the syntactic rules of algebraic language, the multiplicative expression in Fig. 1 relates Amaya's future age and Andrea's current age (second level node on the right-hand side of Fig. 1). Thus, as just two of the four relationships have been correctly expressed,  $RelRatio_{\text{problem}}$  was 0.5.

Similarly, Fig. 4—codified by Fig. 2—shows an example of the resolution for the corresponding task in the concrete value level of the task factor. The solver multiplied Andrea's current age by 2, but not her future age (left side branch of the tree of Fig. 2). The absence of references to Andrea's future age led us to affirm that the solver considered that the value 4 (lowest nodes in Fig. 2) referred to both Andrea's current and future ages ( $MRP_{\text{problem}}=1$ ). On the other hand, the student only expressed one correct relationship ( $4+9=13$ ). Consequently,  $RelRatio_{\text{problem}}$  was 0.25.

Data were analyzed with the R software (Version 4.1.1). We based the statistical analysis on the recommendations and R-functions given by Wilcox (2017a2017b) to select the most suitable version of analysis of variance (ANOVA) type procedures and effect sizes depending on the specific sample distribution features. As it is not possible to know beforehand the sample distribution features, we remit the reader to the “Results” section for the details. We took a  $p$ -value of 0.05 as a decision criterion in every case.

## 5 Results

All the students quickly stated they understood the corresponding task and, generally, they seemed focused on the task during the experiment. On the basis of the students' responses, in-class behavior, and the characteristics of the sample classes (see the “Participants” section), there was no reason to suspect that there were class differences affecting the experiment.

### 5.1 RelRatio variable

Concerning the descriptive statistics, we obtained  $M=0.27$ ,  $SD=0.24$  for the group equation(s) and less explicit,  $M=0.33$ ,  $SD=0.24$  for equation(s) and more explicit,  $M=0.64$ ,  $SD=0.30$  for concrete value and less explicit, and  $M=0.73$ ,  $SD=0.27$  for concrete value and more explicit (Fig. 5).

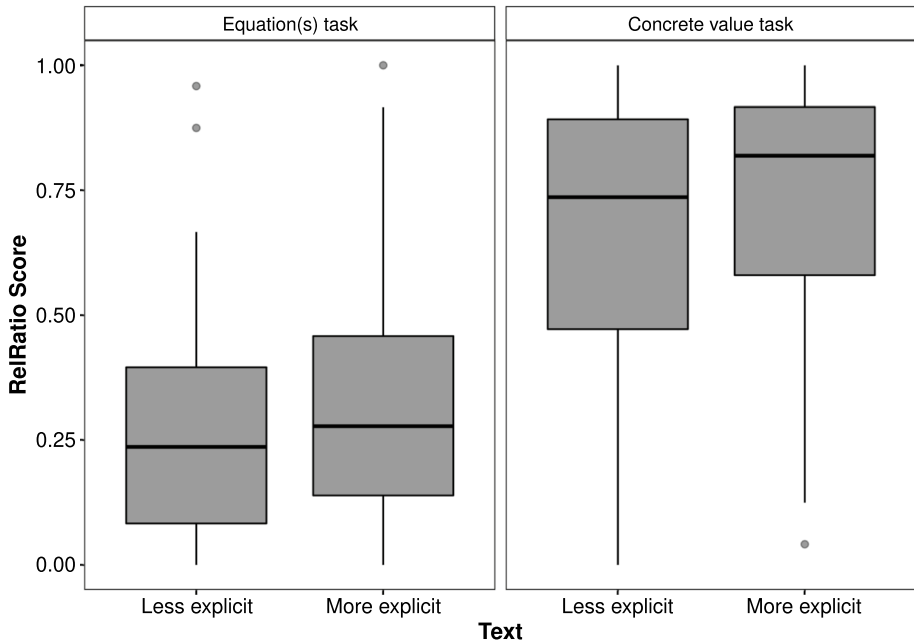


Fig. 5 RelRatio scores grouped by task and text explicitness

Table 3 Results of the robust two-way analysis of variance

Factor	Variable					
	RelRatio			MRP		
	$\Psi^a$	95% CI	$\xi^b$	$\Psi^a$	95% CI	$\xi^b$
Task factor	-0.96***	[-1.09, -0.80]	0.78	-0.92***	[0.72, 1.10]	0.65
Text explicitness factor	-0.17*	[-0.30, -0.007]	0.02	0.40***	[0.18, 0.60]	0.19
Interaction	-0.002	[-0.10, 0.18]	-	0.29*	[0.08, 0.47]	0.36

