

Article

Analyzing Difficulties in Arithmetic Word Problem Solving: An Epistemological Case Study in Primary School

Roberto Capone ^{1,*} , Federica Filiberti ² and Alice Lemmo ³¹ Department of Industrial Engineering, University of Salerno, 84042 Fisciano, Italy² Primary School Matese, 86019 Vinchiaturro, Italy; filiberti.federica@gmail.com³ Department of Human Sciences, University of L'Aquila, 67100 L'Aquila, Italy; alice.lemmo@univaq.it

* Correspondence: rcapone@unisa.it

Abstract: This paper focuses on difficulties that primary school students have in facing mathematical word problems. In particular, we are interested in exploring how they develop in the transition from grade 2 to grade 5. The research basis of the hypothesis is that some difficulties detected in grade 5 are already predictable in grade 2. Starting from the data collected in grade 5 by the National Standardized Assessment, we carry out a quantitative analysis looking for word problems in which students experience difficulties. Subsequently, we conduct a backward analysis of the grade 2 test of the same cohort of students in order to identify a set of word problems linked with those selected in grade 5 test. The analysis shows the presence of many common difficulties in the two grades. We design and carry out specific educational activities concerning word problem-solving in grade 2. These activities produce positive changes in the experimental class compared to the control class. This could suggest that a previous intervention in grade 2 could allow overcoming future difficulties in word problem text comprehension in grade 5.

Keywords: linguistic difficulties; problem-solving; word problems; primary teaching



Citation: Capone, R.; Filiberti, F.; Lemmo, A. Analyzing Difficulties in Arithmetic Word Problem Solving: An Epistemological Case Study in Primary School. *Educ. Sci.* **2021**, *11*, 596. <https://doi.org/10.3390/educsci11100596>

Academic Editors: Giorgio Bolondi, Federica Ferretti and Luis J. Rodríguez-Muñiz

Received: 10 August 2021
Accepted: 25 September 2021
Published: 30 September 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

The problem-solving process is one of the key elements in the design of mathematics teaching/learning activities in all countries [1,2]. According to Schoenfeld [3], solving problems means finding a way out of a difficulty or a goal that is not immediately achievable. In accordance with this thinking, problem-solving is one of the activities in which students and teachers encounter the greatest difficulties [4].

In the teaching practice, mathematical problems are generally shown through a text enriched by other forms of representation (e.g., graphs, tables, images, . . .); for this reason, in literature, they are very often called word problems [5].

In this paper, a teaching experiment focused on the difficulties that students experience in word problems is presented. In particular, our attention focuses on difficulties that grade 5 students experience and how they are predictable, and if it is possible, early intervention in the first primary grades. In this contribution, we propose a quantitative analysis in order to provide material for a reflection on the information that the data collected from the national standardized assessment tests provide. Our aim is to assess whether the data collected allow early identification of difficulties in solving word problems in grade 5 and whether it is possible to structure specific actions in grade 2 in order to prevent them.

For several years, an important research field in national and international mathematics education has focused on the study of the difficulties related to the word problem-solving process; in particular, the focus is on the main factors that can influence students in the choice of implementing or not implementing solving strategies.

Much research shows that most of the difficulties that students encounter in word problem-solving concern, on the one hand, the implementation of a solving algorithm (see, for example, [6,7]); and, on the other hand, the text comprehension (see, for example, [8,9]).

We are interested in this second point: over many decades, authors show that the text has a strong influence on the solver's interpretation and, therefore, on the resolving processes he/she uses [10]. Several international research, for example, the pioneering one presented by Nesher [11], show that in many cases, students try to infer directly from the text the solution steps to solve the problem rather than representing the situation described in the text.

The questions that guided our research are:

1. What kind of difficulties do grade 5 students have in interpreting word problems text?
2. Which of these difficulties are already highlighted in grade 2 students?

2. Conceptual Framework

We chose to adopt various theoretical frameworks to develop the different research phases. Regarding students' difficulties in solving problems, we focused on text comprehension, defined as a process of "horizontal mathematization" [12]. For the quantitative vertical study of the difficulties, we referred to the "chains of questions" [13]. Finally, about the analysis of difficulties and the predisposition for early intervention, we referred to the results of the work of Nesher [11], Gerofsky [14], and Daroczy and colleagues [9].

2.1. Understanding the Text as a Horizontal Mathematization Process

In previous research, author [13] considered the process of understanding the text of word problems as "horizontal mathematization" [12].

Treffers [12] and Freudenthal [15] described the mathematization process as horizontal and vertical, divided into two fundamental phases. According to Treffers [12], "through an empirical approach, that is, through observation, experimentation, inductive reasoning, the problem is changed so that strictly mathematical tools can approach it. The attempting to schematize the problem is called "horizontal" mathematization mathematically. (. . .) The activities related to the mathematical process, problem solution, solution generalization, and further formalization can be described as "vertical" mathematization." [12] (p. 71). In all phases of mathematical activity with word problems, both mathematizations integrate each other [11,16]. In the early definition of horizontal mathematization, the emphasis is placed on the transition from the real world to the mathematical world, just as talking about vertical mathematization refers to the process only within the mathematical world.

Horizontal mathematization requires both a translation into the mathematical language of the real situation through semiotic representations [17] and analysis and interpretation of the mathematical results obtained in the context of the real situation. Therefore, it is the process of transforming the problem in the context into a mathematical problem, i.e., manageable through mathematical tools, concepts, and procedures. A model may be defined as "a system of conceptual frameworks used to construct, interpret, and mathematically describe a situation" [18] (p. VIII). Therefore, horizontal mathematization requires the pupil to identify the mathematical structure within the problem posed [19]. We can therefore consider the comprehension of the word problem text as one of the main phases of the horizontal mathematization process [13].

Consequently, it is the ability to extrapolate the information needed to analyze, set, and solve the problem. Therefore, a priori a deep understanding of the situation and decoding of the transmitted data of the text (even the implicit ones) is shown in several forms (linguistic, arithmetic, algebraic, graphic, etc.). This also requires the ability to extrapolate information from several representations expressed in different semiotic registers [17]. Vertical mathematization, instead, requires a reorganization and reconstruction of the problem within mathematics through the manipulation of mathematical models, the use of procedures and concepts, recognizing recurrent patterns and strategies to be used with known or explored methods [13]. Moreover, it requires checking the problem conditions,

the generalization of solving procedures, and the recognition of a possible application of these procedures in similar problems. By making these mathematization phases explicit, it is possible to analyze pupils' competencies in a more targeted and conscious way and foresee possible specific intervention actions in difficulties [16]. We hypothesize that horizontal mathematization can have an impact on the vertical mathematization process as they are defined.

2.2. The Standardized National Tests

We analyzed students' difficulties in word problems starting from the results of the Italian Standardized Assessment (INVALSI) tests. The questions of these tests are closely related to the Italian Standards [20]. The questions of the tests are divided into 4 areas (Numbers, Space and Figures, Relationships and Functions; Data and Probability), and they relate to specific dimensions (Knowing, Solving Problems, Arguing). The INVALSI tests are administered to grade 2, 5, 8, 10, and 13 students. We were interested in primary school students: grades 2 and 5.

2.2.1. Quantitative Analysis

The results of the INVALSI surveys were analyzed using the Rasch model [21]. It is a logistic model with one parameter that belongs to the question Response Theory (IRT) category and produces a joint estimation of two types of parameters: a difficulty parameter for each test question and an ability parameter for each student. The Rasch model returns the probability that students choose the correct answer to a question, according to the difficulty of the question and the student's ability measured in the whole test. This probability is represented by a curve called the characteristic curve of the questions. In a similar way, it is possible to represent all the empirical data; in particular, there is a characteristic curve for each option in the case of multiple-choice questions (or for wrong or missing answers in open-ended questions). The graph that collects all the curves is the Distractor Plot (DP).

