education sciences

MDPI

Article

Arithmetic Word Problems Revisited: Cognitive Processes and Academic Performance in Secondary School

Gonzalo Duque de Blas [1,*], Isabel Gómez-Veiga [2] and Juan A. García-Madruga [2]

[1] La Escuela Internacional de Doctorado de la UNED (EIDUNED), Department of Developmental and Educational Psychology, Faculty of Psychology, National Distance Education University (UNED), Calle de Juan del Rosal, 10, 28040 Madrid, Spain

[2] Department of Developmental and Educational Psychology, Faculty of Psychology, National Distance Education University (UNED), Calle de Juan del Rosal, 10, 28040 Madrid, Spain; igveiga@psi.uned.es (I.G.-V.); jmadruga@psi.uned.es (J.A.G.-M.)

* Correspondence: gduquedeblas@gmail.com or gduque8@alumno.uned.es

Abstract: Solving arithmetic word problems is a complex task that requires individuals to activate their working memory resources, as well as the correct performance of the underlying executive processes involved in order to inhibit semantic biases or superficial responses caused by the problem's statement. This paper describes a study carried out with 135 students of Secondary Obligatory Education, each of whom solved 5 verbal arithmetic problems: 2 consistent problems, whose mathematical operation (add/subtract) and the verbal statement of the problem coincide, and 3 inconsistent problems, whose required operation is the inverse of the one suggested by the verbal term(s). Measures of reading comprehension, visual–spatial reasoning and deductive reasoning were also obtained. The results show the relationship between arithmetic problems and cognitive measures, as well as the ability of these problems to predict academic performance. Regression analyses confirmed that arithmetic word problems were the only measure with significant power of association with academic achievement in both History/Geography (β = 0.25) and Mathematics (β = 0.23).

Keywords: arithmetic word problems; reading comprehension; reasoning; academic performance; secondary school

check for updates

Citation: Duque de Blas, G.; Gómez-Veiga, I.; García-Madruga, J.A. Arithmetic Word Problems Revisited: Cognitive Processes and Academic Performance in Secondary School. *Educ. Sci.* **2021**, *11*, 155. https://doi.org/10.3390/educsci11040155

Academic Editor: Liudmila Liutsko

Received: 21 February 2021
Accepted: 24 March 2021
Published: 30 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Background

"Understanding a problem and solving it is practically the same thing"—Rumelhart [1].

Mathematics is a type of language through which we construct models that represent the qualities and relationships of measurable reality. Considered the "language of science", the learning of its elementary rules is essential for the development of mathematical thought and constitutes one of the fundamental objectives of formal education. This learning process, not without its difficulties [2,3], often determines the academic and future employability of its students, see [4–8], something well known to students whose level is, in general terms, below the European average according to the Organisation for Economic Co-operation and Development (OECD) [9].

The present paper addresses the relationship that reading comprehension and reasoning processes in arithmetic word problems (AWPs) have with academic achievement in secondary school, focusing not only on classic abstract reasoning, but also on new cognitive measures of verbal deduction. We argue that deduction and reading comprehension are involved in arithmetic word problem solving and most of the complex learning tasks students commonly face at school and constitute an important higher-order cognitive ability that underlies academic achievement.

Different theoretical approaches have been proposed to explain the components and processes involved in AWP solving. Among the most important theories, schema-based

models [1,10], situational models [11,12], and contributions from the semantic field [13–16] stand out. This paper is focused on the schema and situational contributions, Kintsch and Greeno [10] and Kintsch's comprehension model [11], as well as dual process theories of thinking and reasoning [17,18]. Its aim is double. First, to analyze the relationship that AWPs have with reading comprehension and abstract reasoning. To this end, we deepen our understanding of cognitive processes involved in solving a particular kind of mathematical reasoning problem: consistent and inconsistent change problems with two or three arithmetic operations. The second objective is to check the capacity of students to solve these AWPs to predict academic performance in two basic and different subject matters: Mathematics and History/Geography.

Although some of the skills needed to solve these kinds of problems may be different from those used in solving geometric problems or those of probabilistic reasoning, it is possible to determine the existence of certain common features that underlie all mathematical problems. In particular, it is worth highlighting the implications of a kind of general cognitive aptitude based on abstract reasoning, fluid intelligence [19–23], that is identified with processes related to inductive reasoning (the search for rules or patterns that relate stimuli in order to establish more general conclusions), deductive reasoning (those that draw logical conclusions from a sequence of previous premises), and quantitative reasoning (those that involve concepts of a quantitative type, such as AWPs) [24].

1.2. Arithmetic Word Problems (AWPs)

AWPs must be solved by combining the numbers mentioned in the text using basic arithmetic operations (addition, subtraction, multiplication, and division). Solving AWPs is a complex task that demands individuals to activate the underlying working memory executive processes in order to inhibit non-relevant information, semantic biases, overlearned procedures (heuristics) or superficial responses induced by the statement of the problem [25–31].

There are a diversity of mistakes that students commit when solving arithmetic problem [32–35]. Two of them are (1) errors in the calculation required to find the solution (execution phase) and (2) errors derived from a poor understanding of the text and the relationships described. By studying these errors, we can delineate two key aspects regarding these kinds of problems. On the one hand, AWPs represent not only a demonstrative exercise of the calculation skills necessary, but also a way to test a student's textual comprehension and reasoning, as well as an associated measure of student academic performance. This is the main hypothesis considered in the present study. Secondly, given the variety of basic cognitive processes involved in their resolution, AWPs could be used as a screening tool to detect signals of possible learning difficulties in reading and Mathematics.

There is a wide variety of AWPs, which can be classified according to values they can adopt on a set of variables. We have tried to synthetize the most relevant ones in Table 1.

These criteria would affect to the global extension of AWP and the difficulty solving them. The present study was carried out with 5 "change" AWPs (see Materials section for more details), in which only the number of operations and the direction of verbal term used (consistent vs. inconsistent) were manipulated to modulate the difficulty of the task.

1.3. AWP and Reading Comprehension

AWPs have led to numerous studies and explanatory models regarding the processes involved in their resolution. Among the most outstanding is the work of Kintsch and Greeno [10] and Kintsch [11], who applied the notion of schema and the construction–integration model of text comprehension to solving AWPs. These authors showed that, in the same way calculation skills are indispensable for solving these problems, it is equally necessary to consider the different ways these problems can be understood in order to be able to determine the way success occurs. This observation came from the growing evidence that the most frequent errors committed in the resolution of AWPs resulted from

the inadequate construction of the mental representation of the problem [31,43], which could also be derived from semantic biases induced by the statement of the problem [15].

Table 1. Criteria and values used to classify arithmetic word problems (AWPs).

Criteria	Values
General types of AWP	Change, combine, compare or equalize problems [36]
Items named in the problem	
■ Number of elements	The more significant elements, the greater number of steps needed for solving.
■ Context of elements	Familiar or non-familiar for solver [37,38]
■ Semantic Align between elements	Symmetric or asymmetric align (e.g., content vs. container) [12]
Quantities related to elements	
■ Magnitude	Large or small value size [39]
■ Data type	Integers or decimal magnitudes [14]
■ How they are represented	Cardinals (unordered entities) or Ordinals (ordered entities) [40]
■ How they are expressed	Explicitly (e.g., x = 10) and/or relationally (e.g.,: y = x + 2) [40]
■ What they are connected with	Verbal terms: "More than", "Less than", "Equals to", "As many as" [36]
■ Sense of inserted verbal term	Consistent or inconsistent with the suggested operation [34,36]
Problem question	
■ Number of questions	From 1, to above
■ Location	At start or end of the problem [41]
■ Type of question	Referring to the overall outcome of any specific part, or to the whole of the parts involved
■ Demanding data	Numeric or qualitative data (comparative; e.g., "Who has more marbles?")
Operations	
■ Types	Addition, subtraction, multiplication and/or division [42]
■ Number or steps	From 1, to above

Along this same line, numerous authors [34,36,44,45] have studied the resolution of AWPs through the use of non-routine problems. The resolution of an AWP requires an in-depth understanding of what is said in the text in order to realize what the question

refers to, as well as to determine the meaning or direction of the verbal terms used to connect the objects with the quantities and to show the changes or variations narrated in the statement. Both are key conditions to allow the arithmetic formalization stage to be executed following the reliable sense of the relationships reported in the problem statement. However, participants might use two different types of resolution strategies on their way to reach it. The first strategy is the *direct strategy*, one best summarized in the words of Siegler and Jenkins [46]: "calculate first, think later." In the same line, other authors related a "compulsion to calculate" [47] due to children's prior experience with arithmetic [48]. This process would begin by identifying the numbers of each category with which participants, subsequently, would perform the arithmetic operations they deem necessary based on the literal relationships described within the problem. This strategy of direct translation often leads to mistakes in the resolution of problems due to two reasons or errors: (1) that of understanding, and (2) that of metacognition. The first is due to a poor understanding of what the problem represents, that is, by focusing more on its quantitative aspects than on its qualitative ones [49–51]. The second is due to an inadequate estimation of the problem's difficulty, perhaps as a consequence of identifying the problem as an ordinary exercise for which a previous model and resolution plan are already available. This direct strategy is, therefore, a strategy largely dominated by intuitive and automatic Type 1 processes, whereas Type 2 reflexive processes appear absent.

