ACTA DIDACTICA NAPOCENSIA

Volume 16, Number 1, 2023 - DOI: 10.24193/adn.16.1.17

HOW TASK INSTRUCTION INFLUENCES THE PERFORMANCE AND THE TEXT READING PATTERN IN CASE OF INCONSISTENT COMPARE PROBLEMS IN PRIMARY SCHOOL

Ágnes BARTALIS and Iuliana ZSOLDOS-MARCHIȘ

Abstract: Compare word problems are quite difficult for primary school students, especially inconsistent ones when the relational term from the problem text is not consistent with the arithmetic operation required for the solution. The present study investigates the effect of different task instructions on successfulness in solving inconsistent compare word problems and in the reading pattern of the text of these problems. Therefore, in this study eye-tracking was used to monitor participants eye movements while reading and solving compare word problems. Fifty-six 4th graders' eye-movement behavior and responses were collected to analyze the fixation time and fixation number on different data from the text of the tasks and the solution time on each task. In the test inconsistent compared problems were given requiring the use of multiplication/division operations. The problems were divided into three groups with different task instructions. In the first group of problems the requirement was to solve the problems, pupils could freely concentrate on the text elements they wanted as the text of the problem was classically given. In the second case students were forced to spend more time focusing on the context of the problem as they first got the text of the problem with symbols instead of numbers. In the case of the third group of problems, pupils had to create the graphical representation of the word problem considering the value of the given variable and the relation between the two variables. The results show that the instruction type influenced both performance and reading pattern. When students are forced to spend more time reading the entire text (problems given first with symbols), the solution success rate is higher. Graphical representation of the problem's data also contributes to a better performance. There are also differences in percentages of reading times spent on fixation different key elements of the text or different sentences of the problem. The key element fixated for a higher percentage of time depends on the type of instruction.

Key words: compare word problem, inconsistency, eye-tracking, primary school.

1. Introduction

Although solving word problems seems to be an everyday activity in elementary school, this is precisely the part of mathematics education that is closest to everyday life, and perhaps even most likely to provide usable knowledge for students' life and future career. Therefore, it is not by coincidence that much research today deals with this problem area (Andersson, 2007; Blum & Leiß, 2007; Boonen et al., 2016; Daroczy et al., 2015, Hegarty et al., 1992, 1995; Verschaffel & De Corte, 1993).

There are more types of word problems in primary school, among which the compare word problem is one of the most difficult types, with even three times lower success rate than other types of problems, as tasks based on exchange or combination (Stern & Lehrndorfer, 1992). In the compare problems, the value of the second variable must be calculated based on the value of the first variable and the given relation (De Koning et al., 2017), see the examples in Table 1. If the relational expression in the task is consistent with the requested arithmetic operation, then we refer to a consistent task. Conversely, the problem is inconsistent if the opposite arithmetic operation than suggested by the relational word is required for the correct solution (Lewis & Mayer, 1987).

Received June 2023.

Cite as: Bartalis, Á. & Zsoldos-Marchiș, I. (2023). How task instruction influences the performance and the text reading pattern in case of inconsistent compare problems in primary school. *Acta Didactica Napocensia*, 16(1), 252-266, https://doi.org/10.24193/adn.16.1.17

Table 1. Example of consistent and inconsistent compare problems

Consistent compare word problem	Inconsistent compare word problem
A pen costs 15 lei. *A pencil costs 3 times more than a pen.* How much does a pencil and a pen cost in total?	A pen costs 15 lei. *This is 3 times more than the price of a pencil.* How much does a pencil and a pen cost in total?

The question is why these types of tasks cause difficulty. According to Riley & Greeno (1988), reading comprehension and situational understanding make it difficult to solve compare word problems. Situational comprehension in the task may cause difficulties because comparing sets is not a usual or routine action for children. This can also be linked to Piaget's theory, according to which the cognitive function of children of this age is closely linked to their actions, but while actions can be easily linked to addition or exchange, not to comparison. Contrary to this concept, Stern & Lehrndorfer (1992) argues that in many cases small children are used to sharing and thus also to comparison, for example when they have to share a piece of candy with their partner, there is often an opportunity to compare two quantities.

Since many studies have proven that the performance is much lower in the case of solving inconsistent tasks than consistent tasks (Orrantia & Múñez, 2013; Pape, 2003; Riley & Greeno, 1988), in this study we only focus on inconsistent compare problems.

The biggest challenge in solving inconsistent problems is that the relational expression has to be translated, since it is not consistent with the requested arithmetic operation (Lewis & Mayer, 1987). However, for this reason, an adequate mental representation must be constructed based on the problem, since if this step is omitted and only the numbers in the task and the relational expression are taken into account, there is a high chance of a wrong solution. In addition to the construction of the appropriate episodic situational model, it is also a crucial step to construct a magnitude-based mental representation in regard to the variables included in the task (Coquin-Viennot & Moreau, 2003; Orrantia & Múñez, 2013).

The present study is based on these two prominent mental representation models, which, as we will see below, can play a key role when solving inconsistent compare problems. Our aim is to support children's problem-solving process with instructions that, based on well-founded literature, can really contribute to successful problem solving. In addition to our objective of trying to discover how effective the various instructions we give prove to be, our further purpose is to get to know the effect on their problem-solving process from the „inside”, which is why we use Eye-tracking. Several studies have already been based on the possibilities provided by the Eye-tracking device (De Corte et al., 1990; Hegarty et al., 1992; van der Schoot et al., 2009), but no research has yet been published that would have assessed the effect of different instructions on solving inconsistent compare word problems.

