# Superficial Strategies in Solving Compare-Combine Word Problems

Marijana Zeljić*, Milana Dabić Boričić and Svetlana Ilić

Teacher Education Faculty, University of Belgrade, Belgrade, Serbia

*Abstract.* Compare and combine word problems are used in elementary mathematics as a part of standard teaching practice. Their integration enables creation of various types of word problems with different structure and level of difficulty. One of the main obstacles in solving word problems is the use of superficial strategies in which students directly translate words they have recognized as entities and relations into mathematical operations and expressions, without understanding the situational model of the problem. The aim of this paper is to investigate the use of these strategies in solving integrated compare-combine word problems. For this purpose, we posed word problems with varying correspondences between entity keywords and relations given in the text of the problems. One hundred and thirty-four students participated in the study by solving paper and pencil test. Forty-four students were in 2$^{nd}$ grade (7,5 to 8,5 years old students), 48 in 4$^{th}$ grade (9,5 to 10,5 years old students), and 42 in 6$^{th}$ grade (11,5 to 12,5 years old students). Results showed that students did not have different achievement on word problems with different correspondence between entity keywords and relations. The superficial approach they used most often was in identifying relational terms (mathematical operations). As expected, there were differences in achievement and in nature of mistakes regarding students' level of education (2$^{nd}$, 4$^{th}$ or 6$^{th}$ grade).

*Keywords:* compare word problems, combine word problems, problem solving strategies, word problems

*Corresponding author.

## 1. Introduction

Word problems are considered a basis for learning mathematics at the elementary school level. As such, a significant amount of research has been directed towards analyzing different word problem types and identifying the obstacles students face when solving them. Classification of word problems with one mathematical operation on change, combine and compare problems was used as a starting point in many studies. Those studies imply that even though word problems can be solved with the same mathematical operation, they actually belong to different semantical types, trigger different ways of representing and solving, and reveal different types of students' mistakes and misconceptions (Fuson, 1992). More recently, researchers have recognized the importance of problems with higher complexity that integrate change, combine, and compare problems (Nesher et al., 2003). These integrated word problems can be particularly challenging as each sub-problem brings in difficulties that can be attributed to its category, and the integration itself produces problems with varying structures and complexities. According to studies that confirm the relevance of correspondence between the order of information presented in the text of the word problem and the steps in its solving (Daroczy et al., 2015; Vicente et al., 2007), one of the difficulties in integration could also be in the correspondence between the order of entities (in the text of the problem) and order of description (the way in which relations are presented in the text). In this paper, we investigate the importance of this correspondence to understand the use of superficial strategies in students' word problem solving.

## 2. Students' difficulties in solving compare word problems

During the 1980s researchers singled out three groups of word problems: combine, change, and compare problems (e.g., Riley & Greeno, 1988). The values that are unknown in each problem determine its structure and its level of difficulty. Riley and Greeno (1988) identified combine word problems with an unknown total amount (e.g. Joe has 3 marbles. Tom has 5 marbles. How many marbles do they have altogether? (p. 53)) or subset (Joe and Tom have 8 marbles altogether. Joe has 3 marbles. How many marbles does Tom have? (p. 53)) and compare word problems with an unknown difference set (Joe has 8 marbles. Tom has 3 marbles. How many marbles does Tom have less than Joe? (p. 54)), compared set (Joe has 8 marbles. Tom has 5 marbles less than Joe. How many marbles does Tom have? (p. 54)) or referent set (Joe has 3 marbles. He has 5 marbles less than Tom. How many marbles does Tom have? (p. 54)). Various studies, including those by Stern (1993) and Boonen and Jolles (2015), have shown that combine word problems with an unknown total amount are the least difficult for students, while compare word problems with an unknown referent set are the most challenging. Many studies have attempted to explain why compare word problems with an unknown referent set are difficult to solve. One point of view is that these problems can have different language consistencies. Lewis and Mayer (1987) pointed out that compare word problems can have consistent language (CL) or inconsistent language

(IL) formulation. For example, in CL problem the term “more” in the text of the problem can be successfully solved using addition, while in the problems with IL formulation with the term “more” in the text, addition cannot be used to solve the problem. They found that students make more mistakes in IL problems because they automatically activate the rule “add if the relation is more than and subtract if the relation is less than”. This is known as the consistency effect and is confirmed using samples of students from elementary school to college (e.g., Hegarty et al., 1995; Pape, 2003; Stern, 1993). To explain these findings, Stern (1993) and Okamoto (1996) pointed out that students do not understand the symmetrical relationship between relations “more than” and “less than”, which is necessary for solving these types of problems.

