

Can visual aids in representational illustrations help pupils to solve mathematical word problems more realistically?

Tinne Dewolf¹ · Wim Van Dooren¹ · Lieven Verschaffel¹

Received: 9 June 2015 / Revised: 2 April 2016 / Accepted: 16 May 2016 /

Published online: 6 June 2016

© Instituto Superior de Psicologia Aplicada, Lisboa, Portugal and Springer Science+Business Media Dordrecht 2016

Abstract We investigated the effect of two visual aids in representational illustrations on pupils' realistic word problem solving. In part 1 of our study, 288 elementary school pupils received an individual paper-and-pencil task with seven problematic items (P-items) in which realistic considerations need to be made to come to an appropriate reaction. These items were presented together either with representational illustrations, representational illustrations in which an element was added to make the realistic modelling complexity more apparent, or representational illustrations in which this element was cued. In part 2, the pupils received the same P-items together with a realistic and a non-realistic answer option, with the request to choose the best answer. The findings show that there was no positive effect of the visual aids on the number of realistic reactions in part 1 and that when reviewing possible answers to P-items in part 2, there again was no positive effect.

Keywords Mathematical word problems · Line drawings · Cueing · Realistic problem solving

Introduction

In the research literature, there is ample evidence that higher elementary and lower secondary school children, and even higher education students, solve mathematical word problems without much consideration of the realities of everyday life (Verschaffel et al. 2000). According to most mathematics educators, this is considered as problematic, since the historically most important function of word problems is to offer learners the opportunity to learn to apply the mathematics that they have learned at school in real life (=application function) (Verschaffel et al. 2000). This form of applied problem solving—otherwise called mathematical modelling—requires that learners use their mathematical knowledge and skills and their real-world knowledge integratively in the different phases of the solution process,

✉ Lieven Verschaffel
lieven.verschaffel@ppw.kuleuven.be

¹ KU Leuven, Leuven, Belgium

especially in the initial phases wherein they first have to understand the situation and build a situation model, and then construct a mathematical model, and in the final stages wherein they have to derive a mathematical solution and to interpret their solution in relation to the real life situation (Verschaffel et al. 2000). In an attempt to study more systematically learners' approach towards word problem solving, Greer (1993) and Verschaffel et al. (1994) confronted pupils with word problems that are problematic from a realistic modelling perspective (so-called P-items) in the sense that real-life knowledge should be taken into account to come to a correct reaction (i.e., a realistic reaction, RR). They found that most pupils solved these problems as if there was no realistic modelling problem at all; they applied the arithmetic operations suggested by the situation described, under the assumption that the situation could be unproblematically mapped onto these operations. In many countries, replication studies were conducted, and strikingly similar findings were found (see Verschaffel et al. 2000). In addition, researchers tried to encourage learners to use real-life considerations when solving P-items, by adding a warning about the non-routine or problematic nature of some of the problems in the test (Verschaffel et al. 2000; Yoshida et al. 1997). However; the results of these studies were not or only moderately positive. So, in previous studies, we presented the problems together with illustrations that may help to represent the problem situation (Dewolf et al. 2014, 2015). We hypothesized that providing word problems with such representational illustrations would help pupils to mentally imagine the problem situation and, therefore, to react to the word problems in a more realistic way. This expectation was based on the integrated model of text and picture comprehension of Schnotz and Bannert (2003), which assumes that, under certain conditions, reading text accompanied with pictures will lead to a more integrated mental model than when this mental model has to be constructed using the textual channel only. We applied this model of text and picture comprehension to the process of mathematical modelling (Verschaffel et al. 2000), and argued that, when complementing a word problem with an appropriate picture, learners will build a more elaborated mental model of the problem situation (=situation model) (Dewolf et al. 2014, 2015). However, the results of these studies were also disappointing in the sense that there was no positive effect of the illustrations on pupils' realistic reactions (Dewolf et al. 2014, 2015).

In the present study, we built further upon this line of research and investigated the effect of two visual aids in the illustrations, on pupils' realistic reactions (RRs) to these P-items. First, we used representational illustrations in which an extra element was added to make the realistic modelling complexity more apparent and, second, we cued this added element by means of highlighting it. In what follows, we will give a summary of previous research about learners' lack of sense-making when solving mathematical word problems, followed by a short review of the literature about the use of representational illustrations and cueing.

Realistic word problem solving

In their abovementioned pioneering study, Verschaffel et al. (1994) made a distinction between standard word problems (S-items) and problematic word problems (P-items). S-items are classical word problems such as "Pete wins 3 marbles in a game and now has 8 marbles. How many marbles did he have before the game?" (Verschaffel et al. 2000, p. ix), which can be solved by means of a straightforward operation with the numbers in the problem (in this case $8-3=5$ marbles). P-items, on the other hand, require consideration of more subtle aspects

of the real-world situation being described to come to a situationally appropriate reaction. For example, the P-item “A man wants to have a rope long enough to stretch between two poles 12 m apart, but he has only pieces of rope 1.5 m long. How many of these pieces would he need to tie together to stretch between the poles?” cannot be solved correctly simply by dividing 12 m by 1.5. The learner has to take into account somehow that some extra rope is needed to make the knots and to stretch the rope between the poles.