The second strategy for solving arithmetic problems is called the *problem model strategy*. It consists of several stages that, together, guide the participant toward the construction of a mental representation of the problem based mainly on its underlying qualitative aspects [14,50,52]. The first stage begins with understanding the premises of the problem, a process in which each local piece of information is transformed into propositions that represent the elements described in the text of the problem. The result of this transformation forms a latent semantic network known as a "text base", from which a situational model of this problem is then constructed in a second phase [11]. This model is the result of integrating the content of the text base with that of related content and experiences stored in long-term memory. In this way, the characteristics of the objects described in the text, as well as the key words of the problem, are integrated into differentiated entities, forming a mental representation that is clearly more elaborate than that resulting from a direct translation process [11]. In this phase, therefore, the ability to recover information available from memory and relate it to newly acquired data in working memory (WM) takes on special relevance. This phase requires a capacity to reason reflexively and inhibit any automatic responses that may be generated during the process, given that their absence could generate an inadequate representation of the problem.

The situational representation allows a participant to initiate a third phase whose objective is to elaborate a plan to solve the problem posed, determining with exactitude the arithmetical operations that are necessary given the relations between the different objects in the problem. This is an especially relevant component in AWPs, given that one of the elements by which problem difficulty is manipulated is through the use of adverbs of quantity, such as "more", "less" or "equal". It is these adverbs that interrelate the objects of the problem and allow people to infer the meaning of the established relationships. In this way, the operations that are necessary in the process of calculating the response to the problem—addition, subtraction, multiplication, and/or division—are determined. This demonstrates the close relationship between the stage of generating a plan for resolving the problem and the previous stage during which the text base within a situational model is transformed. That is, the adverbs of quantity within the problem statement cannot in themselves serve to identify the arithmetic calculations required for its resolution, a contrary interpretation being similarly possible, the result of which would consequently require the inhibition of any generated automatic response.

Regarding the problems used in this study, change problems are those in which the problem statement narrates an initial state in which the objects/subjects described are associated with some quantities, a transformation process that increases or decreases

the quantities described, and a final question that motivates the purpose of the problem and requires the calculation of final quantities. Looking at the sense of the verbal terms inserted, we can distinguish two kinds of compare problems: (1) problems with a coherent or consistent statement, when the adverb "more" or "less" coincides with the operation of addition or subtraction, respectively, that must be performed to enact resolution; and (2) non-coherent or inconsistent problems, when the adverb suggests performing an inverse arithmetic operation to the literal one. That is, performing subtraction in the presence of the term "more", or addition in the presence of the term "less". In this phase, a higher-order ability would mediate a resolution strategy/plan that establishes the sequences of operations necessary to respond to the problem posed. Ultimately, the resolution process ends with a planned execution stage during which one consciously supervises the process to minimize any calculation errors that may occur.

Although the same stages can be identified in both the "direct strategy" and the "strategy of the problem model", the fundamental difference between the two is found in the second phase, that is, in the way in which information is integrated from the text base. In the "direct strategy", this integration is limited and focuses on the numbers and keywords in the problem text. Conversely, in the other strategy, the situational model is oriented toward the objects of the problem, and this allows for the integration of its relevant characteristics as well as the sense of the relationships established within the problem statement.

In the arithmetic problems of our study, the verbal problem statement may contain two adverbs of quantity when describing the price of the objects in three different places. The description of a greater number of relationships between the terms necessarily involves an increase in the number of propositions in the text base that must be integrated, as well as a greater number of steps or operations that must be planned and executed in the resolution. There are, therefore, two types of problems based on this criterion: (1) problems of a single operation, in which a relationship requires the operation of addition or subtraction in order to solve it (e.g., "Juan has 5 cookbooks. Isabel has two books less than Juan. How many cookbooks does Isabel have?"); and (2) problems of two operations, which describe two chains of relationships that require partial calculations to obtain the final result (e.g., Isabel has two books less than Juan. Marta has three more books than Isabel. How many cookbooks does Marta have?") [53].

The number of relationships described in the problem and the consistency or non-consistency of the adverbs that provide information regarding their meaning are both variables of enormous experimental potential. This is due to the close relationship they have with basic cognitive processes and their high ability to modulate the difficulty of the task.

1.4. AWP and Reasoning

Neither intelligence nor mathematical thought are limited to deduction. However, this reasoning process has been correctly considered the essence of mathematical thought. Deduction implies a process of sequential reflection through a series of steps that lead to a conclusion. This conclusion can be considered the consequence of consolidating or representing a mental model of the premises of the deductive task [54]. The tendency to use deductive reflection, as opposed to intuition in the process of problem solving, constitutes for many theorists one of the fundamental differences between expert and novice arithmeticians [24]. Numerous studies have shown the differences between the intuitive and analytical modes of thought, giving rise to a dual theory of processing in which two types of systems are distinguished [17,18,55]. The first system, Type 1, is intuitive, guided by data, and is therefore automatic. It acts quickly, and only its final product is accessible to consciousness [52]. Thus, the particular modus operandi of the Type 1 system tends to provide answers based on heuristics [56,57] that reduce the complexity involved in evaluating possibilities as well as the need for more cognitive resources, a fact demonstrated by the low correlation this process has with measures of intelligence and

working memory (WM) [58]. The Type 2 system, on the other hand, is slow, costly both cognitively and motivationally, and guided by will and conscience: it is based on WM performance [17,55,59–62]. The fundamental characteristic of the Type 2 system is its ability to generate a cognitive decoupling in the process of solving problems by means of which the automatic responses generated by Type 1 are inhibited and the most characteristic abilities of the human being are enabled; that is to say, hypothetical deductive thought, or mental simulation executed via the generation of different mental models of the problem, and decision making itself [58,63]. The Type 2 system therefore requires an intensive use of WM executive functions, such that tasks which necessitate its use tend to correlate clearly with measures of WM and fluid intelligence.

Fletcher and Carruthers [64] propose three main characteristics that define Type 2 reasoning: (1) it is subject to intentional control, (2) it can be guided by normative beliefs about appropriate reasoning methods (see, [65]), and (3) it cancels out, or in its case confirms, the unreflective responses automatically generated by the Type 1 system. These characteristics describing the supervision of inferential processes are skills that occur later in development, emerge during preadolescence, and culminate their development in youth [66–68]. Before its acquisition, children lack the ability to activate the Type 2 system and control the interaction between both systems.

1.5. Aims

The purpose of this paper is to help to clarify the nature of mathematical thinking by analyzing the relationship between some of the cognitive variables that underlie it and to determine to what extent they are related to the academic achievement of Secondary Obligatory Education. To do this, we compare 5 arithmetic problems of increasing difficulty (consistent or inconsistent, and of either one or two arithmetic operations, plus a multiplication across all cases), as well as two measures of reasoning (Kaufman brief intelligence test (KBIT) [69] and a deductive reasoning test (DRT) [70]) and two of reading abilities (reading processes assessment battery (PROLEC-SE) [71] and The Spelling test for secondary school (Spelling-SE) [72]). A detailed description of the problems used can be seen in Table 2 and in the Supplementary Materials.

Table 2. Description of problems in terms of number of consistent or inconsistent operations (working memory load), inconsistency, and need of inhibition of superficial responses.

	WM Load Number of Operations (Add/Subtract + Mult.)	Inconsistency	Superficial Responses
Problem-1	1 + 1	No	No
Problem-2	1 + 1	Yes	Yes
Problem-3	2 + 1	No–No	No
Problem-4	2 + 1	No–Yes	Yes
Problem-5	2 + 1	Yes–Yes	Yes

The first two problems require two operations, one addition or subtraction, and a multiplication. Problem 2 differs in that the operation that must be applied (addition or subtraction) is inconsistent with the adverb of quantity in the literal statement ("less" or "more", respectively). Consequently, Problem 2 will require the inhibition of a surface representation in the text base created from the literal statement. When this inhibition does not happen, it will produce a superficial response that we could define resulting from the correct performance of a mistaken plan. This inhibition is needed in order to construct a representation that allows the participant to apply the inverse arithmetic operation of that indicated by the adverb of quantity [31]. Apart from correct and superficial responses, any other response is also erroneous and involves either arithmetical mistakes or erroneous answers probably produced by a low commitment of the participant to the task. Problems 3, 4, and 5 are problems of three arithmetic operations, because they describe two relationships

between the terms involved, which implies a greater demand on WM resources, and a final multiplication. These problems involve two addition or subtraction operations and may be consistent (Problem 3), require a consistent and an inconsistent operation (Problem 4), or require two inconsistent operations (problem 5). Consistent problems have no superficial responses, inconsistent problems 2 and 4 have one possible superficial response, while inconsistent problem 5 has three possible superficial responses: in the first operation, in the second or in both.

The number of operations is a source of problem difficulty [34,73,74], and it would be necessary to clarify whether the increase in the number of operations implies any qualitative increase in the problem difficulty, and what kind of mistakes are the most common, superficial or other error responses. In all cases, the same response schema is applied, with each having one additional operation, involving just a temporary increase in WM's updating process required to solve the problem. In the same way, the final multiplication required of all problems is nothing less than trivial, the calculation of which, in principle, poses few difficulties. In this way, Problem 3 should be more difficult than Problem 1, and Problem 4 more difficult than Problem 2.

