ORIGINAL PAPER

Students' suspension of sense making in problem solving

Gemma Carotenuto[1] · Pietro Di Martino[2] · Marta Lemmi[2]

Abstract

Research on mathematical problem solving has a long tradition: retracing its fascinating story sheds light on its intricacies and, therefore, on its needs. When we analyze this impressive literature, a critical issue emerges clearly, namely, the presence of words and expressions having many and sometimes opposite meanings. Significant examples are the terms 'realistic' and 'modeling' associated with word problems in school. Understanding how these terms are used is important in research, because this issue relates to the design of several studies and to the interpretation of a large number of phenomena, such as the well-known phenomenon of students' *suspension of sense making* when they solve mathematical problems. In order to deepen our understanding of this phenomenon, we describe a large empirical and qualitative study focused on the effects of variations in the presentation (text, picture, format) of word problems on students' approaches to these problems. The results of our study show that the phenomenon of suspension of sense making is more precisely a phenomenon of activation of alternative kinds of sense making: the different kinds of active sense making appear to be strongly affected by the presentation of the word problem.

1 Problem solving in mathematics education research

Problem solving is one of the main topics in mathematics education research, ever since its origin as a relatively young field of research. For example, the seminal work of Polya (1945) surely had, and still has, a great influence on research on problem solving, laying the groundwork for further developments. On the other hand, the research on mathematical problem solving before the 1970s had several limits, being a-theoretical, unsystematic and strictly based on a quantitative approach (Kilpatrick 1969).

It is unquestionable that, in line with the development of research in mathematics education as a field of research in its own right, problem solving research has made much progress during the last 50 years, changing perspectives, methods and also goals (Liljedahl et al. 2016). Retracing its evolution can lead to a better understanding of research choices, to identifying the consolidated findings and, ultimately, to making progress.

In order to do that, we can ideally subdivide the 50 years under consideration into two 25-year periods. In his detailed overview of the first of these two periods (1970–1994), Lester (1994) identifies the following main issues:

i. determinants of problem difficulty;
ii. distinction between good and poor problem solvers;
iii. instructions for the teaching of problem solving;
iv. role of metacognitive factors in problem solving.

In this list, we can recognize the three main protagonists of the problem solving activity, namely, the mathematical problem, the students, and the mathematics teacher.

According to Lester "there is no doubt that the nature of the research on problem solving has matured tremendously" (Lester 1994, p. 663). At the same time two significant issues are left unresolved, namely, the need for greater clarity in the meaning of terms, and the need to improve research methods. In particular, concerning the latter issue, Lester brings forward an interesting and pioneering reflection about the need to link the research methods to the research purposes. From the current perspective, we can also identify another significant limitation: the scarce attention to the role of social and affective factors in problem solving.

On the other hand, the general foundations for the reconceptualization of problem solving research were laid in the

✉ Pietro Di Martino
    pietro.di.martino@unipi.it

[1] Department of Mathematics, Università Federico II, Naples, Italy

[2] Department of Mathematics, Università di Pisa, Pisa, Italy

period between the 1980s and 1990s. In this time frame, the gradual affirmation of the interpretive paradigm in the social sciences caused a shift of methods and goals, from explaining phenomena in terms of cause-effect models to making sense of the world (Schoenfeld 1994). In particular, the main interest in problem solving research shifted from the development of effective programs for talented students to the interpretation of students' difficulties in problem solving activities.

In this new frame, several studies described "the apparently nonsensical things that students at all grade levels do when they attempt to solve mathematical problems" (Cobb 1986, p. 2), for example, when they respond with a numerical answer to absurd problems (Baruk 1985) or when they ignore realistic considerations in mathematical school problem solving (Verschaffel and De Corte 1997). To describe these phenomena, Schoenfeld (1991) coined the expression: *students' suspension of sense making* in solving word problems.

This new strand of studies *produced* a double awareness. First, the students' difficulties in problem solving could no longer be explained solely in terms of cognitive limitations (Schoenfeld 1985). The publication of the book "Affect and mathematical problem solving" (Adams and McLeod 1989) represented a sort of manifesto, based on a strong and, by now, consolidated finding: a purely cognitive approach is highly limited when interpreting students' mathematical behaviour in problem solving activities. Second, as Cobb underlines, often students must "resolve problems that are primarily social rather than mathematical in origin" (Cobb 1986, p. 2). In particular, school problem solving can no longer be considered and analysed as a solo activity: it is a complex social activity where explicit and implicit social rules developed in the mathematics classroom play a crucial role (Yackel and Cobb 1996).

Within a socio-constructivist framework, learning, teaching and doing mathematics involve the negotiation of meanings, rules and expectations (Voigt 1994) and the result of these negotiations is a crucial key to interpreting students' behaviours in the activity of problem solving. Insightful studies on school problem solving have been conducted within this framework and, in more recent years, this line of research has led to useful knowledge suggesting conditions for supporting problem solving in school (Liljedahl et al. 2016; Liljedahl 2019).

However, after 50 years of research in this area, we still need to move forward. Problem solving, and in particular word problem solving, is still a source of great difficulties for the majority of mathematics learners, irrespective of students' age (Verschaffel, Schukajlow, Star and Van Dooren 2020). Several recent studies have highlighted the increase over time of the students' suspension of sense making as they develop more experience in problem solving and

modeling in school (Mellone, Verschaffel and Van Dooren 2017). This evolution also involves the affective side: comparing the attitude towards problems of kindergarten and primary school students, Di Martino (2019) showed how the exposure to mathematical problems in primary school has a negative effect on the idea that children have about what a problem is and about how they have to deal with it, as well as on their self-perception and emotional disposition towards mathematical problems.

The phenomenon of students' suspension of sense making appears to be particularly serious with respect to its social and educational consequences. Realistic problem solving and mathematical modeling are nowadays considered central within the educational standards of several countries all over the world, and they are seen as crucial elements in the current understanding of mathematical competence (Hankeln 2020) and for the exercise of effective citizenship (Niss et al. 2016). On the other hand, students' difficulties in problem solving increase over time, sometimes without being evident to students and teachers. Often, they become evident during (school or tertiary) transitions when a specific approach to mathematical problems suddenly no longer works, causing a real crisis in students' mathematical identity (Di Martino and Gregorio 2019).

The expression 'students' suspension of sense making' draws attention to the student-dimension, however there are multiple variables that play a role in this phenomenon. In the case of word problems, one of these variables is the format—text, picture, general presentation—of the problem itself.

In the study on which we report in this paper, we investigate the effect of different variations in the presentation of word problems on students' answers and approaches to the problem. To achieve our research goal, we have designed a methodological cycle where new versions of the original problems are developed on the basis of the data previously collected. Our hypothesis is that even minimal variations in the problem presentation can affect students' approaches to the problem itself and, in particular, the emergence of their realistic considerations.

In the next section, we discuss the meaning of the main terms involved in our research. Despite the extensive literature, the meanings of key terms involved in the problem solving research are still not consensual and, as Schoenfeld (2000) underlines, this is a problematic issue. Mathematics education research has a cumulative nature: in particular new research should build on a critical analysis of previous research, and researchers have an intellectual obligation to push towards greater clarity.