Verschaffel et al. (1994) asked upper elementary school children to solve ten S-items and ten parallel P-items individually in a paper-and-pencil task. Their reactions on the P-items were coded as non-realistic (NR) or realistic (RR) depending on their use of everyday knowledge and realistic considerations in their answer and/or in their possible additional comments. A reaction was scored as a NR when it was the result of a straightforward execution of the mathematical operation, without any further comment about the problematic nature of the problem or real-life consideration that might jeopardize the appropriateness of the executed operation(s). In contrast, when the answer—a precise numerical answer, an answer indicating some kind of estimation, or an answer stating that the problem is unsolvable—was the result of the use of real-world knowledge related to the realistic modelling issue involved in the P-item, it was considered as a RR. Also when a pupil gave a straightforward non-realistic answer but made an additional comment that contained a trace of a realistic consideration about the modelling issue involved in the P-item, it was considered as a RR. For example, for the P-item mentioned above, the answer “ $12 : 1.5 = 8$ ropes” without any remarks about the knots would be considered as a NR, whereas reactions such as the answer “More than 8 ropes, because you need some extra rope to make all the knots and to connect the rope to the poles” but also the non-realistic answer “ $12 : 1.5 = 8$ ropes” followed by the comment “At least, if one would not take into account all the knots” would be considered as RRs. The authors found that pupils generally handled the ten P-items the same as the ten S-items, in the sense that only 17.0 % of all pupils’ reactions to these P-items was considered as realistic. The authors’ conclusion that pupils tend to neglect their everyday knowledge when solving mathematical word problems, was afterwards confirmed in many replication studies with upper elementary and lower secondary school pupils (see Verschaffel et al. 2009) and also with higher education students (Inoue 2005; Verschaffel et al. 1997).

The use of representational illustrations

Another crucial element of the current study’s theoretical and empirical background refers to the literature about the role of illustrations (i.e., visualizations or depictions such as drawings, sketches, paintings, photographs, and other graphical representations) in educational settings in general (Mayer 2005; Schnotz and Bannert 2003) and in word problem solving in particular (e.g., Elia and Philippou 2004). The latter researchers, for instance, differentiated between decorative, representational organizational and informational illustrations in word problem solving. In the present study, we used illustrations that depict the real life context (including the realistic modelling complexity) in the word problems. Referring to the categorisation of Elia and Philippou (2004), these illustrations can be termed as “representational illustrations”, i.e. illustrations that—in contrast to purely decorative illustrations—have a link with the word problem and depict the whole or a part of the content, but that do not provide directions that support the solution and that do not contain information that is essential to solve the word problem—in contrast to respectively

organizational and essential illustrations. These representational illustrations were expected to help pupils to build a situation model and consequently discover the problematic nature of the P-items.

In previous research, we already investigated the influence of such representational illustrations on pupils' RRs to P-items (Dewolf et al. 2014). In two similar studies, one in Turkey and one in Belgium, upper elementary school pupils aged 10–11 received a paper-and-pencil test consisting of 16 word problems (eight S-items and eight P-items) together with representational illustrations, which they had to solve individually. The word problems were presented in four conditions: with or without representational illustrations (see Fig. 1 for an example of the representational illustration for the rope item) and with or without a warning that the test would contain some non-standard problems. It was expected that the representational illustrations, especially in combination with the warning, would help the pupils to mentally imagine the problem situation and consequently help them to solve the problems more realistically. However, no positive effect of the representative illustrations was found, even not in combination with the warning. Most pupils still responded unrealistically to the P-items. The percentages of RRs in all four conditions for the Turkish and the Belgian sample were, respectively, between 9.7 and 13.8 % and between 10.5 and 13.6 %.

In two related follow-up studies, we investigated why these illustrations did not help learners to solve the P-items more realistically (Dewolf et al. 2015). In a first study, the eye movements of higher education students were measured while they were solving P-items. Students who were assigned to the first condition received the word problems together with the representational illustrations from the study of Dewolf et al. (2014). In the second condition, students received the items together with decorative illustrations (that had no link whatsoever with the items), while in the third condition, students received the problems without any illustrations. The study showed that there was no increase in RRs in the first condition as compared to the two other conditions and that this could to a large extent be explained by the fact that students in many cases simply did not look at the illustrations. However, when they did look at the illustrations, they attended significantly more to representational illustrations than to decorative ones. Based on this main finding that students barely looked at the illustrations, a second study was set up in which the presentation of the illustrations was manipulated so that students had to look for five seconds at the illustration before they got the accompanying word problem. The word problems were presented with a decorative illustration, a representational illustration, or a representational illustration with a warning about the usefulness of the illustration for solving the word problem. Again there were no differences between conditions. So, even when students were forced to look at the representational illustrations that accompanied the P-items, and even when they were moreover alerted about the usefulness of these illustrations, no increase in the number of RRs was observed.

A man wants to have a rope long enough to stretch between two poles 12 m apart, but he has only pieces of rope 1.5 m long. How many of these pieces would he need to tie together to stretch between the poles?



Fig. 1 The rope item together with its representational illustration (Dewolf et al. 2014)

Cueing specific elements

Cueing can be defined as “the manipulation of visuospatial characteristics of instructional materials in order to help learners in selecting relevant information, and organizing and integrating the information into a coherent representation” (de Koning et al. 2009, p. 114). Typical examples of visual cues are arrows, circles, boxes and forms of highlighting that rely on visual contrast (Schnotz and Lowe 2008). In the research literature, cueing has been suggested as an effective instructional measure to direct learners’ attention to a key aspect of information. De Koning et al. (2009) for example distinguish three functions of cueing: (1) to guide viewers’ attention towards specific information, (2) to emphasize the major topics of instruction and their organization and (3) to foster integration by making relations between elements more salient. With respect to the first function, Lowe and Boucheix (2010) indicate that the underlying assumption is that the cued elements more likely will be noticed, extracted and incorporated into the viewers’ mental model. The present study relied on this function of directing viewers’ attention towards key aspects in illustrations. More specifically, we directed viewers’ attention towards one or more elements in the representational illustrations that made the realistic modelling complexity involved in the P-item more visible.