Boonen and Jolles (2015) conducted a study with second-grade students which did not confirm the consistency effect. The students who were explicitly taught the relations “more” and “less” performed equally well on IL and CL problems.

However, there are studies where verbal instruction on how to solve CL and IL problems did not reduce the consistency effect (Dewolf et al., 2014; de Koning et al., 2017). In one study (Dewolf et al., 2014) students had verbal instruction on the test that informed them that word problems could have different language constructions. In other words, it was explained that the word problem could have one of several types of relational keywords, and that it is important to pay attention to the use of the correct operation. These instructions did not significantly affect the students’ achievement. In contrast, in the study conducted by de Koning et al. (2017), verbal instructions were given on how to solve CL and IL problems, with an emphasis on the interpretation of relational keywords to avoid mistakes in the choice of operation. The effect of these instructions was significant on problems that included the keywords “less” and “more”, but not on problems that included other keywords such as “higher/lower” or “more expensive/cheaper”.

In previous years, research was also directed to the modeling process – the cyclic process of solving word problems that starts from real situation, goes over the use of real models and mathematical models and ends with the mathematical results and real results (Blum & Leiss, 2007). Situational understanding (“understanding situation described in the problem in order to reduce it to its gist”) plays a vital role in bridging the gap between language comprehension and mathematical problem solving (Stern & Lehrndorfer, 1992, pp. 261). Linguistic models of word problem solving emphasize the role of language in understanding, with situational understanding seen as a process of “going beyond the text’s” (according to Stern and Lehrndorfer, 1992, pp. 261). Students develop an adequate mental representation of the situational model and then translate it into a mathematical model through mathematization (Stern & Lehrndorfer, 1992; Blum & Leiss, 2007). In the modelling process, the mathematical model serves as a foundation for planning and utilizing necessary mathematical operations. Students make mathematical models based on the realistic situation presented in the text of the problem and then they solve the problem in a mathematical context, by performing mathematical operations. In other words, the problem can be formulated in a way that makes semantic relations between the relations in the text more explicit and transparent, and therefore easier for students to solve (Vicente et al., 2007). If the information used to describe the realistic situation follows the order of steps in the solving, students are more likely to succeed (Daroczy et al., 2015). The negative effect of problems formulated this way is that students tend to rely on superficial strategies, such as the "keyword" or "number grabbing" strategy in word problem solving (Briars & Larkin, 1984; Littlefield & Rieser, 1993). Students who transform numbers and keywords directly into arithmetic operations and attempt to "combine" them to find the solution do not construct adequate models for word problem solving (Hegarty et al., 1995). Although one group of problems is formulated in a way that enables students to solve them successfully using this approach, it only promotes the practice of their computational skills and not conceptual understanding and mathematical thinking (Boesen et al., 2014). Our choice to investigate the understanding of compare word problems on problems with complex structures is based on the view that mathematical reasoning could be investigated on the problems that require analysis of the meaning and the structure of a problem, as well as justification of procedures and solving strategies (Stein et al., 2000). Following the authors who use "keywords" for the direct transformation of words "more or less" to mathematical operations, we use "entity keywords" for the direct transformation of entities into numbers without considering the relations between them provided in the text of the problem. For example, in the word problem "Joe has 3 marbles. Tom has 5 marbles. How many marbles do they have altogether?" Joe and Tom are entity key words that students grab and replace them with the numbers 3 and 5.

## 3. Integrating combine and compare word problems

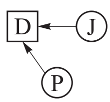
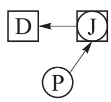
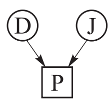
Nesher et al. (2003) investigated the integration of combine and compare word problems, but they used more complex way of integrating, in which the combine problems have unknown subset and compare problems have multiplicative comparisons (Nesher et al., 2003). Their research is conducted with teachers and 15-year-old students and as a result, authors proposed the model of complexity. Two of the variables that Nesher et al. (2003 p. 151) used and that we will use to interpret our results are "number of quantities that are being compared to the number of reference quantities" and "the order of presenting the elementary comparison relations". The variables will be illustrated later through an example (Table 1, rows Reference structure and Order of description). In our paper, we aimed to examine the integration of compare-combine word problems with a simple structure that can be solved by elementary school students. Therefore, we choose compare-combine word problems with an unknown total number of elements (rather than an unknown subset) and additive comparison relations (rather than multiplicative). As we already stated, the semantic structure of these types of problems can be reflected in referent structure and order of description (Nesher et al., 2003) as well as in different language consistency (Lewis & Mayer, 1987; Hegarty et al., 1995; Pape, 2003; Stern, 1993).