2 Problem, word problems, realism and mathematical modeling

What is a 'problem'? According to Duncker: "A problem arises when a living creature has a goal but does not know how this goal is to be reached" (Duncker 1945, p. 1). In this perspective, a situation is *problematic* not in an absolute sense: it depends on the characteristics of the living creature who faces the problem at that particular moment. Moreover, *problem* and *routine* are incompatible: a problem is challenging by its nature, therefore, in general the activation of reproductive and automatic thinking in a problematic situation is not a *good strategy*.

Even if it is true that a routine task can be a departure point for challenging problem solving, this evolution rarely happens in school (Liljedahl et al. 2016). In particular, Duncker's definition induces a characterization of problem solving that appears to be far from what is used in the typical school problem solving session (Pongsakdi et al 2020).

Teachers—in particular, primary school teachers that are not usually mathematics specialists—are often afraid of the unpredictable outcomes of challenging problem solving activities and therefore they mostly promote routine tasks in their mathematics sessions (Russo and Hopkins 2019). It is no coincidence that students' suspension of sense making usually emerges as a result of an external intervention (by researchers, standardized assessments, etc.). The absurd problems (in the style of 'the age of the captain' problem) or the non-routine tasks widely used in literature (Greer 1993; Verschaffel, De Corte and Lasure 1994) would hardly be proposed by a teacher in his or her classroom.

This phenomenon emerges also in a study on pre-service teachers conducted by Verschaffel, Da Corte and Borghart (1997), in which they investigate pre-service teachers' evaluation of students' answers to a list of seven problematic problems. A problematic problem (P-problem) is defined as a non-routine task where realistic considerations need to be taken into account in order to give a meaningful answer (an example of P-problem is the very famous army bus problem: *450 soldiers must be bused to their training site. Each army bus can hold 36 soldiers. How many buses are needed?* quoted the first time by Carpenter et al. 1983). The researchers found that teachers usually consider P-problems as ill-formulated or tricky problems: "it was undesiderable or even totally inappropriate to confront fifth-grade children with such complex, ill-formulated or tricky problems" (Verschaffel, Da Corte and Borghart 1997, p. 357).

P-problems are word problems, intended to be realistic problems. One of the main reasons for the use of word problems in education is to let students experience an occasion in which to apply mathematics for solving real life problems, without the practical inconvenience of having direct contact with real world contexts (Verschaffel et al. 2000). In some sense, the phenomenon of students' suspension of sense making indicates that the school activity with word problems fails to fulfil this *modeling* goal (Mellone et al. 2017).

The relationship between word problems and modeling is controversial and highly debated. Some authors conceive word problems as a specific type of mathematical modeling problems (Verschaffel et al. 2020). Others underline that the word problems do not contain questions that are important in an authentic context: according to this view, modeling and solving word problems are two completely different activities (Kaiser 2017). Gerofsky, referring to the work of several philosophers, even discussed "the constitutional impossibility of 'real-life' word problems according to these theorists" (Gerofsky 2010, p. 61). However, all the quoted scholars underline some significant differences between word problems and problems encountered in daily life.

One of these differences concerns the verification of the reasonableness of the solution: in classroom simulations, once students find a mathematical solution to the problem, they do not use this solution and thus, they have no intrinsic reason to confront their solution with reality. This difference appears to be crucial in the phenomenon of students' suspension of sense making on which we are focusing.

De Franco and Curcio (1997) proposed to 20 sixth graders the following revisited version of the bus problem: *328 senior citizens are going on a trip. A bus can seat 40 people. How many buses are needed so that all the senior citizens can go on the trip?* They obtained 18 out of 20 incorrect responses (12 out of 18 were incorrect interpretation of the remainder). One month later, students were asked to order minivans to take 6th graders to a class party, really making a telephone call to a bus company: We need to transport 32 children to the restaurant so we need transportation. We have to order minivans. Board of Education minivans seat 5 children. These minivans seat 5 children. These minivans have five seats with seatbelts and are prohibited by law to seating more than five children. The two problems proposed by De Franco and Curcio are not mathematically isomorphic (328:40 is not the same as 32:5), however they are *isomorphic* with regard to the interpretation of the remainder. The real difference between the two problems is the request of the telephone call, the need to have a reality check. Indeed, 16 out of 20 students gave a correct response to the second problem, interpreting the remainder correctly.

Another important difference is that, in the school word problems, the person who has to solve the problem (usually the student) is not the same person as the one who administers it (the teacher, the researcher, etc.). For this reason, the students' beliefs about the implicit purposes of the problem

play a crucial role in their approach to the problem itself (Franchini, Lemmo and Sbaragli 2017; Zan 2011). For example, it is not clear how free students actually are to express their considerations about the realism or unrealism of the situation and to accept or refuse the implicit assumptions often needed to solve a word problem. Let us consider the Planks problem used in several studies: Steve has bought four planks each 2.5 m long. How many planks 1 m long can be sawn from these planks? (Verschaffel, Da Corte and Lasure 1994). In a real situation, the first reaction of an external observer would probably be: Why did Steve buy 2.5 m planks if he needs 1 m planks? It is the worst choice one could make.

The overall impression is that, as for the salt in some cooking recipes, the implicit indication for students in solving word problems is 'realism to taste': as a consequence, it is not so strange that students provide unrealistic answers to unrealistic problems with a vague purpose. Students can neglect realistic considerations if the word problem is unrealistic, but also if the *realistic* purpose is not explicit or if they have difficulties in quantifying the 'realism to taste' to be applied.

Even if there is a certain agreement among researchers on what a non-authentic problem is, there is no agreement about what an authentic problem is (Palm 2006). Different meanings of a realistic problem have been used in the literature. Cooper and Dunne (2000) defined a problem as 'realistic' if it contains either people or concrete objects. In our view, the reference to a real person or an object in the text of the problem definitely does not guarantee a realistic interpretation of the described situation. In this regard, Nesher (1980) reported some interesting reactions of first and second graders to the task "Tell a story which would correspond to the mathematical sentence $3+4=7$". A paradigmatic example of pupils' answers is: "I ate three cups and four plates".

Several problems used in literature have been considered as realistic according to the above naïve definition, but actually they are often unrealistic "in the sense that important aspects of the 'real' situations described in the tasks are not well emulated" (Palm 2008, p. 39). If it is evident that a word problem only emulates a real-life task situation, however, it is possible and necessary to take care of its realism. Palm (2008) developed a truly fine framework to assess the degree of realism of a word problem, identifying five crucial aspects, as follows:

- Event (the event has a fair chance of taking place);
- Question (it actually might be posed in the real-life event);
- Purpose (the purpose of the task context needs to be clear and explicit);
- Language (the linguistic difficulties have to be in line with those that can occur in the simulated real situation);

- Information/data (this aspect is subdivided into three subaspects: existence of information/data, realism of information/data, specificity of information/data).

