The role of working memory updating, inhibition, fluid intelligence, and reading comprehension in explaining differences between consistent and inconsistent arithmetic word-problem-solving performance

Maria Chiara Passolunghi [a,*], Gonzalo Duque De Blas [b], Barbara Carretti [c], Isabel Gomez-Veiga [b], Eleonora Doz [a], Juan Antonio Garcia-Madruga [b]

[a] Department of Life Sciences, University of Trieste, 34128 Trieste, Italy
[b] National Distance Education University (UNED), 28015 Madrid, Spain
[c] Department of General Psychology, University of Padova, 35131 Padova, Italy

ARTICLE INFO

Article history:
Received 3 June 2021
Revised 26 May 2022
Available online 25 July 2022

Keywords:
Inhibition
Updating
Word problem solving
Fluid intelligence
Working memory
Reading comprehension
Inconsistent problems

ABSTRACT

Children's performance in arithmetic word problems (AWPs) predicts their academic success and their future employment and earnings in adulthood. Understanding the nature and difficulties of interpreting and solving AWPs is important for theoretical, educational, and social reasons. We investigated the relation between primary school children's performance in different types of AWPs and their basic cognitive abilities (reading comprehension, fluid intelligence, inhibition, and updating processes). The study involved 182 fourth- and fifth-graders. Participants were administered an AWP-solving task and other tasks assessing fluid intelligence, reading comprehension, inhibition, and updating. The AWP-solving task included comparison problems incorporating either the adverb *more than* or the adverb *less than*, which demand consistent or inconsistent operations of addition or subtraction. The results showed that consistent problems were easier than inconsistent problems. Efficiency in solving inconsistent problems is related to inhibition and updating. Moreover, our results seem to indicate that the consistency effect is related to updating processes' efficiency. Path analyses showed that reading comprehension was the most important predictor of AWP-solving accuracy.

* Corresponding author.
E-mail address: passolu@units.it (M.C. Passolunghi).

Moreover, both executive functions—updating and inhibition—had a distinct and significant effect on AWP accuracy. Fluid intelligence had both direct and indirect effects, mediated by reading comprehension, on the overall measure of AWP performance. These domain-general factors are important factors in explaining children's performance in solving consistent and inconsistent AWPs.

Introduction

Mathematical knowledge is a core element of contemporary human cultures and societies, and—along with reading and writing—it is one of the basic components of a literate mind. Therefore, it is unsurprising that arithmetic word problems (AWPs) are highly important at school. Children's performance on AWPs predicts their academic success (Duque de Blas et al., 2021) and their future employment and earnings in adulthood (Gross et al., 2009; Murnane et al., 2001). Therefore, shedding light on the nature and difficulties of understanding and solving AWPs is important for both educational and social reasons.

Investigating how children cope with AWPs is also theoretically relevant. To solve an AWP, children must combine their linguistic and mathematical knowledge in a process characterized by a close relationship between understanding a problem and finding a solution. Solving AWPs involves applying some basic arithmetical operations (addition, subtraction, multiplication, and division) as well as—more important—deciding which kind of operation to apply at a given moment. This demands a correct *problem representation,* which relies on a thorough understanding of the information contained in a complex text including some numerical quantities (De Corte et al., 1985; Lee et al., 2009).

Many studies have investigated the factors underlying difficulties in solving AWPs (Daroczy et al., 2015; Fuchs et al., 2018; Lin, 2021; Pongsakdi et al., 2020) and the use of keywords to identify operations and solve different types of word problems (e.g., Jitendra, 2002; Powell, 2011). It has been clearly demonstrated that several types of mistakes made in solving AWPs stem from comprehension issues, for example, when the problem demands the inhibition of the keyword heuristic for solving (Lin, 2021; Mayer & Hegarty, 1996; Shum & Chan, 2020). This is particularly true for so-called *compare problems,* which are typically more challenging than other simple AWPs (Boonen & Jolles, 2015; Mayer, 1982; Morales et al., 1985; Riley & Greeno, 1988; Riley et al., 1983; Schumacher & Fuchs, 2012).

In the current study, we investigated the relation between primary school children's performance in compare AWPs and basic cognitive variables (reading comprehension, fluid intelligence, inhibition, and updating processes). Below is an example of a compare AWP considered in this study:

> A pair of Adidas sneakers costs €30.00 at Walmart. At Decathlon, the same pair of sneakers costs €6.00 more. At Decathlon, the pair of Adidas sneakers costs €5.00 less than a pair of Nike sneakers. How much would you have to pay for a pair of Nike sneakers at Decathlon?

As we can see, this is a multistep arithmetic word problem because it takes two operations to reach the solution (whereas a one-step arithmetic word problem takes only one operation to reach the solution). The multistep arithmetic word problem of our example takes two additions to find the solution: $30 + 6 + 5 = 41$. The first addition is a *consistent* operation because the adverb *more* fits the operation, whereas the second is *inconsistent* because the adverb *less* does not fit the operation. Therefore, this seemingly easy consistent-then-inconsistent (C-I) problem is deceptive. It demands a thorough understanding of a complex instruction in the text and a complete representation of the problem. According to Kintsch and Greeno (1985), this representation has two levels: a first propositional level accounting for explicit information in the text and a situational model involving a complete integration of this information with the reader's previous knowledge. It is only after constructing an integrated situational model that readers realize the Nike sneakers at Decathlon are more expensive than the Adidas sneakers there.

Many authors have also analyzed comparison problems that included the term *more than* or *less than* (as in our example) and involved applying consistent or inconsistent additions and/or subtractions (Jarosz & Jaeger, 2019; Jiang et al., 2020; Mayer & Hegarty, 1996; Pape, 2003). On the one hand, students can solve problems demanding consistent operations based merely on a *direct translation strategy*, relying on a direct translation of the information in the text. They simply search for linguistic markers and keywords and associate *less* with subtraction and *more* with addition (Hegarty et al., 1995). On the other hand, solving AWPs that demand inconsistent operations requires a *problem model strategy*. Students build an integrated situational mental model of the problem and plan their solution on the basis of this model (Thevenot, 2010; Thevenot & Barrouillet, 2015).

These theoretical approaches have been corroborated by various studies showing that although the arithmetic operations and linguistic structure of compare problems are extremely easy, not only young children but also older students and adults make errors in inconsistent compare problems (see Lubin et al., 2013, 2016). More important, the most common mistakes made in solving inconsistent AWPs are reverse errors—that is, applying the wrong operation suggested by the adverb in the text (see Boote & Boote, 2018; Lewis & Mayer, 1987; Mayer & Hegarty, 1996; Shum & Chan, 2020; Stern, 1993; Verschaffel, 1994; Verschaffel et al., 1992).

Working memory's executive functions and fluid intelligence

Working memory

Multiple processes are involved in solving AWPs. Problem solvers must read the problem, identify and retain relevant information, construct a mental model, plan how to arrive at the solution, and carry out calculations based on this plan. Working memory (WM) resources are involved in all these phases (Friso-van den Bos et al., 2013; Fung & Swanson, 2017; Kintsch, 1998). A large body of research has investigated the role of WM in problem solving by referring to the well-known tripartite WM model introduced by Baddeley and Hitch's (1974) seminal work. Several studies have shown that one of the components of this model, the central executive system, is needed to solve AWPs (Andersson, 2007; Fuchs et al., 2010; Lee et al., 2004; Swanson, 2006; Swanson & Sachse-Lee, 2001). Swanson (2004) reported that the central executive system contributes to accuracy in AWP solving after controlling for phonological processing ability, fluid intelligence, and reading comprehension.

Executive functions

According to Miyake et al. (2000), the central executive system component of Baddeley's model is related to three main executive functions: inhibition, updating, and shifting (see also Miyake & Friedman, 2012). Inhibition involves the ability to suppress dominant or prepotent responses, shifting involves the ability to shift strategies while attending to multiple tasks or mental processes, and updating involves the ability to replace outdated and irrelevant information, retaining only a limited set of elements in WM.

Importantly, for the purposes of this article, several studies have demonstrated the role of these executive processes in AWP solving and clarified that this process is supported by updating and inhibition. Agostino et al. (2010) investigated the extent to which inhibition, updating, shifting, and mental-attentional capacity (M-capacity) contributed to children's ability to solve multiplication word problems. Using a structural equation model, the researchers demonstrated that updating had a more important role than age in predicting performance for multiple-step problems, whereas age and updating were equally important predictors for one-step problems.