## 4. Method

The results presented in this paper are part of a larger study investigating students' understanding of relational terminology and their achievement in solving compare-combine word problems. The aim of this paper is to investigate the differences in students' achievement in solving word problems with diverse correspondence between the order of entities and the order of description (relations given in the text of the problems). The use of superficial strategies by students in solving problems would reveal their obstacles in solving problems in which the entity order and order of description do not correspond. To achieve this aim, we defined three types of problems (presented in Table 1):

— P1, which describes the relation between the first (known) entity (David) and the other two (John and Peter);
— P2, which describes the relation between the first (known) entity (David) and the second entity (John), and then between the second (John) and the third (Peter); and
— P3, which describes the relations between the third (unknown) entity (Peter) and the other two entities (David and John).

*Table 1.* The structure and formulation of compare-combine problems.

| | **Problem 1 (P1)** | **Problem 2 (P2)** | **Problem 3 (P3)** |
|---|---|---|---|
| Text | David has 20 marbles, which is 15 less than Peter and 5 less than John. How many they have altogether? | David has 20 marbles, which is 5 less than John, while John has 10 marbles less than Peter. How many they have altogether? | David has 20 marbles. Peter has 15 more than David, and 10 more than John. How many they have altogether? |
| Given connections |  |  |  |
| Reference structure | 1 (compared) to 2 (referent) sets | 2 (compared) to 2 (referent) sets | 1 (compared) to 2 (referent) sets |
| Entity order | David, Peter, John | David, John, John, Peter | David, Peter, David, John |
| Order of description | D $= f$ (P), D $= g$(J)) | D $= f$ (J), J $= g$(P)) | P $= f$ (D), P $= g$(J)) |
| Language consistency | IL, IL | IL, IL | CL, IL |

In P1 and P3 the order of entities (the order of given names David, John, Peter) does not correspond with the relations described in the text. In problem P1 students have to understand that although the order of entities in the problem text is David, John, Peter, the relations described in the text are between David and John, and David and Peter (D $= f$ (J), D $= g$(P)). Similarly, in problem P3, the entity order is David, Peter, David, John, but the relations described in the text are between Peter and David, and Peter and John $(\mathrm{P} = f(\mathrm{D}), \mathrm{P} = g(\mathrm{J}))$. On the other hand, in P2 the order of the entities (David, John, John, Peter) corresponds with the relations between the sets $(\mathrm{D} = f(\mathrm{J}), \mathrm{J} = g(\mathrm{P}))$. We have formulated all three problems using inconsistent language because problem P3 (which has relations between the third and the second entity) cannot be formulated using exclusively consistent formulations.

Students at different levels of education use various strategies to solve problems. Therefore, we conducted research to investigate the differences in achievement on several levels. We selected 2nd graders because they are familiar with arithmetic strategies for solving word problems and to mathematical texts to some extent, 4th graders because they are more fluent with arithmetic strategies and also have some familiarity with the beginning of algebra, and 6th graders because they are fluent in algebraic strategies. The problems in our study could be solved using arithmetical or algebraic notation and strategies. It is not expected that students fluent with algebraic notation have difficulties in writing and solving system of equations. Hence, we expect older students to be better in solving these problems.

The research sample consists of 134 students from one primary school in Belgrade, which cooperates with the researchers' institution. The students are from two classes of 2nd grade (44 students), two classes of 4th grade (48 students), and two classes of 6th grade (42 students).

Our research questions are:

1) Are there differences and associations in students' achievement in problems with different semantic structures (P1, P2, and P3)?
2) Are there differences in achievement in every type of word problem (P1, P2, and P3) among 2nd, 4th, and 6$^{\text{th}}$ grade students?
3) What are the strategies that students use in solving combine-compare word problems, and what are the most common mistakes they make?

The students were not given a time limit to solve problems P1, P2, and P3. To eliminate the potential influence of task order on their performance, the students were assigned to groups with varying task sequences. For the analysis, we used SPSS and conducted McNemar's exact test, Chi-square test of independence and homogeneity, and phi coefficient to investigate the differences and associations between the achievement in solving word problems P1, P2, and P3. We used 0.05 as level of significance in all tests, and considered association to be moderate if phi is greater than 0.3, and strong if phi is greater than 0.5 (Pallant, 2009). After Chi-square test that compare three variables with significant differences, we performed post hock tests to reveal which pair of variables differ. Additionally, we identified and categorized common mistakes (superficial strategies) made by the students.