Even though 'word problem' literarily means 'problem expressed through a text', in mathematics education research the term is used with several meanings, some of which are less generic:

> Word problems are typically defined as verbal descriptions of problem situations, presented within a scholastic setting, wherein one or more questions are raised the answer to which can be obtained by the application of mathematical operations to numerical data available in the problem statement or on numerical data derived from them. (Verschaffel et al. 2020, p.1).

This acceptation of 'word problems' appears to be questionable: word problems are limited to the arithmetical domain (numerical data) and it is assumed that they always have a solution that can be obtained through the application of mathematical operations. The idea that arithmetical operations are needed to solve a word problem is in line with students' stereotypical view of mathematical problems which is a possible source of difficulties, recreating the students' calculation orientation towards word problems (Verschaffel, Greer & De Corte, 2000).

In our framework, the definition of 'word problem' refers only to how a problem is expressed rather than to how it can be answered. In line with Gerofsky's idea (1996), we consider word problems as a proper literary genre characterized by three recurrent components, as follows:

i. a narrative component that introduces and describes the context and the characters;
ii. an informational component that gives the information needed to address the problem;
iii. the question component.

Gerofsky also introduces terms from linguistics in order to analyse the text of word problems. She underlined the potential role of parenthetical elements in the understanding of a text and, in relation to the suspension of sense making. Gerofsky further introduced the *implicature*, i.e. the reader's inferences beyond the literal meaning of the linguistic expressions uttered in the text. The implicatures in solving word problems can play a role in determining students' beliefs about the implicit purposes of the problem.

All of the three components described by Gerofsky appear to be significant in determining the degree of realism of a word problem. Several studies showed how the rewording of the text can change the students' approach to the problem and their understanding of the situation (Greer 1997; Mellone et al. 2017; Zan 2011).

Word problems can also be accompanied by illustrations. Elia and Philippou ([2004](#)) analyse the different functions of word problem illustrations, developing the following classification:

   i.   decorative—when the illustrations have no link to the context described in the word problem;

   ii.   representational—when the illustrations represent the context (or a detail of the context) of the problem;

   iii.   organizational—when the illustrations suggest directions for the solution procedure;

   iv.   informational—when the illustrations impact the informational component of the word problem, i.e., if they include data not included in the text and needed for the solution.

The effects of these different kinds of illustrations on students' approaches to a problem seem to be limited, since students tend to pay only superficial attention to the illustrations in solving word problems (Dewolf et al. [2015](#)).

So far we discussed different issues related to word problem solving and, in particular, to the suspension of sense making. We addressed the intrinsic distinction between word problems and problems in daily life, students' beliefs about the implicit purposes of the problem, and the degree of realism of the word problems. These issues suggest that rather than using the expression 'suspension of sense making', we should use the expression 'activation of a different sense making', since students' sense making does not seem to be suspended but addressed by other variables.

In our study, we analysed the effects of variations in the presentation of word problems on students' answers in order to better understand the suspension of sense making phenomenon. The variations developed involve all the three components described in Gerofsky's framework.

3 The study

3.1 Aims

The main aim of our study is to understand the effects that variations in the presentation of word problems have on students' approaches to problems. In particular, we are not interested in measuring the change of the students' success rate as in other studies (Vicente, Orrantia and Verschaffel [2007](#)). In line with an interpretative approach (Crespo [2000](#)), our purpose is to overcome the correct or incorrect dualistic approach, accessing, rather than assessing, students' approaches to different versions of a word problem.

3.2 Experimentation cycle

The process of variation of an assigned word problem is a *trial and error method*: some changes might not produce any recognizable effect on students' approach to the problem, others might highlight significant changes, also providing insights for new interpretative hypotheses to be tested with new trials and word problem formulations.

We designed an experimentation cycle (see Fig. [1](#)) where the production of a new version of the word problem used is developed based on the results of a previous experimentation. Adopting a grounded perspective, this experimental cycle is continued until theoretical saturation is achieved (Vollstedt and Rezat [2019](#)).

Phase 0 of the experimentation cycle consists of the selection of the original task. In our research we selected eight word problems within the INVALSI (the Italian National Evaluation Service) survey database. Even if these problems are characterized by multiple choice questions, they usually are challenging word problems and they offer significant quantitative data (Di Martino and Baccaglini-Frank [2017](#)). For the first implementation of the experimentation cycle, we always used the problem in its original version, adding the request: 'Explain your reasoning'.

Phase 1—Data collection. Data collection was subdivided into three steps: the collection of the individual students' written answers, the notes of a classroom discussion developed in a later session, the audio recording of students' individual interviews.

Students had one hour during a classroom period to try to solve one problem. The time allocated for the problem solving activity is one of the main boundary conditions because students need adequate time in order to activate productive reasoning. It is interesting to note that in only few research studies authors reported on the time given to students for solving an array of problematic problems, as if time were not crucial in the experimental setting.

The one hour classroom discussion that followed evolved around the students' comments on the realism of the word problem, and their description and comparison of the different strategies used to solve the problems. The individual interviews with a selection of students were designed to deepen some aspects of the students' answers (written or
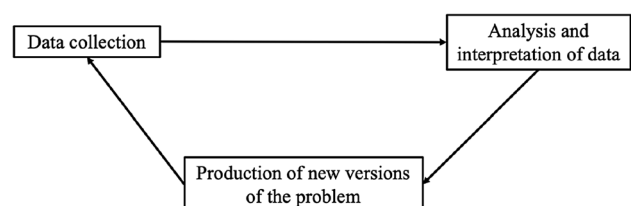

Fig. 1 The experimentation cycle

Fig. 2 The car transporter prob-
lem (translation by the authors)

The car transporter in the picture can carry a maximum of 10 cars.

62 cars in the factory are ready to be delivered.

*What is the minimum number of car transporters - like the one in the picture –
needed to deliver them all?   A- 6   B- 7   C- 6.2   D- 10*

oral). They lasted on average 20 min and were conducted
by the third author.

Phase 2—Analysis and interpretation of data. In line
with recent recommendations for the empirical research
on problem solving and mathematical modeling (Schuka-
jlow, Kaiser and Stillman 2018), we used a mixed methods
approach, integrating quantitative data with a qualitative
approach. We quantitatively compared the students' results
in our experimentation with the results in the Italian national
survey through which the problems were selected. Then, we
developed a qualitative analysis of the written answers to the
prompt 'Explain your reasoning' in order to reconstruct and
classify students' processes. Notes of the classroom discus-
sions and recordings of the individual interviews were used
to deepen some aspects emerging from the students' written
answers.

The analysis was conducted classifying different aspects,
namely, the students' numerical answers, the explicit math-
ematical processes reported in the written answers, and the
students' comments on the context described in the word
problem. This latter aspect was enriched by the notes related
to the class discussions and individual oral interviews.

Phase 3—Production of new versions of the problem. On
the basis of the results of Phase 2, one or more new ver-
sions of the word problem were developed in order to test an
interpretative hypothesis, leaving unaltered the mathematical
structure of the problem itself. These changes were classified
through Gerofsky's framework on word problems.

3.3 Original problem and populations

For reasons of space, we discuss the experimentation cycle
for only one of the eight selected word problems, the car
transporters problem, mainly focusing on students' written
answers. All the names used in the discussion are fictional.