When solving an AWP, we consider four possible scenarios as a response to these problems, which define the three types of mistakes that solvers would incur. First, the ideal scenario consists of correctly solving a correct plan; second, when solvers determine correctly the arithmetic embedded in the problem, but makes computational mistakes, we say that they are wrongly solving a correct plan; the reverse case, when a solver is able to compute correctly the operations, but the type of operations are wrongly determined, we says the solver is solving correctly a wrong plan; and last, when they are unable to determine the arithmetic of the problem and its required computation, solvers are wrongly solving a wrong drawn plan.

In terms of cognitive abilities, these mistakes can be caused, fundamentally, by executive functions' poor combined performance, in particular, the inhibition of superficial responses and the updating of base of text along the whole process. Thus, we might identify four kinds of solvers' patterns of performance, whose outcomes can be also identified by a particular scenario and kind of response (see Table 3).

Table 3. Solvers profiles attending their Executive Functions (EF) performance, scenario, and outcomes.

Profiles	EF Performance		Scenarios		Response
	Inhibition	Updating	Plan	Execution	
1	Good	Good	Correct	Correct	Correct
2	Good	Bad	Correct	Incorrect	Error
3	Bad	Good	Incorrect	Correct	Superficial
4	Bad	Bad	Incorrect	Incorrect	Error

An additional profile can be considered when participants are not willing to face the task. In that case, results would be contaminated by a higher number of errors due to a randomized response.

If our conception is correct, the main difficulty with the problems comes from the process of reflection and inhibition involved in solving their non-consistency, as well as the need to update the base of text with the partial results of each problem according to a new interpretation of the relational magnitudes. Given the scenarios described above and a good participant engagement with the task, it seems reasonable to think that there is a higher probability of finding "errors" as a response (scenarios 2 or 4) than "superficial" responses (3), due to the greater number of them (double) in which an "error" response can be recorded as an outcome. However, we do not know whether such a claim can be empirically substantiated. In any case, the increase in the difficulty in the inconsistent compared to the consistent problems is an effect that has been confirmed by Verschaffel [75], who in his study of 10–11-year-old children found a 71% success rate in the resolution of inconsistent problems, versus 82% in consistent problems. Likewise, Lewis and Mayer [45]

and Hegarty et al. [34] confirmed in adults that the non-consistency of the statements caused a greater number of reversibility errors with two-step problems.

Our hypotheses were:

(1) We expect significant differences in AWP performance between two grades, 2nd and 3rd grade of Compulsory Secondary School courses.
(2) We also expect differences as a function of type of problem. Inconsistent AWPs should be significantly more difficult than consistent problems. In the same way, the arithmetic problems with two add/subtract operations should be significantly more difficult than those of a single operation. The non-consistency effect of the problems should increase the difficulty more than that of the number of operations.
(3) The cognitive measures of reasoning and linguistic comprehension should correlate significantly and moderately with measures of effectiveness in the resolution of AWPs. The correlations should be higher between the cognitive measures and those of the inconsistent problems and the problems of two operations, due to their greater cognitive demand, than with the consistent problems and those of a single operation. All cognitive measures, including the resolution of arithmetic problems, should correlate positively with academic performance.
(4) Given the previous relationship, our aim was to confirm the capacity of association of the cognitive variables measured (reasoning and comprehension) on the score obtained in solving arithmetic problems. Likewise, the various cognitive variables and, in particular, the global arithmetic problems' score, should show their capacity for association with academic performance in History/Geography and Mathematics.

2. Materials and Methods

2.1. Participants

This study was carried out with a sample of 135 native Spanish speaking students (53 female) of compulsory secondary education, without learning difficulties, of a public institute (Madrid), of which 70 2nd grade students (24 female) and 65 3rd grade students (29 female) voluntarily participated. The average age in the 2nd grade was 13.10 (SD = 0.85) and in the 3rd grade 14.18 (SD = 0.80).

2.2. Task and Measures

2.2.1. Arithmetic Word Problems Task (AWP)

To measure the ability to solve arithmetic problems, we used a new test based on the problems of Hegarty et al. [34]. It is formed by 5 AWPs: two consistent and three inconsistent. The problems involve elementary arithmetic calculations consisting of addition or subtraction, as well as multiplication, and varying their number in each problem (one addition or subtraction operation in problems 1 and 2, two addition or subtraction operations in problems 3, 4 and 5, plus a final multiplication operation across all problems). In order to control the possible semantic biases derived from the statement, the elements within the same problem were always the same (semantic align), but varying some of their characteristics: origin, size, color, material, price, and commercial areas where they can be acquired. The question of each problem was always related to the global price that solvers must pay (cardinality) for a specific quantity of objects in a particular store, so we can consider this context as familiar to solvers. The quantities described in the statement were some explicit and some relational. The problem's questions were always located at the end of the statement. The result of the operations necessary to calculate the requested response were always integers and never bigger than three digits. Correct and error responses were collected in all problems, as well as superficial responses for inconsistent problems. Two parallel versions of the problems were used, where the adverbs of quantity "more" and "less" were reversed, as were the place and object names. See the Supplementary Materials for an example of the 5 problems employed. Its measurement range is between 0 and 5. Cronbach's alpha for internal consistency was 0.75.

2.2.2. Reading Processes Assessment Battery (PROLEC-SE)

The participants completed the task number five of Reading processes assessment battery [71]. This evaluates reading processes for secondary students and consist of two expository texts. Questions are divided in two kinds: those with enquiries regarding literal aspects of text and those which require making inferences. Their measurement range is between 0 and 20. Cronbach's alpha for internal consistency was 0.74.

2.2.3. Kaufman Brief Intelligence Test (KBIT)

Intelligence was evaluated through the matrix subtest of the Kaufman Intelligence Brief Test [69], in its Spanish version. This test provides a measure of abstract and visuospatial reasoning, and fluid (non-verbal) intelligence. Its measurement range is between 0 and 48. Cronbach's alpha for internal consistency was 0.81.

2.2.4. Deductive Reasoning Test Simplified (DRTs)

The DRTs is a simplified version of the DRT [70] adapted to suit preadolescent participants (12–15 years old). It consists of nine propositional and syllogistic deductive and meta-deductive reasoning problems, divided into four types of inference problems: propositional deductive, propositional meta-deductive, syllogistic deductive, and syllogistic meta-deductive. The propositional deductive inferences include two inclusive disjunction problems (one affirmative and one negative). Participants must evaluate the possible conclusions of these two inferences. Both require the construction of multiple models. The propositional meta-deductive inferences include three truth-table problems in which participants have to analyze the consistency of three problems, each consisting of a conditional statement and an assertion. In the first problem, the assertion matched the first initial model of the conditional. The second problem requires a participant to construct the second conditional model in which antecedent and consequent are negated. Finally, the last problem requires the construction of the third and most difficult conditional model in which the antecedent is negated but the consequent is affirmed. The syllogistic deductive task requires participants to generate and write the solution to one categorical syllogism. Categorical syllogisms include the combinations of the four kinds of premises: universal affirmative, universal negative, the particular affirmative, and particular negative. Categorical syllogisms can be very difficult, often too difficult for preadolescents. Thus, we used only an easy, single model categorical syllogism. Finally, in the syllogistic meta-deductive necessity/possibility task, reasoners have to decide whether a given conclusion in three syllogistic problems is necessarily true, possible, or impossible. Its measurement range is between 0 and 9. Cronbach's alpha for internal consistency was 0.73.

2.2.5. Spelling-SE for Secondary School

The spelling test for secondary school [72] is a lexical choice test in which participants have 40 words represented orthographically, each next to two pseudo-homophones. For example: abeja, abega, aveja ("abeja" means "bee"—the other two are meaningless words but would be pronounced very similarly). The objective is to select the correct orthographic form for each word. This test is based on the spelling subtest included in the Reading Assessment Battery (BEL, [76]). Distinct levels of usage frequency (high, medium, and low), described in Lexesp—a computerized lexicon of Spanish [77]—were taken into account when selecting the words contained in this test. Its measurement range is between 0 and 40. Cronbach's alpha for internal consistency was 0.76.

2.2.6. Academic Achievement

The final grades in Mathematics and History/Geography (from 1 to 10) at the end of the school year were taken as a measure of academic achievement. As for the latter, it is a discipline whose study requires the student to think critically about events, as well as to contextualize and relativize them in terms of past culture and social organizations in a way that allows drawing reasonable conclusions after evaluating alternative points

of view. The growing effort to promote such critical reflection skills in history teaching is remarkable [78–81] because they can be used as a predictor of academic performance. Moreover, participation of an adolescent sample seemed particularly relevant, as reasoning skills are increasingly important during this developmental period [82], when learning activities become more complex. The final scores are the average result obtained during the three trimesters making up the school year. These scores reference the aptitude and behavioral characteristics of the students.

2.3. Procedure

Each of the two symmetrical versions of the AWPs was randomly assigned to each of the participants. The participants had to solve the tasks in two sessions distributed over a week. The order of test presentation for the first session was PROLEC-SE, DRTs, and arithmetic, while in the second session, they performed KBIT and Spelling-SE, in that order. All the tests were completed digitally via online forms in the computer room during school hours. The protocols of this study were approved by the ethics committee of the National University of Distance Education (UNED).

2.4. Data Analyses

The analyses were carried out using the statistical package for social sciences, SPSS v25. A winsorization procedure [83,84] on global and partial scales of the AWP task, as well as on the global scores of the tasks, was performed in order to control the outliers. Descriptive results on the tasks, ANOVAs, bivariate Pearson correlation coefficient and hierarchical regressions among cognitive measures, AWPs, and academic achievement were run in order to test the hypothesis regarding the relationship between the cognitive measures with the performance on AWP, and the power of association of these variables on the academic achievement.