The above-mentioned study confirmed findings previously reported by Passolunghi and Pazzaglia (2005), who produced evidence of the updating function being involved in the process of solving AWPs. They demonstrated that children who fare well in math also have high updating performance. Other authors have suggested this as well and proved that updating could be a key cognitive process in solving AWPs (Blessing & Ross, 1996; Hammerstein et al., 2019; Iglesias-Sarmiento et al., 2015;

Kotsopoulos & Lee, 2012; Lee et al., 2018). Mori and Okamoto (2016) wrote that updating is a nuclear executive function involved in efficient integration processes that, as suggested by Kintsch and Greeno (1985), are needed to translate each verbal statement into a situational representation to solve a problem. The authors found that individuals with high updating function construct a model of the problem that retains only task-relevant verbal information, whereas those with less effective updating abilities consider extraneous information as well. The study's results were interpreted as reflecting that updating is an underlying executive function essential to activating only the information actually needed to construct an appropriate model of a problem. Constructing an integrated situational model is particularly important to understanding a problem and applying the *model strategy* (Mayer & Hegarty, 1996). An updating failure would produce errors in the integration process, resulting in an inappropriate model and consequently an incorrect solution to the problem (see Kotsopoulos & Lee, 2012; Re et al., 2016).

Only a few studies have investigated shifting's role in mathematics performance, and the outcomes are mixed (Passolunghi & Costa, 2019). Moreover, compared with shifting ability, intelligence showed stronger associations with math performance (Yeniad et al., 2013). However, several studies have shown a relationship between inhibition and mathematical ability (Bull & Scerif, 2001; Khng & Lee, 2009; Passolunghi & Siegel, 2001; Swanson, & Fung, 2016) and support the idea that inhibition might be a core process in solving AWPs. Its contribution was highlighted by Lubin and colleagues (2013, 2016), who demonstrated that successful inconsistent problem solving relies on the ability to inhibit a misleading or overlearned arithmetical strategy such as "add if more, subtract if less." This misleading strategy interferes with AWP-solving performance in childhood and adolescence—and even in adulthood (although experts become more efficient at inhibition and problem solving than nonexperts). In other words, problem solving demands the inhibition of a superficial propositional representation of the problem resulting from a direct translation approach, which would lead to reasoning errors (Hegarty et al., 1992, 1995; Mayer & Hegarty, 1996). Lemaire and Lecacheur (2011) consistently found that children with better inhibitory control made use of efficient strategies to solve arithmetical problems more frequently than children with lower levels of inhibitory control. Further evidence indicated that increasing efficiency in solving inconsistent AWPs from childhood to adulthood (Lewis & Mayer, 1987) is related to the gradual development of cognitive inhibitory control (Lubin et al., 2013, 2016).

Fluid intelligence

The role of fluid intelligence in AWP solving was also considered in the current study. Fluid intelligence comprises processes related to inductive and deductive reasoning, as well as quantitative reasoning (Sternberg & Ben-Zeev, 1996). Previous findings showed that fluid intelligence strongly correlates with measures of executive functions (Conway et al., 2003; Duncan et al., 2008; Kane & Engle, 2002; Salthouse, 2005). More specifically, numerous studies have found moderate to strong relations between intelligence and updating function (Friedman et al., 2006). Some studies have produced empirical support for the importance of intelligence as a predictor of AWP solving. For example, Lee et al. (2004) reported that 10-year-old children scoring higher on intelligence, reading skills, and vocabulary were also better at solving AWPs. Fung and Swanson (2017) found an indirect effect of fluid intelligence on AWP solving in third-graders and found that reading and calculation mediated the influence of the WM executive system on AWP solving in children aged 6 to 10 years. A meta-analysis conducted by Peng et al. (2019) revealed that the relations between fluid intelligence and mathematics tend to increase with age and that fluid intelligence demonstrated stronger relations to complex mathematical skills than to initial and basic mathematical skills.

The current study

Our main objective was to study the role of executive processes (particularly updating and inhibition), comprehension, and fluid intelligence in primary school children solving consistent and inconsistent AWPs. To achieve this aim, participants were administered AWP tasks in which the problem's wording did or did not match the operation needed to solve the problem. In other words, in consistent problems involving one operation the adverbs *more* and *less* used in the problems fitted the arithmetical operation required, whereas in inconsistent problems they did not. We also included problems involving two consistent and two inconsistent operations as well as a consistent request followed by an inconsistent request (C-I) and vice versa (I-C).

The novelty and main objective of the current study was to examine the specific influence of updating, inhibition, and fluid intelligence on AWP-solving accuracy in the same unitary model as well as the mediating role of reading comprehension, accounting for the different types of consistent and inconsistent compare AWPs. Reading comprehension was positioned as a mediator because prior findings indicated that children's reading ability mediated the relationship between executive processes and problem-solving accuracy (Fung & Swanson, 2017; Iglesias-Sarmiento et al., 2015; Passolunghi & Pazzaglia, 2005; Thevenot & Barrouillet, 2015).

Our hypotheses can be outlined as follows. Based on the existing literature (Hegarty et al., 1995; Lewis & Mayer, 1987; Pape, 2003), we expected to find a consistency effect that gives rise to a higher percentage of correct answers for consistent problems than for inconsistent problems (Hypothesis 1).

In addition, we proposed that the accuracy with which participants solve compare problems is more likely to be related to the difficulty in the construction of the AWP's semantic representation (see Lewis & Mayer, 1987) and therefore to the difficulty deriving from the lexical inconsistency (easier when the adverb is consistent and more difficult when the adverb is inconsistent) rather than to the difficulty imposed by the number of arithmetical operations required for the solution (easier when the problem requires one operation and more difficult when it requires two operations). On the basis of this hypothesis, we predicted the consistency effect to be greater than the effect of the number of operations (Hypothesis 2).

The construction of the AWP's semantic representation requires reflection and the controlled application of processes, such as fluid intelligence, inhibition, and updating, mediated by reading comprehension (Hypothesis 3) (see Iglesias-Sarmiento et al., 2015; Passolunghi & Pazzaglia, 2005; Thevenot & Barrouillet, 2015).

Method

Participants

The study involved 203 students aged 10 and 11 years (mean age = 10 years 6 months) attending fourth and fifth grades at a primary school in a city in northeastern Italy. From this initial sample, students with intellectual disabilities (n = 5) or learning disabilities (n = 8) and those with Italian as their second language (n = 8) were excluded, leaving a final sample of 182 participants (95 girls). The sample's arithmetic abilities, as measured by standardized testing, were on the grades' average levels.

Materials

Arithmetic word problem task

In this task, 12 different hypothetical problem situations were presented in two twin versions (called *plus* and *minus* versions; see Table 1) depending on the adverb (*more* or *less*) used in them. Each twin version included the following: two consistent problems, in which the adverb *more* or *less* suggested an operation of addition or subtraction, respectively, and four inconsistent problems, the solutions of which required at least one inverse arithmetical operation to that suggested by the problems' wording. Another variable that can be used to increase the problems' difficulty is the number of relationships described between the problems' elements; more relationships mean more operations are needed to solve the problem. In each version, two of the AWPs involved one operation and four involved two operations. Table 1 shows the six different problems presented in each version as well as their order of presentation. All other components of the word problems (number of sentences, vocabulary, and syntax) are consistent. The average readability, assessed with the Gulpease index (see Dell'Orletta et al., 2011), is above 55 for all texts, suggesting their appropriateness for primary school children (see Tonelli et al., 2012). The AWPs were similar to those assigned at school to students of the considered school grades. However, teachers often fail to emphasize the possibility of

Table 1
Different types and order of arithmetic word problems tested.

Relational term	Type	Number of operations
Plus version	Consistent	1
	Inconsistent	1
	Consistent–consistent	2
	Inconsistent–consistent	2
	Consistent–inconsistent	2
	Inconsistent–inconsistent	2
Minus version	Consistent	1
	Inconsistent	1
	Consistent–consistent	2
	Inconsistent–consistent	2
	Consistent–inconsistent	2
	Inconsistent–inconsistent	2

making mistakes in inconsistent compare problems. Each of the AWP's two versions was presented in sessions on different days to avoid tiring participants. Half the participants were administered the plus version at the first session and the minus version at the second session, whereas the versions were presented in reverse order for the sample's other half. Problems varied in number of operations (one or two) and type of relationship (consistent or inconsistent) between the elements. Cronbach's α for this task in the current sample was.77.

