Do students attend to representational illustrations of non-standard mathematical word problems, and, if so, how helpful are they?

Tinne Dewolf · Wim Van Dooren · Frouke Hermens · Lieven Verschaffel

Abstract During the last two decades various researchers confronted upper elementary and lower secondary school pupils with word problems that were problematic from a realistic modelling point of view (so-called P-items), and found that pupils in general did not use their everyday knowledge to solve such P-items. Several attempts were undertaken to encourage learners to use their everyday knowledge more when solving such problems, e.g., by presenting the P-items together with representational illustrations that represent the problematic situation described in the problem. These illustrations were expected to help learners to mentally imagine the situation and consequently solve the items more realistically. However, no effect of the illustrations was found. In this article we build further on the use of representational illustrations. We report two related experiments with higher education students that investigated whether and how illustrations that represent the problematic situation described in a P-item help to imagine the problem situation and thereby solve the problem more realistically. In Experiment 1 we measured students' eye movements when solving P-items that were accompanied by representational illustrations, to analyse whether the illustrations are processed at all. In Experiment 2 we manipulated the presentation of the illustrations so students could not but look at them, before the word problem appeared. We found that students scarcely looked at the representational illustrations (Experiment 1) and when they did, there was no effect of the illustrations on the

T. Dewolf (✉) · W. Van Dooren · L. Verschaffel
Centre for Instructional Psychology and Technology, KU Leuven, Dekenstraat 2, Box 3773, 3000 Louvain, Belgium
e-mail: tinne.dewolf@ppw.kuleuven.be

W. Van Dooren
e-mail: wim.vandooren@ppw.kuleuven.be

L. Verschaffel
e-mail: lieven.verschaffel@ppw.kuleuven.be

F. Hermens
Centre for Experimental Psychology, KU Leuven, Tiensestraat 102, Box 3711, 3000 Louvain, Belgium
e-mail: hermensf@gmail.com

realistic nature of their solutions (Experiment 2). Possible explanations for these findings are discussed.

Introduction

Word problems as mathematical modelling tasks

In mathematics education pupils are often confronted with word problems. Word problems play an important role in the mathematics lessons: Besides evaluating pupils' mathematical skills, training them to think creatively, motivating them, and helping them to develop new mathematical concepts and skills, word problems are most importantly used to offer practice for situations of everyday life in which mathematics learners will need to use what they have learned in school (=application function). The idea behind this goal is to bring reality into the mathematics classroom, to create occasions for learning and practicing the different aspects of applied problem solving, without the practical inconveniencies of direct contact with the corresponding real-world situation. By means of such alternatives for the real-world situations outside the classroom, students become prepared for the mathematical requirements they will face in their (future) everyday lives (Verschaffel et al. 2000).

The application of mathematics to solve such real world problems, also called "mathematical modelling", can be seen as a complex process involving a number of phases. There are many descriptions of this process (e.g., Blum and Niss 1991; Burkhardt 1994; Verschaffel et al. 2000) but, in essence, they all involve the following components: understanding and defining the problem situation leading to a situation model; constructing a mathematical model of the relevant elements, relations and conditions embedded in the situation; working through the mathematical model using disciplinary methods to derive some mathematical results; interpreting the outcome of the computational work in relation to the original problem situation; evaluating the model by checking if the interpreted mathematical outcome is appropriate and reasonable for its purpose; and communicating the obtained solution of the original real-world problem. This modelling process cannot be described as strictly linear; rather it has to be considered cyclic (Blum and Niss 1991; Burkhardt 1994; Verschaffel et al. 2000).

During the past decades, several scholars have argued and shown that word problems fail to fulfil their application function. In their view, pupils' year-long participation in the practice and culture of word problem solving at school results in the development of a routine, non-realistic approach to these problems. Instead of considering a word problem as a genuine modelling challenge, they treat it as an artificial, puzzle-like task that has to be solved by identifying and executing the mathematical operation that is "hidden" in the problem (Gerofsky 1999; Lave 1992; Reusser and Stebler 1997; Schoenfeld 1991; Verschaffel et al. 2000, 2009).