The simple percentages of answers show the number of students who gave a certain response but not their ability. The DP is useful in analyzing the data collected by the National Assessment Institute because it shows the correlation between the percentages and the student's ability in the test.

Below we proposed an example of DP (Figure 1) to better understand the following analyses.

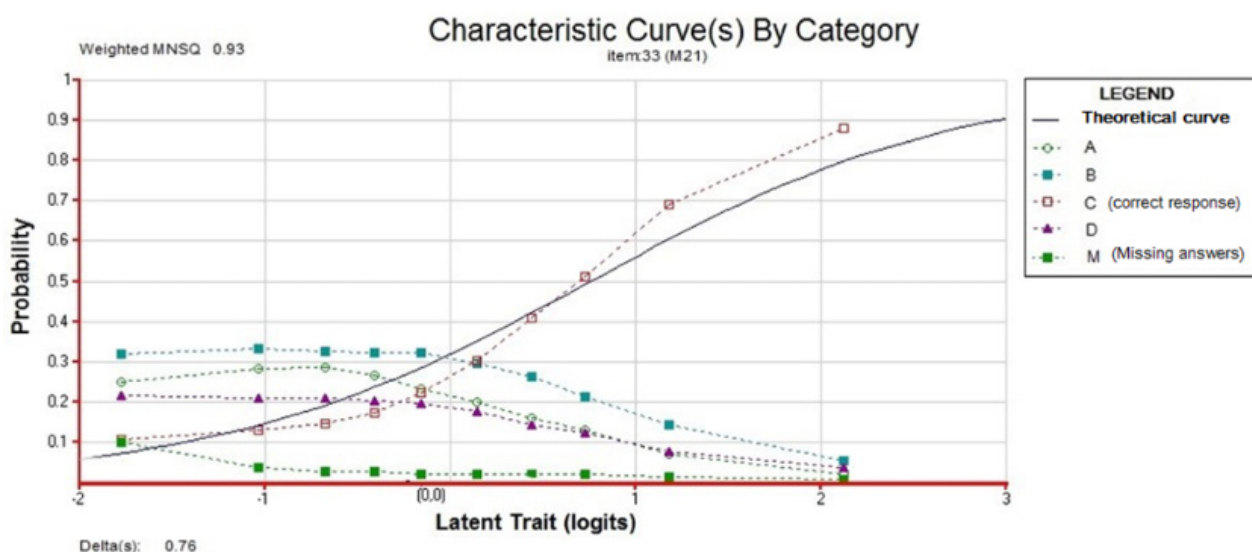


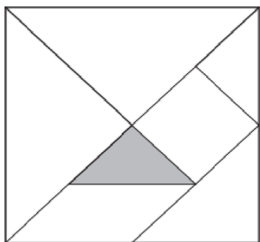
Figure 1. Example of characteristic curve.

In Figure 1, the x -axis reports the Rasch score in terms of student ability in the whole test (-2 represents the ability of a student who do not answer any question of the text; on the contrary, 2 represents the ability of a student who does answer to all the questions of the text).

The continuous line represents the model’s theoretical curve which reveals the theoretical probability of correct answer according to student ability. If the theoretical curve is increasing, it means that as the student’s ability grows, his/her probability of responding correctly increases. The dotted lines represent empirical data collected for each answer to the question (options in the case of a multiple-choice question; correct/wrong/missing answer in the case of open-ended questions).

In Figure 2, we can notice that the most attractive wrong option was B (blue dotted line), which was chosen by a high percentage of students, including those of medium and medium-high level of abilities. The curve began to decrease only beyond medium levels (which in the graph corresponds to 0 on the x -axis) of students’ abilities in the test. The curves of the other two options decreased, and they were chosen by students of low and medium ability because the points of inflection correspond to low values of the x -axis. Finally, it can be noted that only a few students did not answer the question, almost all of whom lowest ability level are.

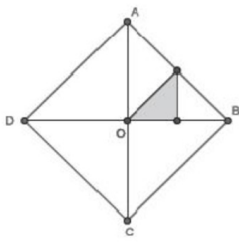
D25. In figure it is represented the game of Tangram with the pieces that compose it



To which fraction of the area of Tangram does the colored piece correspond to?

A. One seventh
 B. One eighth
 C. One fifteenth
 D. One sixteenth

D2 In the square ABCD the middle points of side AB and segment OB are connected.



How many triangles, as the colored one, can we exactly cover the surface of the square ABCD with?

Answer:

Figure 2. Example of “chain of questions”.

2.2.2. The Chains of Questions

Authors [13] use the term “chain of questions” to refer to a set of questions belonging to INVALSI tests of different school grades related to the same areas, mathematical dimensions, and contents.

The questions in the chain are not identical because they change according to the specific guidelines of each school grade. Students’ have to use similar procedures to solve them. In this sense, it is possible to link the questions in a chain because they refer to the same conceptual field [22], i.e., they belong to the same set of reference situations, the same set of invariant operators, and the same set of language representations. Because of this peculiarity, we defined the questions of the chains as “equivalent questions”.

In two works [13,23], the authors showed that the analysis of the chains of questions allows us to link some difficulties that students encounter in particular grades to challenges found in the tests that students have carried out in previous years. This means that the tests carried out in the first grades of primary school have predictive power concerning future difficulties in the following school grades.

In Figure 2. there is an example of a “chain of questions”. The task was similar but not identical. The strategies developed to answer them can be compared. As authors say: “The grade 6 question asks to cover while the grade 8 one to find a fraction, but in both

the cases the operation to carry out the solution is the same: to compose the square using triangles equivalent to the colored one" [23] (p. 1700).

2.3. Difficulties in the Text Comprehension

Since the 1980s, a lot of research has been carried out to study students' difficulties in understanding the text. A pioneering study was Nesher's [11] who analyzed the word problem-solving difficulties by considering the following aspects:

- The logical structure (the type of operation required, the possible presence of extra or missing information, . . .);
- The semantic component (the contextual relationships involved the word suggestions in the problem);
- The syntactic component (the structural variables, i.e., the number of words, position of the parts making up the problem, . . .).

Over the years, research in the field has expanded and developed. Daroczy and colleagues [9] proposed an extensive literature review on the subject. In their work, an analysis of the results obtained in the last thirty years highlighted three main components that can cause difficulties in the process of "horizontal" mathematization and therefore influence the choice of the solving process in word problems:

- The linguistic complexity of the problem text;
- The numerical complexity of the data being presented in the problem;
- The relationship between the linguistic and numerical components of the problem.

However, these three dimensions are not exhaustive; sometimes there are very complex problems: for example, the solver has to decode the problem by interpreting it into more straightforward problems; or there may be irrelevant or missing data that require the solver to decide what numerical and non-numerical data are needed for the resolution and so on. Therefore, solving word problems involves other difficulties because they cannot be solved with routine procedures, and the solver needs specific knowledge and expertise in the disciplinary domain, heuristics, and metacognitive strategies [3]. Gerofsky [14], for example, proposed three fundamental processes to be followed for correct implementation of a word problem-solving strategy:

- The correct contextualization of the problem, with all its implications;
- The identification of all the information necessary for the resolution;
- The implementation of the resolution processes and the resulting mathematical operations.

We also highlight that reading and interpreting a mathematical text does not only mean decoding the question but also implies the ability to find out relationships between the different parts of which it is composed. In this respect, Zan [24] started from the assumption that there is a close link between "context" and "demand" followed by the focus of the problematic situation: the more the two components are connected, the greater will be the understanding. This connection is not always immediate; generally, the student faces a "heteroposto" problem, i.e., formulated by an individual different from him. For this reason, the solver is not the one who spontaneously asks himself the "question" following the reading of the "context", but it is imposed on him [25].