## 5. Results

The students' achievements in solving compare-combine word problems are presented in Table 2, and the results of McNemar's and Chi-square test of independence for investigating the difference and association between problems P1, P2, and P3 are presented in Table 3.

*Table 2.* Number (percentage) of correct answers to every problem.

| **Grade** | **P1** | **P2** | **P3** |
|---|---|---|---|
| $2^{nd}$ ($n = 44$) | 12 (27.3 %) | 14 (31.8 %) | 16 (36.4 %) |
| $4^{th}$ ($n = 48$) | 24 (50.0 %) | 29 (60.4 %) | 31 (64.6 %) |
| $6^{th}$ ($n = 42$) | 32 (76.2 %) | 33 (78.6 %) | 34 (81.0 %) |

*Table 3.* The values of McNemar's test and Chi-square tests.

| **grade** | **Tests** | **P1/P2** | **P1/P3** | **P2/P3** |
|---|---|---|---|---|
| $2^{nd}$ | McNemar's | 0.687 | 0.344 | 0.687 |
| | Chi (44, 1) | 20.184, $p = 0.000$ phi = 0.677 | 10.644, $p = 0.001$ phi = 0.492 | 21.611, $p = 0.000$ phi = 0.701 |
| $4^{th}$ | McNemar's | 0.125 | **0.039** | 0.754 |
| | Chi (48, 1) | 25.176, $p = 0.000$ phi = 0.724 | 20.493, $p = 0.000$ phi = 0.653 | 14.977, $p = 0.000$ phi = 0.559 |
| $6^{th}$ | McNemar's | 1.000 | 0.727 | 1.000 |
| | Chi (42, 1) | 6.364, $p = 0.012$ phi = 0.389 | 8.155, $p = 0.004$ phi = 0.441 | 1.516, $\boldsymbol{p} = \mathbf{0.218}$ |

The results show us that there are no statistically significant differences between any pair of problems, as the values of McNemar's tests are greater than 0.05 (Table 3), except for P1 and P3 in $4^{th}$ grade (McNemars's $p = 0.039$, Table 3) The average success rate for $2^{nd}$ graders was about 32 %, while the average success rate for $6^{th}$ graders was about 79 % (Table 2). The $4^{th}$ graders performed differently on tasks – the highest achievement was on P3 (65 %, Table 2) and the lowest on P1 (50 %, Table 2). Levels of significance of Chi square tests and phi coefficients presented in Table 3 shows us moderate ($p < 0.05$, $0.3 <$ phi $< 0.5$) to strong ($p < 0.05$, phi $> 0.5$) associations between every pair of problems except between problems P2 and P3 where association is missing – Chi $(42, 1) = 1.516, p = 0.218$.

The Chi-square test of homogeneity, which examined the differences in achievement among students of different age (Table 4), showed significant differences among all of them for every problem ($p = 0.000$). The post hoc test also revealed significant differences between almost all pairs of grades across all problems, except for P2 and P3, where 4th and 6th graders demonstrated similar achievement ($p > 0.05$).

*Table 4.* The results of Chi-square test and post hoc tests in comparing success of $2^{nd}$, $4^{th}$, and $6^{th}$ graders in solving problems.

| | | **Problem 1** | **Problem 2** | **Problem 3** |
|---|---|---|---|---|
| $2^{nd}/4^{th}/6^{th}$ grade | Chi (134, 2) | 20.589 | 19.551 | 18.404 |
| | $p$ | 0.000 | 0.000 | 0.000 |
| $2^{nd}$ /$4^{th}$ grade | Chi (92, 1) | 4.978 | 7.542 | 7.316 |
| | $p$ | 0.026 | 0.006 | 0.007 |
| $2^{nd}$ /$6^{th}$ grade | Chi (86, 1) | 20.579 | 18.952 | 17.554 |
| | $p$ | 0.000 | 0.000 | 0.000 |
| $4^{th}$ /$6^{th}$ grade | Chi (90, 1) | 6.537 | 3.445 | 2.992 |
| | $p$ | 0.011 | **0.063** | **0.084** |

To interpret the results, we used types and frequencies of students' incorrect answers (Table 5). The most common mistakes made by students are as follows:

— Incorrect relational term, which occurs when students use the wrong (opposite) mathematical operation. For example, when solving problem P2, "David has 20 marbles, which is 5 less than John. . . " students wrote 20 – 5 for the number of John's marbles.