Reading comprehension

This was measured using two expository texts (appropriate for fourth- and fifth-graders) drawn from the standardized Italian battery for the assessment of reading ability (Cornoldi et al., 2017). Participants were asked to silently read the text and then answer 12 multiple-choice questions (and the text remained available while they did so). The reliability (Cronbach's α) values are.69 and.71 for Grades 4 and 5, respectively.

Fluid intelligence

We used Scale 2, Form A, of the Cattell Culture Fair Intelligence Test (Cattell & Cattell, 1963). This consists of four reasoning subtests involving visuospatial material, each to be completed in 2.5 to 4 min. In the subtest named "Matrices," for example, participants are shown 12 incomplete matrices of four to nine cells containing abstract figures and shapes plus an empty cell and six options from which to choose the item that correctly completes each matrix. The final score is the sum of the correct answers (maximum score = 36). The test–retest reliability coefficients reported in the manual are.84 and.80 for fourth- and fifth-graders, respectively.

Updating

This task, taken from Carretti et al. (2014), comprised six lists of eight nouns each. After listening to each list, participants were asked to remember the three smallest items on each list in the correct order of presentation. All the words were highly familiar to the children and referred to objects that were easy to compare in terms of size. The number of correctly recalled words was the dependent variable (maximum score = 18). Cronbach's α for this task is .68.

Prepotent response inhibition

This was measured with a version of the classic Stroop color task, as used in previous research (Borella et al., 2010). The task was administered in a pencil-and-paper version and comprised 15 trials. Each trial included 20 stimuli, each printed on a separate sheet. In the first 5 trials, participants needed to name the colors of strings of uppercase "X"s (control–color condition). Then, there were 5 trials in which participants were asked to name the color of the ink used for words that were names of colors where the ink colors and the named colors differed. In this condition, to efficiently perform the task, participants needed to inhibit the prepotent response prompted by reading the words in favor of the appropriate but nondominant response prompted by focusing on the ink's color (incongruent condition). In the last 5 trials, participants again needed to name the color of the ink used for words that were names of colors, but in this case the ink colors and named colors were consistent (congruent condition). For each trial, completion times and accuracy were recorded. An interference index was calculated for the response times from the differences between the control–color and incongruent conditions: Higher scores indicated greater difficulty in controlling the prepotent responses in the incongruent condition. The task showed high reliability (control–color: α =.82; incongruent: α =.73; congruent: α =.85).

Procedure

The tasks, except the updating task, were collectively administered in two sessions lasting about 1 h each. During the first session, participants completed the AWP-solving task in the plus or minus version (balanced among participants) and the fluid intelligence test. In the second session 1 or 2 days later, they were administered the other version of the AWP-solving task, the reading comprehension test, and the prepotent response inhibition task. The updating task was administered individually and took about 10 min.

Results

Table 2 shows descriptive statistics for the measures included in the study. There was no difference in accuracy between the two versions (plus and minus) of the AWP, $F(1, 181) = 3.506$, p =.063. The percentage of correct AWP solutions did not differ between boys and girls, $F(1, 181) < .01$, p =.958.

The effect of consistency between the linguistic terms used in the problems was examined. As predicted, differences emerged from the repeated-measures analysis of variance (ANOVA) between problems that were worded consistently (consistent AWP [AWP CON]: M = 3.30, SD = 0.89) and those that were not (Inconsistent AWP [AWP INC]: M = 1.34, SD = 1.25), $F(1, 181) = 415.08$, MSE = 348.18, $p < .01$, η_p^2 =.69. In addition, a repeated-measures ANOVA confirmed the significant differences in response accuracy between one-operation problems (M = 2.70, SD = 0.95) and two-operation problems (M = 1.93, SD = 1.06), $F(1, 181) = 102.43$, MSE = 53.85, $p < .01$, η_p^2 =.36. Consequently, the magnitude of the number-of-operations effect was smaller than that of the consistency effect.

When problems with both consistent and inconsistent requests were analyzed, the effect of their order emerged, $F(1, 181) = 12.97$, MSE = 4.18, $p < .01$, η_p^2 =.07; performance was significantly worse for problems with the inconsistent statement presented later (consistent–inconsistent AWP [AWP C-I]: M = 0.51, SD = 0.70) than for problems with the inconsistent statement presented before

Table 2
Descriptive statistics for measures included in the study.

Measure	*M*	*SD*
AWP CON	3.30	0.89
AWP INC	2.58	2.38
AWP I-C	0.73	0.80
AWP C-I	0.51	0.70
AWP total	5.87	2.79
Updating	10.47	2.39
Inhibition (time difference)	−87.01	33.87
Fluid intelligence	31.02	4.80
Reading comprehension	8.98	2.60

Note. AWP CON, consistent arithmetic word problem; AWP INC, inconsistent arithmetic word problem; AWP I-C, inconsistent-then-consistent arithmetic word problem; AWP C-I, consistent-then-inconsistent arithmetic word problem; AWP total, total arithmetic word problem.

(inconsistent–consistent AWP [AWP I-C]: M = 0.73, SD = 0.80). Nevertheless, what was more interesting was to verify whether the performance differences in responding to AWP C-I and AWP I-C were due to a WM overload. To analyze this issue, we ran a one-way repeated-measures ANOVA using the updating measure as a covariate. After controlling for participants' updating ability, the difference in the difficulty of the two types of problem disappeared, $F(1, 181) = 1.10$, $MSE = 0.36$, $p = .295$. This result further supports the assumption that updating processes' efficiency is strongly related to AWP-solving ability and the ability to build a correct mental representation of the problem.

Correlations

Table 3 shows the correlations between the variables considered in the study. All measures were significantly correlated with each other. In particular, it is interesting to note the higher correlations between reading comprehension and fluid intelligence with overall AWP score and that between updating and inconsistent AWP performance.

Predictors of AWP-solving ability

We computed a series of path analyses as regression equations to clarify the relationships between the variables and the magnitude of their ability to explain performance in AWPs with one operation (consistent or inconsistent) or two operations (one consistent and the next inconsistent [C-I] or one inconsistent and the next consistent [I-C]). These analyses included three basic cognitive measures (fluid intelligence, inhibition, and updating), one mediated measure (reading comprehension), and the outcome variables, which were analyzed separately: total AWP, AWP CON, AWP INC, AWP C-I, and AWP I-C.

General equations for different outcomes were as follows:

$$\begin{aligned}\text{Reading comprehension} = {} & \beta 1 + \alpha 1(\text{inhibition}) + e1 + \alpha 2(\text{fluid intelligence}) + e2 \\ & + \alpha 3(\text{updating}) + e3\end{aligned}$$

$$\text{AWP outcome} = \beta 2 + b1(\text{reading}) + c'1(\text{inhibition}) + c'2(\text{fluid intelligence}) + c'3(\text{updating}) + e4.$$

All path models were computed using the same procedure but changing the predicted AWP outcome (total AWP, AWP CON, AWP INC, AWP C-I, or AWP I-C). Fig. 1 depicts the path model tested.

The first regression equation estimated the effects of fluid intelligence, inhibition, and WM updating on reading comprehension (i.e., the mediator). The second regression equation estimated the

Table 3
Correlations between measures of interest.

	1	2	3	4	5	6	7	8
1. AWP I-C	1							
2. AWP C-I	.44**	1						
3. AWP CON	.26**	.26**	1					
4. AWP INC	.84**	.75**	.31**	1				
5. AWP total	.80**	.72**	.59**	.95**	1			
6. Updating	.20**	.27**	.18**	.27**	.28**	1		
7. Inhibition	.20**	.13*	.18**	.20**	.23**	.15*	1	.1
8. Fluid intelligence	.25**	.23**	.31**	.32**	.37**	.26**	.1	1
9. Reading comprehension	.27**	.30**	.30**	.36**	.40**	.18**	.02	.31**

Note. AWP I-C, inconsistent-then-consistent arithmetic word problem; AWP C-I, consistent-then-inconsistent arithmetic word problem; AWP CON, consistent arithmetic word problem; AWP INC, inconsistent arithmetic word problem; AWP total, total arithmetic word problem.
* p <.05 (one-tailed).
** p <.01 (one-tailed).