2. Theoretical background – Problem solving models

The process of problem solving is modeled in different ways in previous studies (Blum & Leiß, 2007; Csíkos, 2003; Dröse, 2019; Schnotz, 1994), but they mostly agree that this process cannot be described linearly, but cyclically. There is an agreement that when solving a problem, we often return to previous steps, for example after reading the text for the first time we often return to different elements or even to reading the entire text. Csíkos (2003) describes this cyclical process in such a way that the first (reading the task) and the last two steps (creating a solution plan and its implementation) are fixed, however, cycles can develop between the steps in between (a. creating or updating a semantic network, b. selecting numbers and keywords and c. creating or updating a model of the problem). Often, these returns make it possible to solve the task correctly.

Dröse (2019) integrates several problem-solving models. In his theory, she uses the models of Schnotz (1994) and Reusser (1997) as a basis, and based on them, he separates the following steps: identifying the relevant units of the text, structuring, and relating the elements, interpreting the meaning of the units, inferring new situational elements, and integrating them into a mental representation.

Similar steps are distinguished by Blum & Leiß (2007). In this modeling cycle, the starting point is the real situation (the text of the task), and by understanding the problem the situation model can be constructed. This is followed by structuring and simplifying the information, which enables the building of a realistic model. The mathematization of it leads to the mathematical model. In this phase, adequate arithmetic operations will be performed, and its result must be reversed to obtain the real result that can be interpreted based on the situational model. In this concept, the individual problem-solving process is not described in a linear way either, since the process is also influenced by the individual's mathematical tools, so multiple steps forward and backward can occur.

Among the steps of the problem-solving models described above, the creation of the situation model appeared several times. According to some views, this step can be omitted, but in this case, there is less chance of solving the task correctly (Hegarty et al., 1992).

2.1. Episodic situational representation

In general, the first steps of the problem-solving process reflect what kind of procedure the problem solver uses: situational model or direct translation procedure (Hegarty et al., 1995). These two approaches differ mainly from a qualitative point of view.

During the direct translation procedure, the problem solver extracts the content of the numbers and key terms it considers important, and performs the arithmetic operations based on them (Kontra, 2001). All procedural models based on different action rules can be linked to this concept. These rules, algorithms or procedures are usually learned by the problem solver in the classroom, which can be useful in some cases, but in many cases are not enough for the correct solution (Coquin-Viennot & Moreau, 2003). This includes, for example, that all numbers usually found in the text of the task must be used, in the order in which they appear in the text. Many times, these rules can be misleading, on the other hand, they do not improve students' problem-solving skills.

Contrary to this approach, by constructing an episodic situation model by the problem solver, a quality representation is obtained that creates a connection between the text and the mathematization of the problem (Kintsch, 1998). Stern & Lehrndorfer (1992) have a similar opinion about this approach, according to whom the main function of the episodic situation model is to bridge the gap between language understanding and mathematical problem-solving knowledge. Therefore, this model can also be interpreted as a process that helps to see beyond the text.

The episodic situational model significantly contributes to the correct solution of the problem, since it is created even before the construction of the mathematical model, thus the problem solver does not only consider the numbers in the problem and the relational relationships between them (Coquin-Viennot & Moreau, 2003). Although it is just the construction of the situational model that makes solving the tasks difficult (Riley & Greeno, 1988), this may in some cases be the key to the correct solution.

Although many factors can have an influence on the success of solving compare word problems, the situational context also matters to a large extent. Stern & Lehrndorfer (1992) demonstrated that abstract language expressions such as "How much more?" are not the problem in themselves, since the problem task which is embedded in an enriched context also has a great influence. Therefore, in many cases, these tasks are considered difficult because the children lack adequate situational knowledge.

2.2. Magnitude-based mental representation

There is another important aspect of the process of problem solving that is related to the relationship between variables. This is the magnitude-based mental representation that the problem solver constructs based on the text of the problem, or more precisely, the relational structure.

In their study, Orrantia & Múñez (2013) clearly highlighted the aspect of problem solving according to which the magnitude of perceptual information is routinely activated in the solution process. When this representation is activated, what takes place is a mental simulation that is linked to individual, previously lived experiences. Therefore, if we want to help the problem solver in this process, we can use external representations that adequately reflect the magnitude of the variables appearing in the problem. (Vicente et al., 2008)

Support with an external mental representation may be particularly important when dealing with an inconsistent compare word problem. The explanation for this is that the solution of inconsistent tasks requires longer processing and the error rate in the task solution is also higher. In this case, the relation found in the relational sentence must be kept active in the working memory, while the other information in the text is also processed: connecting the known and unknown quantities with the variables and developing the appropriate elaboration plan (Orrantia & Múñez, 2013). Therefore, the correctly constructed magnitude-based mental representation contributes to the correct solution.

Despite the fact that the solution process of the comparative word problems is supported by an external representation, we cannot offer this help in all cases, as it can provide convenience for the problem solver. In certain situations, it may be necessary for the problem solver to create the representation of the relationship between the variables on the basis of how the magnitude-based mental representation was previously constructed based on the task.

3. Method

3.1. Scope and research questions

The scope of the research was to find out if there are differences in solution success and reading patterns of the problems' text based on different instructions when giving the problems.

The research aims to find answers to the following research questions:

1. How does the type of instruction influence the performance?
2. How does the type of instruction influence the fixation durations on different key elements of the text of the problem?
3. How does the type of instruction influence the fixation durations on different sentences of the text of the problem?

3.2. Participants

Convenience sampling was used to obtain a sample of 56 Hungarian speaking fourth grade students (10-11 years old), 20 girls and 36 boys from two regular schools located in Cluj-Napoca, Romania. Since the limitations of the eye-tracking studies (only one person can be recorded at a time, labor resources etc.) enabled only a convenience sampling, however, participants were selected randomly from each class. Therefore, classes participating in the investigation are from urban areas, representing the inner-city schools.