3. Methodology

Working on students' difficulties in the process of solving word problems from a vertical perspective requires working in three main phases:

- Identify common difficulty patterns between the two grades (02 and 05) through a quantitative analysis of the results collected by the mathematical INVALSI to identify the questions' chains;
- Validate this analysis through an experimental quantitative study in a grade 2 class;
- Implement early intervention teaching practices and verify their effectiveness.

3.1. Choosing the Chains of Questions

This research aims to study the difficulties that children have in solving word arithmetic problems. For this purpose, we selected an INVALSI grade 5 test and analyzed all the questions related to the “Numbers” area and the “problem solving” dimension. Starting with the grade 5 test, we explored the grade 2 test looking for possible questions related to the same areas, mathematical content, and conceptual fields. The set of questions identified in the first analysis consists of 4 grade 5 questions (8, 23, 30, 32) and 7 grade 2 questions (2, 4, 8, 9, 14, 16, 22). Below, we present an example of a chain of questions; in particular, question D32 of the INVALSI mathematics grade 5 test (Figure 3) and question D4 of the INVALSI mathematics grade 2 tests (Figure 4).

Antonella parks in a skyscraper parking garage which is located four floors under the ground floor. She gets on the elevator for twenty-four floors. At which floor will Antonella get out from the elevator?

A. 16
B. 20
C. 24
D. 28

Figure 3. D32 question of the INVALSI math grade 5—2011/2012.


	<p>Lucia’s mother is doing grocery shopping. The board shows the number you see here. The mum got the ticket number 39. How many people will be served before here?</p> <p>A. 38 B. 3 C. 5</p>
---	--

Figure 4. D4 question of INVALSI math grade 2—2008/2009.

Both problems present a realistic context, and they have multiple-choice questions. To answering the question, the student must either work on the number line and count in a progressive and regressive way or work through an additive model. Both questions, therefore, relate to the conceptual field of additive structures [22]. The analyses of the following word problems were developed considering the text in the original language and not the text translated into English. Therefore, the following observations may not be found in the English text. For example, subordinate sentences in the same period are commonplace in the Italian language. Question D32 (Figure 4) asks the student to identify the floor where the lift will open, where the student knows the starting floor, and the number of floors to climb.

The question is multiple-choice; only one of the options is correct (B) chosen by 43.5% of the students. Referring to the factors highlighted by Nesher [24], we identified some difficulties:

- In the logical structure: the problem requires students to change their perspective. Usually, when you get on the elevator, you push to button for the floor you want to get to (which is known), and if anything, you choose to count how many floors you have climbed. In this case, the problem asks for reverse reasoning that is far from the child’s experience.
- In the semantic component, the starting plane indicated as “fourth floor below level zero” may not be easily referred to as “−4”.

- The first period is composed of the main sentence and a subordinate one linked by “that” in the syntactic component, which, in this case, has the function of a relative pronoun and refers to the garage. Children could misunderstand and refer to the relative pronoun to both Antonella and the skyscraper itself.

According to Daroczy and colleagues [8], some difficulties can therefore be found due to the linguistic complexity of the problem text (in particular in the syntactic component); the numerical complexity of the data in the problem (one of the plans is shown in word form and refers to a whole number).

A little less than half of the students (43.5%) of the 2012 national sample (30,869 students) provided a correct answer to the question. The others chose the wrong answer options. Specifically, those who answered A (4.2%) presumably identified the number of floors “above” zero ($20 = 24 - 4$) and then carried out the operation that might seem suggested by the text, i.e., “remove 4”: $20 - 4 = 16$. Those who chose option C (20.7%) probably thought that Antonella had parked on the ground floor. In this case, they may not have taken into account that Antonella parked on the floor “−4”. Finally, the students who selected option D (22.8%) probably recognized an additive problem (add the number of floors to go up to the starting floor) but could not manage the operation in the whole number or could not decode the reference “below zero levels” ($24 + 4 = 28$).

The DP of the question (Figure 5) shows the correlation between the probability of choosing an answer and the student’s ability in the test. In this case, we can observe that the probability of choosing the wrong options decreased as the student’s skills increased (option A and C) and vice-versa, the correct answer curve increased (option B). This was not the case for option D, which had a peak of choice for average ability students in the test. This suggests that some average-skill students probably correctly identified the additive model to solve the question but cannot handle the “below zero” information. Reconnecting with the criticality observed in the text components may indicate difficulty in the semantic component.

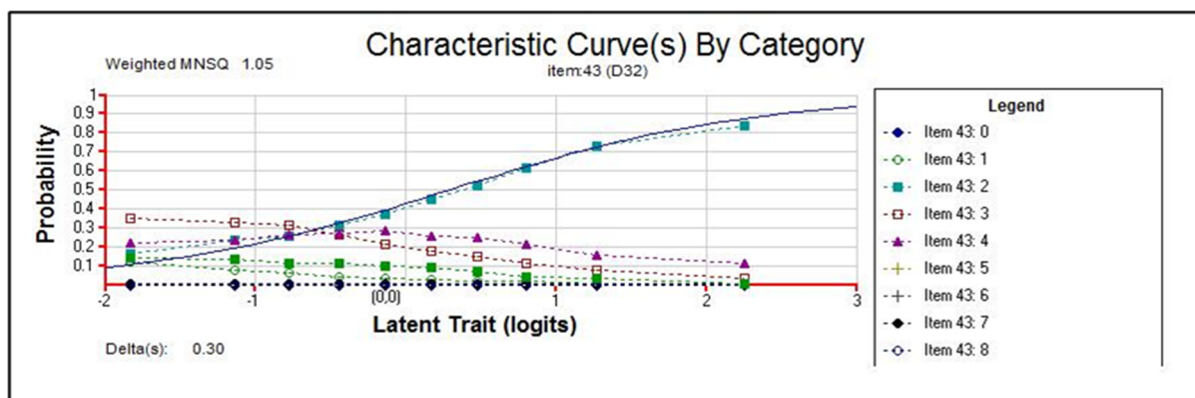


Figure 5. Characteristic curve about question D32 test grade 5—2011/2012.

Question D4 (Figure 5) shows a similar situation but with natural numbers. In addition, this question is a closed answer; among the choice options, only one is correct (B) chosen by 39.3% of the students of the 2009 national sample (43,333 students). We found in the text of the problem some critical aspects highlighted in the previous question; in particular:

- In the logical structure because students may not recognize the situation as familiar;
- In the semantic component, one numerical data is in the text while the other is presented in the image. In addition, it is described as the “number you see” without any reference to the picture.
- In the syntactic component, in a sentence appear two “that” with two different functions: conjunction and relative pronoun in the second case, “that” refers to one of the two numerical data.

In terms of complexity components [8], the linguistic complexity of the text is high: the text is very long. The choice to represent numerical data with two different semiotic registers could increase the numerical complexity, and therefore the relationship between the linguistic and numerical components is less explicit. Regarding the wrong answer options, students who chose option A (33.5%) may have recognized an additive problem but considered 0 as the starting number and not 35. Those who chose option C (22.3%) may have done the counting (35, 36, 37, 38, 39) without also considering the person served (35) and the mother (39).

In the DP (Figure 6), the correct option (B) had an increasing trend; the wrong answer option A had a decreasing trend. The trend of option C was an exception: a slight increase to an average skill level appears again as in the previous problem.