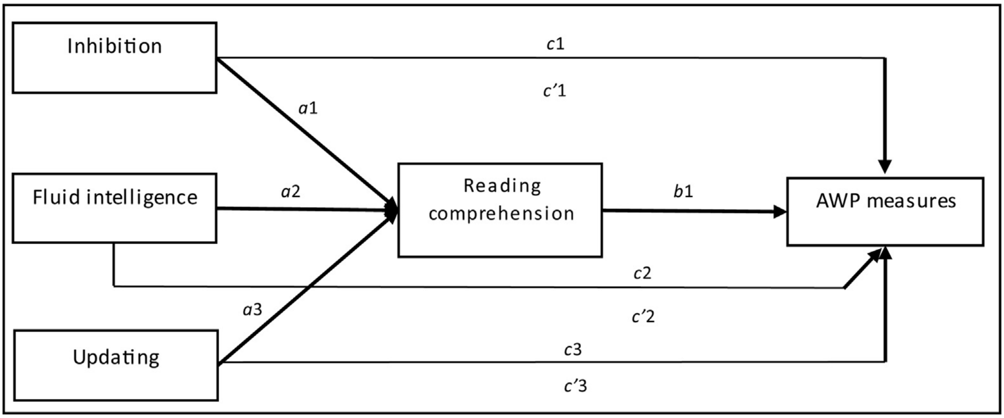

Fig. 1. Path model tested in the study. AWP, arithmetic word problem.

direct effect (c') that fluid intelligence, inhibition, updating, and the mediator (reading comprehension) had on the outcome variables (i.e., AWP performance). The indirect effects of inhibition ($a1 \times b1$), fluid intelligence ($a2 \times b1$), and updating ($a3 \times b1$) through reading comprehension tested these variables' mediation effects on AWP outcome measures. The direct effects (c') of the cognitive measures (inhibition, fluid intelligence, and updating) on AWP outcome measures were independent of the indirect effects. Consequently, the direct effect was equivalent to the difference between the indirect effect and the total effect of cognitive variables on AWP outcome measures. Bootstrapping with 5000 bootstrap samples was used to construct 95 % confidence intervals for indirect effects. Total effect (c) instead represents the overall effects of direct and indirect relations on AWP outcome measures. The indexes for direct and indirect effects and p values for regression coefficients are presented in the tables available in the online supplementary material. Effect sizes (Cohen's f^2) are also reported to account for the overall regressors' effect on each dependent variable: small effect (f^2 =.02), medium effect (f^2 =.15), and large effect (f^2 =.35).

Regarding the total AWP-solving performance (total AWP), results are presented in Fig. 2 (and Table 4 in the supplementary material). The total amount of variance explained by the whole model was R^2 =.28 (f^2 =.38). The effects of executive functions and fluid intelligence on the mediator (a paths)

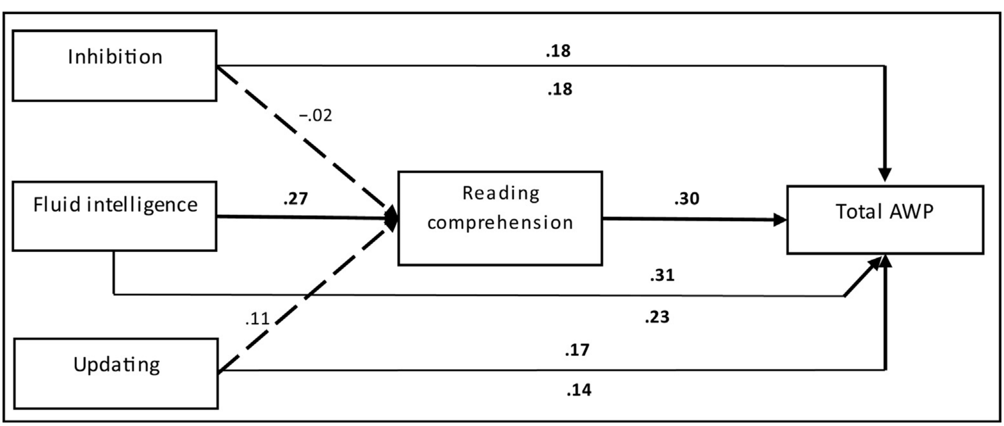

Fig. 2. Path model for total arithmetic word problem (AWP)-solving performance. Lines in bold indicate statistically significant relations, whereas dashed lines indicate nonsignificant relations.

were significant for only fluid intelligence (B =.16, SE =.05, β =.27, p <.001, *95 % CI* [confidence interval] =.07 to.25). This result was common among all AWP outcome models. The effect of the mediator, reading comprehension, on total AWP (b path) was significant (B =.33, SE =.08, β =.30, p <.001, *95 % CI* =.17 to.49). The total effects of executive functions on total AWP (c paths) were significant for inhibition (B =.02, SE =.01, β =.18, p <.012, *95 % CI* =.01 to.04), fluid intelligence (B =.19, SE =.05, β =.31, p <.001, *95 % CI* =.01 to.28), and updating (B =.21, SE =.08; β =.17, p =.006, *95 % CI* =.05 to.36). When direct effects on total AWP (c' paths) were considered, significant relations were found for inhibition (B =.02, SE =.01, β =.18, p =.007, *95 % CI* =.01 to.04), fluid intelligence (B =.14, SE =.04, β =.23, p =.001, *95 % CI* =.05 to.23), and updating (B =.17, SE =.08, β =.14, p <.027, *95 % CI* =.01 to.32). As for indirect effects of predictors on output (total AWP) through the mediator variable (reading comprehension), only intelligence was significant (B =.05, SE =.02, β =.08, p =.012, *95 % CI* =.02 to.10) after bootstrapping.

Regarding the consistent-AWP-solving performance (AWP CON), results are presented in Fig. 3 (and Table 5 in the supplementary material). The total amount of variance explained by the whole model was R^2 =.16 (f^2 =.19). The effect of the mediator, reading comprehension, on AWP CON (b path) was significant (B =.08, SE =.03, β =.23, p =.006, *95 % CI* =.03 to.14). The total effects of executive functions and fluid intelligence on AWP CON (c paths) were significant for inhibition (B <.01, SE <.01, β =.16, p =.037, *95 % CI* =.00 to.01) and fluid intelligence (B =.05, SE =.02, β =.26, p <.008, *95 % CI* =.02 to.09), but not for updating (B =.03, SE =.03, β =.09, p =.22). When direct effects on total AWP CON (c' paths) were considered, significant relations were found for inhibition (B <.01, SE <.01, β =.16, p =.027, *95 % CI* =.00 to.01) and fluid intelligence (B =.04, SE =.02, β =.20, p =.039, *95 % CI* =.00 to.08), but not for updating (B =.02, SE =.03, β =.06, p =.403). As for indirect effects of predictors on output (AWP CON) through the mediator variable (reading comprehension), only intelligence was significant (B =.01, SE =.01, β =.06, p =.046, *95 % CI* =.00 to.03) after bootstrapping.

Regarding Inconsistent-AWP-solving performance (AWP INC), results are presented in Fig. 4 (and Table 6 in the supplementary material). The total amount of variance explained by the model was R^2 =.22 (f^2 =.28). The effect of the mediator, reading comprehension, on AWP INC (b path) was significant (B =.25, SE =.07, β =.27, p <.001, *95 % CI* =.12 to.38). The total effects of executive functions on AWP INC (c paths) were significant for inhibition (B =.02, SE =.01, β =.15, p <.035, *95 % CI* =.00 to.03), fluid intelligence (B =.14, SE =.04, β =.27, p <.001, *95 % CI* =.07 to.21), and updating (B =.18, SE =.07, β =.17, p =.012, *95 % CI* =.04 to.31). When direct effects on AWP INC (c' paths) were considered, significant relations were found for inhibition (B =.02, SE =.01, β =.16, p =.026, *95 % CI* =.00 to.03), fluid intelligence (B =.11, SE =.04, β =.20, p =.004, *95 % CI* =.03 to.27) and updating (B =.15; SE =.07; β =.14; p =.037; *95 % CI* =.01 to.27). As for indirect effects of predictors on output (AWP INC) through the mediator

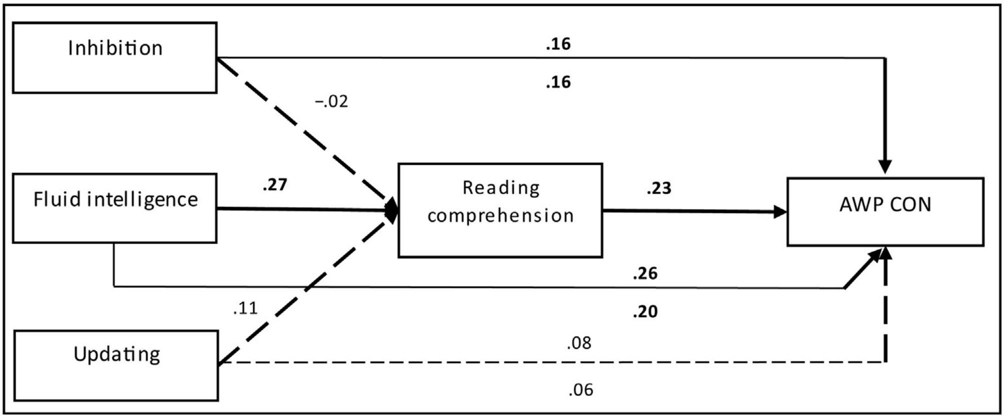

Fig. 3. Path model for consistent arithmetic word problem (AWP CON)-solving performance. Lines in bold indicate statistically significant relations, whereas dashed lines indicate nonsignificant relations.