Prior to the study being carried out, parents provided informed consent about the purpose of the study. In addition, the details of the study were discussed with the teachers of pupils involved in the study. Vertical drift and poor reading ability were the exclusion criteria.

3.3. Material

The material (the test) consisted of 7 two-steps inconsistent compare word problems (tasks), containing one warm-up task and 6 test tasks. The problem sheet was written in Hungarian. The students got the problem sheet on the computer with eye-tracking technology included and solved the problems on a sheet of paper.

The tasks are based on the same scheme as in previous research (Hegarty et al., 1992, 1995; Lewis & Mayer, 1987; van der Schoot et al., 2009a), but in different familiar contexts, for example:

> The price of a diary is 20 lei.
> This is 5 times less than the price of a book.
> How much does a book and a diary cost in total?

As the example shows, each task is based on three sentences: assignment sentence, relational sentence, and the question. In addition, when creating the tasks, it was also considered that the words with similar role in different problems should be of the same length and consist of the same number of syllables.

According to the pupils' cognitive level and the national curriculum requirements, the solutions required multiplication/division of a two-digit number with a one-digit number and addition of a one/two-digit number with a one/two-digit number.

The structure of the test can be seen in Figure 1. After the warming-up task the problems were divided into three groups with different task instructions. In case of each instruction type two tasks were given, their order was randomly selected. In addition, after each task an empty slide was included to present the instruction and to answer some incidental questions that may arise during the test.

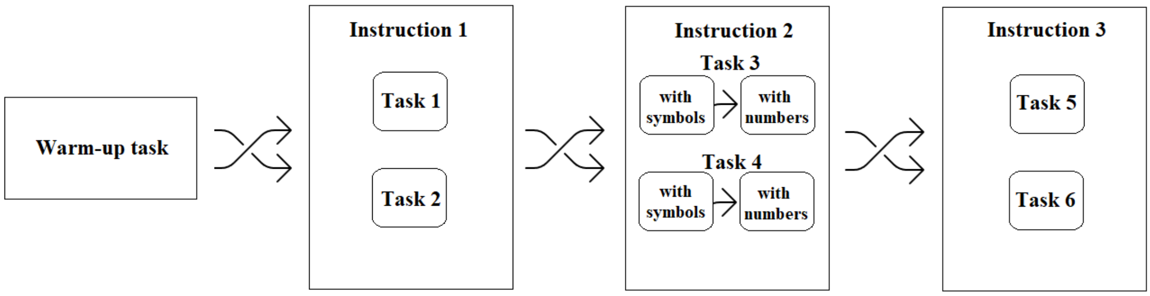

Figure 1. Structure of the test

In the following the different instruction types are presented.

In the first case (task 1 and task 2) the problem text is given without any additional instruction. Thus, pupils could freely concentrate on the text elements they wanted. We will call these problems *simple problems* in the following.

> Solve the following problem!
>
> The price of a pen is 15 lei.
> This is 3 times less than the price of a pencil.
> How much does a pen and a pencil cost in total?

In the second case (*task 3 and task 4*) students were forced to spend more time focusing on the context of the problem. For this purpose, they got the task without any number or data, the numbers were replaced by different symbols. An example of such a task is given below:

> What operation would you use if instead of symbols numbers are given? Write down the signs of those operations (+ / - / x / :) which lead to the correct solution!
>
> The price of a plate is □ lei.
> This is ♦ times less than the price of a pot.
> How much does a pot and a plate cost in total?

Consequently, they could not actually solve the problem, however, they had time to think about it, to understand the context, to familiarize themselves with the situation and to start elaborating the solution strategy. Pupils were asked only one question: what arithmetic operation/ operations would they perform to solve the problem correctly? We will call these problems as *problems with symbols* in the following. After answering this question, the same problem with numbers instead of symbols appeared, so they could solve the problem properly:

> Now you can see the numbers. Solve the problem!
>
> The price of a plate is 11 lei.
> This is 3 times less than the price of a pot.
> How much does a pot and a plate cost in total?

We will call these problems in the following as *problems with numbers after symbols*. Actually, in the research problems with symbols and problems with numbers after symbols are treated separately, but they are two variants of task 3 respectively task 4.

In the third case (task 5 and task 6), pupils got the problem, and they were asked not only to solve the problem, but also had to create the graphical representation of the word problem taking into account the known values of the variables and the relation between the variables:

> Represent the problem for the solution. Then write down the solution!
>
> The price of a scissors is 18 lei.
> This is 6 times less than the price of a rubber.
> How much does a scissors and a rubber cost in total?

We will call these problems as *problems with graphical representation.*

As observable above, even if the focus was on three different instruction types, the given problems can be grouped into four types: simple problems, problems with symbols, problems with numbers after symbols, and problems with graphical representation.

3.4. Apparatus

Although the majority of eye-tracking research takes place in a laboratory environment, in this research the pupils were assessed at their school. We chose this location because it is particularly important in this age group that pupils are assessed in the environment they are used to. Accordingly, it was possible to create a room within the school that met the lighting conditions required for the assessment, as well as maintaining the focus of attention by limiting distraction factors. Furthermore, it was easier to assess them in their own school environment than to take them away from there to a foreign place.

For data collection, Tobii Pro Fusion hardware and Tobii Pro Lab Screen-Based Edition software were used that samples pupil location at the rate of 250 Hz. The pupils' heads were set at a distance of about 60 cm from the screen according to the accurate data collection. The Eye-tracker was located under the laptop screen (15,6-inch) and the tasks appeared on it.

3.5. Procedure

Before starting data collection, some information was given to the pupils about: type of tasks, number of tasks, lack of time constraints, different task instructions that will be given later, position on the chair (relatively comfortable and stable), distance between the screen and the eyes, avoiding covering his/her eyes, calibration procedure with 5 calibration and 4 validation points, how to use the answer sheet, when it is allowed to talk and how to move from one problem to another.