The present study

Overall rationale and design

Previous studies showed that adding representational illustrations to P-items did not lead to an increase in the number of RRs to these items (Dewolf et al. 2014). Two follow-up studies (Dewolf et al. 2015) revealed that learners scarcely looked at the representational illustrations next to the P-problems and that when they did look at them (because they were experimentally forced to do so), there was still no positive effect on the number of RRs. Two possible (complementary) explanations for the absence of a positive effect of these representational illustrations are (1) that the design of the representational illustrations used so far was too weak to draw the attention to the modelling complexity underlying the accompanying P-item and (2) that the learners’ previous histories with illustrations accompanying school word problems prevented them from processing these representational illustrations in a sufficiently attentive and detailed way to allow them to significantly profit from them. To counter these two possible causes, we included in the present study two visual aids. First, we designed new representational illustrations in which a specific element was added with the aim to make the realistic modelling complexity in the P-item more apparent. Second, we cued this specific element by highlighting it, with a view to attract the learners’ attention to that specific element. We investigated the effect of these two visual aids on the number of RRs on P-items. The study consisted of two parts. In part 1, pupils were asked to solve seven P-items. In condition 1 (RI condition), the P-items were presented together with representational line drawings. In condition 2, the P-items were presented together with the same line drawings as in condition 1, but a specific element was added so that the realistic modelling complexity was made more apparent (RI+ condition). In condition 3, pupils received the same representational illustrations as in condition 2, but now the added specific element was highlighted with an orange marker with a view to attract the pupils’ attention to that specific element even more (RI++ condition). After pupils had solved all seven P-items in one of the three conditions, they were confronted, in part 2 of the study, with two possible answers (a RR and a NR) to each P-item, together with the request to

choose the best alternative and to explain why. At the end of part 2, pupils got a general question about the experienced helpfulness of the illustrations.

Hypotheses

First, we hypothesized that there would be a positive effect on pupils' realistic word problem solving of a first visual aid consisting of adding a specific element to the representational illustrations so that the realistic modelling complexity was made more apparent, and an even greater effect of a second visual aid of cueing this specific element by means of highlighting it (*hypothesis 1*). We expected that the addition of a specific element to make the realistic modelling complexity more apparent would help pupils to construct a richer and more realistic mental representation of the situation of the problem and, thus, to solve it more realistically. Furthermore, the extra cue on that specific element of the illustration was expected to draw pupils' attention to this element even more and therefore to help them even more to represent and solve the item more realistically. So, we predicted that pupils would give more RRs to the P-items in part 1 of the study in the RI+ condition than in the RI condition and even more in the RI++ condition (*prediction 1*).

Second, and analogously to hypothesis 1, we hypothesized a positive impact of the first and even a greater impact of the second visual aid on pupils' appreciation for the realistic response when confronted with both the realistic and the non-realistic response alternative in part 2 (*hypothesis 2*). More specifically, we first expected that, compared to the RI condition, more pupils in the RI+ condition would choose the RR as the better alternative, and even still more would do so in the RI++ condition (*prediction 2a*). However, we decided to test hypothesis 2 also for a specific and particularly interesting subset of trials, namely those trials wherein the pupil had reacted with a NR during the first part of the study. So, we further predicted that, for the trials being solved with a NR in part 1 of the study, the RI+ condition would yield more choices for the RR alternative than the RI condition and that even more RRs would be selected in the RI++ condition (*prediction 2b*). If we would observe this positive effect of the RI+ and RI++ manipulation for this particular subgroup of trials wherein the P-item had initially been solved non-realistically, it would support the claim that the two visual aids may bear a positive impact on pupils' appreciation for the RR, without necessarily manifesting themselves also in pupils' own actual response behaviour (as assessed in part 1 of the study).

Lastly, we hypothesized that there would be a positive effect of the two visual aids on pupils' experienced helpfulness of the illustrations (*hypothesis 3*). It was expected that, because of the addition of a specific element in the illustration that made the realistic modelling complexity more apparent in the RI+ condition, and because of the additional cueing of this element by means of highlighting it in the RI++ condition, more pupils in the RI+ condition and even more in the RI++ condition would report that the illustrations were helpful to solve the items, as compared to the RI condition (*prediction 3*).

Method

Participants

Participants were 288 upper elementary school pupils (girls = 134, boys = 154) between 9 and 12 years old ($M = 10.67$, $SD = 0.67$) from the 5th and 6th grade. These pupils, coming from 16

classes from four different schools, were within each class randomly assigned to one of the three conditions (the RI condition, the RI+ condition, or the RI++ condition). In RI condition were 91 pupils (girls = 43, boys = 48), in the RI+ condition 100 pupils (girls = 47, boys = 53) and in the RI++ condition 97 pupils (girls = 44, boys = 53).