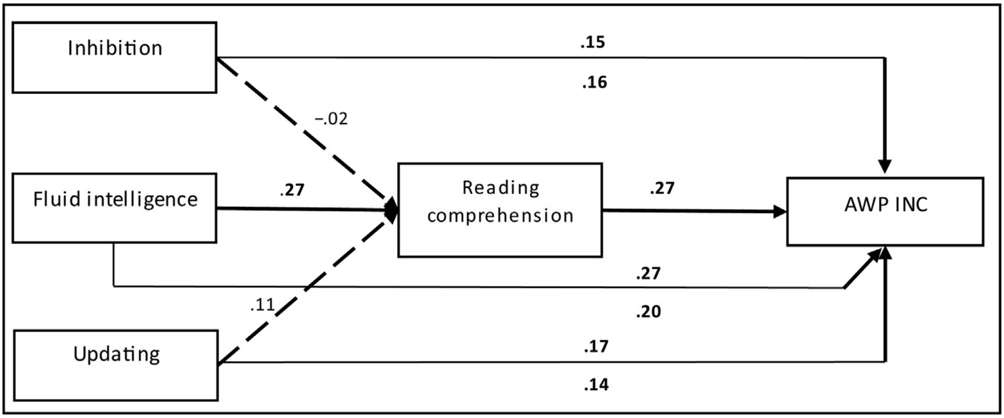

Fig. 4. Path model for Inconsistent arithmetic word problem (AWP INC)-solving performance. Lines in bold indicate statistically significant relations, whereas dashed lines indicate nonsignificant relations.

variable (reading comprehension), only intelligence was significant (*B* =.04, *SE* =.02, β =.07, *p* =.019, *95 % CI* =.02 to.08) after bootstrapping.

Regarding the two-operation inconsistent-then-consistent AWP (AWP I-C), results are presented in Fig. 5 (and Table 7 in the supplementary material). The total amount of variance explained by the whole model was R^2 =.15 (f^2 =.18). The effect of the mediator, reading comprehension, on AWP I-C (*b* path) was significant (*B* =.06, *SE* =.02, β =.20, *p* =.005, *95 % CI* =.02 to.10). The total effects of executive functions on AWP I-C (*c* paths) were significant for inhibition (*B* <.01, *SE* <.01, β =.17, *p* =.017, *95 % CI* =.00 to.01) and fluid intelligence (*B* =.04, *SE* =.01, β =.22, *p* =.002, *95 % CI* =.01 to.06), but not for updating (*B* =.04, *SE* =.03, β =.12, *p* =.100). When direct effects on AWP I-C (*c′* paths) were considered, significant relations were found for inhibition (*B* =.01, *SE* <.01, β =.17, *p* =.015, *95 % CI* =.00 to.01) and fluid intelligence (*B* =.03, *SE* =.01, β =.17, *p* =.022, *95 % CI* =.00 to.06), but not for updating (*B* =.04, *SE* =.03, β =.10, *p* =.173). As for indirect effects of predictors on output (AWP I-C) through the mediator variable (reading comprehension), only intelligence was significant (*B* =.01, *SE* =.01, β =.05, *p* =.036, *95 % CI* =.00 to.02) after bootstrapping.

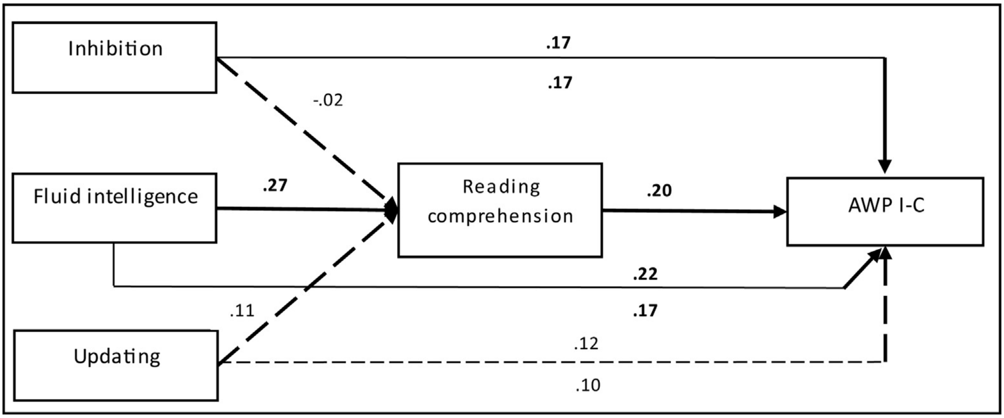

Fig. 5. Path model for inconsistent-then-consistent arithmetic word problem (AWP I-C)-solving performance. Lines in bold indicate statistically significant direct relations, whereas dashed lines indicate nonsignificant direct relations.

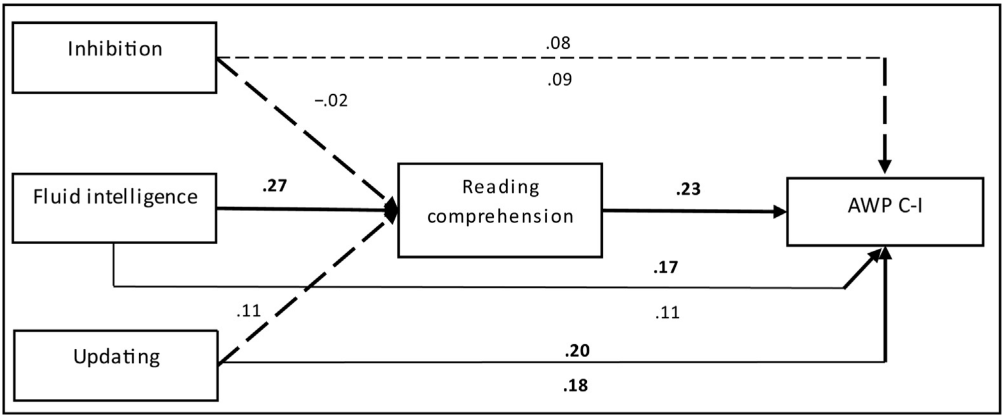

Fig. 6. Path model for consistent-then-inconsistent arithmetic word problem (AWP C-I)-solving performance. Lines in bold indicate statistically significant direct relations, whereas dashed lines indicate nonsignificant direct relations.

Finally, the two-operation consistent-then-inconsistent AWP (AWP C-I) results are presented in Fig. 6 (and Table 8 in the supplementary material). The total amount of variance explained by the model was R^2 =.15 (f^2 =.18). The effect of the mediator, reading comprehension, on AWP C-I (*b* path) was significant (*B* =.06, *SE* =.02, *β* =.23, *p* =.003, *95 % CI* =.02 to.11). The total effects of executive functions on AWP C-I (*c* paths) were significant for fluid intelligence (*B* =.03, *SE* =.01, *β* =.17, *p* =.020, *95 % CI* =.00 to.05) and updating (*B* =.06, *SE* =.02, *β* =.20, *p* =.005, *95 % CI* =.02 to.10), but not for inhibition (*B* <.01, *SE* <.01, *β* =.08, *p* =.277). When direct effects on AWP C-I (*c′* paths) were considered, significant relations were found for updating (*B* =.05, *SE* =.02, *β* =.18, *p* =.013, *95 % CI* =.01 to.09), but not for inhibition (*B* <.01, *SE* <.01, *β* =.09, *p* =.235) or fluid intelligence (*B* =.02, *SE* =.01, *β* =.11, *p* =.141). As for indirect effects of predictors on output (AWP I-C) through the mediator variable (reading comprehension), only intelligence was significant (*B* =.01, *SE* =.01, *β* =.06, *p* =.036, *95 % CI* =.00 to.02) after bootstrapping.