$\Psi^a$ =estimator of the difference between factor level means calculated by the R-function *bbmcpb* because it controls type 1 errors—with the Rom’s method—and provides confidence intervals; *CI*, confidence interval;  $\xi^b$ =measure of the effect size calculated by the R-function *ESmainMCP* for the main effects and *esImcp* for the interaction effects. The R- function *bbmcpb* performs multiple comparisons for a between-between ANOVA design with a percentile bootstrap method, while *ESmainMCP* and *esImcp* compute the corresponding effect size for main effects and interactions, respectively, so the names of the R-functions allude to what they compute (see Wilcox, 2017a, for details)

\* $p < 0.05$ . \*\*\* $p < 0.001$

Regarding testing H01, we could assume homogeneity of variances, but the qq-norm plots indicated severe deviations from normality for the concrete value and less explicit text subgroup and the concrete value and more explicit text subgroup. Moreover, we had unequal sample sizes and distinct skewed distributions for the equation(s) ( $s=0.78$ ) and concrete value ( $s=-0.91$ ) groups, in addition to outliers in the concrete value groups. For this reason, we used a robust version of a 2-way ANOVA based on trimmed means and bootstrapping—this offers advantage over no trimming and the median (Wilcox, 2017b), and a percentile bootstrap method appears to be effective (Wilcox, 2017a). This revealed that both the task and the text explicitness factors had a statistical effect, while the factors did not interact (Table 3), leading us to reject H01.

## 5.2 MRP variable

The descriptive statistics came out as  $M=0.69$ ,  $SD=0.33$  for the group equation(s) and less explicit,  $M=0.45$ ,  $SD=0.34$  for equation(s) and more explicit,  $M=0.22$ ,  $SD=0.25$  for concrete value and less explicit, and  $M=0.13$ ,  $SD=0.15$  for concrete value and more explicit (Fig. 6).

Concerning testing H02, all four subgroups were heavily tailed, and the qq-norm plots indicated severe deviations from normality for all the subgroups. Moreover, we were dealing with unequal sample sizes, differing skewness between the equation(s) ( $s=-0.08$ ) and concrete value ( $s=1.15$ ) subgroups, and outliers in the more explicit text subgroups. Hence, following Wilcox (2017a2017b), we used a robust version of a