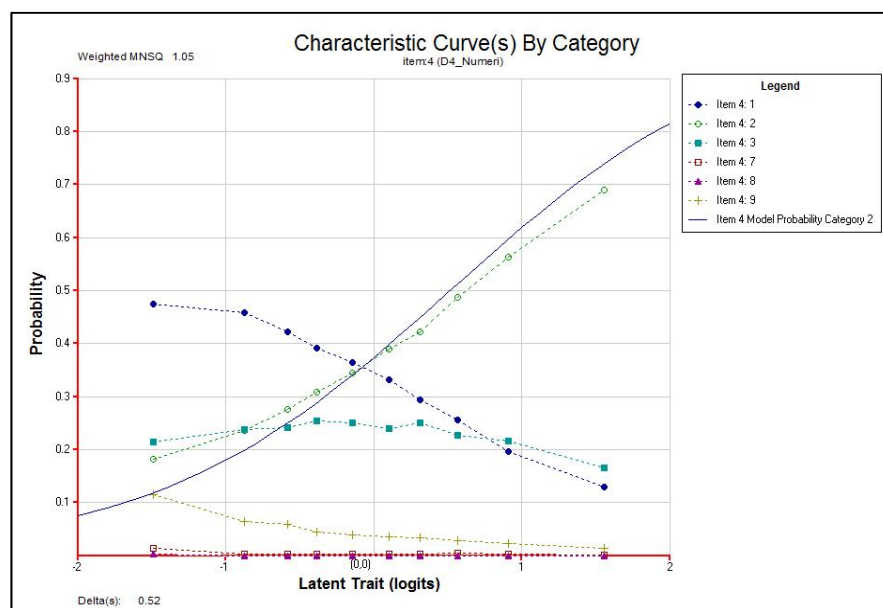


Figure 6. Characteristic Curve and DP of the question D4 of the mathematics INVALSI test of the school year 2008/2009.

The curves for options D (Figure 6, grade 5 item) and C (Figure 5, grade 2 item) had a similar trend. In addition, considering the previous analysis of the two items, we can hypothesize that there is a similar difficulty behind the choice of the two options. Students who chose D probably recognized the additive structure in the elevator item but had difficulty using the information compared to the mathematical model selected. Likewise, in the grade 2 item, students who chose option C maybe identified the appropriate additive structure, but they probably could not adapt it to the model.

Therefore, we can hypothesize that the difficulties that emerged from the INVALSI grade 2 test may be linked to the difficulties found in the future grade 5. In both cases, we identified possible causes related to critical aspects in the logical structure and the text's semantic component. In the same way, we selected other items from grade 5 (listed in Appendix A) and grade 2 (listed in Appendix B).

In the following table (Table 1), the selected items from grade 5 are classified according to the conceptual field and the critical points present in the components shown by Nesher [24].

Table 1. List of the selected items in the 2008/2009 mathematics INVALSI test. “X” indicates the components where complexities appear in the text of the item.

Item	Conceptual Field	Complexity in the Logical Structure	Complexity in the Semantic Component	Complexity in the Syntactic Component
D2	Multiplicative structures	X	X	
D4	Additive structures	X	X	X
D8	Additive structures	X		X
D9	Multiplicative structures		X	X
D14	Multiplicative structures	X	X	X
D16	Multiplicative structures		X	X
D22	Additive structures		X	

3.2. Participants

The sample of students involved in the experimentation consists of two grade 2 classes of the school “Colozza” in Campobasso (south of Italy). Class 2 A was the experimental class, while class 2 B was the control class. The experimental class was composed of 22 students (10 males and 12 females), of which 21 students were seven years old, and one was six years old; the control class was also composed of 22 students (13 males and 9 females), all seven years old. The methodological premise for carrying out the experimentation was the homogeneity of the sample: the number of pupils involved for each class, the ratio between the number of male and female pupils, their age.

3.3. Experimental Plant

We administered to both classes a pre-test consisting of the seven selected items from the 2008/2009 INVALSI test of grade 2 (listed in Appendix B). In order to enrich the data collection, we chose to leave a blank space between the items, asking the students to motivate their answers. They could freely choose how to do it; for example, students could use arithmetic expressions, drawings, diagrams, or others. This test has two main aims:

- The analysis of students’ answers allowed us to validate the a priori assumption described before on grade 2 students’ difficulties;
- The comparison between the results of this first test and the one we administered at the end of the didactic intervention allowed us to measure its effectiveness.

We dedicated the first part of the didactic intervention to a guided discussion on the items with a high percentage of wrong answers to deepen the analysis of the pre-test answers. This discussion was proposed first to pairs of students chosen by the experimental team and the teacher and then the whole class. The purpose of this choice was to allow cooperative work among the students. We administered the test carried out by the partner to each student to enhance discussion and sharing, first in the couple and then in the class. We led the discussion with guiding questions:

1. What does the question ask?
2. What is the correct answer?
3. How did you answer?
4. What difficulties did you encounter?
5. Rewrite the text of the question more simply and quickly; how would you do it?

Then a second didactic intervention was carried out through a laboratory methodology, also in pairs, and focused on the solution of new items. (Listed in Appendix C). The items were constructed ad hoc; they were equivalent to those of the pre-test concerning the previous definition and the same critical aspects in the textual components highlighted in the theoretical framework [8,24]. Throughout the activity, we observed the situation, giving only a few simple guidelines on approaching the tasks. Still, without giving procedural or other indications, the students had to proceed independently and rely on the comparison with their partners. At the end of the experimentation, a post-test (listed in Appendix D) was administered to both sample classes; it was composed of new items, also created

equivalent to those of the pre-test, in order to verify if the didactic interventions produced some changes in the experimental class, compared to the control class. The experiment lasted just over a month (from 19 February to 1 April 2019).

4. Results

In the previous paragraph, we described the three experimental phases: administration of the pre-test to the whole sample; guided discussion and subsequent workshops with the experimental class; administration of the post-test to the whole sample. In this paragraph, we will describe the results collected at each phase. The first phase of the experimentation was carried out in both sample classes; it consisted of administering the seven items selected from the grade 2 INVALSI test (No. 2, 4, 8, 9, 14, 16, 22). The following graphs show the percentage of correct answers in the pre-test items (Figure 7).

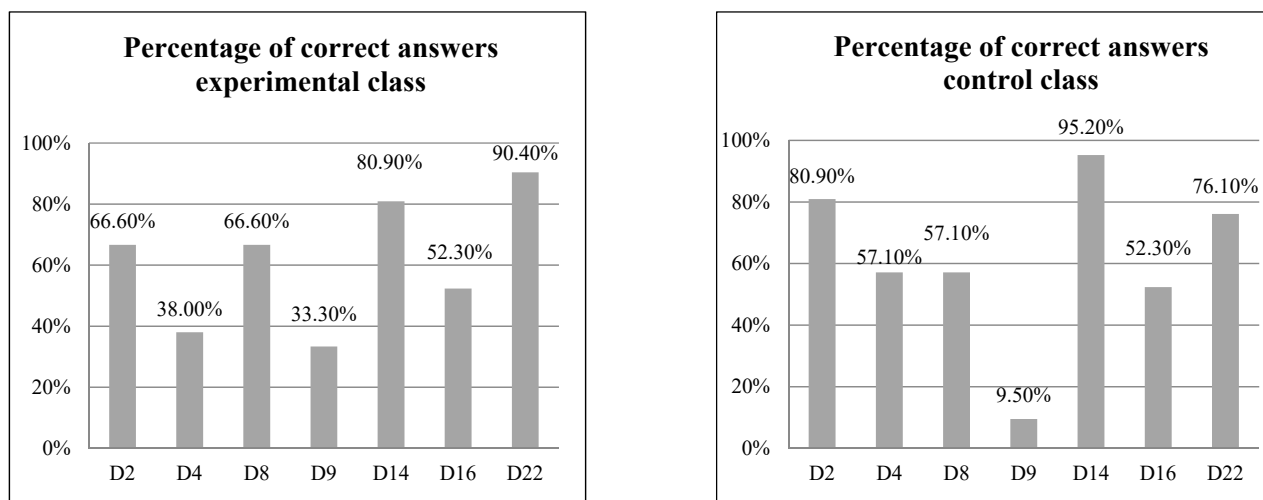
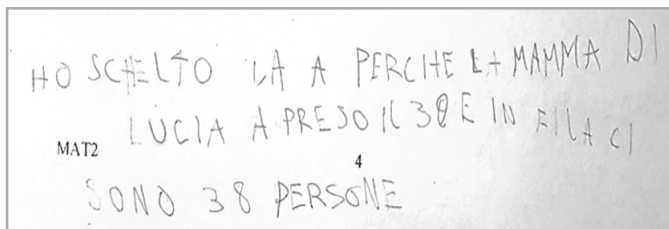


Figure 7. Percentages of correct answers to pre-test items in the sample classes.