— Incorrect entity, which occurs when students use the wrong entity in relations. For instance, when solving problem "David has 20 marbles, which is 15 less than Peter and 5 less than John. . . " students made relations between Peter and John instead of David and John.

— Incorrect relational term and entity, which is when students make both mistakes.

— "Keyword" approach, which is when students directly translate words from the problem text into mathematical operations between the numbers in the text. For example, when solving a problem "David has 20 marbles, which is 15 less than Peter and 5 less than John. How many do they have altogether?" students wrote $20 - 15 - 5$ as the answer.

*Table 5.* Types and frequencies of students' mistakes on P1, P2, and P3, and number of students who made a certain type of error at least once (No).

| **Grade** | **$2^{nd}$** | | | | **$6^{th}$** | | | | **$4^{th}$** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P1 | P2 | P3 | No | P1 | P2 | P3 | No | P1 | P2 | P3 | No |
| Incorrect relational term | 4 | 8 | 4 | **11** | 8 | 10 | 5 | **16** | 3 | 5 | 5 | **11** |
| Incorrect entity | 5 | 2 | 2 | **8** | 1 | 0 | 4 | **5** | 1 | 1 | 1 | **3** |
| Incorrect relational term and entity | 5 | 1 | 4 | **9** | 1 | 0 | 0 | **1** | 0 | 0 | 1 | **1** |
| "Keyword" approach | 9 | 11 | 9 | **12** | 1 | 2 | 1 | **2** | 1 | 0 | 1 | **1** |
| Total | 23 | 22 | 19 | **40** | 11 | 12 | 10 | **24** | 5 | 6 | 8 | **16** |

## 6. Discussion

Our first research question aimed to explore the differences and associations in students' achievement on compare-combine word problems with diverse correspondence between the order of entities and the order of descriptions. Contrary to our expectations, the results showed that students in $2^{nd}$ and $4^{th}$ grade solved problems equally successfully, regardless of this correspondence. There were no statistically significant differences between pairs of problems in $2^{nd}$ and in $6^{th}$ grade (Table 3) and the students' average achievement was about 32 % and 79 % in $2^{nd}$ and $6^{th}$ grade, respectively (Table 2). These results are to a certain extent opposite to our expectations based on previous research (Daroczy et al., 2015; Vicente et al., 2007) which showed the relevance of correspondence between the order of information and steps in problem solving. We anticipated that students would achieve the highest scores on problem P2, in which the entity keyword and the order of description correspond. However, it seems that language consistency disrupted students' word problem solving. If the problem was formulated in a consistent way, students could solve it by using keyword approach as a superficial strategy (Briars & Larkin, 1984; Littlefield & Rieser, 1993). However, problem P2 was formulated with inconsistent language formulation, so students' use of keyword resulted in an incorrect understanding of the relational term and incorrect choice of mathematical operation. This is in accordance with many studies that investigated inconsistent language word problems (Lewis & Mayer, 1987; Stern, 1993; Okamoto, 1996; Hegarty et al., 1995; Pape, 2003). Students' difficulties with inconsistent formulation of P2 were also confirmed in the analysis of students' incorrect responses. The largest number of incorrect relational term answers were on problem P2 (8 in $2^{nd}$, 10 in $4^{th}$, and 5 in $6^{th}$ grade, Table 5), while the smallest number of incorrect answers on P2 was related to an incorrect entity (2, 0 and 1 respectively in $2^{nd}$, $4^{th}$ and $6^{th}$ grade, Table 5).

In addition, difficulties in solving tasks with no correspondence between entity word and order of description (problems P1 and P3) are reflected in students' mistakes in which they used the wrong entity when solving word problem. The frequencies of these mistakes are provided in Table 5 as "Incorrect entity" and "Incorrect relational term and entity". These incorrect answers were made by seventeen $2^{nd}$ graders ($8+9$ in the "No column", Table 5), six $4^{th}$ graders ($5+1$), and four $6^{th}$ graders ($3+1$). As expected, the mistakes are made on P1 and P3, but they were not frequent enough to make the differences in achievement on problems P1, P2, and P3 in $2^{nd}$ and $6^{th}$ grade.