Informing the pupil was followed by the calibration and the warm-up task. After that, he/she was given the opportunity to ask a question, and only then received the first instruction. He/she was then instructed to solve the first two tasks in any way he/she wished. Therefore, there were no expectations other than that he/she had to write the result on the answer sheet. According to the second instruction, the requirement is to read the problem that appears on the screen (here the text of the problem appears without data with symbols instead of the numbers) and decide what arithmetic operation should be performed in order to solve the problem correctly. When the answer was written down, the same task appeared in front of him/her once again with numbers instead of symbols. The requirement was to determine the result of the problem. Finally, for the last two tasks, pupils were instructed to draw a graphical representation of the word problem next to the result.

3.6. Data analysis

For analyzing the reading pattern of pupils' while solving inconsistent compare problems, some Area of Interests (AOIs) were determined. These AOIs were associated with those key data from the problem text, the use of which were essential in order to solve the problem. These data were: the two numbers, the relational terms (more or less) and the pronominal reference word (this). In addition, different sentences of the problem were also analyzed individually: the statement sentence, the relational sentence, and the question of the problem. For the established AOIs fixation duration and number of fixations were recorded, as well as solution time for every problem. Furthermore, in order to make the numerical data easier comparable, we converted them into percentage data. Descriptive statistics (mean, standard deviation) and comparisons with repeated-measure ANOVA were performed in data analysis.

4. Results

4.1. Scores, solution times, and reading times of different type of problems

In case of all the four types of problems each correctly solved task was counted (1 point for each correctly solved problem, no fractional points were assigned). The possibly obtained highest score for each problem type was 2. The highest average score was obtained in case of the problems with numbers after symbols (M =1.18, SD = 0.81), and the lowest average score in case of the problems with symbols (M =0.84, SD = 0.80), see Table 2. A repeated-measures ANOVA was performed to compare scores obtained for different problem types which indicated significant differences ($F(3) = 5.318$, $p = .002$). The solution success in case of problems with numbers after symbols is significantly higher than in case of the simple problems ($p = .006$) or problems with symbols ($p = .004$) and higher, but not significantly, than in case of problems with graphical representation ($p = .302$).

Solution times were also analyzed, in Table 2 the means (M) and standard deviations (SD) are presented in case of each problem type. A repeated-measures ANOVA was performed to compare solution times in case of different problem types. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi 2(5) = 42.304$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .757$). The differences on solution times for different problems are statistically significant ($F(2.27) = 74.072$, $p < .001$). The solution time for problems with graphical representation is significantly higher than the solution times of the other three types of problems ($p < .001$ in all three cases).

Table 2. Means and standard deviations of score, solution time, and fixation durations in case of different type of problems

Problem type	Score (number of correct solutions)		Solution time (ms)		Fixation duration on the task's text (ms)		Percentage of solution time spent fixating on the text
	M	SD	M	SD	M	SD	
Simple problem	0.86	0.82	58450.30	4515.26	15540.97	9081.54	58.44
Problem with symbols	0.84	0.80	82185.54	8481.75	23461.58	16125.39	68.73
Problem with numbers after symbols	1.18	0.81	79210.68	7325.99	18366.44	13499.95	73.50
Problem with graphical representation	1.02	0.84	193345.90	11841.46	26105.84	17607.02	67.44

As regards reading times (fixation durations on the tasks' text), Table 2 contains the means (M) and standard deviations (SD) in case of each problem type. A repeated-measures ANOVA was performed to compare reading times in case of different problem types. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi 2(5) = 12.932$, $p = .024$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = .887). The differences on reading times for different problems are statistically significant ($F(2.66) = 8.756$, $p < .001$). The reading time for simple problems is significantly lower than the reading time for problems with symbols (p = .003) and problems with graphical representation (p < .001). Also, reading time for problems with graphical representation is significantly higher than reading time for problems with numbers after symbols (p = .004).

Table 2 also contains the percentage of the solution times spent with reading the text of the problems. To calculate these percentages the solution times and the fixation durations on the task's text is taken into consideration.

Solution times and fixation duration on tasks (reading times) are strongly positively correlated in case of all type of problems (r = 0.660, p < .001 in the case of simple problems, r = 0.721, p < .001 in the case of problems with symbols, r = 0.689, p < .001 in the case of problems with numbers after symbols, and r = 0.614, p < .001 in the case of problems with graphical representation). The solution time and performance are strongly negatively related only in the case of the problems with numbers after symbols. There is not any strong correlation between reading time and performance in any of the problem types. There is a mild negative correlation in the case of problems with numbers after symbols (r = - 0.286, p = .033).

4.2. Percentage of time fixating on each key element in case of different types of problems

In this section the fixation durations on each text element are studied comparing different types of problems. As reading times (total fixation duration) for different types of problems were different, to compare fixation durations on different text elements, percentages from the total fixation durations were calculated (for example, to calculate the percentage of reading number 1 in case of the simple problems the fixation duration on number 1 and the total fixation duration on the text in case of simple problems were used). The means (*M*) and standard deviations (*SD*) are presented in Table 3 for 4 text elements: number 1 (the number from the first sentence of the problems, the value of the first variable), number 2 (the number from the second sentence, the value which describes how many times the first variable is bigger/smaller than the second variable), relational word (which is in the second sentence), and the pronominal reference word (also from the second sentence, the word which refers to the first variable from the first sentence).