The car transporters problem (CTP). This problem
(Fig. 2), administered by INVALSI in the 2016 grade 5

Table 1 Results of the Italian national sample

Omitted answer (%)	A (6) (%)	B (7) (%)	C (6.2) (%)	D (10) (%)
0.8	11.7	35.8	40.5	11.2

Table 2 Quantitative results—N_1 version

Omitted answer (%)	A (6) (%)	B (7) (%)	C (6.2) (%)	D (10) (%)
3.7	7.5	52.6	32.5	3.7

national assessment, has the following explicit aim: to test
students' control of division with remainder (INVALSI
assigns an explicit mathematical aim to each proposed
task).

On purpose we decided to test a grade 5 problem with
middle school students, because we wanted it to be math-
ematically accessible to as many students as possible. We
chose this problem because of the worrying results of fifth
grade Italian students (Table 1) and for the clear inspira-
tion provided by the already mentioned army buses prob-
lem (Carpenter et al. 1983). Schoenfeld coined the expres-
sion 'suspension of sense making' discussing students'
answers to the army bus problem; moreover, this problem
is frequently used in the literature because the realistic
aspects that need to be considered for a correct solution
appear rather visibly (Palm 2008).

Population. A total of 480 middle school students (in
grades 6–8) were involved in the four experimental cycles
developed for the CTP. The different versions of the prob-
lems were given to four non-overlapping groups of stu-
dents, which varied in their absolute number, maintaining
the same percentages of students from each school grade.

4 Results and discussion

4.1 N₁ version: the CTP in its original version

A total of 134 students from 13 different classes were involved in the first cycle. Although, unlike the INVALSI national sample, our sample is not representative, the quantitative results (Table 2) confirm the high appeal of the *unrealistic* option C, there is also a significant increase in the percentage of correct answers.

Two differences between the national survey and our experimentation are evident: our sample is made up of middle grade students, while students involved in the INVALSI national survey were fifth graders; and there was a significant difference in the time limitation because the INVALSI national survey asks students to answer 33 problems in 75 min. The different conditions concerning time also have a clear effect: students given sufficient time have the time for reflecting on their first answer and, eventually, for going beyond this first attempt. The corrections we found on the students' papers confirm that several students took advantage of this opportunity (Fig. 3).

Focusing on students who chose B, the analysis of the reported mathematical processes identifies four categories (Table 3). In contrast to the explicit aim of the problem (related to division with remainder), we observed that the approach students used the most involves multiplication.

However, our main focus (related to the suspension of sense making) is on students who chose option C. Multiplication is the most commonly used operation also in this case, however a peculiarity emerges: the process often starts with the numbers given in the four options, as in the case of Edo (Fig. 4). The answer options are a crucial part in the presentation of multiple choice word problems, having a clear impact on the students' approach. Considering Gerofsky's framework, the answer options seem to be considered by the students as a sort of appendix to the informational component.

Four other significant aspects emerged from the qualitative analysis of the students' written answers.

(i) *The car that disappeared and the role of illustration*. As often happens in mathematics education research, unexpected data may reveal a hidden phenomenon. We found

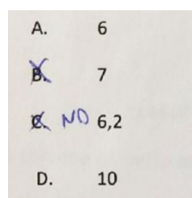**Table 3** The four processes for the choice of the correct option—N₁ version

Explicit process	Percentage (%)
Computation of a division and interpretation of the numerical results (students in this group usually used the *exact* division 62:10 = 6.2, deducing the need for 7 car transporters)	26.9
Computation of a multiplication and interpretation of the numerical results (i.e., $6 \times 10 = 60$, then 6 car transporters are needed plus another one for the transportation of the two remaining cars)	52.5
Elimination of incorrect answers	14.9
No written justifications	5.7

that several students used the number 9 in their calculations (see Fig. 5).

This is very strange because the number 9 is not included in the CTP text, however the discussion and individual interviews sections allowed us to understand the origin of this '9'. Students explained they did not use the number in the text, but they counted the cars in the illustration and many of them did not consider the less visible black car behind the driver' seat.

This students' attention towards the illustration is not in line with Dewolf et al.'s results (2015): Why did this happen?

The illustration used is hardly classifiable in Elia and Philippou's categories; it seems to be more representational than informational, since it does not contain information that is not represented in the text, but it has been interpreted as informative by several students. On the other hand, if it is true that the textual informational component includes the number 10 (the maximum load of the trucks), analysing the CTP text within Gerofsky's framework, it also includes a parenthetical element ("like the one in the picture") that seems to have led students' attention towards the illustration. We also observe that the INVALSI authors made the choice of representing a fully loaded truck.
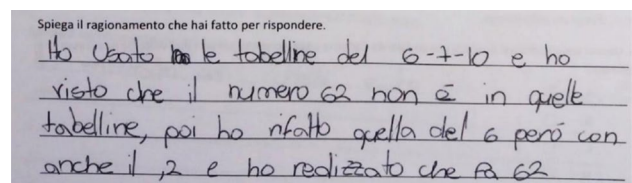
Fig. 3 Example of changed answer


```
A.      6
B.      7
C. NO  6,2
D.      10
```


Fig. 4 Edo's answer (grade 6)—N₁ version ("I used the 6, 7 and 10 times tables and I realized that the number 62 does not belong to any of those, then I did the 6 times table again but with also the ,2 and I realized it is 62")

Fig. 5 Sara's answer (grade 6)—N_1 version ("Knowing that the cars in the truck are 9, I made the computation (9×7=63); the minimum number of trucks is 7 because in each truck there are 9 cars")
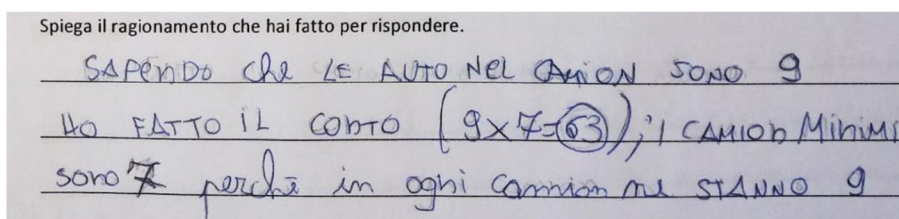
(ii) *The different meanings related to the answer 6.2*. As our quantitative data show, many students seem to ignore the realistic consideration, using the words of Verschaffel and colleagues (1994), about the *indivisibility of a truck*, choosing the option C (6.2). This trend—recurrent in all studies involving *the bus problem* in the literature—is usually interpreted as a significant sign of the students' tendency to ignore real-world considerations in word problem solving. The reference to the *indivisibility of a truck* is actually the main difference in the students' answers distinguishing between those who choose the correct answer and those who choose the option C.