Realistic word problem solving in the mathematics class

To investigate pupils' tendency to exclude real world considerations when solving word problems at school, Greer (1993) and Verschaffel et al. (1994) gave students (respectively

13–14-years and 10–11 years old) pencil-and-paper tests consisting of matched pairs of word problems. In each pair one problem, called a standard item or S-item, was such that the straightforward application of an arithmetic operation on the given numbers was reasonable (in the authors' judgment). The other problem, called a problematic item or P-item, required serious consideration of more subtle aspects of the situation described. For example, the S-item "A man cuts a rope of 12 m into pieces of 1.5 m each. How many pieces does he get?" can be solved correctly by dividing 12 m by 1.5 m without any further consideration, whereas the corresponding 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 simply by dividing 12 m by 1.5 m, because one has to take into account that the pieces of rope need to be knotted together and that the rope needs to fit around the poles, so therefore more than eight pieces are needed.

The 10–11-year olds in the study by Verschaffel et al. (1994) received ten such pairs of an S-item and a matched P-item randomly ordered in the test. The pupils written responses on the P-items were coded as realistic or non-realistic, depending on their answer and possible additional comments. Non-realistic reactions (NRs) were replies where the numerical answer was the result of a straightforward execution of the operation with the numbers given in the problem, without any comment about the problematic nature of the problem from a realistic modelling perspective. Realistic reactions (RRs) were replies that did take into account realistic considerations, by giving a realistic answer, a non-realistic answer followed by a realistic comment, or a statement that it was not possible to give a (precise) numerical answer to the problem due to the realistic modelling complexity. For the rope item, for example, the straightforward answer "12 ÷ 1.5 = 8 ropes" without any reference to the "knotting issue" was considered as a NR, whereas the answer that more than eight pieces of rope are needed (because the ropes need to be knotted together to stretch between the poles) or the answer that it is impossible to give a precise numerical answer (because one does not know how much rope is needed for all the knots), were considered as RRs. Verschaffel et al. found that pupils' barely included real-world knowledge. Only 17.0 % of the reactions to the ten P-items were realistic. Similar percentages of RRs were obtained in numerous replications in different countries involving (translations of) essentially the same items and pupils from a similar age range (for an overview see Verschaffel et al. 2009).

P-items have also been given to higher education students. For example, Verschaffel et al. (1997) presented pre-service elementary school teachers (18–21 years) a subset of P-items from Verschaffel et al.'s (1994) study, and found an overall percentage of 48.0 % RRs. Inoue (2005) presented college students another set of word problems with realistic modelling complications and found that only 30.0 % of the responses reflected reality. So, both studies demonstrate that, like upper elementary school children, higher education students tend to exclude realistic considerations when solving P-items, even though at some lower rates.

In a first line of follow-up studies, researchers have set up intervention studies wherein they developed, implemented and evaluated experimental instructional programs more or less directly aimed at the enhancement of students' conceptions about and skills in mathematical modelling and applied problem solving. For instance, starting from the findings of the studies reported above, Verschaffel and De Corte (1997) set up a teaching experiment in which they subjected pupils to a series of five focused lessons, in order to change their conceptions of the role of real-world knowledge in mathematical modelling and to develop in them a more realistic approach towards mathematical modelling. The

results of that intervention study, as a well of a replication study by Renkl (1999) were moderately positive (see Verschaffel et al. 2000, for a discussion of these studies).

Another line of follow-up studies investigated whether it was possible to increase the number of RRs to P-items by manipulating the way in which these P-items are presented or administered to the pupils (without subjecting these pupils to new forms of instruction). Examples of such manipulations aimed at encouraging and/or helping pupils to activate and include real-world knowledge and considerations are, for instance, explicitly warning pupils about the non-standard nature of the P-items (Yoshida et al. 1997) or making the problem formulations more authentic (Palm 2008). Another manipulation, which is the focus of the present article, involves adding illustrations to the word problems aimed at making the real-world context more salient and available to the solver, and, in doing so, helping them to create a rich mental representation of the problem situation (Dewolf et al. 2014).

In the remainder of this introduction, we will first discuss the literature about the use of illustrations in learning and problem solving in general and in realistic word problem solving in particular, and then formulate the general research questions.