The experimental class and the control class were chosen a priori. We noted that the correct response percentage in the experimental class was often lower than in the control class. However, the comparison between the response percentages allowed establishing a certain homogeneity in the answers given by the students of the sample classes (difference not exceeding 20%). They were consistent with the national sample of the INVALSI test in 2009.

The second experimentation phase only involved the experimental class and was divided into two consecutive didactic developments: a guided discussion on the pre-test items and new laboratory activity. The first didactic intervention highlighted the main difficulties encountered by the students and their proposals for solutions to the items. As an example, we showed some evidence observed in the discussion of question number 4 proposed as an example in the previous paragraph (Figure 3). Some students explicitly stated that they did not fully understand the problematic situation. For example, students who answered A stated that they had “*forgotten the number 35*” (student 3) because it was not explicitly present in the text but only in the picture. In this case, therefore, their attention was focused only on the text and not on the other representations present. It, thus, became clear that the semantic difficulty was due to the choice to represent parts of the text not in verbal form, which caused the students’ difficulties. Some pointed out that they were not familiar with questions that present information through an image. Precisely for this reason, although they understood the situation described in the text, they imagined that the first customer had been served and counted the customers still to be served from 1 (Figure 8).



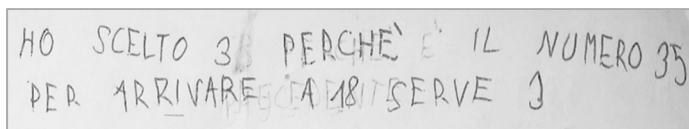
HO SCELTO LA A PERCHE LA MAMMA DI
LUCIA A PRESO IL 30 E IN FILA CI
SONO 38 PERSONE

MAT2

"I chose the A because Lucia's mom got the number 39, and there are 38 people in the queue".

Figure 8. Answer given by student 22 of the experimental class.

Students who chose option C showed difficulties in decoding the logical structure of the problem. They stated that they counted the natural numbers from 35 to 39, counting also the extremes. Students who responded correctly applied various strategies. Some students said they simply trusted the numbers between 35 and 39 (e.g., student 15 says: "I counted: 36, 37, 38; it's 3"). Similarly, someone transcribed the numbers in succession on the paper and then counted them. Other students, looking at the response options, proceeded by exclusion. Student 2 stated that it was impossible to choose answer A because "38 is too large a number and cannot be between 35 and 39". Similarly, option C was excluded; for example, student 8 claimed that: "between the number 35 and the number 39 there is a number smaller than 5". In another way, some students argued for exclusion by adding the answer choices to 35; student 7 said the numbers are "too big to get to number 39". Finally, some students tried to find the difference between numbers. Student 21 said: "the right number between 35 and 39 is exactly 3"; he referred to the range between 35 and 39 without considering 35 and 39. (Figure 9).

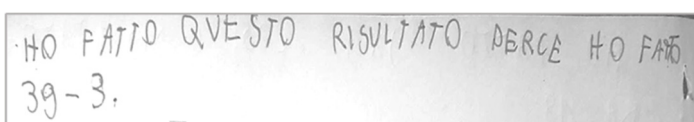


HO SCELTO 3 PERCHE' E' IL NUMERO 35
PER ARRIVARE A 18 SERVE 3

"I chose three because starting from 35; it needs 3 to get 18 (the student misspelled 18 to mean 38)"

Figure 9. Answer given by student 18 of the experimental class.

Some students performed the difference between 39 and 36. Other students performed the difference: "39-3" (Figure 10), "36 is correct because it is the successor of 35".



HO FATTO QUESTO RISULTATO PERCA HO FATTO
39-3.

"I found this result because I did 39-3".

Figure 10. Answer given by student 19 of the experimental class.

The guide question that aroused the most interest among the students was the one asking to reformulate the text of the question (guide question number 5). Below are two examples of answers given by the students (question number 4, Figure 3). Response to question-guide 5 of the couple 6: "Lucia's mom wants to go shopping. The scoreboard indicated number 35. She has the number 39. How many people are in front of her in the queue?"

Answer to question-guide 5 of couple 9: "Lucia's mum is at the supermarket. Her mother has ticket number 39. They are serving number 35. How many more people have to wait?"

The students' proposals noted that the reformulations did not present any criticality in the semantic component (both numerical data are shown in the text) and the syntactic component (the periods are concise: all the subordinate sentences are eliminated).

In the laboratory activity, researchers and the teacher did not intervene in the children's discussions but simply collected the answers given to each item. In the following graph (Figure 11), there is the correct response percentage to the items of the laboratory activity.

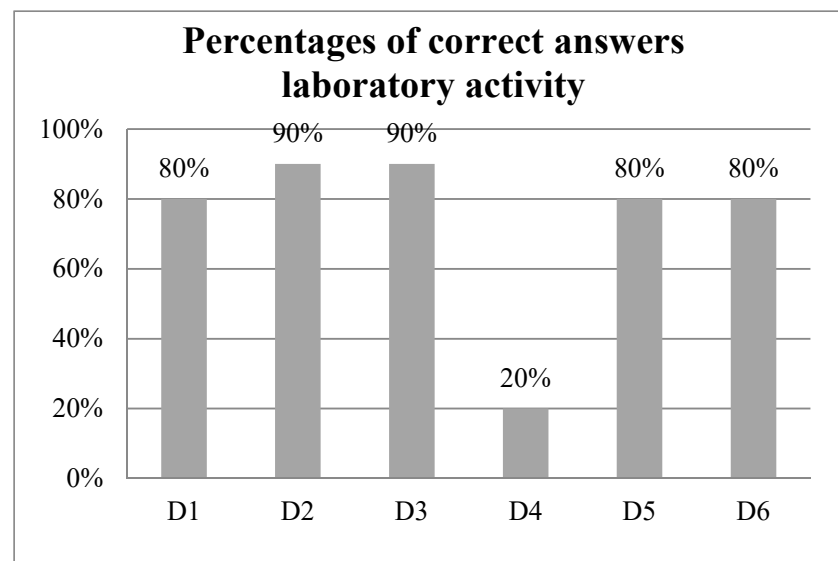


Figure 11. Percentages of correct answers to the items of the laboratory activity carried out by the experimental class.

From the charts, it can be seen that the correct answer percentages were very high except for question no. 4, which received a low percentage of correct answers (20% of the pairs). The purpose of the workshop activity was to leave students free to confront and explore problems independently. For this reason, no intervention was made even when a negative result was found in one of the items. In conclusion, we administered the post-test to both sample classes. The graphs (Figure 12) allow comparing the performance of the sample: the experimental class obtained, in most cases, higher or equivalent results to the control class.