However, the qualitative analysis of the collected data offers another interpretation of this phenomenon, related to the students' use and understanding of the decimal part of the number 6.2. Some students chose the answer 6.2 not because they ignored the real-world context, but because they gave a personal interpretation to the digit 0.2. Some students interpret 0.2 using a sort of proportional vocabulary: 1 is the truck represented in the figure, 0.2 is a smaller truck. Leonardo (grade 6) wrote: "We cannot represent 6.2 trucks, unless one of these trucks is one fifth smaller than the others". For other students, the 0.2 means a smaller cargo. Lorenzo (grade 6) wrote: "6 trucks transport 60 cars, then the 6.2 means that another truck transports the last 2 cars". This phenomenon is also reported in Palm's study: "A few students giving the answer 7.5 buses said that they meant 7 buses with students on every seat and 1 bus that was only half filled with students" (Palm 2008, p. 53).

(iii) *Uncertainty about purpose*. As Gerofsky (1996) described, the question component in a word problem is often interpreted by the students beyond its literal meaning. The students' beliefs about the implicit purposes of the problem can play a significant role in their attempts; however, it is not simple to collect information about such beliefs without the explicit request 'Explain your reasoning' (Di Martino and Baccaglini-Frank 2017). Mattia's written answer (Fig. 6) is particularly interesting in this sense.

Mattia marked two different options (7 and 6.2) showing awareness of the fact that the answer to the contextualized problem is 7. On the other hand, 6.2 appears to be the answer he sees as appropriate to show that he controls the division algorithm very well. Mattia seems to be undecided about the *real* purpose of the CTP, leaving both answers, as if to say: "whatever your goal is, I have the answer".

(iv) *The implicit constraints*. Matteo (grade 7) wrote: "I did 62:10. It is 6.2 but we cannot divide a single trip. For this reason, I choose 7". Matteo calculated the number of trips needed rather than the number of car transporters needed as

Fig. 6 Mattia's answer (grade 6)—N_1 version ("At first I did 62:10=6.2, so I find the number of trucks, but I realized that one half of a truck does not exist. Therefore, this problem has two answers. They are: 7 trucks or, as I did initially, 6.2 trucks")
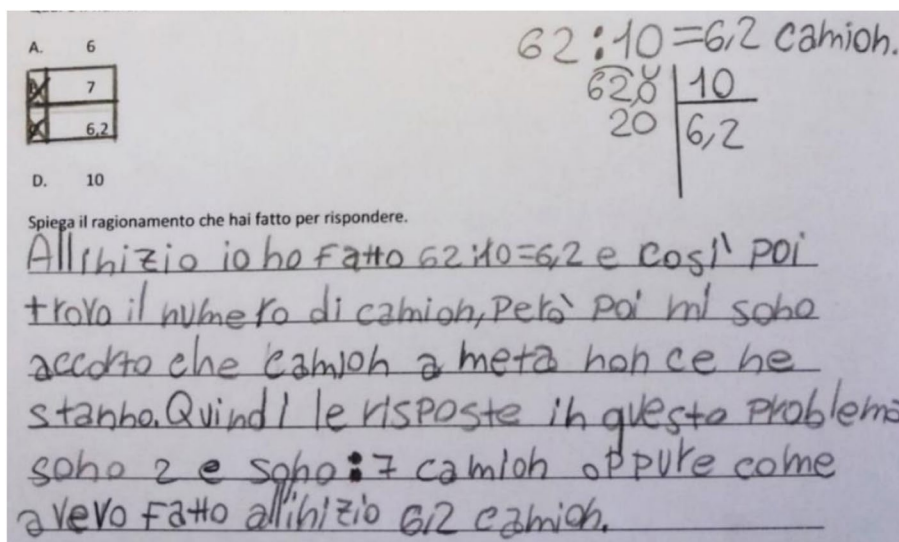
Table 4 Quantitative results—N$_2$ version

Omitted answer	6 (%)	7 (%)	6.2 (%)	Other answers (%)
4.3	13.7	67.6	5.8	8.6

requested. Actually, the correct answer to N$_1$ is 7 if and only if we assume as implicature one of the following constraints: each car transporter can be used for a single trip, or all 62 cars must be transported simultaneously. The assumption of one of these implicatures in modeling the situation is forced by the possible answer choices because they do not include the answer 1 (a unique car transporter making 7 trips). Once again, the model of the situation is constrained by aspects unrelated to the real-world situation described in the text and this seems to be the case for all *bus problems* used in the literature: going beyond the realism 'to taste', the correct answer to all bus problems could always be 1.

On the basis of this analysis, we developed and experimented with three new versions of the CTP.

4.2 N$_2$ version: removing the answer choices

The presence of answer choices acted as a sort of appendix to the informational component. As we described, many students choosing the option 6.2 in the N$_1$ version of CTP based their processes on the numbers given in the four options. In order to better understand the extent of the effect of this extra informational component, we developed and experimented with the N$_2$ version of the CTP, removing the multiple choice from the N$_1$ version.

A total of 139 students from 13 different classes participated in the experimentation of N$_2$. The quantitative effect of the absence of the answer choices is quite evident: the answer 6.2 becomes residual (Table 4).

From a qualitative point of view, the most evident effect is the greater variety of approaches to solving the N$_2$ version of the CTP than those reported in Table 3 for the N$_1$ version. The absence of the multiple choice has effects not only on the students' answers, but also on the variety and quality of their processes. Obviously, it is no longer possible to proceed with the elimination of the alternatives in the N$_2$ version, while the computation of a multiplication or the computation of a division remain the two most commonly reported processes. In the case of division, students computed the exact division between 62 and 10, developing realistic considerations to give the right answer, 7, which is in line with the findings of other studies (Palm 2008). However, two other strategies emerged in the N$_2$ data, namely, the computation of a repeated sum, and the *draw a picture* strategy.

In the first case, the students calculated $10 + 10 + 10 + 10 + 10 + 10$ adding a seventh not fully loaded truck. In her oral interview Martina (grade 6) explained: "a truck can transport up to 10 cars, but it can also transport fewer cars. To conclude: a truck can travel with two cars, and then we need seven trucks". Martina's clarification about the load of the seventh truck is common to almost all the students using the repeated sum strategy.

This clarification is arithmetically needed because otherwise a seventh 'plus ten' would exceed 62. However, it is also related to the process of modeling. During different classroom discussions, students underlined how it is unusual to see a car transporter travelling not fully loaded: they usually are empty or full. Luigi (grade 6) stated in his individual interview: "I have never seen a car transporter with two cars! The car transporters are fully loaded or unloaded when they come back. For me the right answer is six car transporters: the two cars left will be transported later". Also in this case, overcoming the 'realism to taste', Luigi's position can be considered a valid argumentation for the answer 6, and his explanation is anything but suspension of sense making.

One of the recurrent foci of the classroom discussions for N$_2$ was exactly the best distribution of the 62 cars on the 7 trucks needed. Students debated the *adequacy* of the required answer, underlying how the purely numeric solution (7) leaves room for different organizations of the truck trips. In particular, a general agreement emerged about the fact that the solution corresponding to the sum $10 + 10 + 10 + 10 + 10 + 10 + 2$ is not optimal. Several students suggested a more balanced car distribution on the seven trucks for *reasons of symmetry* or reasons of savings: the balanced solutions guarantee a lower consumption of fuel and tires.