Table 3. Means and standard deviations on the percentages of time spent fixating on different text elements

Problem type	**Number 1 (%)**		**Number 2 (%)**		**Relational word (%)**		**Pronominal reference word (%)**	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Simple problem	21.30	12.24	25.70	8.96	12.02	4.56	6.29	4.42
Problem with symbols	11.25	7.97	37.54	10.52	15.12	5.99	6.22	4.20
Problem with numbers after symbols	24.09	15.99	31.59	10.61	14.31	5.57	4.97	4.41
Problem with graphical representation	18.12	9.70	30.02	10.33	16.17	4.84	4.29	3.51

A repeated-measures ANOVA was performed to compare the percentages of reading different text elements in case of different types of tasks.

In case of number 1 the Mauchly's test indicated that the assumption of sphericity had been violated ($\chi 2(5) = 18.736$, $p = .002$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .824$). The differences in percentages of time spent fixating on number 1 in case of different types of problems are statistically significant ($F(2.47) = 20.137$, $p < .001$). Students fixated a significantly lower percentage of time on number 1 in case of the problem with symbols than in the case of the other three problem types ($p < .001$ in all cases). The highest percentage of time fixating on number 1 is in the case of problems with numbers after symbols, which is significantly higher than in the case of problems with symbols ($p < .001$) and problems with graphical representation ($p = .002$). In Figure 2 the problem types are given in the increasing order according to the percentage of time fixating on number 1 and the *p*-values obtained with Holm post hoc comparisons.

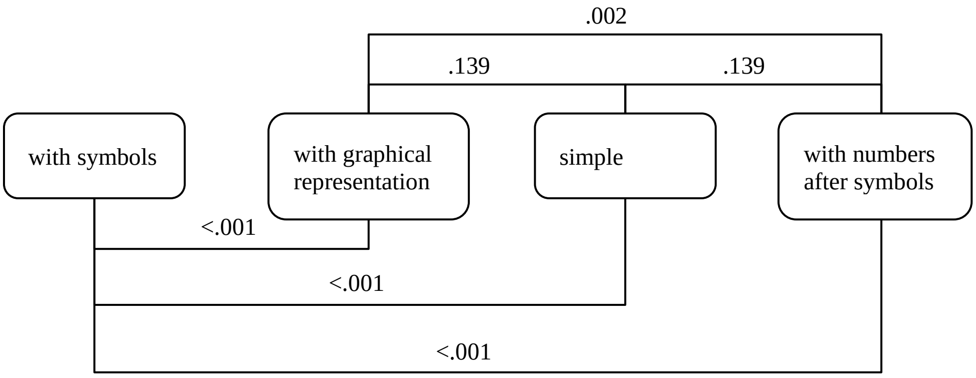

Figure 2. Increasing order according to the percentage of time fixating on number 1

The differences in percentages of time spent fixating on number 2 in case of different types of problems are statistically significant ($F(3) = 16.525$, $p < .001$). Students fixated a significantly lower percentage of time on number 2 in case of simple problems than in the case of the other three problem types ($p < .001$ for comparison with problems with numbers after symbols and problems with symbols, $p = .024$ for problems with graphical representation). The highest percentage of time fixating on number 2 is in the case of problems with symbols, which is significantly higher than in the case of other three types of problems ($p < .001$ for comparison with problems with graphical representation and simple problems, $p = .002$ with problems with numbers after symbols). In Figure 3 the problem types are given in increasing order according to the percentage of time fixating on number 2 and the *p*-values obtained with Holm post hoc comparisons.

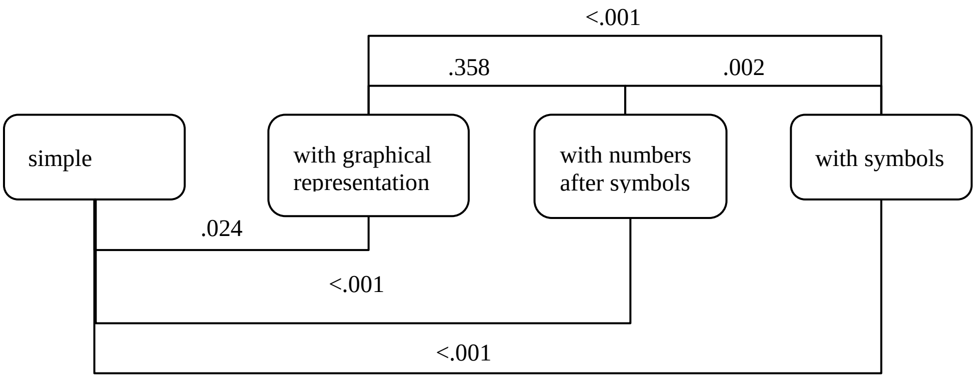

Figure 3. Increasing order according to the percentage of time fixating on number 2

The differences in percentages of time spent fixating on the relational word in case of different types of problems are statistically significant ($F(3) = 7.660$, $p < .001$). Students fixated a significantly lower percentage of time on the relational word in case of simple problems than in the case of the other three problem types ($p < .001$ for comparison with problems with graphical representation, $p = .024$ for problems with numbers after symbols and $p = .004$ for problems with symbols). The highest percentage of time fixating on the relational word is in the case of problems with graphical representations, but this mean is significantly higher only in comparison with the mean in the case of simple problems. In Figure 4 the problem types are given in increasing order according to the percentage of time fixating on the relational word and the *p*-values obtained with Holm post hoc comparisons.