In summary, the effect sizes of the presented models are medium (except for total AWP), demonstrating that all models included relevant cognitive factors. In the case of the total AWP, the effect size is large (probably due to a large number of AWPs) and gives additional support to all the cognitive variables considered in the current study being relevant in explaining the problem-solving performance.

Discussion

The current study sheds new light on the relationship between consistent- and inconsistent-AWP-solving performance and a few cognitive variables (i.e., inhibition, updating, and fluid intelligence), demonstrating the mediating role of reading comprehension. First, as expected and in line with the literature (Hegarty et al., 1995; Lewis & Mayer, 1987; Pape, 2003), the results confirmed a *consistency effect* on children's AWP-solving performance (Hypothesis 1). Participants were more efficient at solving consistent versus inconsistent AWPs. It should be noted that the sample considered in the current study obtained a very high percentage of correct answers for consistent AWPs and that two-operation consistent AWPs were easier than either AWP I-C or AWP C-I, although the number of operations was a factor contributing to the problems' difficulty (Castro-Martínez & Frías-Zorilla, 2013; Quintero, 1983). The pattern of correlations supports comprehension-related abilities' crucial role (Kintsch, 1998; Peng et al., 2019; Thevenot, 2010), which emerged as a stronger correlate of performance for inconsistent AWPs than for consistent AWPs. These results suggest that the differences observed in how accurately participants solved consistent AWPs versus inconsistent AWPs might not relate to the number of arithmetical operations per se (Hypothesis 2). Instead, an AWP's difficulty would be related to the particular requirements of the representation of inconsistent problems linked to the construction of an integrated mental model of the problem (see Fuson, 1992; Hasanah et al., 2017; Thevenot & Barrouillet, 2015). We know that semantic elements of AWPs influence their difficulty and the strategies children use to solve them (e.g., Boonen et al., 2016; De Corte et al., 1985; Pape, 2003). As mentioned in the Introduction, a key aspect differentiating inconsistent AWPs from consistent AWPs concerns the conflict between the relational term (e.g., *more than*) and the arithmetical operation required (e.g., subtraction) to solve an inconsistent problem. Instead of using a superficial direct strategy or a well-established association (e.g., *more* with "addition" and "increases in amount"; Schumacher & Fuchs, 2012), the child must use a problem-model strategy. This involves translating the statement of the inconsistent problem into a situational model representing the appropriate relations between the variables (de Koning et al., 2017; Hegarty et al., 1995; Thevenot, 2010) and then planning and executing the required arithmetical operations.

Such nonroutine thinking to solve inconsistent AWPs requires reflection and a controlled application of executive fluid processes such as inhibition and updating the information about the problem in WM (Hypothesis 3). Failure to inhibit the strong association between the keyword in the literal statement (e.g., *more*) and the operation (e.g., addition) or to update the operation, switching from the wrong one to the right one (e.g., subtraction), would tend to prompt a superficial erroneous response. For an adequate situational representation of an inconsistent step, the problem solver must reason reflexively and inhibit any inappropriate automatic response (e.g., Jiang et al., 2020; Passolunghi & Siegel, 2004) or misleading overlearned strategy (Lubin et al., 2013, 2016).

Our results indicated that AWP C-I tests were more difficult than AWP I-C tests. However, this result must be interpreted cautiously because the order of presentation was not controlled. Nevertheless, it is highly interesting that after controlling for participants' updating abilities, the differences in accuracy between AWP C-I and AWP I-C disappeared. This result gives additional support to the assumption that updating ability is strongly related to the ability to build a correct mental representation of a problem and then solve it (Iglesias-Sarmiento et al., 2015; Passolunghi & Pazzaglia, 2005; Thevenot & Barrouillet, 2015). Moreover, this updating ability seems to be more involved in the construction of an integrated situational model in AWP C-I than in AWP I-C.

As concerns the specific contributions of updating, inhibition, reading, and fluid intelligence to AWP solving, their roles depended on the problem type. The model confirmed that both executive functions—updating and inhibition—had a distinct and significant effect on AWP-solving accuracy. We also found that fluid intelligence had both direct and indirect effects (mediated by reading comprehension) on the overall measure of AWP-solving performance. It contributed a unique part of the variance in AWP-solving accuracy even after accounting for the effect of reading comprehension. The magnitude of the β weights suggests that fluid intelligence had only a moderate effect, however, and that updating and inhibition had even lower effects on AWP-solving performance. Our results likewise confirmed that reading comprehension is the most important predictor of AWP-solving ability in the final grades of primary school (fourth and fifth grades), by which time basic reading skills have become relatively automated. Reading comprehension had a mediating role in AWP-solving accuracy. The model explained a moderate proportion of variance (30 %) in total AWP accuracy, although its weights varied for different types of AWP, being lowest for AWP I-C and highest for total AWP.

Moreover, an interesting finding concerns the differential roles of inhibition and updating on different problem types. Regarding inhibition, results demonstrated its contribution to explaining accuracy on consistent, inconsistent, and I-C problems. As suggested by Passolunghi and Siegel (2001), inhibition is a core ability necessary to select and disregard irrelevant and unneeded information and therefore is required in the solving process of various problem types. However, our results also showed that inhibition lost relevance in C-I problems in favor of updating abilities. The updating process is a more complex cognitive process because it involves comparison processes, inhibition of no longer relevant information and its substitution with new information, and resistance to interference (see e.g., Linares et al., 2016). It must be noted that C-I problems were the most difficult problem type, and thus it could be speculated that a problem's difficulty may increase the demand on the solver's ability to update and integrate information (Agostino et al., 2010) in order to create a coherent mental representation of the problem. In fact, we found updating abilities to explain accuracy on more difficult problem types (i.e., C-I and inconsistent problems), but not on the easier problem types (i.e., I-C and consistent problems). These results highlight the importance of considering problem type and problem complexity when investigating the role of executive functions in word problem solving.

To sum up, the findings presented in this article provide new insight into the relationships among fluid cognition, executive functions, and AWP-solving performance, particularly in the case of consistent and inconsistent compare problems. Moreover, to our best knowledge, no study has examined the relations between cognitive variable and consistent and inconsistent problems considering inhibition, updating, and fluid intelligence in the same unitary model. We demonstrated that AWP solving relies not only on the acquisition of arithmetical skills but also on reading comprehension and, importantly, on reasoning and the central executive processes involved in actively processing information, updating relevant information, and inhibiting irrelevant information and inappropriate strategies. These domain-general factors are important, each in its own right, in explaining children's performance in consistent and inconsistent AWP solving.

The results of the current study should be interpreted in the light of some limitations. A first possible limitation arises from the fact that a shifting measure is lacking. Some authors (e.g., Toll et al., 2010) have suggested, however, that the role of shifting in mathematics would relate to the complexity of the task at hand and the required knowledge. This being the case, shifting might be minimal in AWPs such as those analyzed in the current study. A second limitation concerns other factors not studied here that might influence and explain the variation in AWP-solving performance (see e.g., Fung & Swanson, 2017). For instance, individual differences and other domain-specific factors, such as processing speed, attention ability, and visuospatial ability, might contribute to performance in AWP solving (see Boonen et al., 2013; Cragg et al., 2017; Passolunghi & Lanfranchi, 2012). It may be that the relations described in this article would change if other variables were examined, and future research should aim to refine the proposed model.

Nevertheless, this study's findings could contribute to highlighting the importance of teaching students to integrate a problem's textual information into an adequate mental representation, which is the basis for a solution strategy (Jimenez & Verschaffel, 2014; Thevenot & Oakhill, 2005; Verschaffel et al., 2020). Indeed, solving an AWP is not a simple translation of problem sentences into arithmetic operations; rather, it involves other factors such as solution strategies and establishing a correct mental model of the problem. Educational intervention focused on comprehension skills, controlled thinking, and WM processes might be an effective way to improve AWP-solving skills (e.g., Boonen et al., 2016; Cornoldi et al., 2015; Fuchs et al., 2020). Understanding the role of these cognitive and linguistic factors is essential to promoting tailored training programs to enhance word-problem-solving abilities.