The will to balance out the cargo of the seven trucks also emerged in the data of the students using the draw a picture strategy. As in the case of Martina, the redistribution of the cars in the seven trucks implies a deep understanding of the information "can carry a maximum of ten cars". For example, Camilla (grade 7, Fig. 7) developed
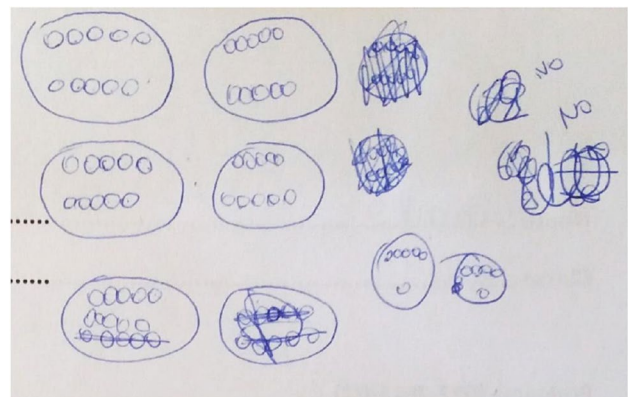
Fig. 7 Camilla's drawing (grade 6)—N$_2$ version

Fig. 8 Andrea's answer—N2 version ("6 trucks and one half of a truck. I thought that we need 6 trucks in order to transport 60 cars, then the front of the truck can transport the other two cars")
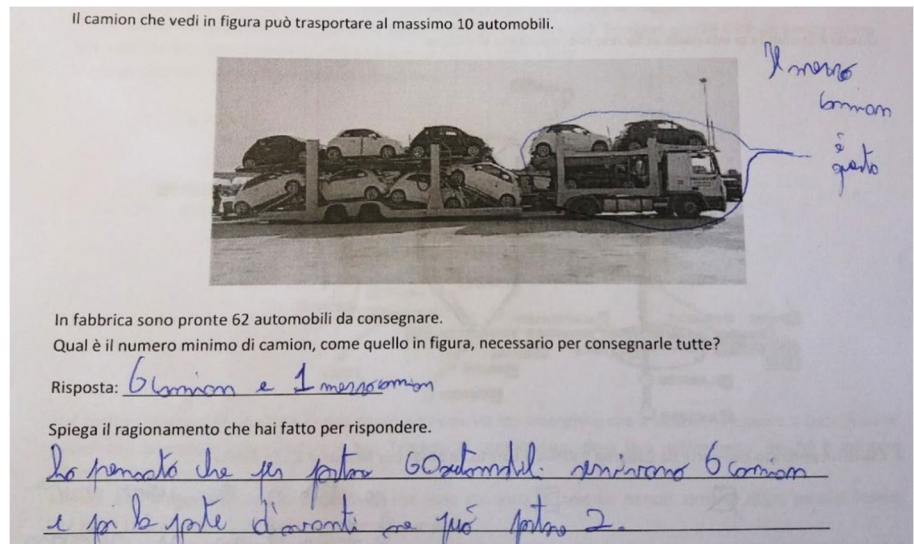
Fig. 9 N$_3$ version of the CTP (translation by the authors)

The car transporter in the picture can carry a maximum of 10 cars.

62 cars in the factory are ready to be delivered. The factory has a unique car transporter like the one in the picture.

What is the minimum number of trips that the car transporter has to make in order to deliver all of them?   A- 6   B- 7   C- 6.2   D- 10

several graphic attempts finally coming to the decision to equally distribute the last 12 cars on 2 trucks (from an arithmetic point of view this corresponds to the sum 10 + 10 + 10 + 10 + 10 + 6 + 6). Camilla explained that "Each truck can transport a maximum of 10 cars. This does not imply that each truck has to transport ten cars. Therefore five trucks transport ten cars and the other two trucks transport six cars".

To conclude the analysis of the data from the N$_2$ version, we mention the very interesting answers of Andrea (grade 7, Fig. 8): "six trucks and one half". Andrea's interpretation of a fraction of truck is partly in line with Leonardo's view: he marked a part of the truck in the

Table 5 Quantitative results—N$_3$ version

Omitted answer (%)	A (6) (%)	B (7) (%)	C (6.2) (%)	D (10) (%)
1.9	6.9	61.8	26.5	2.9

illustration of his A4 paper, handwriting in the margin: "this is the half truck". The truck in the picture is actually composed of two parts, namely, the tractor (the one half for Andrea) and the trailer.

4.3 N₃ version: from the number of trucks to the number of trips

Reflecting on Matteo's answer to the version N_1 of the CTP, we administered a reworded version, N_3 (Fig. 9), modifying the question component, replacing the term 'trucks' with the term 'trips'.

A total of 102 students from 9 different classes were involved in the experimentation of this N_3 version. Unlike the data of the N_2 version, the quantitative data of the N_3 version (Table 5) are similar to those of the N_1 version, with a small increase of the percentage of correct answers (61.8% vs 52.6%).

The analysis of students' processes also confirms a substantial analogy with the data reported in Table 3 for the version N_1; however, the rewording of the question component has two interesting effects.

The first one is the growth in the number of explicit references to the indivisibility of the *object* (the trip) in the written answers of the students who chose the correct option. This may sound strange: the indivisibility of a truck should be more evident, instead it would seem possible to imagine a part of a trip. Nevertheless, several comments in the classroom discussions underlined that the trucks have to complete their trips in order to deliver the cars.

The rewording also caused a debate about the meaning of 'trip'. During the classroom discussions the following question was discussed: does the trip include the outward and the return or does the trip correspond to a single direction? As in version N_1, the multiple choice format introduces an external constraint, because the option '14' is absent. Fabio (grade 6) is the only one that explicitly refers to the conflict between these external constraints and his representation of the situation: "None of the given options is correct! We have also to consider the return trip, therefore the minimum number of trips is 14".

4.4 N₄ version: the division with remainder

The results of the versions N_1 and N_3, as well as the INVALSI national survey, confirm the appeal of the *unrealistic* option 6.2. As we discussed, this choice is the result of two very different reasonings by the students. In some cases, the choice follows a purely computational approach, thus falling within the well-known phenomenon of students' tendency to ignore realistic considerations in problem solving. In other cases, there is a use of the mathematical symbol 0.2 (the decimal part of 6.2) with a non-mathematical meaning: it represents a truck transporting only two cars. These two very different approaches produce the same numerical choice because ten is the maximum number of cars a truck can transport; thus the decimal part of the quotient of the division corresponds to the remainder of the integer division.

Table 6 Quantitative results—N_4 version

Omitted answer (%)	A (7.6) (%)	B (7) (%)	C (7.75) (%)	D (8) (%)
1.9	12.4	10.5	33.3	41.9

We developed version N_4 (from the N_1) in order to differentiate the numerical answers resulting from these two different approaches. In this version, the maximum load of the trucks passes from 10 to 8 (the illustration was edited accordingly) and the proposed options are as follows:

- A. 7.6 (6 is the remainder of the integer division 62:8);
- B. 7 (the lower integer part of 62:8);
- C. 7.75 (the quotient of the exact division 62:8);
- D. 8 (the correct answer).