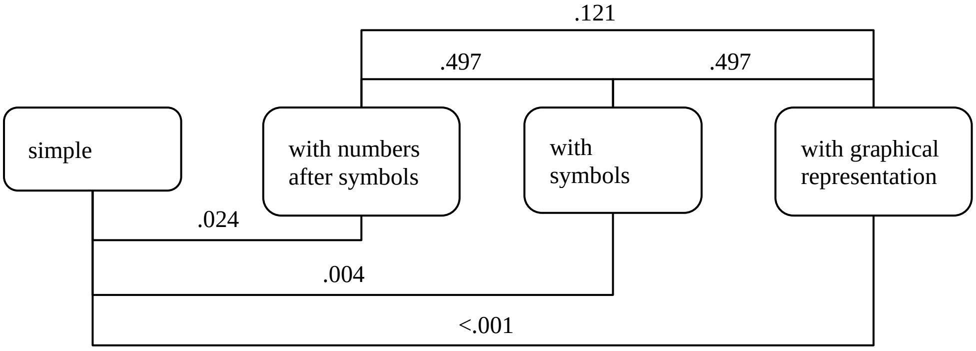

Figure 4. Increasing order according to the percentage of time fixating on the relational word

The differences in percentages of time spent fixating on the relational word in case of different types of problems are statistically significant ($F(3) = 5.389$, $p < .001$). Students fixated the lowest percentage of time on the pronominal reference word in case of problems with graphical representation and the highest percentage of time in the case of simple problems. The Holm post hoc comparison test indicates significant differences only between problems with graphical representation and problems with symbols ($p = .004$) respectively simple problems ($p = .008$), see Figure 5.

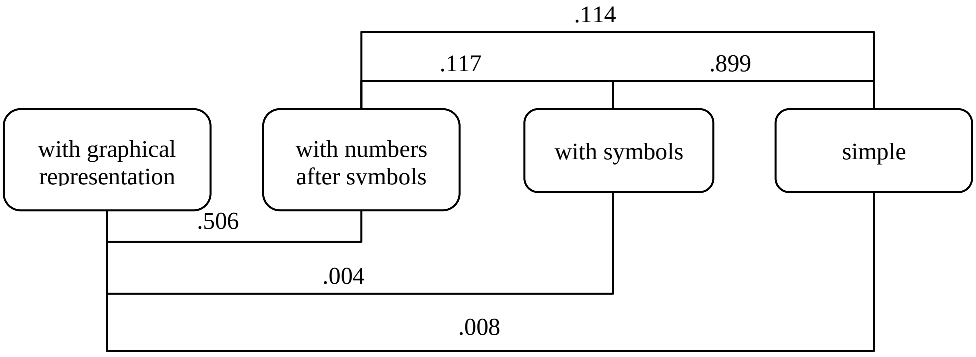

Figure 5. Increasing order according to the percentage of time fixating on the pronominal reference word

Calculating the percentage of time fixating on the four key elements of the text in case of each problem type, students fixated the highest percentage of time on the key element in case of the problems with numbers after symbols (74.96%), then in case of problems with symbols (70.13%), problems with graphical representation (68.6%), and simple problems (65.31%).

4.3. Percentages of time fixating different key elements in case of each type of problems

In this section the percentages of time fixating different key elements of the text are compared in case of each problem type using repeated-measure ANOVA.

In case of simple problems, the Mauchly's test indicated that the assumption of sphericity had been violated ($\chi 2(5) = 43.755$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .694$). The differences in percentages of times spent fixating on different key elements are statistically significant ($F(2.08) = 63.269$, $p < .001$). Students fixated the highest percentage of time on number 2, then on number 1, then on the relational word, and the lowest percentage of time on the pronominal reference word. The Holm post hoc comparison test indicates significant differences in all of the comparisons.

In case of problems with symbols, the Mauchly's test indicated that the assumption of sphericity had been violated ($\chi 2(5) = 19.702$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .827$). The differences in percentages of times spent fixating on different key elements are statistically significant ($F(2.48) = 174.808$, $p < .001$). Students fixated the highest percentage of time on number 2, then on the relational word, then on number 1, and the lowest percentage of time on the pronominal reference word. The Holm post hoc comparison test indicates significant differences in all of the comparisons.

In case of problems with numbers after symbols the Mauchly's test indicated that the assumption of sphericity had been violated ($\chi2(5) = 55.348$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .637$). The differences in percentages of times spent fixating on different key elements are statistically significant ($F(1.91) = 70.143$, $p < .001$). Students fixated the highest percentage of time on number 2, then on number 1, then on the relational word, and the lowest percentage of time on the pronominal reference word. The Holm post hoc comparison test indicates significant differences in all of the comparisons.

In case of problems with graphical representation the Mauchly's test indicated that the assumption of sphericity had been violated ($\chi2(5) = 28.649$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .795$). The differences in percentages of times spent fixating on different key elements are statistically significant ($F(2.39) = 112.806$, $p < .001$). Students fixated the highest percentage of time on number 2, then on number 1, then on the relational word, and the lowest percentage of time on the pronominal reference word. The Holm post hoc comparison test indicates that the only not significant difference is between percentages of time spent fixating on number 1 and the relational word ($p = .166$).

4.4. Percentage of time fixating on each sentence of the text in case of different types of problems

In this section the fixation durations on each sentence of the text are studied comparing different types of problems. As reading times (total fixation duration) for different types of problems were different, to compare fixation durations on different sentences, percentages from the total fixation durations were calculated. The means (*M*) and standard deviations (*SD*) are presented in Table 4 for the three sentences. A repeated-measures ANOVA was performed to compare the percentages of reading different text elements in case of different types of tasks.

Table 4. Means and standard deviations on the percentages of time spent fixating on different sentences of the text

Problem type	Sentence 1 (%)		Sentence 2 (%)		Sentence 3 (%)		Repeated-measure ANOVA	
	M	SD	M	SD	M	SD	F	p
Simple problem	32.18	8.79	47.58	8.18	20.24	8.57	96.717	< .001
Problem with symbols	30.80	7.78	54.37	7.31	14.83	5.85	298.985	< .001
Problem with numbers after symbols	36.27	9.99	50.18	9.41	13.55	8.62	145.811	< .001
Problem with graphical representation	34.64	8.33	50.73	8.07	14.63	7.45	192.726	< .001

The Holm post hoc comparison test indicates significant differences between the percentages of times fixating on different sentences in case of each type of problems (with $p < .001$ in all cases).