Research on the use of illustrations in general

Learners are frequently confronted with different kinds of illustrations in their textbooks. Some of these illustrations just decorate the pages, while others are linked to the content of the textbook. Illustrations can for example be included with the aim to attract attention, enhance enjoyment, or facilitate learning the text content (Levie and Lentz 1982). Regarding learning text content, it is generally assumed that people learn better from words accompanied by relevant pictures than from words alone. This assumed positive effect of illustrations added to text—the so-called multimedia effect (Mayer 1997)—has elicited a lot of theoretical and empirical research (e.g., Carney and Levin 2002; Hannus and Hyönä 1999; Lenzner et al. 2013; Mayer 2005; Mayer and Moreno 2003; Schnotz and Bannert 2003).

Among the most influential theories of the impact of illustrations are the dual coding theory (Paivio 1986), the cognitive theory of multimedia learning (Mayer 2005), and the integrated model of text and picture comprehension (Schnotz and Bannert 2003; see also Schnotz 2005). In the current article, we will rely on the latter model, which assumes that an integrated mental model is built while reading text accompanied with pictures. In this process, reading a text leads to a representation of the text surface, after which a propositional representation is built and finally a mental model is constructed. In contrast, when looking at an illustration, a visual image is created first and afterwards a mental model is built (or the visual image is added to the already available mental model). Thus information from text and pictures that are presented together is combined in an integrated mental model. This mental model can be defined as "a mental representation of a subject matter by an internal structure that is analogous to the subject matter" (Schnotz 2005, p. 67). Applying this model of text and picture comprehension to mathematical problem solving, it can be 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), which is the starting point of the construction of a mathematical model, from which a mathematical outcome and the ultimate answer to the problem are derived. A positive effect of text combined with pictures, however, is not predicted under all conditions. To maximize the chances of a positive effect, text and pictures should be semantically related (coherence principle), they should be presented closely to each other (split-

attention principle), they should not contain redundant information (redundancy principle), and spoken text should be used instead of written text when animation is used (modality principle) (Schnotz 2005).

As mentioned above, the facilitative effect of illustrations depends on the nature of the illustrations in relation to the text. Elia and Philippou (2004) described four different functions of illustrations that apply to mathematical word problem solving based on the general categorisation of functions that pictures may serve in text processing by Carney and Levin (2002). The first three of these functions correspond more or less to those of Carney and Levin. Decorative illustrations are defined as illustrations that have no link to the word problem. Representational illustrations represent either the whole content or a part of the content of the problem. Organizational illustrations provide directions that support the solution procedure. Lastly, informational illustrations are illustrations that contain information that is not represented in the text but that is essential for the solution of the problem. Elia and Philippou investigated the effect of these four types of illustrations on students' mathematical word problem solving. They found that decorative illustrations had no effect on children's problem solving success, whereas representational, organizational and informational illustrations were conducive to word problem solving.

Up to now, little or no research has been done about the influence of illustrations on the solution of non-standard word problems such as P-items, except for a study by Dewolf et al. (2014), which acted as the point of departure for the present study. Hereafter we briefly report that study of Dewolf et al. (2014).

The use of illustrations in realistic word problem solving

Dewolf et al. (2014) investigated whether (a) representational illustrations and (b) a warning that some problems are not as easy as they seem, would increase the number of RRs on the P-items that were used previously in the research of Verschaffel et al. (1994). The researchers conducted two similar data collections, one in Turkey with 402 elementary school pupils (10–11 years) and one in Belgium with 233 elementary school pupils (10–11 years). The pupils were confronted with the word problems from the study of Verschaffel et al., together with or without a representational illustration, and with or without a warning (depending on the condition to which they belonged). These representational illustrations were expected to help pupils to create an elaborated mental representation of the situation, involving the realistic modelling complexity, which would ultimately help them to construct a more appropriate mathematical model and ultimately respond to the P-items in a more realistic way (for an example of a representational illustration see Table 1 infra). Firstly, the authors found that, as in Verschaffel et al.'s original study, both Turkish and Belgian pupils tended to largely exclude realistic considerations when solving mathematical word problems; the overall mean percentages of RRs were, respectively, 12.6 and 11.9 %. Secondly, for both Turkish and Belgian pupils there was no positive effect of either the representational illustrations or of the warning. There even was no effect of the combination of the representational illustrations and the warning.