Material and procedure

As explained before, all pupils were presented with seven P-items in a paper-and-pencil task. These problems were presented together with illustrations that, depending on the condition, represented the global situation of the problem (RI condition), the global situation of the problem but with the realistic modelling complexity being made more apparent by adding a specific element (RI+ condition) or the situation of the problem with that added element being cued by means of highlighting it (RI++ condition). In part 1 of the study, pupils had to solve the seven P-items individually. In part 2, pupils received the same P-items and the accompanying illustrations together with two possible reactions, a RR and a NR, from which they had to indicate which alternative was better and why. At the end of part 2, they also had to answer a general question about the helpfulness of the illustrations.

Part 1: solving word problems Table 1 contains an overview of the seven P-items used in the study. Five of these items were adopted from the research of Verschaffel et al. (1994), and were also used in our previous studies about the role of illustrations (Dewolf et al. 2014, 2015), while the other two were new. Of the group of five, three items remained unchanged, the planks, the buses and the rope item, while the other two, i.e. the school and the flask item, were somewhat reformulated. The pond and the truck item were new.

In all three conditions, these P-items were presented together with a line drawing in black and white that represented the situation of the word problem. These illustrations were the same

Table 1 Overview of the seven P-items used in the present study

Item	Statement
Pond	Sven and Karen are nature enthusiasts. They counted how many ducks there are in the pond in the woods. Sven has counted 12 ducks and Karen has counted 18 ducks. How many ducks are there in the pond?
Planks	Steve has bought 4 planks of 2.5 m each to make a bookshelf. How many planks of 1 m can he get out of these planks?
Buses	450 soldiers must be bussed to their training site. Each army bus can hold 36 soldiers. How many buses are needed?
Truck	A truck with two pallets of bricks is being weighed on a weighbridge and weighs 6800 kg. A third pallet of bricks is being loaded on the truck. How much does the truck with the pallets of bricks weigh now?
School	The sports centre is located 17 km from the school, and the station is located 8 km from the school. How far are the sports center and the station located from each other?
Rope	A man wants to have a rope long enough to stretch between two poles 12 m apart, but he has only pieces of rope 1.5 m long. How many of these pieces would he need to tie together to stretch between the poles?
Flask	Kim fills a conic shaped flask with water. The height of the flask is 12 cm. Kim looks how long it takes till the flask is completely full. The depth of the water is 4 cm after 10 s. After how many seconds shall the flask be full?

as the ones in the previous studies (Dewolf et al. 2014), except that line drawings instead of coloured illustrations were used because of the modifications we had to make for the RI+ condition and the RI++ condition. As shown in Table 2, the illustrations in the RI+ condition were identical to those in the RI condition, except that an element was added in order to make the critical realistic modelling complexity more apparent. The illustrations were still considered representational, because they represented the situation described in the problem, without containing essential information that is not given in the problem to come to the solution. For instance, in the illustration of the planks P-item in Table 2 two boxes are added: one for the planks for the bookshelf and another one for the remaining pieces of plank that are useless for making the bookshelf. In the RI++ condition these added elements in the illustrations were highlighted with an orange marker. For the planks item, for instance, the box for “waste” was highlighted. These line drawings were presented at the right side of the text (as in Fig. 1).

Every P-item was printed on a separate page, with ample room for pupils’ answer, their calculations and possible additional comments. The seven P-items were presented in two different orders. The experimenter introduced the task to the pupils as a mathematics class assignment that would not be graded. The pupils had to solve the task individually and were not allowed to ask questions about the task or specific items. They were told that if they had a question or query they could note it on their sheet. They were allowed to use a calculator.

With a view to maximise the chance that the visual aids would have an effect, two kinds of warnings were included in all three conditions. Both warnings were provided orally by the experimenter at the end of the general introduction to the task and were also written on the first page of the booklet. The first warning was as follows: “Pay attention, some problems in the task are difficult to solve because they are unclear. If you encounter such a problem, write as precisely as possible why it is unclear”. The second warning stated: “Look closely at the illustrations! They can help you to understand and solve the problem”.

Part 2: choosing between alternative responses When all pupils were finished with part 1 and all booklets were collected, the paper-and-pencil task of part 2 was distributed. As mentioned above, in part 2, pupils first received the seven-word problems from part 1 once again but this time together with two (and in two cases three) possible reactions, without any further calculation or explanation. These reactions were either RRs or NRs and pupils were asked which reaction they thought was the best and to explain why. For example, for the rope item, the RR and NR were, respectively, “More than 8 pieces of 1.5 m are needed” and “There are 8 pieces of 1.5 m needed”.

After pupils had indicated their preferential reaction for the seven P-items, they were invited to indicate how strongly they agreed or disagreed with the following general statement “The illustrations have helped me to find the answer”.

Results

Part 1: solving word problems

To test our first prediction, pupils’ answers on the seven P-items were coded as RRs or NRs, in exactly the same way as in the previous studies of Verschaffel et al. (1994) and Dewolf et al. (2014). The percentage of RRs was calculated on a total of 2008 trials (since eight items were not solved). Across all conditions, only 23.8 % of the reactions was scored as realistic. The percentage of RRs

was more or less the same in all three conditions: 25.0 % RRs in the RI condition, 20.2 % in the RI+ condition and 26.2 % in the RI++ condition.