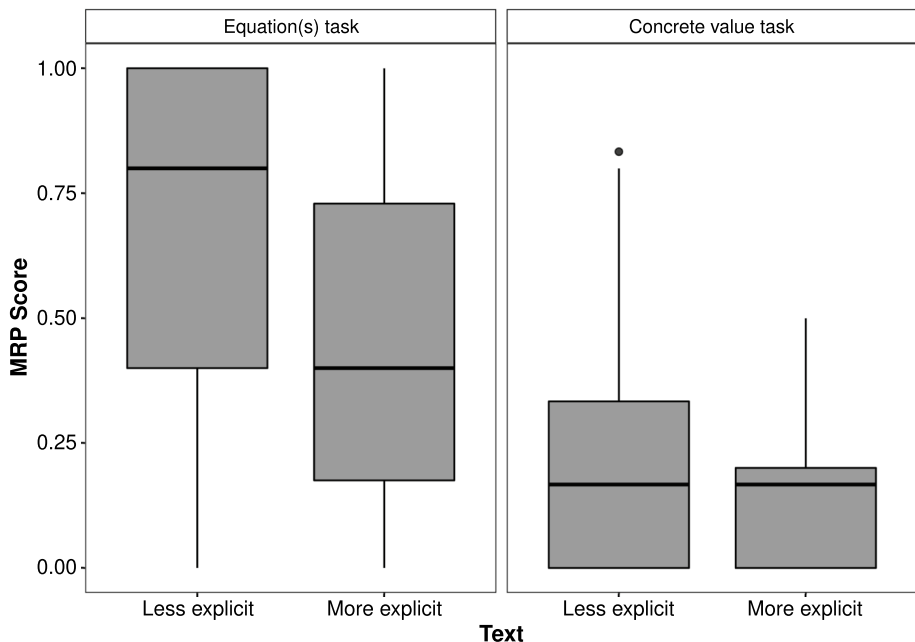


Fig. 6 MRP scores grouped by task and text explicitness

2-way ANOVA based on trimmed means and bootstrapping. We found that both factors had an effect and that they interacted (Table 3). Thus, we rejected  $H_{02}$ .

Regarding the interpretation of the interaction, the boxplot suggests that the text explicitness factor had an effect within the equation(s) task, but not within the concrete value task (Fig. 3). This was confirmed by using trimmed means with bootstrapping, whose values were ( $Y_t = 3.89$ ,  $p < 0.001$ ,  $CI = [0.162, 0.508]$ ,  $r = 0.47$ )— $Y_t$  denoting the statistical estimator and  $r$  the effect size—and ( $Y_t = 1.14$ ,  $p = 0.26$ ,  $CI = [-0.041, 0.15]$ ), respectively.

## 6 Discussion and conclusions

In this section, we will answer our research questions, provide some plausible explanations for the results, and discuss their implications. We start with the first research question (RQ1). Given a task in which secondary school students must construct a problem model corresponding to a problem text, RQ1 asks whether there are differences in the accuracy of the problem model depending on two task factors (RQ1a, RQ1b). We base the answer on the results for the *RelRatio* variable as it measures the degree of accuracy of the problem model.

We did not find evidence of an interaction between factors, so the effect of the text explicitness factor was analogous in both task conditions. Furthermore, although we obtained a statistical effect, the effect size of the text explicitness factor was so small ( $\xi = 0.02$ ) that it would lack practical importance. Thus, the presence of expressions whose meaning evolves as one moves through the sentences, increasing the difficulty with connecting information provided in different sentences, has very limited impact on the accuracy of the problem model constructed by solvers (RQ1b). In addition, the averages for *RelRatio* in the concrete value task groups were acceptable ( $M = 0.64$  for less explicit text;  $M = 0.73$ , for more explicit text), while in the subgroups which performed the equation(s) task, the averages were low ( $M = 0.27$  for less explicit text and  $M = 0.33$  for more explicit text). Indeed, the main effect of the task factor was very large ( $\xi = 0.78$ ), so we rejected  $H_{01}$ .

Summarizing our answer to RQ1a, we conclude that there are differences in problem model accuracy depending on whether a problem model is constructed in arithmetic or algebraic language. It is also important to note that the declarative section of the text expressing the relationships between quantities to be formalized in the problem models is the same in both task levels. Hence, we interpret the large effect of the task factor as the weight of the difficulties in problem model accuracy derived from the student being asked to use letters instead of hypothetical concrete values when formalizing the relationships between quantities. This would provide statistical support for some descriptive results concerning the differences in the student's capacity to solve word problems depending on the approach—algebraic or arithmetic—they use, as proposed by Stacey and MacGregor (1999) and Filloy et al. (2008).

Concerning the plausible explanations for the preceding results about the accuracy of the problem model, the use of algebraic language would increase the task's cognitive demand. This is because working memory must store, in addition to the relationships between quantities, the algebraic symbols which represent them, and the structural rules of natural and algebraic languages involved in the translation process. However, once the solver gains command of algebraic language and the translation process, all these elements can be managed as a single operation in working memory, so as not to overload it (Chen

et al., 2017; Sweller et al., 2019). In addition, when trying to find equations to represent word problems, secondary school students seldom follow the steps sequentially. They usually start writing an equation (step 4) without having inferred all the algebraic relationships involved in the latter (Filloy et al., 2008; Soneira et al., 2018). As previously stated, when this happens, the construction of the equation implies a higher loading of working memory. This could lead to an overload which causes the omission of necessary relationships between the quantities of the problem, resulting in an incorrect problem model.