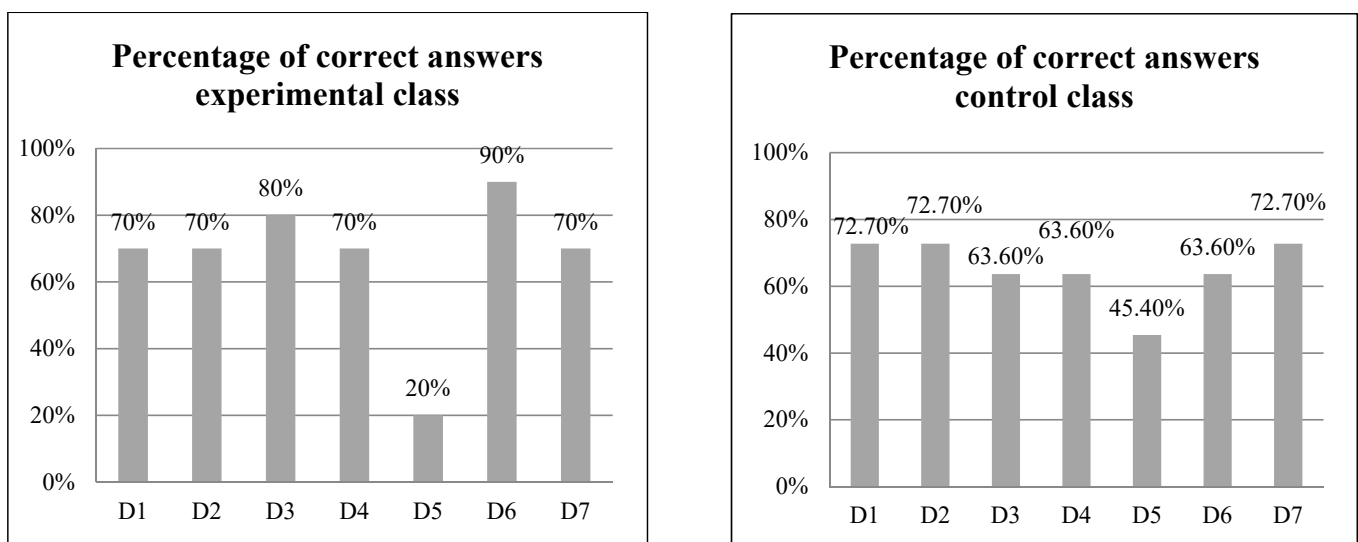


Figure 12. Percentages of correct answers to post-test items in the sample classes.

The data collected showed that in both classes, there was little difference in performance in items 1, 2, and 7. In items 3, 4, and 6, however, we noted that the performance of the experimental class was higher than the performance of the control class. Only in question 5, the experimental class's correct answers were lower than that of the control class. Similar to the pre-test, the post-test items also related to the conceptual field of additive or multiplicative structures and had some critical points in some text components. In Table 2, the items are shown concerning this classification. In the last column, a comparison

between the performance of the students of the two classes is shown: “=” indicates that the corrected response percentages were equal (with an error of 0.05); “+” means that the corrected response percentages of the experimental class were higher than those of the control class; “–” indicates that the corrected response percentages of the experimental class were lower than those of the control class.

Table 2. List of the selected items in the post-test. “X” indicates the components where complexities appear in the text of the item.

Item	Conceptual Field	Complexity in the Logical Structure	Complexity in the Semantic Component	Complexity in the Syntactic Component	Comparison between the Performance of the Sample Classes
1	Multiplicative structures	X	X		=
2	Additive structures	X	X	X	=
3	Additive structures	X		X	+
4	Multiplicative structures		X	X	+
5	Multiplicative structures	X	X	X	–
6	Multiplicative structures		X	X	+
7	Additive structures		X		=

The table shows that the experimentation seems to have had a good effect on additive word problems regardless of the text component where complexities are present. Students show good results also in multiplicative word problems except for one item: this was the only multiplicative word problem that had complexity in all components.

5. Conclusions

This research focused on difficulties that students experience with word problems. Both ideographic and nomothetic analyses were carried out: the purpose was to analyze difficulties that primary school students have in word problems at the national level from a quantitative point of view and then analyze the individual protocols. The main focus of the study was to develop a methodology for early detection of students’ difficulties in their school career and to take action to overcome them.

For the early detection of difficulties, we were inspired by the chain of questions of the INVALSI tests of mathematics [13] of the problem-solving dimension. The chains of questions are tasks belonging to different mathematical tests, administered in different years to different school grades students. Such tasks are not identical, but they are related to the same mathematical content, referred to the same dimension, and afferent to the same conceptual field [22]. The chains of questions analyzed in this paper were related to word problems belonging to the conceptual field of additive and multiplicative structures. This analysis was useful to highlight the difficulties encountered by the children in grade 5 and previously in grade 2. In this way, we showed that most of the difficulties encountered in grade 5 could be related to previous difficulties that emerged in grade 2. From this perspective, we can hypothesize that the results of the grade 2 INVALSI tests can be predictive of the future results of grade 5 tests. The predictive power of the chain of questions allowed us to hypothesize possible didactic suggestions to intervene immediately and avoid the difficulties observed in a certain school grade propagate or even increase in subsequent school grades.

In our specific case, the 2012 grade 5 INVALSI test was compared backward with the 2009 grade 2 test, administered to the same cohort of students. The comparison allowed us to point out that some difficulties in grade 5 had already emerged in grade 2 tasks. In particular, we focused on the process of horizontal mathematization [12]; we analyzed the texts of word problems to highlight possible criticalities in their components among those highlighted by Nesher [10] logical structure, semantic component, syntactic component.

Based on this evidence, the experimental work was set up and involved a sample of grade 2 students from “Colozza” school in Campobasso in the south of Italy. A pre-test

was initially administered to both classes; it was constructed using the word problems selected from the grade 2 INVALSI test to assess skills and knowledge *ex-ante*. Subsequently, the activity was carried out only on the experimental sample. We started from a guided discussion with the students on the pre-test results to investigate whether the difficulties encountered were, in fact, in agreement with the experimental hypotheses. Then we did a workshop in pairs to solve new word problems. These problems were designed ad hoc to be equivalent and maintained the same complexity in terms of the textual components highlighted in the theoretical framework [9,10]. At the end of the experimentation, we administered a post-test to both sample classes, with new word problems equivalent to those of the laboratory activity and the pre-test, to verify if the didactic interventions had produced some changes in the experimental class compared to the control class.

The data collected showed a fair improvement in the experimental sample, probably due to the didactic interventions made in the period between the pre-test and post-test administration. The guided discussion after the administration of the pre-test was useful to show that the difficulties encountered by the students were due to the lack or misinterpretation of the text of the problems caused by the observed complexities in the textual components. The experimentation and, in particular, the discussion highlighted the coherence between the a priori hypotheses and the actual obstacles students encountered in the test. The weaknesses in the semantic and syntactic components played a crucial role in the process of horizontal mathematization, which subsequently influenced vertical mathematization. The discussion also compared the various strategies adopted by the students to make them helpful material to be exploited in the next workshop activity.

The experiment had a positive effect on grade 2 students. This fact allowed us to hypothesize that early identification of the difficulties and a specific intervention could help to overcome future obstacles in understanding the text of word problems in the arithmetic field. Only for one question, the student's performance in the experimental class was lower than in the control class. However, it was the only question with weaknesses in all three text components and related to the conceptual field of multiplicative structures. This may suggest that early intervention has been successful in the conceptual additive field but still too early in the multiplication field. This is a conceptual field not yet widely explored by students at this school level, and for this reason, word problems in that field may be difficult to manage in the presence of critical aspects in the text.