4.4. Percentages of time fixating on each sentence of the text in case of different types of problems

In this section the percentages of time fixating different sentences of the text are compared in case of each problem type using repeated-measure ANOVA.

In case of sentence 1 the Mauchly's test indicated that the assumption of sphericity had been violated ($\chi2(5) = 13.229$, $p = .021$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .867$). The differences in percentages of time spent fixating on sentence 1 in case of different types of problems are statistically significant ($F(2.60) = 6.361$, $p < .001$). Students fixated a significantly higher percentage of time on sentence 1 in case of the problem with numbers after symbols than in the case of the simple problems ($p = .017$) or problems with symbols ($p < .001$).

In case of sentence 2 the Mauchly's test indicated that the assumption of sphericity had been violated ($\chi 2(5) = 12.092$, $p = .034$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .864$). The differences in percentages of time spent fixating on sentence 2 in case of different types of problems are statistically significant ($F(2.59) = 7.411$, $p < .001$). Students fixated a significantly higher percentage of time on sentence 2 in case of the problem with symbols than in the case of the simple problems ($p < .001$) or problems with numbers after symbols ($p = .022$).

The differences in percentages of time spent fixating on sentence 3 in case of different types of problems are statistically significant ($F(3) = 11.681$, $p < .001$). Students fixated a significantly higher percentage of time on sentence 3 in case of the simple problem than in the case of other three types of problems (p < .001 in all three cases).

5. Discussion and conclusions

The highest score was obtained for the problems which were first given with symbols instead of numbers (Table 2). This result could be explained by the fact that students were forced to read the text more carefully as they got it without numbers first. In this way they spent more time reading all the element of the text, they understood better the context of the problem, and this helped them to construct the situational model. The construction of the situational model is important for the successful solution of the word problem (Hegarty et al., 1995; Kintsch,1998). The second highest score was obtained for the problems where students were asked to make a graphical representation of the data from the problem. Processing the numbers involves a mental representation of numerical magnitude (Orrantia & Múñez, 2013). In case of word problems not only the numerical magnitude of each number is important, but also the relation between the variables. While working on the graphical representation students read the text of the problem more carefully and they concentrated on the magnitude-based representation of the variables' values and the relation between these variables. The results show that creating the magnitude-based model helps them in selecting correctly the operation needed for the solution. The score for correct graphical representations is lower than the score for the correct solution, which shows that even if the student couldn't correctly draw the representation, it still helped him/her in elaborating the solution. Previous research also proved that self-generated drawing has a positive effect on performance for solving word problems (Boonen et al., 2016).

Solution time was the highest in case of the problem with graphical representation (Table 2), which was expected as during this time students not only solved the problem but also created a magnitude-based model. The second highest solution time was obtained for problems where symbols are given instead of numbers. This type of problem was unusual for students thus they needed more time to read the problem and think about the answer. However, it is surprising that students also need a significantly longer time than in case of simple problems to solve the same problem as read before given with numbers instead of symbols now. The hypothesis would be that they already know the context of the problem, they already constructed the situational model, they just need to process the numbers for solving the problem.

Fixation durations on the problem's text in case of different type of problems follow the same pattern as solution times (Table 2). The fixation duration on the problem with graphical representation was the highest, then follows the problem with symbols, then the problem with numbers after symbols, and finally, with the lowest fixation duration, the simple problem. However, analyzing the percentages of solution times spent with reading the text of the problems, the order is different. The highest percentage of time spent reading the text was in the case of problems with numbers after symbols, then in the case of problems with symbols. The third in decreasing order is the percentage in case of the problems with graphical representation, and the lowest percentage in case of the simple problem. High percentage of solution time spent for reading the problem means in general that the student spent less time with thinking and writing down the solution. As in the case of the problems with symbols students only need to think about the required operation, the high percentage of solution time spent with reading the problem is explained. When solving the problem with numbers after symbols, students already know the context of the problem, also the percentage of solution time spent with reading the problem is high. However, it is surprising that the lowest percentage of solution time spent with reading the problem is in case of simple problems. Seems these problems needed longer time for thinking about the solution and writing it down. Maybe the fact that these problems were the first had an influence on this.

In the case of all types of problems the solution time is strongly positively correlated with the reading time of the problem's text. But solution time didn't predict success, the only strong correlation is between performance and solution time in the case of problems with numbers after symbols. Here it seems that students who spent less time on the solution had a greater success rate. This could be explained by the fact that those students already understand the problem when they got it with symbols only, they already identified the needed operation, they just filled in the numbers in the solution plan when they got them.

As regards percentage of times fixating on key elements of the text, different key elements were fixated in higher percentage in case of different type of problems: number 1 in the case of problems with numbers after symbols (Figure 2), number 2 in the case of problems with symbols (Figure 3), the relational word in the case of problems with graphical representation (Figure 4), and the pronominal reference word in the case of simple problems (Figure 5). It is not surprising that the percentage of time fixating on number 1 is high in case of problems with numbers after symbols, as students already read the text of the problem before and when got the problem with numbers, fixated more on them. The number 2 (actually, the symbol which replaced number 2) was fixated the highest percentage of time in case of problems with symbols. Number 2 is part of the relational sentence (sentence 2), which was also fixated the highest percentage of time in the case of this type of problem. Sentence 2 also contains the relational word and the pronominal reference word. As the requirement was to identify the operation needed for the solution, it is not surprising that in the case of problems with symbols students spent the highest percentage of time on sentence 2 in general, and, in particular, on number 2. The second highest percentage of time fixating on number 2 is in the case of problems with numbers after symbols, which can be explained, as in case of number 1, with the fact that students already knew the problem, they just needed the numerical data for the solution. The relational word is fixated the highest percentage of time in case of problems with graphical representation. In this type of problem students needed to represent in a drawing the relations between variables, which made them concentrate more on the relational word. The second highest percentage of time was in case of problems with symbols, where students needed to determine the operation required for the solution thus, they concentrated on the relational word.