On the contrary, when using concrete—although merely hypothetical—values, it is less likely that the solver will jump to step 4 before having completed the previous ones. This is because relationships between concrete values are not likely to be left without computing the result of the operations. Moreover, even if the solver skips some of the steps, the relationships are successively unloaded from working memory right after each computation is conducted. In this case, as Sweller (1988) asserts, when the sequential order of steps is not followed, an increase in working-memory load would be observed with the equation(s) task rather than the concrete value task.

The second research question (RQ2) asks about the effect of students' inability to connect information provided in different sentences by means of expressions with evolving meaning (RQ2b), and the command of the algebraic language (RQ2a), in the rate of multiple referents. To answer RQ2, we use the *MRP* variable. Overall, the analysis showed a significant interaction between factors, so we rejected  $H_{02}$ . Text explicitness had a particularly moderate to large effect within the equation(s) task ( $r=0.47$ ), but no effect within the concrete value task. Hence, the use of expressions with evolving meaning itself does not increase the difficulty of connecting information provided in different sentences. In fact, when students are asked to base reasoning in hypothetical values, the presence of those expressions does not increase the rate of multiple referents. Our outcomes aid in clarifying the results from Bloedy-Vinner (1996) and Soneira et al. (2018) concerning the source of the error by multiple referents when the problem text includes expressions with evolving meaning.

Furthermore, according to the criteria for the explanatory measures of effect sizes (Wilcox, 2017a, 2017b), the task factor had a large effect ( $\xi=0.65$ ), while the text explicitness factor had a moderate to large effect and only when the task implied using the algebraic language. Although the error by multiple referents is directly linked to the use of expressions with evolving meaning, the presence of these expressions only had effect when algebraic language was used. We know that the goal of the reader affects the reading process (Kintsch, 1998) and that the goal of posing equations is specific to the equation(s) task. Thus, it seems that the difficulties in using the algebraic language interfere with the regular verbal processing of expressions with evolving meaning, which otherwise would be correctly processed. This fact, together with the large effect of the task factor, reveals greater difficulties related to finding the equations than those related to understanding the problem text. In addition, as RQ2 can be seen as a particular case of RQ1, the plausible explanations for RQ1 would also apply to RQ2.

Letters—in algebraic language—can represent any varying quantity (e.g., Amaya's age at any time), but when solving word problems once a letter is attached to a specific unknown quantity, the meaning remains invariable; otherwise, multiple referents will arise. This is another fundamental issue of the algebraic method which could contribute to explaining our results in the equation(s) tasks. Also, the low scores in the equation(s) tasks correspond with prior research about the algebraic method (e.g., Filloy et al., 2008; Koedinger & Nathan, 2004; Stacey & MacGregor, 1999). In summary, our results suggest



that students are fairly able to infer the relationships between quantities, but face difficulties coordinating these inferences when required to use letters to formalize relationships between unknown quantities.

Our results solidify certain implications in regard to the teaching process. The idea is to reduce the cognitive demands in the early stages of the learning process until the student achieves expertise in the algebraic method. Based on our results, word problems containing expressions with evolving meaning, combined with algebraic problem solving, increase problem model inaccuracy, and the rate of multiple referents. Nevertheless, in school algebra, solving word problems through algebra is a learning objective and non-artificial language in the problems' texts includes expressions with evolving meaning. Despite the greater challenge posed to students, these expressions should not ultimately be removed from the instructional design.

However, through instructional design, the cognitive demand can be regulated during the initial phases of the teaching process. Specifically, instruction can focus on students' application of the algebraic method to word problems relying on scaffolding. In particular, students must finish assigning algebraic expressions to all the unknown quantities in the problem before beginning to construct the equation. This could be imposed by means of a simple computer program. By doing so, working memory load is reduced, because only one relationship between quantities needs to be stored at each step. For example, in steps 2 and 3, the algebraic representations serve for unloading the working memory and step 4 is reduced to simply determining which ones represent the same quantity.