In the $4^{th}$ grade, there were no significant differences between success in problems P1 (50 %, Table 2) and P2 (60 %, Table 2) and between P2 and P3 (65 %, Table 2), while the difference between P1 and P3 was significant. The number of incorrect answers that students in $4^{th}$ grade made on P1 and P3 was practically equal (11 on P1, and 10 on P3, Table 5). Therefore, the reason for the significant difference could be found in other factors that define the structure of the problem. The inconsistent language formulation of problem P1 leads to a mathematical model in which two unknown sets are referent, and the other (known) set becomes compared. This is, according to previous research, the most complex compare problem (Hegarty et al., 1995; Pape, 2003; Stern, 1993; Okamoto, 1996). In the third problem (P3) there are also two referent sets that are compared to one (unknown) set, but the relations are presented in the way that only one inconsistent language formulation is used. Considering that the 4$^{th}$ graders had the most incorrect answers that included incorrect relational term (Table 5), we hypothesize that the consistency effect is (at least) one of the reasons for students' higher achievement on problem P3 (with one CL and one IL formulation) than on P1(with two IL formulations), as it was case with studies we mentioned previously in the paper (Boonen & Jolles 2015; Dewolf et al., 2014; de Koning et al., 2017; Pape, 2003; Stern, 1993; Okamoto, 1996).

Moderate to strong associations between P1, P2, and P3 (Table 3) confirm the finding that students are equally successful in solving compare-combine word problems, regardless of the correspondence between entities and the order of description. This result was expected as we used problems with a more complex structure that are more reliable for investigating students' mathematical thinking, which was recommended by Stein et al. (2000). The lowest phi coefficients are between P1 and P2, and P2 and P3 in the 6$^{th}$ grade (phi $< 0.5$, Table 3), while the association between P2 and P3 is even missing. This implies that the correspondence of entity and order of description is less important for the students in 6$^{th}$ grade than for lower graders (2$^{nd}$ and 4$^{th}$). An analysis of students' incorrect responses confirms this result. Incorrect response that includes incorrect entity reveals the use of superficial strategies. Students in 6$^{th}$ grade had a low number of these incorrect responses (4 incorrect answers, Table 5). The most frequent obstacle for the students in 6$^{th}$ grade was using incorrect relational terms (12 incorrect answers, Table 5).

Our second research task aimed to investigate the differences in students' achievement across different age. The results of Chi-square homogeneity tests results (Table 4) showed significant differences in performance across all problems. As expected, 6$^{th}$ graders were the most successful, followed by 4$^{th}$ graders and 2$^{nd}$ graders (Table 2). However, there were some exceptions found in the post hoc tests, specifically on problems P2 and P3, where 4$^{th}$ and 6$^{th}$ graders achieved similar scores (Table 4). We suppose that this is due to the lack of a clear gradation of achievement on P1, P2, and P3 in the 6$^{th}$ grade, in which achievement rise from 76.2 % on P1, over 78.6 % on P2, to 81.0 % on P3. This growth is steeper in the 4$^{th}$ grade (from 50.0 % on P1, over 60.4 % on P2 to 64.6 % on P3), hence the achievements of 4$^{th}$ and 6$^{th}$ graders got close enough on problems P2 and P3 to make insignificant difference. Nonetheless, there was still a difference in achievement between 4$^{th}$ and 6$^{th}$ graders on P1, which we attribute to its more complex mathematical structure, involving two unknown referent sets and one known compared set. Only on this problem, with the most complex structure, we have a significant difference in achievement between students of different age.

Finally, the analysis of students' incorrect answers and strategies gives us a closer look at their process of solving compare-combine word problems. Students made the biggest number of mistakes by using the incorrect (opposite) relational term (11 students in 2$^{nd}$ grade, 16 students in 4$^{th}$ grade, and 11 students in 6$^{th}$ grade,

Table 5). In other words, 25 % of 2$^{nd}$ graders, 33 % of 4$^{th}$ graders, and 26 % of 6$^{th}$ graders made relational term mistake. This implies that the consistency effect, that was investigated and confirmed in many studies (Hegarty et al., 1995; Pape, 2003; Stern, 1993), seems to be present at all levels of education, and it is still a significant obstacle for solving these kinds of word problems. It is surprising to note that the 4$^{th}$ graders made more incorrect relational term responses than the 2$^{nd}$ and 6$^{th}$ graders.