To determine whether there were significant differences in the number of RRs between conditions, we analysed the data with a repeated measures logistic regression analysis (using

Table 2 Overview of the illustrations per condition



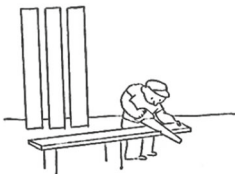
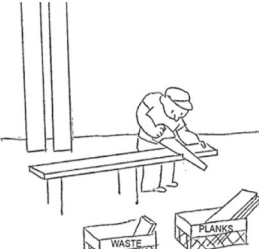



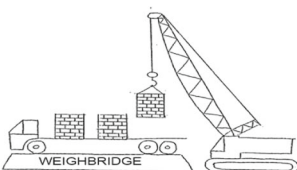
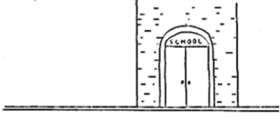


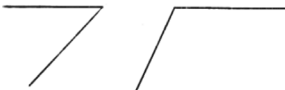
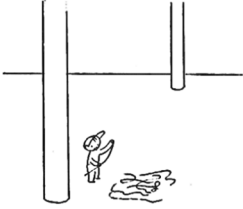
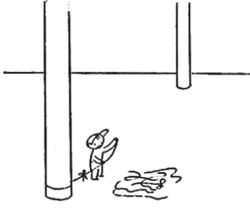


RI	RI+	RI++ *
Pond item		
		<p>The looking lines going from the two children's eyes to the ducks in the pond</p>
Planks item		
		<p>The box of waste filled with some remaining pieces of planks</p>
Buses item		
		<p>The empty seats in the last bus</p>
Truck item		
		<p>The truck on the weighbridge</p>

Table 2 (continued)

School item		
		The unfixed arrow on the man's shoulder
		
Rope item		
		The rope around the first pole and the knot linking the two pieces of rope
Flask item		
		The bottles on the table

^aIn the RI++ condition, participants were confronted with exactly the same illustrations as in the RI+ condition except that the part of the illustration mentioned in the RI++ column was highlighted in colour

the Generalised Estimating Equations (GEE) module in SPSS) with RRs as dependent variable and condition as independent variable. This analysis showed no significant effect of condition, $Wald X^2(2,2008) = 5.07, p = .079$. So, contrary to prediction 1, the two visual aids did not lead to an increase in pupils' RRs on the P-items.

Additionally, when looking at the number of RRs given by individual pupils (see Table 3), it turns out that more than two thirds of the pupils gave fewer than three RRs and that very few pupils solved five or more P-items realistically.

Part 2: choosing between alternative responses

Before testing predictions 2a, 2b and 3, we explored the number of times the RR was chosen as the best alternative, and compared it to the number of self-generated RRs during part 1 of the study.

Table 3 Percentage of pupils giving a specific number of realistic reactions (RRs), per condition and in total

Number of RRs	RI	RI+	RI++	Total
0	15.4	29.0	16.5	20.5
1	35.2	30.0	32.0	32.3
2	24.2	22.0	24.7	23.6
3	14.3	12.0	13.4	13.2
4	7.7	5.0	8.2	6.9
5	2.2	1.0	4.1	2.4
6	1.1	1.0	0.0	0.7
7	0.0	0.0	1.0	0.3

Because answers for 39 trials were missing, 1977 answers were analysed. Overall, pupils still preferred the NR option in 61.0 % of the cases, implying that they chose the RR in only 39.0 % of the trials. This was significantly more than in the first part of the study, where only 23.8 % RRs were given, $Wald X^2(1,3985)=209.81$, $p<.000$. So, although the NR remained the most preferential answer in the majority of cases, the confrontation with the RR in part 2 anyhow resulted in an increase of RRs, as compared to part 1 wherein the pupils had to generate a response themselves. This increase is also found when only including the trials for which a NR was given in the first part of the study. Indeed, when looking only at the trials being answered with a NR during part 1 of the study (76.2 % of all trials), it was found that in 18.0 % of all these NR trials, pupils preferred the realistic response when confronted with both response alternatives.

In relation to prediction 2a, we compared, for all trials from part 2 of the study, the percentages of RRs per condition. They were, again, very similar: in the RI condition, there were 38.8 % preferences for the RR, in the RI+ condition 36.9 % and in the RI++ condition 41.5 %. The GEE analysis showed again no significant difference between conditions, $Wald X^2(2,1977)=2.20$, $p=.332$, rejecting prediction 2a that claimed a positive effect of the visual aids on pupils' appreciation for the RR in part 2 of the study. When only including the trials for which a NR was given in part 1 of the study, the percentages for the RI condition, RI+ condition and RI++ condition were again very similar, namely, respectively, 22.7, 23.7 and 24.4 %. The GEE analysis showed again no significant differences $Wald X^2(2,1504)=0.322$, $p=.851$, indicating that prediction 2b had to be rejected too.

Furthermore, to test prediction 3, pupils' reactions to the statement that the illustrations had helped them to solve the problems showed that in general, 10.2 % of the pupils strongly agreed, 31.8 % agreed, 31.4 % had no opinion, 18.0 % did not agree and 8.5 % strongly disagreed. So, only 42.0 % of the pupils indicated that the illustrations had been of any help. When looking at the percentages per condition, more pupils in the RI+ condition and the RI++ condition than in the RI condition (strongly) agreed that the illustrations had helped to find the answer (see Table 4).

Table 4 Percentage of pupils' reactions on the question about the helpfulness of the illustrations, per condition and in total

	RI	RI+	RI ++	Total
I strongly disagree	12.5	7.1	6.3	8.5
I disagree	23.9	16.2	14.6	18.0
I have no opinion	33.0	32.3	29.2	31.4
I agree	28.4	33.3	33.3	31.8
I strongly agree	2.3	11.1	16.7	10.2

Although there was a clear trend in the expected direction, a chi-square test showed that the effect of condition on pupils' reactions on the statement was not significant, $Wald \chi^2(8283) = 15.07, p = .058$. So, prediction 3 was also rejected.