In addition, our study suggests that, in the context of the algebraic method, trial-and-error method could be used as an auxiliary tool to trigger the inference of the network of relationships between quantities in the case of difficulties constructing a problem model in algebraic language. This practice involves the assertions from Filloy et al. (2001). Trial-and-error method could also be utilized in the final years of primary school and the early years of secondary school as a means of introducing the logic of the algebraic method through familiar mathematical language. In particular, we could consider a teaching sequence that begins with an introduction to the logic of the algebraic method, followed by the scaffolding process mentioned in the previous paragraph, and finally the free use of the algebraic method. Teachers could also use concrete value tasks to identify students who have difficulties inferring relationships between quantities in problem texts.

Regarding the limitations of this study and possible future research, two remarks should be made. First, we found that when tackling the equation(s) task, students produced less accurate problem models and made more errors of multiple referents when the problem text contained expressions with evolving meaning. However, we do not know each individual student's thought process, nor do we know what specific way of posing the equations triggers these phenomena. This could be addressed through qualitative comparisons between the students' behavior when performing the equation(s) and the concrete value tasks. Second, we only considered age problems and a relevant but specific type of error linked to the features of text wording in our study. Future studies should delve into whether our results are restricted to this case, or if they are transferable to problem texts with other characteristics. Further studies might also attempt to overcome the limitations concerning the assignment of subjects to the experimental conditions. In our case, as usually happens in educational research, whole class groups were assigned to the different experimental conditions, though future designs should advocate for random assignment of individuals to avoid possible biases.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was supported by the Spanish Government through the project PGC2018-096463-B-I00, and by the Castilla-La Mancha Regional Administration under grant SBPLY/19/180501/000278.

**Data availability** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Code availability** Not applicable.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bloody-Vinner, H. (1996). The analgebraic mode of thinking and other errors in word problem solving. In A. Gutierrez & L. Puig (Eds.), *Proceedings of the Conference of the International Group for the Psychology of Mathematics Education (PME 20) (20th, Valencia, Spain, July 8-12, 1996)* (Vol. 2, pp. 105–112). International Group for the Psychology of Mathematics Education. <https://www.igpme.org/publications/current-proceedings/>
- Boero, P., Douek, N., & Ferrari, P. L. (2008). Developing mastery of natural language: Approach to theoretical aspects of mathematics. In L. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (pp. 262–295). Routledge. <https://doi.org/10.4324/9780203930236.ch12>.
- Boonen, A. J. H., Van Der Schoot, M., Van Wesel, F., De Vries, M. H., & Jolles, J. (2013). What underlies successful word problem solving? A path analysis in sixth grade students. *Contemporary Educational Psychology, 38*(8), 271–279. <https://doi.org/10.1016/j.cedpsych.2013.05.001>.
- Boonen, A. J. H., de Koning, B. B., Jolles, J., & Van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Frontiers in Psychology, 7*, 191. <https://doi.org/10.3389/fpsyg.2016.00191>.
- Chen, O., Kalyuga, S., & Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review, 29*(2), 393–405. <https://doi.org/10.1007/s10648-016-9359-1>.
- Clinton, V., & van den Broek, P. (2012). Interest, inferences, and learning from texts. *Learning and Individual Differences, 22*(6), 650–663. <https://doi.org/10.1016/j.lindif.2012.07.004>.
- Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology, 6*, 348.
- Drouhard, J.-P., & Teppo, A. R. (2004). Symbols and language. In K. Stacey, H. Chick, & M. Kendal (Eds.), *The Future of the Teaching and Learning of Algebra. The 12th ICMI Study. New ICMI Study Series* (Vol. 8, pp. 227–264). Springer. [https://doi.org/10.1007/1-4020-8131-6\\_9](https://doi.org/10.1007/1-4020-8131-6_9).
- Filloy, E., Rojano, T., & Rubio, G. (2001). Propositions concerning the resolution of arithmetical-algebraic problems. In R. Sutherland, T. Rojano, A. Bell, & R. Lins (Eds.), *Perspectives on school algebra* (pp. 155–175). Kluwer Academic Publishers. [https://doi.org/10.1007/0-306-47223-6\\_9](https://doi.org/10.1007/0-306-47223-6_9).
- Filloy, E., Rojano, T., & Puig, L. (2008). *Educational algebra*. Springer. <https://doi.org/10.1007/978-0-387-71254-3>.
- Filloy, E., Rojano, T., & Solares, A. (2010). Problems dealing with unknown quantities and two different levels of problems dealing with unknown quantities and two different levels of representing unknowns. *Journal for Research in Mathematics Education, 41*(1), 52–80. <https://doi.org/10.5951/jresmetheduc.41.1.0052>.