Such a study emphasizes the formative value of standardized assessment. Indeed, a central issue in epistemological and educational debates is integrating the results, methods, and theoretical tools of standardized assessment into the formative assessment process [26]. These results showed that standardized assessments can provide teachers with tools and benchmarks for diagnostic assessment [27], which is about finding out what students have and have not achieved, as well as their strengths and weaknesses related to different contents. All this is already known; what is new is that test results from later grades (in our case, grade 5) can also be a diagnostic and formative assessment resource for the lower grades (grade 2) and vice versa. Therefore, we can state that students' results in previous grades tests can be interpreted as predictive of possible performance in the subsequent grades. Hence, our analysis can provide information for designing timely educational interventions to support students in difficulty.

However, the positivity of the experimentation should be consolidated by proposing subsequent experimentation to the same cohort of students when attending grade 5. A critical point of the experiment was the too-small size of the sample. Although this study started from a purely quantitative basis, the interviews and the subsequent educational intervention had a qualitative perspective and involved a tiny group of students. In order to overcome this limitation of the research, in the future, we intend to carry out more similar experiments on a numerically significant sample in order to validate or not what emerged here also in grade 5 students.

A possible future perspective for this study could be to extend such research to the other dimensions of the INVALSI tests in Mathematics, such as "Knowing" or "Arguing",

to see whether grade 2 tests can also be a tool to highlight early difficulties and intervene accordingly. In another way, one could consider the same chain of questions by extending it to school grades above 5, organizing, also in this case, specific educational interventions.

Author Contributions: The authors contributed equally to the writing of the article. F.F. oversaw the experiments in the classes. R.C. and A.L. oversaw the research and analysis of the data. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and according to the Ethical code of the Industrial Engineering Department of University of Salerno.

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

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy reasons.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. INVALSI Test Grade 5 of A.S. 2011/2012

Question 8

Uncle Elena goes to the pastry shop; he buys a chocolate cake and a cream cake. The total price for the two cakes is 24 euros. The chocolate cake costs 6 euros extra than the cream cake.

- How much money is the cream cake?
- How did you find the answer?

Question 23

Carla decided to go to England for six months. Before leaving, she exchanges 2000 euros in pounds. At the bank, the exchange rate is: 1 euro = 0.95 pounds. How many pounds will Carla receive? Write calculations to find the correct answer and then write the result.

Result: pounds.

Question 30

Marta loves comic strips. Her grandmother gave her 20 euros, and Marta decided to spend them on buying some comics that cost 2.20 euros. How many comics can she buy at most?

... ..

Question 32

Antonella parks in a skyscraper parking garage which is located four floors under the ground floor. She gets on the elevator for twenty-four floors. At which floor will Antonella get out from the elevator?

- 16
- 20
- 24
- 28

Appendix B. INVALSI Test Grade 5 A.S. 2008/2009

Question 2



Giovanni buys some pencil boxes like this one: [. . .]
He has 18 pencils overall. How many boxes has he bought?

- A. 2
- B. 3
- C. 6

Question 4



Lucia's mum is doing grocery shopping. The board shows the number you see here. The mum got ticket number 39. How many people will be served before her?

- A. 38
- B. 3
- C. 5

Question 8

Carlo thinks of a number; he adds 26 and obtains 41. Who will guess the number Carlo thought about between Matteo, Laura, and Chiara?

- A. Matteo: is the number 25
- B. Laura: is the number 15
- C. Chiara: is the number 67

Question 9

Do the double of 14, and then double the solution that you have obtained again. What number do you arrive at?

- A. 28
- B. 42
- C. 56

Question 16

Daniele has three packs of 40 cards each. Nicola has 120 cards. How many cards has Nicola to buy to have the same cards number as Daniele?

- A. 80
- B. 40
- C. 0

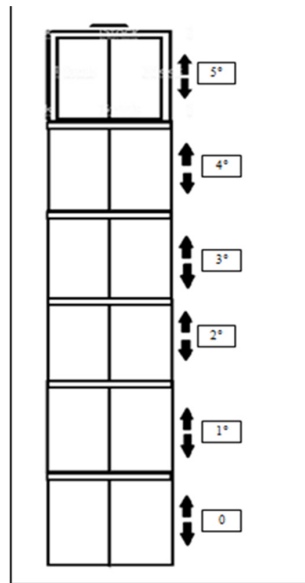
Question 22

Giovanni buys 2 pencil boxes like this in the picture.
But then he loses 3 pencils. How many pencils are left?

- A. 9
- B. 15
- C. 3

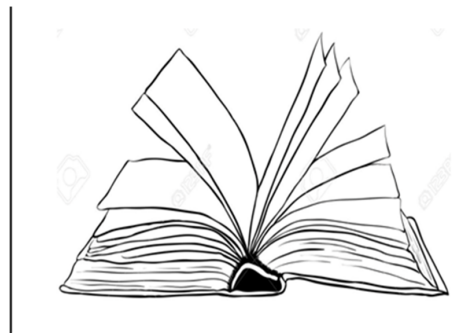


Appendix C. Items about the Laboratory Activity Carried Out in the Experimental Class



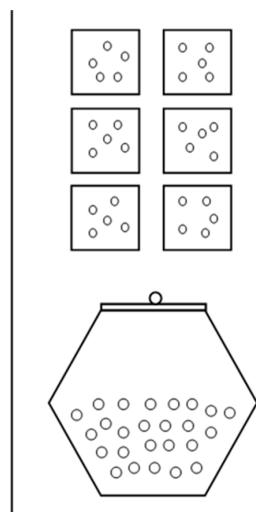
Question 1

Antonio is on the first floor of his apartment building. He has to come up to the fifth floor to his grandmother. How many floors will Antonio come up to go to his grandmother?



Question 2

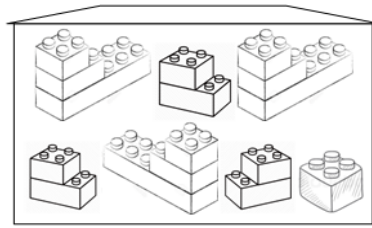
These days Emanuele is reading "The Jungle Book", and he has read to page 74. There are 12 pages remaining to the end. How many pages does the book have?



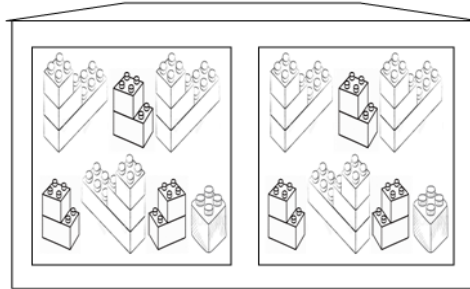
Question 3

Mum gives some fruit candies to her children. Laura has six boxes with 5 candies each. His brother Paolo has 25 candies organized in a big can. How many more candies has Paolo than Laura?

Laura's candies
Paolo's candies



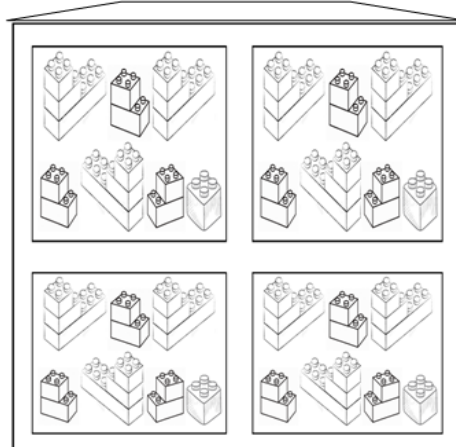
A Natale riceve un'altra scatola contenente il doppio delle costruzioni che gli ha regalato la nonna



Quante costruzioni possiede ora Damiano?

Scrivi l'operazione che hai svolto qui →

All'Epifania riceve il doppio delle costruzioni totali che possiede.



Quante costruzioni ha adesso in tutto Damiano? Scrivi l'operazione che hai svolto qui →

Question 4

For his birthday, Damiano receives a box that contains 16 lego from his grandmother. At Christmas, he receives another box that contains twice as many lego as his grandmother gave him.