Conclusion and discussion

The aim of the present study was to investigate the effect of two visual aids on the realism of elementary school pupils' reactions to mathematical word problems that involve problematic modelling assumptions from a realistic point of view (P-items): (1) enriching representational illustrations with a specific element that makes the realistic modelling complexity involved in the P-item more apparent and (2) cueing this added element by means of highlighting it.

Our first prediction was that pupils would give more RRs in the RI+ condition than in the RI condition, and even more in the RI++ condition, because of the expected positive impact of the two visual aids (*prediction 1*). The results of part 1 revealed that neither of these two visual aids helped to answer the P-items realistically.

Secondly, we predicted that in part 2, for all the trials (*prediction 2a*) as well as for the subset of trials answered with a NR during part 1 (*prediction 2b*), the RR would be chosen least frequently in the RI condition and most frequently in the RI++ condition. When all trials were included, the data showed again no significant differences between the three conditions, and also when focusing on the trials answered with a NR in part 1 of the study, there were no significant differences between the conditions in pupils' preference for the realistic alternative. So, both predictions 2a and 2b had to be rejected too. Though there was again no effect of condition in part 2, there was a modest increase in the number of RRs from part 1 to part 2.

Thirdly, we predicted that, compared to the RI condition, more pupils in the RI+ condition and even more in the RI++ condition would indicate that the illustrations had been helpful (*prediction 3*). There was a trend in the expected direction, but the effect of condition was again not significant.

These findings show once more that elementary school pupils have a very strong and resistant tendency to exclude their everyday knowledge from solving word problems and that the representational illustrations as used in previous studies (Dewolf et al. 2014, 2015) are of little or no help to increase pupils realistic reactions. However, the findings from part 1 additionally reveal that altering the representational illustrations from the studies of Dewolf et al. (2014) so that the realistic modelling complexity is depicted more visibly (=condition 2) and cueing the added element by means of highlighting it with a view to attract pupils' attention to this element more strongly (=condition 3) did not seem to provide pupils any help to answer P-items more realistically (even not when explicitly warning them about the tricky nature of the problems and about the helpfulness of the illustrations). But also when being confronted with given RRs on these P-items, in part 2 of the study, the majority of the pupils stuck to their NRs, and the two visual aids (and the warnings) continued to be of little or no help. The fact that these two manipulations were ineffective in encouraging pupils to answer the P-items realistically (in part 1) or choosing the realistic answer option (in part 2) demonstrates that pupils seem to be very determined to keep solving P-items as if they were S-items. As in the older studies in the Einstellung effect (Luchins 1942), pupils' repetitious previous experiences with classical word problems (of the S-type) seem to have brought them into a mental set that makes them behave routinely vis-à-vis all word problems and to prevent them from addressing P-items in a thoughtful way.

Furthermore, these findings leave us with the question why the two visual aids that focused on the realistic modelling complexity (and even the explicit warnings) did not elicit more RRs to the P-items themselves (part 1) or more preferences for the RR alternatives (part 2). Based on previous research and the present findings, we believe that the combination of two kinds of beliefs—i.e. beliefs about word problem solving and about the use of illustrations accompanying word problems—lie at the basis of the lack of significant effects of the two experimental manipulations. First is the beliefs about word problem solving that pupils have developed and cultivated through their yearlong participation in the mathematics classroom practice and culture. Examples are the beliefs that there is always one correct numerical solution when solving word problems, that a correct solution always requires one or more arithmetic operations, that the problem contains all the information needed to find the solution of the problem, that all given numbers should be used to come to the solution and that no information extraneous to the problem can be sought (Jimenez and Verschaffel 2014; Reusser and Stebler 1997; Verschaffel et al. 2000). A second possible explanation why pupils did not profit from the visual aids is because they also may have developed certain beliefs about the role and importance of the illustrations next to mathematical word problems, as a result of their yearlong experiences in the school mathematics practice and culture. In a discussion on perceived informativeness of pictures in general, Weidenmann (1989) stated that pictures “are susceptible to being undervalued (...) because the subjective ease of encoding them at a superficial level may lead the learner to the illusion of a full understanding. As a consequence, the subject may stop the information processing after a short glance.” (pp. 162–163). Besides these explanations focusing on pupils’ beliefs, one can also look for an explanation for the finding that the two visual aids did not elicit more realistic reactions or preferences by looking at these aids themselves. As explained in Dewolf (2014), it is probable that even more explicit pictures of the realistic modelling complexities than the ones used in the present study (e.g., a picture that zooms in on the person tying together two pieces of rope in the rope item) or other types of external representations (such as confronting pupils with informational rather than representational pictures, which necessarily have to be processed in order to be able to solve the problem, Elia and Philippou 2004) or with dynamic instead of static illustrations (Höffler and Leutner 2007), might be more helpful in increasing the realism of pupils’ reactions to P-items. As far as the broader theoretical implications of our studies are concerned, it seems hard to account for our findings in terms of Schnotz and Bannert’s (2003) model of text and picture comprehension that was referred to in the introduction. Based on that model, we expected that the extra visual aids in the representational illustrations would help participants to create an elaborated mental model of the problem situation of the P-items (=situation model), leading to a more appropriate mathematical model and ultimately, to a realistic reaction. But this was not what we found. Apparently, the cognitive-psychological model of Schnotz and Bannert does not pay sufficient attention to factors that determine whether a picture is attended to at all and how the importance of processed information coming from the two different channels is balanced and valued by the subject in the later stages of the comprehension process, such as the socio-cultural setting wherein the comprehension task is situated and participants’ beliefs about the importance of textual versus pictorial information in that particular setting. Integrating these socio-cultural and affective factors into Schnotz and Bannert’s model is an important challenge for future research.