3. Results

3.1. Descriptive Statistics and Comparisons

The first analysis aimed to determine if there was any difference between versions (more or less) of AWPs. The results confirmed the non-existence of statistically significant differences as a function of which arithmetic test version was carried out (F (1, 133) = 0.365, p = 0.547).

The results of the tasks, including mean number of correct responses, standard deviation, range, number of participants, and differences between scholar grades, are shown in Table 4. The most difficult task in terms of percentage of correct responses was the AWP task.

Table 4. Global descriptive statistics (mean of number of correct responses, range, standard deviation, and mean differences between 2nd and 3rd grades).

N = 135	Mean	Range	Percentage	SD	Difference between 2nd (N = 56) and 3rd (N = 57)			
					M (SD)–M (SD)	*F*	*p*	*Cohen's d*
AWP-Global	2.22	0–5	44	1.47	1.66 (1.18)–2.83 (1.52)	25.41	0.001	0.80
PROLEC-SE	11.18	0–20	56	3.89	10.01 (4.38)–12.43 (2.81)	14.3	0.001	0.62
Spelling-SE	30.74	0–40	77	2.68	30.34 (2.68)–31.17 (2.64)	3.26	0.073	0.31
KBIT	35.10	0–48	73	4.55	34.24 (4.91)–36.02 (3.96)	5.28	0.023	0.39
DRTs-Global	4.74	0–9	53	1.37	4.6 (1.36)–4.89 (1.38)	1.54	0.217	0.21

Likewise, we can observe that the differences in mean correct response between the two academic years were statistically significant, except for the DRTs and Spelling-SE. Results in KBIT and PROLEC-SE were as expected for their age and educational level.

The descriptive statistics and results obtained for each of the AWPs, including mean of correct and superficial responses for each problem, can be seen in detail in Table 5. The results showed the expected pattern, depending on the number of operations involved

and the consistency or non-consistency of the verbal statement. Likewise, results confirm an increase in the number of correct answers in 3rd grade students compared to that of students in the 2nd grade. The results showed a high standard deviation on each problem, typical of a categorical variable where hit or miss was recorded, and similar to other studies on AWP.

Table 5. Global descriptive statistics (range, mean, standard deviation) of arithmetic word problem (AWP) measures (correct, superficial, and error responses).

	Range	2nd (N = 70)	3rd (N = 65)	Total (N = 135)
		M (SD)	M (SD)	M (SD)
AWP-1	0–1	0.63 (0.49)	0.82 (0.39)	0.72 (0.45)
AWP-2	0–1	0.27 (0.45)	0.45 (0.50)	0.36 (0.48)
AWP-2-sup.	0–1	0.17 (0.38)	0.22 (0.41)	0.19 (0.40)
AWP-2-error	0–1	0.56 (0.50)	0.34 (0.48)	0.45 (0.50)
AWP-3	0–1	0.41 (0.50)	0.66 (0.48)	0.53 (0.50)
AWP-4	0–1	0.24 (0.43)	0.52 (0.50)	0.38 (0.49)
AWP-4-sup.	0–1	0.10 (0.30)	0.11 (0.31)	0.10 (0.31)
AWP-4-error	0–1	0.66 (0.48)	0.37 (0.49)	0.52 (0.50)
AWP-5	0–1	0.20 (0.40)	0.38 (0.49)	0.29 (0.46)
AWP-5-sup. (any)	0–1	0.09 (0.28)	0.12 (0.33)	0.10 (0.31)
AWP-5-error	0–1	0.71 (0.46)	0.49 (0.50)	0.61 (0.49)
AWP-one-op.	0–2	0.9 (0.73)	1.26 (0.64)	1.07 (0.71)
AWP-two-op.	0–3	0.77 (0.82)	1.57 (1.07)	1.16 (1.03)
AWP-Consistent	0–2	1.04 (0.73)	1.48 (0.66)	1.25 (0.73)
AWP-Inconsistent	0–3	0.61 (0.79)	1.35 (1.19)	0.97 (1.06)
AWP-Superficial	0–3	0.36 (0.61)	0.45 (0.73)	0.40 (0.67)
AWP-Error	0–3	2.03 (1.60)	1.20 (1.08)	1.63 (1.27)
AWP-Global	0–5	1.66 (1.18)	2.83 (1.52)	2.22 (1.47)

Sup.: superficial response. (any): all superficial responses (first, second or both operations).

A mixed ANOVA for repeated measures was performed in order to check the differences among the problems' difficulty considering, also, the students' grades. The five AWPs were englobed in a factor called "difficulty of problems", while school grade was introduced as an intersubject factor. Results confirmed an overall significant difference in the difficulty of the problems (F (3.6, 135) = 24.12, MSE = 4.39, $p < 0.001$, $\eta 2p = 0.15$), as well as a significant difference in performance regarding the student's grade, easier for 3rd grade than 2nd grade (F (1, 135) = 18.86, MSE = 7.77, $p < 0.001$, $\eta 2p = 0.24$). However, factors did not interact significantly (F (3.6, 135) = 0.44, MSE = 0.08, $p = 0.764$). A post-hoc analysis of the problems (Table 6) identified differences in difficulty among all the problems (Bonferroni $p < 0.05$) except when comparisons were between AWP-2 (one inconsistent operation) and AWP-4 (consistent then inconsistent operations), AWP-2 and AWP-5 (two inconsistent operations), and AWP-4 and AWP-5; that is, there were no significant differences among inconsistent problems in terms of the number of operations.

Table 6. Post-hoc (Bonferroni) for performance comparison in AWP tasks.

(I) AWP	(J) AWP	Mean Differences (I-J)	Dev. Error	Sig.	*Cohen's d*
1	2	0.363	0.052	<0.001	0.77
	3	0.184	0.053	0.007	0.40
	4	0.339	0.051	<0.001	0.72
	5	0.430	0.051	<0.001	0.94
2	3	−0.179	0.057	0.019	0.35
	4	−0.024	0.044	>0.05	
	5	0.066	0.045	>0.05	
3	4	0.155	0.052	0.032	0.30
	5	0.246	0.051	<0.001	0.50
4	5	0.091	0.039	>0.05	

Comparison between AWP-1 with AWP-2 problems revealed the differences in performance between one consistent operation and one inconsistent operation problems. This result could also be confirmed with the comparison between the problems with two operations. Thus, as expected, a consistent two-operation problem (AWP-3) was less difficult to solve than an inconsistent two-operation problem (AWP-5). Regarding the number of operations, we confirmed also significant differences in difficulty between one operation problems and two operations problems, only when they were consistent operations (AWP-1 with AWP-3).

Regarding mistakes, we can observe that the proportion of superficial answers in inconsistent problems 2, 4, and 5 is always lower than the proportion of other erroneous responses in each course. Likewise, the decrease in age and school level for the other erroneous responses, but not the superficial ones, is remarkable (F (9, 135) = 2.06, MSE = 0.48, p = 0.038, $\eta 2p$ = 0.13). Additionally, we looked at the differences among superficial responses in problem 5. An ANOVA for repeated measures revealed non-differences regarding the probability to commit a superficial mistake in the first step/operation of the problem (superficial 1), or in the second step/operation of the problem (superficial 2) or both (F (1.8, 135) = 0.99, MSE = 0.04, p = 0.366).

3.2. Interrelationships among Variables

All of the intercorrelations between the four cognitive measures are significant, even if slight (see Table 7). The correlations between the global mean for arithmetic problems and that of the cognitive variables were significant. A clearly increasing pattern of correlations between the cognitive measures and the one- and two-operation problems can be observed, as well as between the consistent and inconsistent problems, respectively. The correlations are higher in the two-operation and inconsistent measures—for example, with KBIT and the two-operation measures, as opposed to one-operation, and with inconsistent problems versus consistent ones. We can find a similar effect with PROLEC-SE and two-operation problems, as opposed to one-operation, and inconsistent versus consistent.

Regarding academic performance, Spelling-SE and KBIT showed slight correlational relationships with performance on History/Geography, whereas DRTs did not reach significance with academic performance. Academic performance on History/Geography significantly correlated with the diverse AWP measures, except consistent problems; particularly, it correlated with inconsistent problems and the global mean measure. As for Mathematics, the highest correlations were found with the global measure of arithmetic problems and the fluid intelligence measure of KBIT. Academic achievement in Mathematics and the AWP measures showed a relevant relationship, especially on the global measure and the inconsistent problems. In addition, it was confirmed that Arithmetic problems require both comprehension and reasoning abilities in the same mathematical domain, and KBIT requires an abstract visuospatial reasoning ability also demanded in mathematical learning.