Question 4

How many lego does Damiano have now? Write the calculation you did At Epiphany, he receives twice as many lego that he owns. How many lego does Damiano have overall? Write the calculation you did

Question 5

Choose a number. Subtract 5 from this number. What do you get? Add 27 to the result. What number do you get?

Question 6

Anastasia has 14 euros in her pouch, and the father gives her another 5 euros. While Anastasia was going to the library to buy a book, she lost 4 euros. How many euros remains for her to buy the book?

Appendix D. Post-Test

Question 1



Francesca buys some tempera boxes like this in the picture. She has 35 temperas overall. How many boxes has Francesca bought?

- A. 5
- B. 6
- C. 10

Question 2



In the football rank, Juventus is in position number 1. If it lost the next game, Juventus moved to position number 6. How many positions is the football team relegated to?

- A. 1
- B. 5
- C. 7

Question 3

An airplane is heading to Rome. When it lands at the airport, 28 people get in. Now on the airplane, there are 70 people overall. How many passengers were the airplane carrying before landing in Rome?

Question 4

Rachele has split a pizza into 8 slices, and she ate half. Then her sister Alice wanted to eat half of the number of slices Rachele ate. How many pizza slices are left?

- A. 8
- B. 4
- C. 2

Question 5

Guido has 83 books in the basement. His brother Marco has 4 bookcases. Each one contains 20 books. How many books does Marco have to add in his bookcase to own the same number of books as Guido?

- A. 3
- B. 20
- C. 63

Question 6

Three teachers talk about children of their classroom:

Teacher Tiziana says: today in 3^A, there were 10 children.

Teacher Irma says: today in 3^B, there were half the children in 3^A!

Teacher Maria says: today in 3^C, there were twice as 3^A.

How many children there are in 3^B and in 3^C?

- In 3^B there are 20 children; in 3^C 5
- In 3^B there are 5 children; in 3^C 20
- In 3^B there are 7 children; in 3^C 24

Question 7

Little Red Riding Hood's grandmother is bringing 3 full baskets of apples to her grandchild, like you see in the picture. While she is walking into the woods, the grandmother loses 10 apples. How many apples does the grandmother have still to give to Little Red?



- A. 9
- B. 26
- C. 36

References

- Ball, D.; Bass, H. Making mathematics reasonable in school. In *A Research Companion to Principles and Standards for Mathematics*; Kilpatrick, J., Martin, W.G., Schifter, D., Eds.; NCTM: Reston, VA, USA, 2003; pp. 27–44.
- Kilpatrick, J.; Swafford, J.; Findell, B. *Adding It up: Helping Children Learn Mathematics*; National Academy Press: Washington, DC, USA, 2001.
- Verschaffel, L.; Schukajlow, S.; Star, J.; Van Dooren, W. Word problems in mathematics education: A survey. *ZDM* **2020**, *52*, 1–16. [[CrossRef](#)]
- Schoenfeld, A.H. *Mathematical Problem Solving*; Academic Press: New York, NY, USA, 1985.
- Verschaffel, L.; Greer, B.; De Corte, E. *Making Sense of Word Problems*; Swets & Zeitlinger: Lisse, The Netherlands, 2000.
- Jupri, A.; Drijvers PH, M. Student difficulties in mathematizing word problems in algebra. *EURASIA J. Math. Sci. Technol. Educ.* **2016**, *12*, 2481–2502. [[CrossRef](#)]
- Wijaya, A.; Van den Heuvel-Panhuizen, M.; Doorman, M.; Robitzsch, A. Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. *Math. Enthus* **2014**, *11*, 555–584.
- Vilenius-Tuohimaa, P.M.; Aunola, K.; Nurmi, J. The association between mathematical word problems and reading comprehension. *Educ. Psychol.* **2008**, *28*, 409–426. [[CrossRef](#)]
- Daroczy, G.; Wolska, M.; Meurers, W.D.; Nuerk, H.C. Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Front. Psychol.* **2015**, *6*, 348. [[CrossRef](#)] [[PubMed](#)]
- Nesher, P.; Hershkovitz, S.; Novotna, J. Situation model, text base and what else? Factors affecting problem solving. *Educ. Stud. Math.* **2003**, *52*, 151–176. [[CrossRef](#)]
- Nesher, P. Levels of description in the analysis of addition and subtraction word problems. In *Addition and Subtraction: A Cognitive Perspective*; Carpenter, T.P., Moser, J.M., Romberg, T.A., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1982; pp. 25–38.
- Treffers, A. Integrated column arithmetic according to progressive schematisation. *Educ. Stud. Math.* **1987**, *18*, 125–145. [[CrossRef](#)]
- Franchini, E.; Lemmo, A.; Sbaragli, S. Il ruolo della comprensione del testo nel processo di matematizzazione e modellizzazione. *Didattica Della Matematica. Dalla Ricerca Alle Pratiche D'aula* **2017**, *1*, 38–63. [[CrossRef](#)]
- Gerofsky, S. A linguistic and narrative view of word problems in mathematics education. *Learn. Math.* **1996**, *16*, 36–45.
- Freudenthal, H. *Revisiting Mathematics Education. China Lectures*; Kluwer Acad. Publ.: Dordrecht, The Netherlands, 1991.
- Capone, R.; Adesso, M.G.; Del Regno, F.; Lombardi, L.; Tortoriello, F.S. Mathematical competencies: A case study on semiotic systems and argumentation in an Italian High School. *Int. J. Math. Educ. Sci. Technol.* **2021**, *52*, 896–911. [[CrossRef](#)]
- Duval, R. *Interaction des Différents Niveaux de Représentation dans la Compréhension de Textes. Annales de Didactique et de Sciences Cognitives*; IREM: Strasbourg, France, 1991; pp. 136–193.
- Richardson, K. A Design of Useful Implementation Principles for the Development, Diffusion, and Appropriation of Knowledge in Mathematics Classrooms. Unpublished Doctoral Dissertation, Purdue University, West Lafayette, IN, USA, 2004.
- English, L.D.; Watters, J.J. Mathematical Modeling in the Early School Years. *Math. Educ. Res. J.* **2004**, *16*, 59–80.
- MIUR. *Indicazioni Nazionali per il Curricolo Della Scuola Dell'infanzia e del Primo ciclo D'istruzione*; Ministero dell'Istruzione, dell'Università e della Ricerca: Roma, Italy, 2012.
- Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*; Danmarks Paedagogische Institut: Copenhagen, Denmark, 1960.
- Vergnaud, G. The theory of conceptual fields. *Hum. Dev.* **2009**, *52*, 83–94. [[CrossRef](#)]
- Bolondi, G.; Branchetti, L.; Ferretti, F.; Lemmo, A.; Maffia, A.; Martignone, F.; Matteucci, M.; Mignani, S.; Santi, G. *Un Approccio Longitudinale per L'analisi Delle Prove INVALSI di Matematica: Cosa ci Può Dire Sugli Studenti in Difficoltà?* Falzetti, P., Ed.; Concorso di idee per la ricerca; Cleup: Padova, Italy, 2016; pp. 81–102.
- Zan, R. The crucial role of narrative thinking in understanding problems. *Dalla Ricerca Alle Pratiche D'aula.* **2017**, *1*, 45–57.
- Zan, R. *I Problemi di Matematica: Difficoltà di Comprensione e Formulazione del Testo*; Carrocci: Roma, Italy, 2016.
- Looney, J.W. *Integrating Formative and Summative Assessment: Progress Toward a Seamless System?* OECD Education Working Papers, No. 58; OECD Publishing: Paris, France, 2011.
- Harlen, W. *Teaching, Learning and Assessing Science 5–12*, 3rd ed.; Paul Chapman Publishing: London, UK, 2000.