The present study has some limitations. First, space restrictions did not allow to provide a detailed analysis of the results in the three conditions for each of the P-items involved in the study. As several authors have pointed out (Greer 1993; Verschaffel et al.

2000), P-items differ with respect to their solvability. Whereas for some items a precise realistic answer can be given (e.g. the planks item), for others, one can only answer realistically in terms of a number range (e.g. the school item), and for still others, only an approximate number (e.g. the rope item). Arguably, these differences between P-items may have a serious impact on the realism of pupils' reactions to these items and, more importantly, on the effectiveness of the extra visual aids in the illustrations accompanying them. However, a more detailed analysis at item level yielded no significant differences between the three conditions, except for one item, namely the buses item, for which the R+ condition appeared to contain fewer RRs than the other two. Second, the experimental design of the study did not yield direct evidence on pupils' thinking processes vis-à-vis the P-items and their accompanying illustrations in the different conditions. So, future research should aim at a better understanding of *why* these illustrations did not yield the expected effect, by looking in a more fine-grained way at how pupils actually perceive the illustrations, how they handle them, and how they think about their importance and function. This could be done by means of in-depth individual interviews wherein pupils are confronted with a number of illustrated P-items and different types of illustrations and explicitly questioned about their understanding and solution processes as well as about their underlying beliefs about word problems and illustrations accompanying them. In a complementary line of research one could try to uncover the origins of these pupil beliefs, by systematically investigating the number, function and nature of illustrations next to word problems in mathematics textbooks and by analysing the way teachers handle these illustrations in their mathematics lessons. This kind of studies may also help to explain why pupils react differently to P-items and to various kinds of efforts to increase the realism of their reactions via illustrations.

In anticipation of the results of this future research, we give already some tentative recommendations for the practice of word problem solving education and the place of illustrations in particular. First, we recommend teachers to be aware that, due to the current practice and culture of school word problems, pupils seem to gradually develop a superficial approach and associated beliefs towards word problem solving in which there is little or no room for realistic world knowledge and realistic considerations and that the repetitious presentation of classical S-items may bring pupils into a mental set in which they keep solving P-items as if they were S-items. So, teachers but also textbook writers may try to stimulate and guide pupils in approaching and solving word problems more thoughtfully and more realistically by confronting them with word problems for which their routine-based approaches and their inaccurate beliefs do not hold and by engaging them in discussions wherein these approaches and beliefs are explicitly addressed (see Verschaffel and De Corte 1997, for an illustrative intervention study). Second, we recommend teachers and textbook writers also to keep in mind that pupils may not always interpret and use the illustrations next to the word problems as is intended by their designers (Elen 2013). They for example may perceive these illustrations as purely decorative and therefore ignore them or only scan them very superficially, or they may consider these illustrations as informationally inferior to the textual input and treat them accordingly in their problem-solving endeavours. So, as educators we need to look for ways to teach pupils how to handle these illustrations in relation to the problem text.

Acknowledgments This research was funded by grant *** from the Research Fund **. This research was partially supported by Grant GOA 2012/10 “Number sense: analysis and improvement” from the Research Fund KU Leuven, Belgium.

References

- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: guidelines for research and design. *Educational Psychology Review*, *21*, 113–140. doi:10.1007/s10648-009-9098-7.
- Dewolf, T. (2014). Get the picture? Are representational illustrations effective in helping pupils to solve mathematical word problems realistically? (Unpublished doctoral dissertation). KU Leuven, Leuven, Belgium.
- Dewolf, T., Van Dooren, W., Ev Cimen, E., & Verschaffel, L. (2014). The impact of illustrations and warnings on solving mathematical word problems realistically. *Journal of Experimental Education*, *82*, 103–120.
- Dewolf, T., Van Dooren, W., Hermens, F., & Verschaffel, L. (2015). Do students attend to representational illustrations of nonstandard mathematical word problems, and, if so, how helpful are they? *Instructional Science*, *43*, 147–3171.
- Elen, J. (2013). “Instructional disobedience”: Challenging instructional design research. *Plenary lecture given at the 15th European Conference for Research on Learning and Instruction*, Munich, Germany.
- Elia, I., & Philippou, G. (2004). The functions of pictures in problem solving. In M. J. Hoines & A. B. Fuglestad (Eds.), *Proceedings of the 28th conference of the international group for the psychology of mathematics education* (Vol. 2, pp. 327–334). Bergen, Norway: University College.
- Greer, B. (1993). The mathematical modeling perspective on wor(l)d problems. *Journal of Mathematical Behavior*, *12*, 239–250.
- Höfler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: a meta-analysis. *Learning and Instruction*, *17*, 722–738. doi:10.1016/j.learninstruc.2007.09.013.
- Inoue, I. (2005). The realistic reasons behind unrealistic solutions: the role of interpretive activity in word problem solving. *Learning and Instruction*, *15*, 69–83. doi:10.1016/j.learninstruc.2004.12.004.
- Jimenez, L., & Verschaffel, L. (2014). Development of children’s solutions of non-standard arithmetic word problem solving. *Revista de Psicodidáctica*, *19*, 93–123. doi:10.1387/RevPsicodidact.7865.
- Lowe, R., & Boucheix, J.-M. (2010). Attention direction in static and animated diagrams. In A. K. Goel, M. Jamnik, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 250–256). Berlin Heidelberg, Germany: Springer.
- Luchins, A. S. (1942). Mechanization in problem solving: the effect of Einstellung. In Dashiell J. F. (Ed), *Psychological Monographs*, *54* (Whole No. 248).
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge: Cambridge University Press.
- Reusser, K., & Stebler, R. (1997). Every word problem has a solution—the social rationality of mathematical modeling in schools. *Learning and Instruction*, *7*, 309–327. doi:10.1016/S0959-4752(97)00014-5.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*, 141–156. doi:10.1016/S0959-4752(02)00017-8.
- Schnotz, W., & Lowe, R. K. (2008). A unified view of learning from animated and static graphics. In R. K. Lowe & W. Schnotz (Eds.), *Learning with animation: research and design implications*. New York, NY: Cambridge University Press.
- 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*, 577–601.
- Verschaffel, L., De Corte, E., & Borghart, I. (1997). Pre-service teachers’ conceptions and beliefs about the role of real-world knowledge in mathematical modelling of school word problems. *Learning and Instruction*, *7*, 339–359. doi:10.1016/S0959-4752(97)00008-X.
- Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction*, *4*, 273–294. doi:10.1016/0959-4752(94)90002-7.
- Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, The Netherlands: Swets & Zeitlinger.