Table 7. Pearson correlations among academic achievement (History/Geography, Mathematics), reading comprehension (PROLEC-SE), reading decoding processes (Spelling-SE), visuospatial reasoning (KBIT), Deductive Reasoning Test simplified (DRTs), and arithmetical word problems (AWPs; one and two-operation problems, consistent and inconsistent problems, and global score).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. History-Geo	1	0.61 **	0.33 **	0.32 **	0.22 **	0.04	0.26 **	0.27 **	0.14	0.35 **	0.33 **
2. Mathematics		1	0.30 **	0.19 *	0.39 **	0.22 **	0.34 **	0.33 **	0.24 **	0.36 **	0.38 **
3. PROLEC-SE			1	0.26 **	0.38 **	0.23**	0.20 **	0.35 **	0.24 **	0.32 **	0.35 **
4. Spelling-SE				1	0.17 *	0.11	0.19 *	0.23 **	0.20 **	0.23 **	0.27 **
5. KBIT					1	0.20 *	0.22 **	0.35 **	0.19 *	0.34 **	0.34 **
6. DRTs						1	0.21 **	0.21 **	0.08	0.27 **	0.24 **
7. AWP-one-op.							1	0.42 **	0.58 **	0.65 **	0.76 **
8. AWP-two-op.								1	0.61 **	0.82 **	0.90 **
9. AWP-Consistent									1	0.32 **	0.73 **
10. AWP-Inconsistent										1	0.88 **
11. AWP-Global											1

** $p < 0.01$; * $p < 0.05$. PROLEC-SE: reading processes battery test; Spelling-SE: spelling test for secondary education; KBIT: Kaufman Brief Intelligence Test; DRTs: Deductive Reasoning Test simplified; AWP: arithmetic word problem.

Once the relationship between cognitive measures and academic achievement on the sample had been confirmed, a linear regression analysis was run to check the power of association of the cognitive measures on AWP (see Table 8). Results confirmed that cognitive tasks were able to explain about 20% of the variance of the performance in the AWP-Global in this sample. KBIT, PROLEC-SE, and Spelling-SE, in order of higher to lower explanatory capacity, were the significant variables included in the model. DRTs was not significant.

Table 8. Multiple linear regression among AWPs, cognitive measures (Spelling-SE, PROLEC-SE, AWP Global or AWP Inconsistent, DRTs and KBIT), and academic achievement in secondary school (hierarchical regression).

Dependent Variable	R^2	R^2 Adj	F	B	Beta	
AWP Global	0.22	0.20	9.101 ***			
KBIT				0.07	0.22	$t = 2.56$, $p = 0.012$
PROLEC-SE				0.07	0.19	$t = 2.24$, $p = 0.027$
DRTs				0.14	0.13	$t = 1.64$, $p = 0.103$
Spelling-SE				0.09	0.17	$t = 2.10$, $p = 0.038$
History & Geography	0.23	0.20	7.774 ***			
KBIT				0.02	0.04	$t = 0.42$, $p = 0.673$
PROLEC-SE				0.11	0.21	$t = 2.34$, $p = 0.021$
WP-Inconsistent				0.27	0.25	$t = 2.92$, $p = 0.004$
DRTs				−0.16	−0.11	$t = -1.30$, $p = 0.195$
Spelling-SE				0.16	0.22	$t = 2.68$, $p = 0.008$
Mathematics Achievement	0.24	0.21	8.261 ***			
KBIT				0.11	0.25	$t = 2.93$, $p = 0.004$
PROLEC-SE				0.04	0.09	$t = 1.01$, $p = 0.314$
AWP Global				0.31	0.23	$t = 2.67$, $p = 0.009$
DRTs				0.13	0.09	$t = 1.15$, $p = 0.254$
Spelling-SE				0.04	0.05	$t = 0.61$, $p = 0.541$

*** $p < 0.001$.

Similarly, a new hierarchical regression analysis allowed determining which variables were associated with academic performance. Regression analysis on History/Geography was performed by including the previous cognitive measures with the inconsistent AWPs. The results indicated that the model was able to explain the 20% of the variance in History/Geography, of which AWP-Inconsistent, Spelling-SE, and PROLEC-SE, in order of higher to lower explanatory power, were significant. KBIT and DRTs were not significant. When looking at academic performance in Mathematics, the model explained 21% of the variance in the Mathematics performance for both secondary grades levels; only KBIT and AWP-Global, in this order, were significant.

4. Discussion

The goal of this work was to dig more deeply into the relationship between reasoning and reading comprehension processes in AWPs and their relationship to academic performance. The results confirm our first hypothesis: We found a reliable increase in the number of correct responses between 2nd and 3rd grade participants. Participants in the 2nd year of secondary school experienced greater difficulty solving the problems than did those of 3rd, who obtained a number of hits more than double that of younger students across the various problems posed. As expected, adolescents are in the process of acquiring necessary curriculum knowledge in diverse subjects at school while developing basic cognitive and formal thinking abilities, as well as reading comprehension abilities.

As has been shown in other studies [34,53,73], the number of calculation operations influences the difficulty in solving AWPs. This is because the increase in the number of steps or operations required in solving the problem involves constructing a text base with a greater number of propositions and a more complex situational model. The process of understanding two-operation problems is therefore quantitatively more complex than those of one consistent operation. This process requires, therefore, a greater number of

resources dedicated to updating and integrating any new contents that constitute the situational representation of the problem [29,85,86]. However, contrary to our predictions, the effect on the difficulty generated by two-operation problems disappears in the presence of inconsistences, reinforcing the idea that the difficulty core falls on the way solvers interpret the key elements inserted on the problems. That is why Mathematics instruction for problem solving must be focused not only on computational procedures (quantitative skills), but also on reasoning abilities (qualitative skills) [87].

Our data also confirm the importance of comprehension processes that determine the arithmetic operations that must be applied when resolving AWPs. Thus, the inconsistent problems, in which the linguistic comparative does not coincide with the arithmetic operation necessary for its resolution, are significantly more difficult to solve than problems in which the linguistic expression and the operation coincide. This increase in difficulty is related to the need to inhibit the propositional microstructure referring to the comparative and replacing it with the inverse term. This requires explicit reflection during the process of constructing the situational model of the problem. The greater the number of non-consistencies contained in the problem statement, the greater the number of interpretation errors that can lead to an incorrect arithmetic formalization and, therefore, errors in calculation during the resolution phase. The results of our study are similar to those found in other studies on consistent and inconsistent AWPs [34,44,45,75,88,89]. Inconsistent problems require a greater number of sentence readings and, therefore, greater dedication in terms of time and resources to be able to integrate the propositions relating the elements of the text, and to inhibit the initial representational models that arise from a superficial reading of the statements. As other studies have pointed out [31,85,86,90], a deficit in inhibitory control capacity during the process of solving arithmetic problems predisposes a person to adopt inadequate strategies for solving arithmetic problems.

We found that inconsistent one-operation problems were significantly more difficult to solve than the same consistent one-operation problems, and even more than consistent two-operation problems. These results confirm the importance of inhibition and updating processes in reasoning [18,31,58,63,64]. Inconsistent problems require, unlike consistent problems, the resolver's ability to suppress one's intuitive, automatic, and superficial response tendency that is guided by the literal representation of utterances, and which leads to the construction of an erroneous situational representation.

The analysis of mistakes allows us to confirm that the most common are not superficial errors, defined as the result of the right performance of a mistaken plan, but other error responses. We also observe that the number of superficial responses is low and shows a stable ratio in both ages. As a matter of fact, correct responses increase in the same proportion that error responses decrease, whereas superficial responses keep constant. Other errors constitute a set of responses formed by calculation errors and other wrong answers due to various factors, such as a low commitment of the participants with the task, and the inability to understand the statement of the problem and difficulties in updating the base of text. It will be very interesting in future studies to check participants' metacognitive assessment of problems difficulty.

Regarding superficial responses in inconsistent problems, they were never higher than error responses, which were always the most frequent answer in the whole sample. Although there is a significant increase in response accuracy between grades, this result is evidence that even in 14-year-olds, there is a poor performance in the skills needed to inhibit superficial responses and those that allow updating of the base of text content as the problem-solving process takes place. Given the type of response recorded by the participants, it is not possible to determine which of the two factors, inhibition or updating, contributed more to the performance produced. However, recent studies would point to a greater effect of updating as a main cause of poor performance for solving these problems [89].

With regard to our second hypothesis, we have confirmed the relationship between reasoning (KBIT and DRTs) and reading comprehension (PROLEC-SE and Spelling-SE)

measures with arithmetic problems. We can observe, also, that the highest correlations are found in two-operation and inconsistent problems, as opposed to single operation and consistent problems. This confirms our prediction about the relationship between the degree of complexity of a task and the processing requirements necessary to carry it out successfully. The inconsistent problems require reflection and the controlled application of WM executive processes. That is, they require the activation of Type 2 reasoning processes. The processes of reflexive reasoning when solving inconsistent arithmetic problems take on a special importance, and this is revealed in their relationship with measures of deductive (DRTs) and visuospatial (KBIT) reasoning, both clearly superior with inconsistent problems than with consistent ones.

The relationships between cognitive measures, arithmetic problems, and academic performance are especially evident in the results obtained through regression analysis, which allowed us to establish the associative capacity of the various variables (hypothesis 4). Thus, we verified that the reasoning and text comprehension measures explain about 20% of the variance toward the effective resolution of AWPs. The measure with the most explanatory capacity was intelligence with KBIT, followed by the reading comprehension test PROLEC-SE. These data confirm the importance of the cognitive processes of reading comprehension and reasoning underlying the generation of situational models which enables the discovery of correct answers in arithmetic problems.

On the other hand, we found that AWPs show a significant ability to explain academic performance not only in Mathematics, but also in History/Geography. The measures of performance in History/Geography were Spelling-SE, PROLEC-SE, and AWP-Inconsistent; that is, two measures of reading abilities, one of superficial decoding and the other of comprehension, and a measure of deep comprehension and reasoning in arithmetic. These results confirm the basic linguistic nature of this subject and also show the sensitivity and usefulness of AWPs to investigate in areas of knowledge not directly related to Mathematics.