Acknowledgment

This work was carried out within the scope of the "Use-Inspired Basic Research" project for which the Department of General Psychology at the University of Padova has been recognized as a "Dipartimento di Eccellenza" by the Italian Ministry for the University and Research.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2022.105512.

References

Agostino, A., Johnson, J., & Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning: Problem type matters. *Journal of Experimental Child Psychology.*, *105*, 286–305. https://doi.org/10.1016/j.jecp.2009.09.006.

Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Applied Cognitive Psychology*, *21*, 1201–1216. https://doi.org/10.1002/acp.1317.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.). *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1.

Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 792–810. https://doi.org/10.1037/0278-7393.22.3.792.

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*. https://doi.org/10.3389/fpsyg.2016.00191. Article 191.

Boonen, A. J. H., & Jolles, J. (2015). Second grade elementary school students' differing performance on combine, change and compare word problems. *International Journal of School and Cognitive Psychology, 2*, 1–6. https://doi.org/10.4172/2469-9837.1000122.

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*, 271–279. https://doi.org/10.1016/j.cedpsych.2013.05.001.

Boote, S. K., & Boote, D. N. (2018). ABC problem in elementary mathematics education: Arithmetic before comprehension. *Journal of Mathematics Teacher Education, 21*, 99–122. https://doi.org/10.1007/s10857-016-9350-2.

Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities, 43*, 541–552. https://doi.org/10.1177/0022219410371676.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*, 273–293. https://doi.org/10.1207/S15326942DN1903_3.

Carretti, B., Caldarola, N., Tencati, C., & Cornoldi, C. (2014). Improving reading comprehension in reading and listening settings: The effect of two training programmes focusing on metacognition and working memory. *British Journal of Educational Psychology, 84*, 194–210. https://doi.org/10.1111/bjep.12022.

Castro-Martínez, E., & Frías-Zorilla, A. (2013). Two-step arithmetic word problems. *The Mathematics Enthusiast, 10*, 379–406. https://doi.org/10.54870/1551-3440.1272.

Cattell, R. B., & Cattell, A. K. S. (1963). *Culture Fair Intelligence Test: Scale 3*. Champaigne, IL, USA: Institute for Personality and Ability Testing.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 12*, 547–552. https://doi.org/10.1016/j.tics.2003.10.005.

Cornoldi, C., Carretti, B., & Colpo, C. (2017). *Prove MT-Kit Scuola: Dalla valutazione degli apprendimenti di lettura e comprensione al potenziamento [MT-Kit School Test: From the assessment of the reading to the intervention]*. Florence, Italy: Giunti Editore.

Cornoldi, C., Carretti, B., Drusi, S., & Tencati, C. (2015). Improving problem solving in primary school students: The effect of a training programme focusing on metacognition and working memory. *British Journal of Educational Psychology, 85*, 424–439. https://doi.org/10.1111/bjep.12083.

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition, 162*, 12–26. https://doi.org/10.1016/j.cognition.2017.01.014.

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*. https://doi.org/10.3389/fpsyg.2015.00348. Article 348.

De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solution. *Journal of Educational Psychology, 77*, 460–470. https://doi.org/10.1037/0022-0663.77.4.460.

de Koning, B. B., Boonen, A. J. H., & van der Schoot, M. (2017). The consistency effect in word problem solving is effectively reduced through verbal instruction. *Contemporary Educational Psychology, 49*, 121–129. https://doi.org/10.1016/j.cedpsych.2017.01.006.

Dell'Orletta, F., Montemagni, S., & Venturi, G. (2011). READ-IT: Assessing readability of Italian texts with a view to text simplification. In *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies (SLPAT '11)* (pp. 73–83). Stroudsburg, PA, USA: Association for Computational Linguistics.

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., ... Nimmo-Smith, I. (2008). Goal neglect and Spearman's *g:* Competing parts of a complex task. *Journal of Experimental Psychology: General, 137*, 131–148. https://doi.org/10.1037/0096-3445.137.1.131.

Duque de Blas, G., Gómez-Veiga, I., & García-Madruga, J. A. (2021). Word arithmetic problems revisited: Cognitive processes and academic performance in secondary school. *Education Science, 11*. https://doi.org/10.3390/educsci11040155. Article 155.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*, 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x.

Friso-van den Bos, I., van der Ven, S. H., Kroesbergen, E. H., & van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational research review, 10*, 29–44. https://doi.org/10.1016/j.edurev.2013.05.003.

Fuchs, L., Fuchs, D., Seethaler, P. M., & Barnes, M. A. (2020). Addressing the role of working memory in mathematical word-problem solving when designing intervention for struggling learners. *ZDM Mathematics Education, 52*, 87–96. https://doi.org/10.1007/s11858-019-01070-8.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., ... Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology, 46*, 1731–1746. https://doi.org/10.1037/a0020662.

Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Martin, B. N. (2018). Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension. *Scientific Studies of Reading, 22*, 152–166. https://doi.org/10.1080/10888438.2017.1398259.

Fung, W., & Swanson, L. (2017). Working memory components that predict word problem solving: Is it merely a function of reading, calculation, and fluid intelligence? *Memory & Cognition, 45*, 804–823. https://doi.org/10.3758/s13421-017-0697-0.

Fuson, K. C. (1992). Research on whole number addition and subtraction. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 243–275). MacMillan. https://doi.org/10.3389/fpsyg.2014.00381.

Gross, J., Hudson, C., & Price, D. (2009). *The long term costs of numeracy difficulties*. East Sussex, UK: National Numeracy.

Hammerstein, S., Poloczek, S., Lösche, P., Lemaire, P., & Büttner, G. (2019). Effects of working memory updating on children's arithmetic performance and strategy use: A study in computational estimation. *Journal of Experimental Child Psychology, 184*, 174–191. https://doi.org/10.1016/j.jecp.2019.04.003.

Hasanah, N., Hayashi, Y., & Hirashima, T. (2017). An analysis of learner outputs in problem posing as sentence-integration in arithmetic word problems. *Research and Practice in Technology Enhanced Learning, 12*. https://doi.org/10.1186/s41039-017-0049-5. Article 9.

Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology, 84*, 76–84. https://doi.org/10.1037/0022-0663.84.1.76.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*, 18–32. https://doi.org/10.1037/0022-0663.87.1.18.

Iglesias-Sarmiento, V., Carriedo-López, N., & Rodríguez Rodríguez, J. L. (2015). Updating, executive function and performance in reading comprehension and problem solving. *Anales de Psicología, 31*, 298–309. https://doi.org/10.6018/analesps.31.1.158111.

Jarosz, A. F., & Jaeger, A. J. (2019). Inconsistent operations: A weapon of math disruption. *Applied Cognitive Psychology, 33*, 124–138. https://doi.org/10.1002/acp.3471.

Jiang, R., Li, X., Xu, P., & Lei, Y. (2020). Do teachers need to inhibit heuristic bias in mathematics problem-solving? Evidence from a negative-priming study. *Current Psychology*. https://doi.org/10.1007/s12144-020-01209-x. Advance online publication.

Jimenez, L., & Verschaffel, L. (2014). Development of children's solutions of non-standard arithmetic word problem solving. *Revista de Psicodidáctica, 19*, 93–123. https://doi.org/10.1387/RevPsicodidact.7865.

Jitendra, A. (2002). Teaching students math problem-solving through graphic representations. *Teaching Exceptional Children, 34* (4), 34–38. https://doi.org/10.1177/004005990203400405.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*, 637–671. https://doi.org/10.3758/BF03196323.

Khng, K. H., & Lee, K. (2009). Inhibiting interference from prior knowledge: Arithmetic intrusions in algebra word problem solving. *Learning and Individual Differences, 19*, 262–268. https://doi.org/10.1016/j.lindif.2009.01.004.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*, 109–129. https://doi.org/10.1037/0033-295X.92.1.109.

Kotsopoulos, D., & Lee, J. (2012). A naturalistic study of executive function and mathematical problem-solving. *Journal of Mathematical Behavior, 31*, 196–208. https://doi.org/10.1016/j.jmathb.2011.12.005.

Lee, K., Ng, S. F., & Bull, R. (2018). Learning and solving algebra word problems. The roles of relational skills, arithmetic, and executive functioning. *Developmental Psychology, 54*, 1758–1772. https://doi.org/10.1037/dev0000561.