A total of 105 students from 9 different classes were involved in the experimentation of version N_4. From a quantitative point of view, there is a clear reduction in the percentage of correct answers, however all the options were chosen by at least 10% of the students (Table 6).

From a qualitative point of view, the analysis of students' written comments shows how students answering C applied an abstract arithmetic scheme without realistic or contextual consideration: having the number of objects (X) and the number of objects a container can contain (Y), the result of the division of X by Y is the number of containers needed to contain all the objects. Alessandro wrote as follows: "In order to find the number of needed trucks I did 62 (total number of cars) divided by 8 (maximum number of cars for each trucks). It is 7.75".

Vice versa, students answering A considered the realistic context: their numerical answer is not a natural number, rather it refers to two integral numbers of trucks and cars. For example, Giuly (grade 6) justified her choice (7.6) with the following words: "I did 62:8 (a fully loaded truck). It is 7 and remainder 6. That is, the total number of trucks is 7 plus 1 truck with 6 cars". Despite computing a multiplication, Gaia (Fig. 10) used the symbol 7.6 with the same meaning as Giuly gave it: "First, I did the times table of eight in order to understand which result was closer to the number of cars to be delivered. The closer result that does not exceed that number was 8 (maximum number of cars for each truck) times 7 (car transporters). But, it still was not 62, 6 cars were still missing. Therefore, the answer is 7.6".

Students answering A seem to feel the need that the answer comes from numbers directly obtained by applying some arithmetical operations to the numerical data in the

Fig. 10 Gaia's answer (grade 6)—N$_4$ version
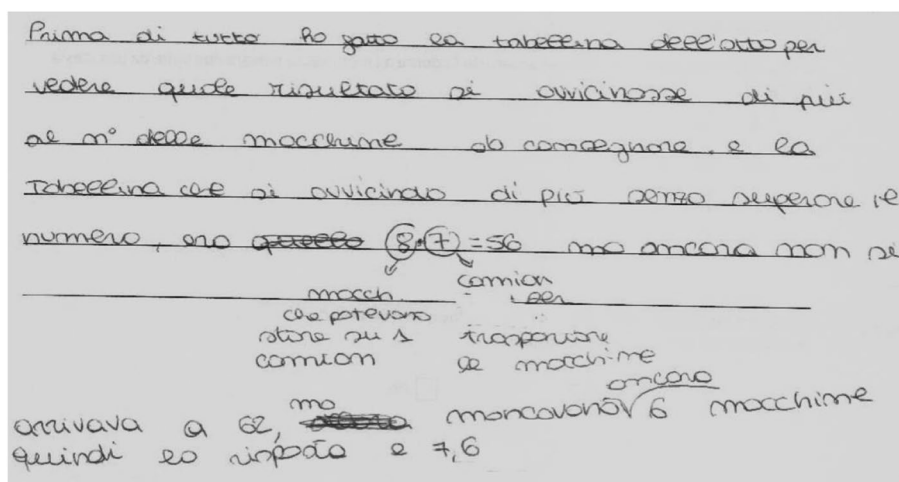
text. However, are we sure that their behavior is a manifestation of the phenomenon of suspension of sense making? We do not think so. These students consider the context of the CTP, in particular bearing in mind the indivisibility of a truck but adopting an alternative meaning for the decimal part of a non-integer number, a meaning that is different from the mathematical one. The proposed rewording of the CTP gave insight into this last approach, allowing us to distinguish it from the purely computational one.

5 Conclusions

The main aim of our study was to investigate the effects of variations in the presentation of word problems on students' answers and approaches to the problems.

First, we discussed the differences between realistic word problems and daily life problems. In line with Kaiser's view (2017), we consider word problems and modeling as two different worlds, with different rules and constraints. In this view, the main initial hypothesis was that 'suspension of sense making' is actually the 'activation of a different kind of sense making'. This hypothesis is confirmed by the effects of the variations in CTP presentation on students' approaches. As we discussed, the presentation of the word problem is surely not the only variable influencing the phenomenon of suspension of sense making in problem solving. However, it is significant that we observed relevant changes in students' processes and answers by experimenting with the variations in the presentations of CTP, variations that substantially do not modify the context described.

The results of the variations also suggest some refinements to Gerofsky's framework. In particular, all the three components she considers in a realistic word problem seem to have an informational role. As we have seen, the options in the multiple choice format and the illustration describing

the type of trucks (see Sara and Andrea's answers, reported in Figs. 5 and 8) can have an informational role, adding information to the text of the word problem and strongly affecting the students' processes. Especially what Gerofsky identities as the narrative component has a crucial, although underestimated, informational role. Goulet-Lyle, Voyer and Verschaffel, describing three categories of inferences during word problem solving, state: "Inferences deemed unnecessary are those which, while providing a richer understanding of the story around the problem, are not specifically oriented toward its resolution" (Goulet-Lyle et al. 2020, p. 142). This kind of inferences is highly necessary, determining the quality of students' realistic considerations (Zan 2011).

As our data show, all the elements included in the word problem presentation can have a strong informational component for the students, because such elements can resolve some contextual aspects such as implicit and boundary conditions. Students' comprehension of this broadly understood informational component is rarely investigated. This understanding is usually inferred through the students' numerical answers. Our data show that this approach is questionable.

From a methodological point of view, our aim challenged us to develop an experimentation cycle to test several interpretative hypotheses related to a *fixed* context for the word problem. We made the choice to collect mainly qualitative data because, in our opinion, the focus on the process is crucial to developing an interpretative approach to the phenomenon of suspension of sense making. As our data show, an a priori and absolute classification of a class of problems appears to be questionable, since students' approaches to the same problem can be quite different and sometimes unpredictable. For example, students' answers to the bus problem are often explained in the literature in terms of difficulties in interpreting the remainder in a division problem (Greer 1997). Our results showed a much more complex picture.

If it is true that we have no direct access to students' ideas, we believe that the reflective narratives we collected are particularly informative. According to Bruner (1990), we are interested in what the individual thinks he has done, rather than in an *objective* report, which is hard to imagine.

In particular, the multiple data collected and analysed in the experimental cycle highlight some specific mathematical issues: for example, the alternative meanings of the decimal part of a number and, at a more general and interesting level, the role of the 'realism to taste' in the appearance of the phenomenon of suspension of sense making. Using terms from the field of medical testing, our study seems to confirm that there are a lot of *false positive* students to whom the phenomenon of suspension of sense making is attributed, i.e., students who actually activate an alternative kind of sense making.

References

Adams, V., & McLeod, D. (Eds.). (1989). *Affect and mathematical problem solving: A new perspective*. New York: Springer Verlag.

Baruk, S. (1985). *L'age du capitaine. De l'erreur en mathématiques*. Paris: Seuil.

Bruner, J. (1990). *Acts of meaning*. Cambridge: Harvard University Press.

Carpenter, T., Lindquist, M., Matthews, W., & Silver, E. (1983). Results of the third NAEP mathematics assessment. Secondary school. *Mathematics Teacher, 76*, 652–659.

Cobb, P. (1986). Contexts, goals, beliefs, and learning mathematics. *For the Learning of Mathematics, 6*(2), 2–9.