Another mistake is using the keyword approach, where students follow words in the problem and transform them into mathematical operations, like in other studies (Briars & Larkin, 1984; Hegarty et al., 1995). We observed that 12 students in 2$^{nd}$ grade made this kind of mistake, which is about one-fourth of 2$^{nd}$ graders, while only a few students in 4$^{th}$ and 6$^{th}$ grade (about 5 % of students, Table 5). This means that the keyword approach is present in 2$^{nd}$ grade, while 4$^{th}$ and 6$^{th}$ graders realize that this strategy will not take them to the correct solution. Littlefield and Rieser (1993) used the term number grabbing for keyword approach. Interestingly, students in our research also used number grabbing – they tend to single out numerical and relational data from the text, without considering entities in the text, nor the meaning of the situational model which researchers (Stern & Lehrndorfer, 1992; Blum & Leiss, 2007) find crucial for word problem solving. For these students, the result of the problem is simply the result of the operation between "grabbed" numbers, which was also the case in other studies (Boesen et al., 2014; Briars & Larkin, 1984; Littlefield & Rieser, 1993).

Surprisingly, none of the students used algebraic strategies for problem solving. This is surprising because the curriculum in our country requires arithmetical and algebraic strategies in problem solving on elementary level (first four grades) and use of mathematical modeling. However, the types of the problems, their extent, complexity and strategies in solving are left for the teachers' choice. In our research, only a few students used algebraic syntax to write down the relations between entities, and then continued with an arithmetic strategy for solving the problem. One 4$^{th}$ grade student used algebraic symbols to represent relations in problem P1, three students in P2 (one 2$^{nd}$ grader, one 4$^{th}$, and one 6$^{th}$ grader), and one (4$^{th}$ grader) in P3, but none of them managed to set up and solve the equation. We expected that 6th grade students, who are familiar with solving equations and algebraic syntax, would use algebraic strategies to solve complex problems, but this was not the case. Khng and Lee (2009) already noted that many students return to arithmetic strategies in problem solving even if it is explicitly stated that the problem should be solved using equations. They found that using algebra is a step forward to higher mathematics and that students should practice algebra even if they know how to solve a problem using arithmetic strategies. In this context, the persistence in using arithmetic strategies could be seen as an inhibition for further learning. Therefore, we could pose the question of whether the algebraic knowledge of 6$^{th}$ graders is only formal since students did not see the equation as a suitable model for solving problems with complex structures. Students also did not use geometric models or visual representations for representing problems, which implies that they are not eager to use them in problem solving.

## 7. Conclusion

It is essential to understand the obstacles that students face at each level of education when solving compare-combine word problems and suggest ways to overcome them. Findings form the literature presented in the theoretical part of the paper imply that students have difficulties in understanding the simple compare word problems with inconsistent formulations. Our research showed that the difficulties are present also on the problems with more complex structure that are generated by integrating compare and combine problems. Considering that these problems also reflect students' understanding of mathematical terminology and situational understanding, they serve as a foundation for solving more complex routine and non-routine problems in mathematics education, making it essential for students to solve them correctly and efficiently, without relying on superficial strategies.

In our study, we analyzed the use of superficial strategies in solving compare-combine word problems with varying correspondence between entity order and order of description. Interestingly, we found no significant differences in problem solving with respect to the correspondence. Students often used a "shortened" model by converting the problem text into a mathematical model, which involved choosing an arithmetic operation based on a quick and superficial analysis of the data, relying on keywords in the text.

Interestingly, younger students ($2^{nd}$ and $4^{th}$ graders) showed a higher association in achievement on problems with varying correspondence than older students ($6^{th}$ graders). Students in $2^{nd}$ and in $4^{th}$ grade used superficial strategies (keywords for operations and entities) more frequently than $6^{th}$ grade students. On the other hand, regardless of the students' age, incorrect relational terms were the most frequent type of mistake made. Thus, the main obstacle in solving compare-combine word problems was the consistency effect – the use of the opposite operation due to a misunderstanding of relational terminology.

Based on our findings, we suggest that instructions aimed at developing a conceptual understanding of relations should be incorporated into the curriculum starting from the first grade. Conceptual understanding includes perceiving the structure of compare-combine word problems and understanding both consistent and inconsistent formulations. The literature offers three possible guidelines for improving students' achievement. First, Boonen and Jolles (2015) showed that instructions directed at developing the meaning of relations can eliminate the consistency effect. Second, a series of research is focused on the advantages of graphically representing the structure of the problem using diagrams (Boonen & Jolles, 2015; De Koning et al., 2022). These representations could be used for improving the understanding and achievement of students and reducing the consistency effect on higher levels of education. The use of representations for solving problems is related to the third guideline – the use of phases of mathematical modeling. Numerous studies suggest that students do not use phases of mathematical modeling during the problem-solving process (Stern & Lehrndorfer,1992; Blum & Leiss, 2007). The results of our study support the idea that word problem solving with the use of modeling process should be included in the curriculum in the early years of mathematics education. The ability of students to understand the problem situation and construct relations between elements of the problem determines their success and eliminates the use of superficial strategies in problem solving.