Lee, K., Ng, E. L., & Ng, S. F. (2009). The contributions of working memory and executive functioning to problem representation and solution generation in algebraic word problems. *Journal of Educational Psychology, 101*, 373–387. https://doi.org/10.1037/a0013843.

Lee, K., Ng, S. F., Ng, E. L., & Lim, Z. Y. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *Journal of Experimental Child Psychology, 89*, 140–158. https://doi.org/10.1016/j.jecp.2004.07.001.

Lemaire, P., & Lecacheur, M. (2011). Age-related changes in children's executive functions and strategy selection: A study in computational estimation. *Cognitive Development, 26*, 282–294. https://doi.org/10.1016/j.cogdev.2011.01.002.

Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology, 79*, 363–371. https://doi.org/10.1037/0022-0663.79.4.363.

Lin, X. (2021). Investigating the unique predictors of word-problem solving using meta-analytic structural equation modeling. *Educational Psychology Review, 33*, 1097–1124. https://doi.org/10.1007/s10648-020-09554-w.

Linares, R., Bajo, M. T., & Pelegrina, S. (2016). Age-related differences in working memory updating components. *Journal of Experimental Child Psychology, 147*, 39–52. https://doi.org/10.1016/j.jecp.2016.02.009.

Lubin, A., Rossi, S., Lanoë, C., Vidal, J., Houdè, O., & Borst, G. (2016). Expertise, inhibitory control and arithmetic word problems: A negative priming study in mathematics experts. *Learning and Instruction, 45*, 40–48. https://doi.org/10.1016/j.learninstruc.2016.06.004.

Lubin, A., Vidal, J., Lanoë, C., Houdé, O., & Borst, G. (2013). Inhibitory control is needed for the resolution of arithmetic word problems: A developmental negative priming study. *Journal of Educational Psychology, 105*, 701–708. https://doi.org/10.1037/a0032625.

Mayer, R. E. (1982). Different problem-solving strategies for algebra word and equation problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 448–462. https://doi.org/10.1037/0278-7393.8.5.448.

Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 29–53). Lawrence Erlbaum. https://doi.org/10.4324/9780203053270.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*, 8–14. https://doi.org/10.1177/0963721411429458.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wagner, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100. https://doi.org/10.1006/cogp.1999.0734.

Morales, R. V., Shute, V. J., & Pellegrino, J. W. (1985). Developmental differences in understanding and solving simple mathematics word problems. *Cognition and Instruction, 2*(1), 41–57. https://doi.org/10.1207/s1532690xci0201_2.

Mori, K., & Okamoto, M. (2016). The role of the updating function in solving arithmetic word problems. *Journal of Educational Psychology, 109*, 245–256. https://doi.org/10.1037/edu0000134.

Murnane, R. J., Willett, J. B., Braatz, M. J., & Duhaldeborde, Y. (2001). Do different dimensions of male high school students' skills predict labor market success a decade later? Evidence from the NLSY. *Economics of Education Review, 20*, 311–320. https://doi.org/10.1016/S0272-7757(00)00056-X.

Pape, S. J. (2003). Compare word problems: Consistency hypothesis revisited. *Contemporary Educational Psychology, 28*, 396–421. https://doi.org/10.1016/S0361-476X(02)00046-2.

Passolunghi, M. C., & Costa, H. M. (2019). Working memory and mathematical learning. In A. Fritz, V. G. Haase, & P. Räsänen (Eds.), *International handbook of mathematical learning difficulties: From the laboratory to the classroom* (pp. 407–421). Springer International. https://doi.org/10.1007/978-3-319-97148-3_25.

Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: A longitudinal study from kindergarten to first grade. *British Journal of Educational Psychology, 82*, 42–63. https://doi.org/10.1111/j.2044-8279.2011.02039.x.

Passolunghi, M. C., & Pazzaglia, F. (2005). A comparison of updating processes in children good or poor in arithmetic word problem-solving. *Learning and Individual Differences, 15*, 257–269. https://doi.org/10.1016/j.lindif.2005.03.001.

Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology, 80*, 44–57. https://doi.org/10.1006/jecp.2000.2626.

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology, 88*, 384. https://doi.org/10.1016/j.jecp.2004.04.002.

Peng, P., Wang, T., Wang, C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin, 145*, 189–236. https://doi.org/10.1037/bul0000182.

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM Mathematics Education, 52*, 33–44. https://doi.org/10.1007/s11858-019-01118-9.

Powell, S. R. (2011). Solving word problems using schemas: A review of the literature. *Learning Disabilities Research & Practice, 26*, 94–108. https://doi.org/10.1111/j.1540-5826.2011.00329.x.

Quintero, A. H. (1983). Conceptual understanding in solving two-step work problems with a ratio. *Journal of Research in Mathematics Education, 14*, 102–112. https://doi.org/10.2307/748578.

Re, A., Lovero, A., Cornoldi, C., & Passolunghi, M. C. (2016). Difficulties of children with ADHD symptoms in solving mathematical problems when information must be updated. *Research in Developmental Disabilities, 59*, 186–193. https://doi.org/10.1016/j.ridd.2016.09.001.

Riley, M. S., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction, 5*, 49–101. https://doi.org/10.1207/s1532690xci0501_2.

Riley, M. S., Greeno, J. G., & Heller, J. L. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). Academic Press.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19*, 532–545. https://doi.org/10.1037/0894-4105.19.4.532.

Schumacher, R., & Fuchs, L. (2012). Does understanding relational terminology mediate effects of intervention on compare word problems? *Journal of Experimental Child Psychology, 111*, 607–628. https://doi.org/10.1016/j.jecp.2011.12.001.

Shum, H. Y., & Chan, W. W. L. (2020). Young children's inhibition of keyword heuristic in solving arithmetic word problems. *Human Behaviour and Brain, 1*(2), 43–48. https://doi.org/10.37716/HBAB.2020010202.

Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children? *Journal of Educational Psychology, 85*, 7–23. https://doi.org/10.1037/0022-0663.85.1.7.

Sternberg, R. J., & Ben-Zeev, T. (Eds.). (1996). *The nature of mathematical thinking.* Lawrence Erlbaum. https://doi.org/10.4324/9780203053270.

Swanson, H. L. (2004). Working memory and phonological processing as predictors of children's mathematical problem solving at different ages. *Memory & Cognition, 32*, 648–661. https://doi.org/10.3758/BF03195856.

Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology, 98*, 265–281. https://doi.org/10.1037/0022-0663.98.2.265.

Swanson, H. L., & Fung, W. (2016). Working memory components and problem-solving accuracy: Are there multiple pathways? *Journal of Educational Psychology, 108*, 1153–1177. https://doi.org/10.1037/edu0000116.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79*, 294–321. https://doi.org/10.1006/jecp.2000.2587.

Thevenot, C. (2010). Arithmetic word problem solving: Evidence for the construction of a mental model. *Acta Psychologica, 133*, 90–95. https://doi.org/10.1016/j.actpsy.2009.10.004.

Thevenot, C., & Barrouillet, P. (2015). Arithmetic word problem solving and mental representations. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 158–179). Oxford University Press.

Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representation in arithmetic word problem solving. *Quarterly Journal of Experimental Psychology, 58*, 1311–1323. https://doi.org/10.1080/02724980443000593.

Toll, S. W., van der Ven, S. H., Kroesbergen, E. H., & Van Luit, J. E. (2010). Executive functions as predictors of math learning disabilities. *Journal of Learning Disabilities, 44*, 521–532. https://doi.org/10.1177/0022219410387302.

Tonelli, S., Tran, K. M., & Pianta, E. (2012). Making readability indices readable. In S. Williams, A. Siddharthan, & A. Nenkova (Eds.), *Proceedings of the First Workshop on Predicting and Improving Text Readability for Target Reader Populations* (pp. 40–48). Stroudsburg, PA, USA: Association for Computational Linguistics.

Verschaffel, L. (1994). Using retelling data to study elementary school children's representations and solutions of compare problems. *Journal for Research in Mathematics Education, 25*, 141–165. https://doi.org/10.2307/749506.

Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology, 84*, 85–94. https://doi.org/10.1037/0022-0663.84.1.85.

Verschaffel, L., Schukajlow, S., & Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM Mathematics Education, 52*, 1–16. https://doi.org/10.1007/s11858-020-01130-4.

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences, 23*, 1–9. https://doi.org/10.1016/j.lindif.2012.10.004.