- Fossard, M., Garnham, A., & Cowles, H. W. (2012). Between anaphora and deixis... The resolution of the demonstrative noun phrase "that N." *Language and Cognitive Processes*, 27(9), 1385–1404. <https://doi.org/10.1080/01690965.2011.606668>.
- Fuchs, L. S., Fuchs, D., Compton, D. L., Hamlett, C. L., & Wang, A. Y. (2015). Is word-problem solving a form of text comprehension? *Scientific Studies of Reading*, 19(3), 204–223. <https://doi.org/10.1080/10888438.2015.1005745>.
- Kieran, C. (2007). Learning and teaching algebra at the middle school through college levels: Building meaning for symbols and their manipulation. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 707–762). National Council of Teachers of Mathematics. [https://doi.org/10.1007/978-94-007-4978-8\\_6](https://doi.org/10.1007/978-94-007-4978-8_6).
- Kintsch, W. (1998). *Comprehension*. Cambridge University Press.
- Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences*, 13(2), 129–164. [https://doi.org/10.1207/s15327809jls1302\\_1](https://doi.org/10.1207/s15327809jls1302_1).
- Molina, M., Rodríguez-Domingo, S., Cañadas, M. C., & Castro, E. (2017). Secondary school students' errors in the translation of algebraic statements. *International Journal of Science and Mathematics Education*, 15(6), 1137–1156. <https://doi.org/10.1007/s10763-016-9739-5>.
- Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, 9(4), 329–389. [https://doi.org/10.1207/s1532690xci0904\\_2](https://doi.org/10.1207/s1532690xci0904_2).
- Soneira, C., González-Calero, J. A., & Arnau, D. (2018). Indexical expressions in word problems and their influence on multiple referents of the unknown. *International Journal of Science and Mathematics Education*, 16(6), 1147–1167. <https://doi.org/10.1007/s10763-017-9824-4>.
- Stacey, K., & MacGregor, M. (1997). Multiple referents and shifting meanings of unknowns in students use of algebra. In E. Pehkonen (Ed.), *Proceedings of the Conference of the International Group for the Psychology of Mathematics Education (PME 21) (21<sup>st</sup>, Lahti, Finland, July 14-19, 1997)* (Vol. 3, pp. 190–197). International Group for the Psychology of Mathematics Education. <https://www.igpme.org/publications/current-proceedings/>
- Stacey, K., & MacGregor, M. (1999). Learning the algebraic method of solving problems. *Journal of Mathematical Behavior*, 18(2), 149–167. [https://doi.org/10.1016/S0732-3123\(99\)00026-7](https://doi.org/10.1016/S0732-3123(99)00026-7).
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7).
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>.
- Walkington, C., Clinton, V., & Sparks, A. (2019). The effect of language modification of mathematics story problems on problem-solving in online homework. *Instructional Science*, 47(5), 499–529. <https://doi.org/10.1007/s11251-019-09481-6>.
- Wilcox, R. (2017a). *Introduction to robust estimation and hypothesis testing* (4<sup>th</sup> edn.). Academic Press. <https://doi.org/10.1016/b978-0-12-804733-0.00001-9>.
- Wilcox, R. (2017b). *Modern statistics for the social and behavioral sciences. A practical introduction* (2<sup>nd</sup> ed.). Taylor & Francis. <https://doi.org/10.1201/9781315154480>.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.