The most significant measure in terms of explaining performance in Mathematics was KBIT, followed by AWP-Global—that is, a measure of abstract reasoning and a measure of comprehension, reasoning, and calculation in Arithmetic. This KBIT beta value is somewhat lower than other studies on the predictive capacity of Mathematics performance in primary school [91], and similar to others in secondary education using RAVEN as an intelligence measure (β = 0.26; [92]).

A result deserving some comment is the lack of a significant involvement of DRTs in both regression models. In fact, DRTs correlations are low with all variables. The highest DRTs correlation is with AWP-Inconsistent (r = 0.27). Its correlation with Mathematics performance is significant but low, and with History/Geography is very low and not significant. A likely related result is the lack of reliable developmental differences between 2nd and 3rd grade students in DRTs: Deductive problems are equally difficult for students in both grades. These results pose some doubts regarding the use of this simplified version of DRT to assess preadolescents' linguistic reasoning.

On the whole, regression analysis results account for the diverse but related nature of the academic subjects, as well as the function of cognitive abilities, different in History/Geography than in Mathematics, required to be successful in both subjects. It is important to emphasize that the AWP measures were the unique ones present significantly in the regressions of both academic achievement measures. This interesting result corroborates the capability and usefulness of arithmetic word problems at school.

A possible limitation of our work arises from the use of teacher evaluations as a measure of academic performance. Being a summary of student academic achievement, it may prove a biased measure when reporting progress made by the students during the school year, thus compromising its validity and reliability [93]. However, the numerical scores the educational center provided were individualized data obtained by students in each of the subjects (History/Geography and Mathematics) as a result of the achievements obtained over a prolonged period in time (i.e., an entire school year), and which included other characteristics of the student themselves, such as their motivation regarding learning

or their behavior in the classroom. For this reason, we consider it a good measure of academic performance [94], something that we can also confirm based on its widespread use in other studies [92,95].

There were some more limitations in this study. One of them stems from the number of problems the test contains. This measure was designed to be a short screening test. However, it would be advisable to consider increasing the number of problems in future studies in order to give greater statistical strength to the measure provided by this test. On the other hand, in problem number 4 (three operations and one inconsistency), we did not take into account whether the performance of the participants could vary due to the fact that the inconsistency was located in the second operation of the three necessaries to solve it, instead of the first as in the present study. Therefore, these limitations should be taken into account not only for future research projects, but also for the educational implications they might have for student learning.

Concerning educational implications, the different stages of the resolution process involve the need for relevant updating and inhibition resources. The results of this study invite us to consider the AWP task as a valid and apparently simple instrument to measure comprehension and reasoning aptitudes of secondary school students as a valid screening tool, particularly in the field of Mathematics. However, for this purpose, it would be necessary to clarify in detail the role played by working memory (WM) and the executive functions that arbitrate this kind of task and introduce in future studies specific measures of updating, inhibition, and WM span. These measures have frequently been related to academic success and its associative capacity in both Mathematics and other subjects [91,96–98].

As we have demonstrated, the intelligence measure KBIT provides a relevant but limited explanatory role, and its explanatory capability is similar to the arithmetic problems in Mathematics, and not reliable in History/Geography. This fact holds special relevance, serving as an example to show that, apart from other contextual, motivational, and emotional considerations, the academic success of students depends on cognitive processes that are not reduced to classic fluid intelligence. Our AWP task might be included in the educational world as a particularly useful and complementary tool for measuring general capacity or screening learning difficulties. Its use would detect specific deficits in the cognitive abilities involving comprehension and reasoning.

5. Conclusions

The resolution of AWPs continues to be an educational milestone that must be acquired throughout a student's academic life. In addition to being an educational mechanism for the training and development of mathematical skills, AWPs stand out as an excellent tool with enormous experimental potential due to the high and diverse number of cognitive skills involved in solving them. This paper contributes to showing some of them. Another aim of this work was to demonstrate how these problems could also be used as predictors of academic performance, not only in closely related subjects such as Mathematics, but also in other apparently distant subjects such as History/Geography. However, it would be interesting to explore other psychoeducational applications as a measure for detecting learning difficulties in order to prevent students' academic failure.

Supplementary Materials: A detailed description of AWPs is provided in this section. It contains the problem statements, arithmetic to solve them, the number of operations, the inconsistences involved and the possible superficial responses. Details are available online at https://www.mdpi.com/article/10.3390/educsci11040155/s1.

Author Contributions: Conceptualization, G.D.d.B., I.G.-V. and J.A.G.-M.; methodology, G.D.d.B., I.G.-V. and J.A.G.-M.; software, G.D.d.B., I.G.-V. and J.A.G.-M.; validation, G.D.d.B., I.G.-V. and J.A.G.-M.; formal analysis, G.D.d.B., I.G.-V. and J.A.G.-M.; investigation, G.D.d.B., I.G.-V. and J.A.G.-M.; resources, G.D.d.B., I.G.-V. and J.A.G.-M.; data curation, G.D.d.B., I.G.-V. and J.A.G.-M.; writing—original draft preparation, G.D.d.B., I.G.-V. and J.A.G.-M.; writing—review and editing, G.D.d.B., I.G.-V. and J.A.G.-M.; visualization, G.D.d.B., I.G.-V. and J.A.G.-M.; supervision, G.D.d.B., I.G.-V. and

J.A.G.-M.; project administration, I.G.-V. and J.A.G.-M.; funding acquisition, I.G.-V. and J.A.G.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was conducted with financial support of the research project EDU2014-56423-R from the Spanish Ministry of Economy and Competitiveness and the project PID2019-110476RB-I00 from the Spanish Ministry of Science and Innovation.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the National University of Distance Education (UNED).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data will be available upon reasonable request to the corresponding authors.

Acknowledgments: We appreciate the collaboration of our participants for the commitment shown with our research and the interest of their teachers in seeking improvements in the teaching of their students.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Rumelhart, D.E. Schemata: The building blocks of cognition. In *Theoretical Issues in Reading Comprehension: Perspectives from Cognitive Psychology, Linguistics, Artificial Intelligence and Education*; Spiro, R.J., Bruce, B.C., Brewer, W.F., Eds.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1980; ISBN 978-1-315-10749-3.
2. Nelson, G.; Powell, S.R. A Systematic Review of Longitudinal Studies of Mathematics Difficulty. *J. Learn. Disabil.* **2018**, *51*, 523–539. [CrossRef]
3. Peng, P.; Wang, T.; Wang, C.; Lin, X. A Meta-Analysis on the Relation between Fluid Intelligence and Reading/Mathematics: Effects of Tasks, Age, and Social Economics Status. *Psychol. Bull.* **2019**, *145*, 189–236. [CrossRef]
4. Caspi, A.; Wright, B.R.E.; Moffitt, T.E.; Silva, P.A. Early Failure in the Labor Market: Childhood and Adolescent Predictors of Unemployment in the Transition to Adulthood. *Am. Sociol. Rev.* **1998**, *63*, 424. [CrossRef]
5. Marjoribanks, K. Family Environments and Children's Outcomes. *Educ. Psychol.* **2005**, *25*, 647–657. [CrossRef]
6. Orrantia, J. Dificultades En El Aprendizaje de Las Matemáticas: Una Perspectiva Evolutiva. *Rev. Psicopedag.* **2006**, *23*, 158–180.
7. Sánchez-Pérez, N.; Fuentes, L.J.; Pina, V.; López-López, J.A.; González-Salinas, C. How Do Different Components of Effortful Control Contribute to Children's Mathematics Achievement? *Front. Psychol.* **2015**, *6*. [CrossRef] [PubMed]
8. Watt, H.M.G.; Shapka, J.D.; Morris, Z.A.; Durik, A.M.; Keating, D.P.; Eccles, J.S. Gendered Motivational Processes Affecting High School Mathematics Participation, Educational Aspirations, and Career Plans: A Comparison of Samples from Australia, Canada, and the United States. *Dev. Psychol.* **2012**, *48*. [CrossRef] [PubMed]
9. OECD. *PISA 2018 Results: What Students Know and Can Do*; OECD Publishing: Paris, France, 2019; Volume 1. [CrossRef]
10. Kintsch, W.; Greeno, J.G. Understanding and Solving Word Arithmetic Problems. *Psychol. Rev.* **1985**, *92*, 109–129. [CrossRef]
11. Kintsch, W. *Comprehension: A Paradigm for Cognition*; Cambridge University Press: Cambridge, UK, 1998; ISBN 978-0-521-58360-2.
12. Reusser, K. *Textual and Situational Factors in Solving Mathematical Word Problems*; University of Bern: Bern, Switzerland, 1989.
13. Bassok, M.; Wu, L.L.; Olseth, K.L. Judging a Book by Its Cover: Interpretative Effects of Content on Problem-Solving Transfer. *Mem. Cognit.* **1995**, *23*, 354–367. [CrossRef] [PubMed]
14. Bassok, M.; Chase, V.; Martin, S. Adding Apples and Oranges: Alignment of Semantic and Formal Knowledge. *Cognit. Psychol.* **1998**, *35*, 99–134. [CrossRef]
15. Gros, H.; Thibaut, J.P.; Sander, E. Semantic Congruence in Arithmetic: A New Conceptual Model for Word Problem Solving. *Educ. Psychol.* **2020**, *55*, 1–19. [CrossRef]
16. Thevenot, C. Arithmetic Word Problem Solving: Evidence for the Construction of a Mental Model. *Acta Psychol.* **2010**, *133*. [CrossRef] [PubMed]
17. Evans, J.B.T. *Hypothetical Thinking*; Psychology Press: London, UK, 2007; ISBN 978-0-367-42362-9.
18. García-Madruga, J.A.; Gutiérrez, F.; Carriedo, N.; Luzón, J.M.; Vila, J.O. Mental Models in Propositional Reasoning and Working Memory's Central Executive. *Think. Reason.* **2007**, *13*, 370–393. [CrossRef]
19. Cattell, R.B. Theory of Fluid and Crystallized Intelligence: A Critical Experiment. *J. Educ. Psychol.* **1963**, *54*, 1–22. [CrossRef]
20. Engle, R.W.; Tuholski, S.W.; Laughlin, J.E.; Conway, A.R.A. Working Memory, Short-Term Memory, and General Fluid Intelligence: A Latent-Variable Approach. *J. Exp. Psychol. Gen.* **1999**, *128*, 309–331. [CrossRef] [PubMed]
21. Fung, W.; Swanson, H.L. Working Memory Components That Predict Word Problem Solving: Is It Merely a Function of Reading, Calculation, and Fluid Intelligence? *Mem. Cognit.* **2017**, *45*, 804–823. [CrossRef] [PubMed]
22. Horn, J.L.; Cattell, R.B. Age Differences in Fluid and Crystallized Intelligence. *Acta Psychol.* **1967**, *26*, 107–129. [CrossRef]