In conclusion we can say that the instruction type influenced both performance and reading pattern. When students are forced to spend more time reading the entire text (problems given first with symbols), the solution success rate is higher. Graphical representation of the problem's data also contributes to a better performance. There are also differences in percentages of reading times spent on fixation different key elements of the text or different sentences of the problem. The key element fixated for a higher percentage of time depends on the type of instruction.

The results of this paper have an important message to the primary school teachers: when solving mathematical word problems more approaches should be used for a better understanding of the problem's context and of a more accurate modelling of the problem's data.

References

Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. Applied Cognitive Psychology, 21(9), 1201–1216. https://doi.org/10.1002/acp.1317

Blum, W., & Leiß, D. (2007). How do Students and Teachers Deal with Modelling Problems? In Mathematical Modelling (pp. 222–231). Elsevier. https://doi.org/10.1533/9780857099419.5.221

Boonen, A. J. H., de Koning, B. B., Jolles, J., & van der Schoot, M. (2016). Word Problem Solving in Contemporary Math Education: A Plea for Reading Comprehension Skills Training. Frontiers in Psychology, 7. https://doi.org/10.3389/fpsyg.2016.00191

Coquin-Viennot, D., & Moreau, S. (2003). Highlighting the role of the episodic situation model in the solving of arithmetical problems. European Journal of Psychology of Education, 18(3), Article 3. https://doi.org/10.1007/BF03173248

Csíkos, C. (2003). Matematikai szöveges feladatok megértésének problémái 10-11 éves tanulók körében. Magyar Pedagógia, 103(1), 35–55.

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. Frontiers in Psychology, 06. https://doi.org/10.3389/fpsyg.2015.00348

De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. Journal of Educational Psychology, 82(2), Article 2. https://doi.org/10.1037/0022-0663.82.2.359

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

Dröse, J. (2019). Comprehending mathematical problem texts – Fostering subject-specific reading strategies for creating mental text representations. In U. T. Jankvist, M. van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), Eleventh Congress of the European Society for Research in Mathematics Education (Vol. TWG06, Issue 10). Freudenthal Group. https://hal.science/hal-02408765

Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. Journal of Educational Psychology, 84(1), Article 1. https://doi.org/10.1037/0022-0663.84.1.76

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. Journal of Educational Psychology, 87(1), Article 1. https://doi.org/10.1037/0022-0663.87.1.18

Kintsch, W. (1998). Comprehension: A paradigm for cognition. (pp. xvi, 461). Cambridge University Press.

Kontra, J. (2001). Nyelvi és strukturális tényezők befolyása a szöveges feladatok. Magyar Pedagógia, 101(1), 5–45.

Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. Journal of Educational Psychology, 79(4), 363–371. https://doi.org/10.1037/0022-0663.79.4.363

Orrantia, J., & Múñez, D. (2013). Arithmetic word problem solving: Evidence for a magnitude-based mental representation. Memory & Cognition, 41(1), 98–108. https://doi.org/10.3758/s13421-012-0241-1

Pape, S. J. (2003). Compare word problems: Consistency hypothesis revisited. Contemporary Educational Psychology, 28(3), 396–421. https://doi.org/10.1016/S0361-476X(02)00046-2

Reusser, K. (1987). From text to situation to equation. In H. Mandl, E. de Corte, N. S. Bennett, &

H. F. Friedrich (Eds.), Learning and instruction (pp. 477–498). New York, NY: Pergamon.

Riley, M. S., & Greeno, J. G. (1988). Developmental Analysis of Understanding Language About Quantities and of Solving Problems. Cognition and Instruction, 5(1), 49–101. https://doi.org/10.1207/s1532690xci0501_2

Schnotz, W. (1994). Aufbau von Wissensstrukturen: Untersuchungen zur Kohärenzbildung beim Wissenserwerb mit Texten. Beltz, Psychologie-Verlag-Union. https://books.google.ro/books?id=HXEQAQAAIAAJ

Stern, E., & Lehrndorfer, A. (1992). The role of situational context in solving word problems. Cognitive Development, 7(2), 259–268. https://doi.org/10.1016/0885-2014(92)90014-I van der Schoot, M., Bakker Arkema, A. H., Horsley, T. M., & van Lieshout, E. C. D. M. (2009). The consistency effect depends on markedness in less successful but not successful problem solvers: An eye movement study in primary school children. Contemporary Educational Psychology, 34(1), Article 1. https://doi.org/10.1016/j.cedpsych.2008.07.002

Verschaffel, L., & De Corte, E. (1993). A decade of research on word problem solving in Leuven: Theoretical, methodological, and practical outcomes. Educational Psychology Review, 5(3), 239–256. https://doi.org/10.1007/BF01323046

Vicente, S., Orrantia, J., & Verschaffel, L. (2008). Influence of mathematical and situational knowledge on arithmetic word problem solving: Textual and graphical aids. Infancia Y Aprendizaje, 31(4), 463–483. https://doi.org/10.1174/021037008786140959

Acknowledgement

The work of the first author was partially supported by the Collegium Talentum Programme of Hungary.

Authors

Ágnes BARTALIS, Babeș-Bolyai University, Cluj-Napoca (Romania). E-mail: agnes.bartalis@ubbcluj.ro.

Iuliana ZSOLDOS-MARCHIȘ, Babeș-Bolyai University, Cluj-Napoca (Romania). E-mail: iuliana.marchis@ubbcluj.ro.