*Acknowledgements*

This work was financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia under Contract 451-03-1/2022-14/4, which was signed with the Faculty of Teacher Education, University of Belgrade.

## References

[1] BLUM, W., & LEISS, D. (2007). How do students and teachers deal with modelling problems. In C. Haines, P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling: Education, engineering, and economics* (pp. 222–231). Horwood.

[2] BOESEN, J., HELENIUS, O., LITHNER, J., BERGQVIST, E., BERGQVIST, T., PALM, T., & PALMBERG, B. (2014). Developing mathematical competence: From the intended to the enacted curriculum. *Journal of Mathematical Behavior*, 33(1), 72–87.

[3] BOONEN, A. J., & JOLLES, J. (2015). Second grade elementary school students' differing performance on combine, change and compare word problems. *International Journal of School and Cognitive Psychology*, 2(122), 1–6.

[4] BRIARS, D. J., & LARKIN, J. H. (1984). An integrated model of skill in solving elementary word problems. *Cognition and Instruction*, 1(3), 245–296.

[5] DAROCZY, G., WOLSKA, M., MEURERS, W. D., & NUERK, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, 6, 348.

[6] DE KONING, B. B., BOONEN, A. J., JONGERLING, J., VAN WESEL, F., & VAN DER SCHOOT, M. (2022). Model method drawing acts as a double-edged sword for solving inconsistent word problems. *Educational Studies in Mathematics*, 111(1), 29–45.

[7] 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.

[8] DEWOLF, T., VAN DOOREN, W., CIMEN, E. E., & VERSCHAFFEL, L. (2014). The impact of illustrations and warnings on solving mathematical word problems realistically. *Journal of Experimental Education*, 82(1), 103–120.

[9] FUSON, K. C. (1992). Research on whole number addition and subtraction. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 243–275). Macmillan Publishing Co, Inc.

[10] 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), 18–32.

[11] KHNG, K. H., & LEE, K. (2009). Inhibiting interference from prior knowledge: Arithmetic intrusions in algebra word problem solving. *Learning and Individual Differences*, 19(2), 262–268.

[12] LEWIS, A., & MAYER, R. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology*, 79(4), 199–216.

[13] LITTLEFIELD, J., & RIESER, J. J. (1993). Semantic features of similarity and children's strategies for identification of relevant information in mathematical story problems. *Cognition and Instruction*, 11(2), 133–188.

[14] NESHER, P., HERSHKOVITZ, S., & NOVOTNA, J. (2003). Situation model, text base and what else? Factors affecting problem solving. *Educational Studies in Mathematics*, 52(2), 151–176.

[15] OKAMOTO, Y. (1996). Modelling children's understanding of quantitative relations in texts: A developmental perspectives. *Cognition and Instruction*, 14(4), 409–440.

[16] PALLANT, J. (2009). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th ed.). Allen & Unwin.

[17] PAPE, S. J. (2003). Compare word problems: Consistency hypothesis revisited. *Contemporary Educational Psychology*, 28(3), 396–421.

[18] RILEY, M. S., & GREENO, J. G. (1988). Developmental analysis of language about quantities and of solving problems. *Cognition and Instruction*, 5(1), 49–101.

[19] STEIN, M., SMITH, M., HENNINGSEN, M., & SILVER, E. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development*. Teacher College.

[20] STERN, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children. *Journal of Educational Psychology*, 85(1), 7–23.

[21] STERN, E., & LEHRNDORFER, A. (1992). The role of situational context in solving word problems. *Cognitive Development*, 7(2), 259–268.

[22] VICENTE, S., ORRANTIA, J., & VERSCHAFFEL, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology*, 77(4), 829–848.

*Contact address:*

Marijana Zeljić
Teacher Education Faculty, University of Belgrade
Kraljice Natalije 43, 11000 Belgrade, Serbia
e-mail: `marijana.zeljic@uf.bg.ac.rs`

Milana Dabić Boričić
Teacher Education Faculty, University of Belgrade
Kraljice Natalije 43, 11000 Belgrade, Serbia
e-mail: `milana.dabic@uf.bg.ac.rs`

Svetlana Ilić
Teacher Education Faculty, University of Belgrade
Kraljice Natalije 43, 11000 Belgrade, Serbia
e-mail: `svetlana.ilic@uf.bg.ac.rs`