23. Jaeggi, S.M.; Buschkuehl, M.; Jonides, J.; Perrig, W.J. Improving Fluid Intelligence with Training on Working Memory. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6829–6833. [CrossRef]
24. Sternberg, R.J.; Ben-Zeev, T. *The Nature of Mathematical Thinking*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1996; ISBN 978-0-8058-1799-7.
25. de Koning, B.B.; Boonen, A.J.H.; van der Schoot, M. The Consistency Effect in Word Problem Solving Is Effectively Reduced through Verbal Instruction. *Contemp. Educ. Psychol.* **2017**, *49*, 121–129. [CrossRef]
26. Jiang, R.; Li, X.; Xu, P.; Lei, Y. Do Teachers Need to Inhibit Heuristic Bias in Mathematics Problem-Solving? Evidence from a Negative-Priming Study. *Curr. Psychol.* **2020**. [CrossRef]
27. Lubin, A.; Rossi, S.; Lanoë, C.; Vidal, J.; Houdé, O.; Borst, G. Expertise, Inhibitory Control and Arithmetic Word Problems: A Negative Priming Study in Mathematics Experts. *Learn. Instr.* **2016**, *45*, 40–48. [CrossRef]
28. Passolunghi, M.C.; Pazzaglia, F. Individual Differences in Memory Updating in Relation to Arithmetic Problem Solving. *Learn. Individ. Differ.* **2004**, *14*. [CrossRef]
29. Passolunghi, M.C.; Pazzaglia, F. A Comparison of Updating Processes in Children Good or Poor in Arithmetic Word Problem-Solving. *Learn. Individ. Differ.* **2005**, *15*. [CrossRef]
30. Passolunghi, M.C.; Siegel, L.C. Short-Term Memory, Working Memory, and Inhibitory Control in Children with Difficulties in Arithmetic Problem Solving. *J. Exp. Child. Psychol.* **2001**, *80*. [CrossRef]
31. Shum, H.Y.; Chan, W.W.L. Young Children's Inhibition of Keyword Heuristic in Solving Arithmetic Word Problems. *Hum. Behav. Brain* **2020**, *1*, 43. [CrossRef]
32. Abdullah, A.H.; Abidin, N.L.Z.; Ali, M. Analysis of Students' Errors in Solving Higher Order Thinking Skills (HOTS) Problems for the Topic of Fraction. *Asian Soc. Sci.* **2015**, *11*, 133. [CrossRef]
33. Hasanah, N.; Hayashi, Y.; Hirashima, T. An Analysis of Learner Outputs in Problem Posing as Sentence-Integration in Arithmetic Word Problems. *Res. Pract. Technol. Enhanc. Learn.* **2017**, *12*, 9. [CrossRef]
34. Hegarty, M.; Mayer, R.E.; Green, C.E. Comprehension of Arithmetic Word Problems: Evidence from Students' Eye Fixations. *J. Educ. Psychol.* **1992**, *84*, 76–84. [CrossRef]
35. Rohmah, M.; Sutiarso, S. Analysis Problem Solving in Mathematical Using Theory Newman. *EURASIA J. Math. Sci. Technol. Educ.* **2018**, *14*, 671–681. [CrossRef]
36. Riley, M.S.; Greeno, J.G.; Heller, J.L. Development of children's problem-solving ability in arithmetic. In *The Development of Mathematical Thinking*; Ginsburg, H.P., Ed.; Academic Press: New York, NY, USA, 1984; pp. 153–196.
37. Bassok, M.; Pedigo, S.F.; Oskarsson, A.T. Priming Addition Facts with Semantic Relations. *J. Exp. Psychol. Learn. Mem. Cogn.* **2008**, *34*, 343–352. [CrossRef] [PubMed]
38. Guthormsen, A.M.; Fisher, K.J.; Bassok, M.; Osterhout, L.; DeWolf, M.; Holyoak, K.J. Conceptual Integration of Arithmetic Operations with Real-World Knowledge: Evidence from Event-Related Potentials. *Cogn. Sci.* **2016**, *40*, 723–757. [CrossRef] [PubMed]
39. Thevenot, C.; Oakhill, J. The Strategic Use of Alternative Representations in Arithmetic Word Problem Solving. *Q. J. Exp. Psychol. Sect. A* **2005**, *58*, 1311–1323. [CrossRef] [PubMed]
40. Gros, H.; Sander, E.; Thibaut, J.P. When Masters of Abstraction Run into a Concrete Wall: Experts Failing Arithmetic Word Problems. *Psychon. Bull. Rev.* **2019**, *26*, 1738–1746. [CrossRef]
41. Thevenot, C.; Fanget, M.; Fayol, M. Retrieval or Nonretrieval Strategies in Mental Arithmetic? An Operand Recognition Paradigm. *Mem. Cognit.* **2007**, *35*, 1344–1352. [CrossRef] [PubMed]
42. Fischbein, E. Tacit Models and Mathematical Reasoning. *Learn. Math.* **1989**, *9*, 9–14.
43. Sternberg, R.J. What is mathematical thinking? In *The Nature of Mathematical Thinking*; Sternberg, R.J., Ben-Zeev, T., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1996; pp. 303–318. ISBN 978-0-8058-1799-7.
44. Hegarty, M.; Mayer, R.E.; Monk, C.A. Comprehension of Arithmetic Word Problems: A Comparison of Successful and Unsuccessful Problem Solvers. *J. Educ. Psychol.* **1995**, *87*, 18–32. [CrossRef]
45. Lewis, A.B.; Mayer, R.E. Students' Miscomprehension of Relational Statements in Arithmetic Word Problems. *J. Educ. Psychol.* **1987**, *79*. [CrossRef]
46. Siegler, R.; Jenkins, E.A. *How Children Discover New Strategies*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1989; ISBN 978-0-8058-0472-0.
47. Stacey, K.; MacGregor, M. Learning the Algebraic Method of Solving Problems. *J. Math. Behav.* **1999**, *18*, 149–167. [CrossRef]
48. Khng, K.H.; Lee, K. Inhibiting Interference from Prior Knowledge: Arithmetic Intrusions in Algebra Word Problem Solving. *Learn. Individ. Differ.* **2009**, *19*, 262–268. [CrossRef]
49. Boote, S.K.; Boote, D.N. ABC Problem in Elementary Mathematics Education: Arithmetic before Comprehension. *J. Math. Teach. Educ.* **2018**, *21*, 99–122. [CrossRef]
50. Chi, M.T.H.; Glaser, R.; Farr, M.J. *The Nature of Expertise*; Psychology Press: New York, NY, USA, 1988. [CrossRef]
51. Sternberg, R.J.; Frensch, P.A. On being an expert: A cost-benefit analysis. In *The Psychology of Expertise*; Hoffman, R., Ed.; Springer: New York, NY, USA, 1992; pp. 191–203.
52. Kahneman, D. *Thinking Fast, and Slow*; Farrar, Straus and Giroux: New York, NY, USA, 2011; ISBN 978-1-84614-606-0.
53. Mayer, R.E.; Hegarty, M. The Process of Understanding Mathematical Problems. *Nat. Math. Think.* **1996**, *12*, 24–59.