Cooper, B., & Dunne, M. (2000). *Assessing children's mathematical knowledge: Social class, sex and problem-solving*. Buckingham: Open University Press.

Crespo, S. (2000). Seeing more than right and wrong answers: Prospective teachers' interpretations of students' mathematical work. *Journal of Mathematics Teacher Education, 3,* 155–181.

De Franco, T., & Curcio, F. (1997). A division problem with a remainder embedded across two contexts: Children's solution in restrictive vs real-world settings. *Focus on Learning Problems in Mathematics, 19*(2), 58–72.

Dewolf, T., van Dooren, W., Hermens, F., & Verschaffel, L. (2015). Do students attend to representational illustrations of non-standard mathematical word problems, and if so, how helpful are they? *Instructional Science, 43,* 147–171.

Di Martino, P. (2019). Pupils' view of problems: The evolution from kindergarten to the end of primary school. *Educational Studies in Mathematics, 100,* 291–307.

Di Martino, P., & Baccaglini-Frank, A. (2017). Beyond performance results: Analyzing the informational and developmental potentials of standardized mathematics tests. *For the Learning of Mathematics, 37*(3), 39–44.

Di Martino, P., & Gregorio, F. (2019). The mathematical crisis in secondary-tertiary transition. *International Journal of Science and Mathematics Education, 17,* 825–843.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 58*(5), 1–113.

Elia, I., & Philippou, G. (2004). The functions of pictures in problem solving. In M. J. Hoines & A. B. Fuglestad (Eds.), *Proc. of the 28th IGPME conference* (Vol. 2, pp. 327–334). Bergen: University College.

Franchini, E., Lemmo, A., & Sbaragli, S. (2017). The role of text comprehension in the process of mathematizing and modeling. *Didattica della matematica, dalle ricerche alle pratiche d'aula, 2017*(1), 38–64.

Gerofsky, S. (1996). A linguistic and narrative view of word problems in mathematics education. *For the Learning of Mathematics, 16*(2), 36–45.

Gerofsky, S. (2010). The impossibility of 'real-life' word problems (according to Bakhtin, Lacan, Zizek and Baudrillard). *Discourse: Studies in the Cultural Politics of Education, 31*(1), 61–73.

Goulet-Lyle, M., Voyer, D., & Verschaffel, L. (2020). How does imposing a step-by-step solution method impact students' approach to mathematical word problem solving? *ZDM Mathematics Education, 52,* 139–149.

Greer, B. (1993). The modeling perspective on wor(l)d problems. *Journal of Mathematical Behavior, 12,* 239–250.

Greer, B. (1997). Modeling reality in mathematics classrooms: the case of word problems. *Learning and Instruction, 7*(4), 293–307.

Hankeln, C. (2020). Mathematical modeling in Germany and France: A comparison of students' modeling processes. *Educational Studies in Mathematics, 103,* 209–229.

Kaiser, G. (2017). The teaching and learning of mathematical modeling. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 267–291). Reston: NCTM.

Kilpatrick, J. (1969). Problem solving and creative behavior in mathematics. In J. W. Wilson & L. R. Carey (Eds.), *Reviews of recent research in mathematics education. Studies in Mathematics Series* (Vol. 19, pp. 153–187). Stanford: School Mathematics Study Group.

Lester, F. (1994). Musings about mathematical problem solving research: 1970–1994. *Journal for Research in Mathematics Education, 25*(6), 660–675.

Liljedahl, P. (2019). Conditions for supporting problem solving: Vertical non-permanent surfaces. In P. Liljedahl & M. Santos-Trigo (Eds.), *Mathematical problem solving. ICME-13 Monographs* (pp. 289–310). Cham: Springer.

Liljedahl, P., Santos-Trigo, M., Malaspina, U., & Bruder, R. (2016). *Problem solving in mathematics education. ICME-13 topical surveys*. Cham: Springer.

Mellone, M., Verschaffel, L., & Van Dooren, W. (2017). The effect of rewording and dyadic interaction on realistic reasoning in solving word problems. *The Journal of Mathematical Behaviour, 46,* 1–12.

Nesher, P. (1980). The stereotyped nature of school word problems. *For the Learning of Mathematics, 1*(1), 41–48.

Niss, M., Bruder, R., Planas, N., Turner, R., & Villa-Ochoa, J. A. (2016). Survey team on: Conceptualisation of the role of

competencies, knowing and knowledge in mathematics education research. *ZDM Mathematics Education, 48,* 611–632.

Palm, T. (2006). Word problems as simulations of real-world situations: A proposed framework. *For the Learning of Mathematics, 26*(1), 42–47.

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 67*(1), 37–58.

Polya, G. (1945). *How to solve it*. Princeton: Princeton University Press.

Pongsakdi, N., Kajamies, A., Veermans, K., Hannula-Sormunen, M. M., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM Mathematics Education, 52,* 33–44.

Russo, J., & Hopkins, S. (2019). Teachers' perceptions of students when observing lessons involving challenging tasks. *International Journal of Science and Mathematics Education, 17,* 759–779.

Schoenfeld, A. (1985). *Mathematical problem solving*. Orlando: Academic Press.

Schoenfeld, A. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce of formal and informal mathematics. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 311–343). Hillsdale: Erlbaum.

Schoenfeld, A. (1994). A discourse on methods. *Journal for Research in Mathematics Education, 25*(6), 697–710.

Schoenfeld, A. (2000). Purposes and methods of research in mathematics education. *Notices of the AMS, 47*(6), 641–649.

Schukajlow, S., Kaiser, G., & Stillman, G. (2018). Empirical research on teaching and learning of mathematical modeling: A survey on the current state-of-the-art. *ZDM Mathematics Education, 50,* 5–18.

Verschaffel, L., & De Corte, E. (1997). Teaching realistic mathematical modeling in the elementary school: A teaching experiment with fifth graders. *Journal for Research in Mathematics Education, 28*(5), 577–601.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 22*(4), 273–294.

Verschaffel, L., De Corte, E., & Borghart, I. (1997). Pre-service teachers' conceptions and beliefs about the role of real-world knowledge in mathematical modeling of school/word problems. *Learning and Instruction, 7,* 339–360.

Verschaffel, L., De Corte, E., & Greer, B. (2000). *Making sense of word problems*. Lisse: Swets and Zeitlinger.

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM - Mathematics Education, 52,* 1–16.

Vicente, S., Orrantia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology, 77,* 829–848.

Voigt, J. (1994). Negotiation of mathematical meaning and learning mathematics. *Educational Studies in Mathematics, 26,* 275–298.

Vollstedt, M., & Rezat, S. (2019). An introduction to grounded theory with a special focus on axial coding and the coding paradigm. In G. Kaiser & N. Presmeg (Eds.), *Compendium for early career researchers in mathematics education. ICME-13 monographs* (pp. 81–100). Cham: Springer.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458–477.

Zan, R. (2011). The crucial role of narrative thought in understanding story problems. In K. Kislenko (Ed.), *Proc. of MAVI 16* (pp. 287–305). Tallin: University of Applied Sciences.