- Verschaffel, L., Greer, B., Van Dooren, W., & Mukhopadhyay, S. (2009). *Words and worlds: modeling verbal descriptions of situations*. Rotterdam, The Netherlands: Sense Publishers.
- Weidenmann, B. (1989). When good pictures fail: an information-processing approach to the effect of illustrations. *Advances in Psychology*, 58, 157–170. doi:10.1016/S0166-4115(08)62153-6.
- Yoshida, H., Verschaffel, L., & De Corte, E. (1997). Realistic considerations in solving problematic word problems: Do Japanese and Belgian children have the same difficulties? *Learning and Instruction*, 7, 329–338. doi:10.1016/S0959-4752(97)00007-8.

Tinne Dewolf. Finalized her PhD in Educational Sciences at the KU Leuven, Belgium, in 2014, in the domain of realistic mathematical modelling and applied problem solving.

Current themes of research

Her major research interest is psychology of mathematics education, with special attention to realistic word problem solving and text and picture comprehension.

Most relevant publications in the field of Psychology of Education:

Dewolf, T., Van Dooren, W., & Verschaffel, L. (2011). Upper elementary school children's understanding and solution of a quantitative word problem inside and outside the mathematics class. *Learning and Instruction*, 21, 770–780.

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.

Dewolf, T., Van Dooren, W., & Verschaffel, L. (2015). Mathematics word problems illustrated: An analysis of Flemish mathematics textbooks. *Mediterranean Journal for Research in Mathematics Education*, 14, 17–42.

Wim Van Dooren. He is an associate professor in educational sciences at the K.U. Leuven.

Current themes of research

His major re-search interest is psychology of mathematics education, with special attention to number sense and estimation, arithmetic word problem solving, proportional reasoning and rational number knowledge.

Most relevant publications in the field of Psychology of Education:

Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2013). The heuristic interpretation of box plots. *Learning and Instruction*, 26, 22–35.

Obersteiner, A., Van Dooren, W., Van Hoof, J., & Verschaffel, L. (2013). The natural number bias and magnitude representation in fraction comparison by expert mathematicians. *Learning and Instruction*, 28, 64–72.

Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2013). Educated adults are still affected by intuitive beliefs about the effect of arithmetical operations: Evidence from a reaction-time study. *Educational Studies in Mathematics*, 82, 323–330.

Van Hoof, J., Vandewalle, J., Verschaffel, L., & Van Dooren, W. (2015). In search for the natural number bias in secondary school students' interpretation of the effect of arithmetical operations. *Learning and Instruction*, 37, 30–38.

Van Hoof, J., Verschaffel, L., & Van Dooren, W. (2015). Inappropriately applying natural number properties in rational number tasks: characterizing the development of the natural number bias through primary and secondary education. *Educational Studies in Mathematics*, 92(1), 2–8.

Lieven Verschaffel. He is a full professor in educational sciences at the K.U. Leuven.

Current themes of research

His major research interest is psychology of mathematics education, with special attention to number sense and estimation, mental and written arithmetic, arithmetic word problem solving, and rational number knowledge

Most relevant publications in the field of Psychology of Education:

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, 24, 335–359.

Dewolf, T., Van Dooren, W., & Verschaffel, L. (2011). Upper elementary school children's understanding and solution of a quantitative word problem inside and outside the mathematics class. *Learning and Instruction*, 21, 770–780.

Fernández, C., Llinares, S., Van Dooren, W., De Bock, D., Verschaffel, L. (2012). The development of students' use of additive and proportional methods along primary and secondary school. *European Journal of Psychology of Education, 27*, 421–438.

Obersteiner, A., Van Dooren, W., Van Hoof, J., & Verschaffel, L. (2013). The natural number bias and magnitude representation in fraction comparison by expert mathematicians. *Learning and Instruction, 28*, 64–72.

Linsen, S., Verschaffel, L., Reynvoet, B., & De Smedt, B. (2015). The association between numerical magnitude processing and mental versus algorithmic multi-digit subtraction in children. *Learning and Instruction, 35*, 42–50.