54. Johnson-Laird, P.N. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*; Harvard University Press: Cambridge, MA, USA, 1983; ISBN 978-0-89859-242-9.
55. Evans, J.B.T.; Over, D.E. Rationality in the Selection Task: Epistemic Utility Versus Uncertainty Reduction. *Psychol. Rev.* **1996**, *103*, 356–363. [CrossRef]
56. Kahneman, D.; Tversky, A. Subjective Probability: A Judgment of Representativeness. *Cognit. Psychol.* **1972**, *3*, 430–454. [CrossRef]
57. Tversky, A.; Kahneman, D. Availability: A Heuristic for Judging Frequency and Probability. *Cognit. Psychol.* **1973**, *5*, 207–232. [CrossRef]
58. Stanovich, K.E.; Toplak, M.E. Defining Features versus Incidental Correlates of Type 1 and Type 2 Processing. *Mind Soc.* **2012**, *11*. [CrossRef]
59. Evans, J.S.B.T. In Two Minds: Dual-Process Accounts of Reasoning. *Trends Cogn. Sci.* **2003**, *7*, 454–459. [CrossRef] [PubMed]
60. Kahneman, D.; Frederick, S. Representativeness revisited: Attribute substitution in intuitive judgment BT—Heuristics and biases: The psychology of intuitive judgment. In *Heuristics and Biases: The Psychology of Intuitive Judgment*; Cambridge University Press: New York, NY, USA, 2002; ISBN 978-0-521-79260-8.
61. Sloman, S.A. The Empirical Case for Two Systems of Reasoning. *Psychol. Bull.* **1996**, *119*. [CrossRef]
62. Stanovich, K.E.; West, R.F. Individual Differences in Reasoning: Implications for the Rationality Debate? *Behav. Brain. Sci.* **2000**, *23*, 645–655. [CrossRef] [PubMed]
63. Evans, J.B.T.; Stanovich, K.E. Dual-Process Theories of Higher Cognition. *Perspect. Psychol. Sci.* **2013**, *8*, 223–241. [CrossRef] [PubMed]
64. Fletcher, L.; Carruthers, P. Metacognition and Reasoning. *Philos. Trans. R. Soc. B Biol. Sci.* **2012**, *367*, 1366–1378. [CrossRef] [PubMed]
65. Gómez-Chacón, I.; García-Madruga, J.A.; Vila, J.O.; Elosúa, M.R.; Rodríguez, R. The Dual Processes Hypothesis in Mathematics Performance: Beliefs, Cognitive Reflection, Working Memory and Reasoning. *Learn. Individ. Differ.* **2014**, *29*, 67–73. [CrossRef]
66. Pillow, B.H. Children's and Adults' Evaluation of the Certainty of Deductive Inferences, Inductive Inferences, and Guesses. *Child. Dev.* **2002**, *73*. [CrossRef]
67. Moshman, D. *Adolescent Rationality and Development*; Taylor & Francis: Abingdon, UK, 2004; ISBN 978-1-135-61658-8.
68. Santamaría, C.; Tse, P.P.; Moreno-Ríos, S.; García-Madruga, J.A. Deductive Reasoning and Metalogical Knowledge in Preadolescence: A Mental Model Appraisal. *J. Cogn. Psychol.* **2013**, *25*. [CrossRef]
69. Kaufman, A.S.; Kaufman, N.L. *K-BIT, Test Breve de Inteligencia de Kaufman*; Pearson Clinical: Madrid, Spain, 2013; ISBN 978-84-938825-2-5.
70. García-Madruga, J.A.; Santamaría, C.; Moreno-Ríos, S.; Vila-Chaves, J.O.; Gómez-Veiga, I.; Orenes, I. A Test of Deductive Reasoning. Unpublished work. 2014.
71. Ramos, J.L.; Cuetos, F. *Batería de Evaluación de Los Procesos Lectores En Los Alumnos Del Tercer Ciclo de Educación Primaria y Educación Secundaria Obligatoria, PROLEC-SE*; TEA Ediciones: Madrid, Spain, 1999; ISBN 84-7174-953-X.
72. Duque de Blas, G.; Gómez-Veiga, I. The Spelling Test for Secondary School. Unpublished work. 2018.
73. Castro-Martínez, E.; Frías-Zorilla, A. Two-Step Arithmetic Word Problems. *Math. Enthus.* **2013**, *10*, 379–406.
74. Viterbori, P.; Traverso, L.; Usai, M.C. The Role of Executive Function in Arithmetic Problem-Solving Processes: A Study of Third Graders. *J. Cogn. Dev.* **2017**, *18*, 595–616. [CrossRef]
75. Verschaffel, L. Using Retelling Data to Study Elementary School Children's Representations and Solutions of Compare Problems. *J. Res. Math. Educ.* **1994**, *25*. [CrossRef]
76. López-Higes, R.; Mayoral, J.A.; Villoria, C. *Batería de Evaluación de La Lectura (BEL)*; Psymtec: Madrid, Spain, 2002; ISBN 84-931200-1-4.
77. Sebastián, N.; Cuetos, F.; Martí, M.A.; Carreiras, M.F. *LEXESP: Léxico Informatizado Del Español*; Edicions de la Universitat de Barcelona: Barcelona, Spain, 2000; ISBN 978-84-8338-187-8.
78. McLaughlin, A.C.; McGill, A.E. Explicitly Teaching Critical Thinking Skills in a History Course. *Sci. Educ.* **2017**, *26*. [CrossRef]
79. Reed, J.; Kromrey, J. Teaching Critical Thinking in a Community College History Course: Empirical Evidence from Infusing Paul's Model. *Coll. Stud. J.* **2001**, *35*, 201.
80. Williams, R.L.; Worth, S.L. The Relationship of Critical Thinking to Success in College. *Inq. Crit. Think. Discip.* **2001**, *21*, 5–16. [CrossRef]
81. Yang, S.C. E-Critical/Thematic Doing History Project: Integrating the Critical Thinking Approach with Computer-Mediated History Learning. *Comput. Hum. Behav.* **2007**, *23*. [CrossRef]
82. Barrouillet, P. Dual-Process Theories and Cognitive Development: Advances and Challenges. *Dev. Rev.* **2011**, *31*, 79–85. [CrossRef]
83. Field, A. *Discovering Statistics Using IBM SPSS Statistics*; SAGE Publications: Thousand Oaks, CA, USA, 2013.
84. Wilcox, R.R. *Introduction to Robust Estimation and Hypothesis Testing*; Academic Press: Cambridge, MA, USA, 2011.
85. Bull, R.; Lee, K. Executive Functioning and Mathematics Achievement. *Child. Dev. Perspect.* **2014**, *8*, 36–41. [CrossRef]
86. Friso-van den Bos, I.; Van der Ven, S.H.G.; Kroesbergen, E.H.; Van Luit, J.E.H. Working Memory and Mathematics in Primary School Children: A Meta-Analysis. *Educ. Res. Rev.* **2013**, *10*, 29–44. [CrossRef]
87. Mayer, R.E.; Hegarty, M.; Lewis, A.B. Mathematical Misunderstandings: Qualitative Reasoning About Quantitative Problems. *Adv. Psychol.* **1992**, *91*. [CrossRef]
88. Pape, S.J. Compare Word Problems: Consistency Hypothesis Revisited. *Contemp. Educ. Psychol.* **2003**, *28*, 396–421. [CrossRef]

89. Passolunghi, M.C.; Duque de Blas, G.; Carretti, B.; Gómez-Veiga, I.; García-Madruga, J.A. The Role of Working Memory Updating, Inhibition, Fluid Intelligence and Reading Comprehension in Explaining Differences between Consistent and Inconsistent Arithmetic Word Problem Solving Performance. 2021, in press.
90. St. Clair-Thompson, H.L.; Gathercole, S.E. Executive Functions and Achievements in School: Shifting, Updating, Inhibition, and Working Memory. *Q. J. Exp. Psychol.* **2006**, *59*, 745–759. [CrossRef]
91. García-Madruga, J.A.; Vila, J.O.; Gómez-Veiga, I.; Duque, G.; Elosúa, M.R. Executive Processes, Reading Comprehension and Academic Achievement in 3rd Grade Primary Students. *Learn. Individ. Differ.* **2014**, *35*, 41–48. [CrossRef]
92. Gómez-Veiga, I.; Vila Chaves, J.O.; Duque de Blas, G.; García Madruga, J.A. A New Look to a Classic Issue: Reasoning and Academic Achievement at Secondary School. *Front. Psychol.* **2018**, *9*. [CrossRef] [PubMed]
93. Allen, J.D. Grades as Valid Measures of Academic Achievement of Classroom Learning. *Clear. House J. Educ. Strateg. Issues Ideas* **2005**, *78*, 218–223. [CrossRef]
94. Roth, B.; Becker, N.; Romeyke, S.; Schäfer, S.; Domnick, F.; Spinath, F.M. Intelligence and School Grades: A Meta-Analysis. *Intelligence* **2015**, *53*. [CrossRef]
95. Kuhn, J.T.; Holling, H. Gender, Reasoning Ability, and Scholastic Achievement: A Multilevel Mediation Analysis. *Learn. Individ. Differ.* **2009**, *19*, 229–233. [CrossRef]
96. Re, A.M.; Lovero, F.; Cornoldi, C.; Passolunghi, M.C. Difficulties of Children with ADHD Symptoms in Solving Mathematical Problems When Information Must Be Updated. *Res. Dev. Disabil.* **2016**, *59*. [CrossRef] [PubMed]
97. Passolunghi, M.C.; Mammarella, I.C. Spatial and Visual Working Memory Ability in Children with Difficulties in Arithmetic Word Problem Solving. *Eur. J. Cogn. Psychol.* **2010**, *22*. [CrossRef]
98. García-Madruga, J.A.; Elosúa, M.R.; Gil, L.; Gómez-Veiga, I.; Vila, J.O.; Orjales, I.; Contreras, A.; Rodríguez, R.; Melero, M.A.; Duque, G. Reading Comprehension and Working Memory's Executive Processes: An Intervention Study in Primary School Students. *Read. Res. Q.* **2013**, *48*, 155–174. [CrossRef]