The present study

In this article we build further on the study of Dewolf et al. (2014) by reporting two experiments in which we investigated why these representational illustrations that make the real-world context more salient and available to the students, did not have any positive

effect on the learners' realistic solutions of P-items. We generally hypothesised that learners did not profit from these representational illustrations because they neglected or discounted them in their problem-solving process. To test this hypothetical explanation, students [1] were confronted with a set of P-items with either a representational, a decorative or no illustration, while their eye movements were recorded. The results of this first experiment essentially showed that learners paid little or no attention to these illustrations so we set up a second experiment in which students received a set of P-items together with a representational or a decorative illustration to which they had to look for 5 s before it was accompanied by a P-item. As such, both experiments fit into the above-mentioned line of research aimed at better understanding why learners fail to solve P-items realistically and if these failures can be resolved by specific changes in the way these problems are presented, without deviating too much from the classroom practice, rather than in the research line wherein the effects of alternative instructional environments are investigated.

In both experiments, we also tested another possible explanation for the absence of any effect of the representational illustrations in the study of Dewolf et al. (2014). According to this alternative explanation, the number of RRs to P-items was low—and remained low even with the representational illustrations and/or warnings—because of learners their deeply entrenched beliefs about (solving) mathematical word problems. Learners may have discovered the problematic nature of a P-item and considered to include the realistic modelling complexity in their mathematical model and their final answer, but ultimately may have given a NR because of these beliefs. Indeed, research has shown that many learners believe that all word problems have a single numerical solution, that all numbers given in the problem should be used, and that all necessary (numerical) information is given (Caldwell 1995; Reusser and Stebler 1997; Schoenfeld 1991; Verschaffel et al. 2000). When a possible answer conflicts with those beliefs, it may prevent learners to give a realistic reaction. Therefore, in both studies, students received a questionnaire at the end of the test session that asked, for each item, whether they had hesitated about their answer and if so, why.

Experiment 1

Research question, hypotheses and overall design

The central question of Experiment 1 was why providing P-items with representational illustrations that aim at representing the problem situation and at evoking the solver's real-world knowledge about that situation, does not result in more RRs to these P-items (as found by Dewolf et al. 2014). As stated above, we hypothesised that these representational illustrations will not help because learners will only pay scarce attention to illustrations next to P-items (*Hypothesis 1a*). Additionally, we hypothesised that in the rare cases that learners will look at an illustration accompanying a P-item, they will look more at a representational than at a decorative illustration, because of the greater informative value of representational illustrations (*Hypothesis 1b*). To shed light on these two closely related hypotheses, we collected and analysed students' responses and eye movements on a subset of word problems from the study of Verschaffel et al. (1994). Students were confronted

---

[1] Previous research has shown that also higher education students tend to exclude realistic considerations when solving P-items, even though at some smaller rates (Inoue 2005; Verschaffel et al. 1997). So we expected to find similar findings with higher education students as with upper elementary school pupils in our previous studies.

with a set of P-items as well as a set of parallel S-items under one experimental and two control conditions. In the experimental condition, they received the word problems together with the same representational illustrations as in Dewolf et al.'s (2014) study (RI-condition). In the first control condition the problems were accompanied with decorative illustrations (DI-condition), which have no link at all with the problem situation (Elia and Philippou 2004), while in the second control condition the problems were presented without illustrations (NI-condition). The DI-condition was included to allow comparison of eye movements and problem solving performance when given an illustration that had nothing to do with the problem situation as compared to a representational illustration that depicted the situation described in the problem. The NI-condition made it possible to determine whether participants looked at the "illustration area" (i.e., the area that was reserved for the illustration in the other two conditions but which was empty in the NI-condition), due to a natural shift of the eyes. Moreover, it allowed for establishing the baseline performance.

We also tested a second possible explanation for the absence of a positive effect of representational illustrations on the realism of students' reactions to P-items (as found by Dewolf et al. 2014), which we considered as complementary to the first: namely that students who did attend to and profit from a representational illustration may nevertheless ultimately give a NR because of their beliefs about word problems. Reacting realistically to a P-item requires violating these beliefs about word problems, because a RR to a P-item is typically a reaction that does *not* consist of a single precise numerical solution based on one or more arithmetic operations on the numbers given in the problem. So, according to this alternative explanation, it is possible that Dewolf et al. (2014) did not find the expected positive effect of the representational illustrations on students' solutions of P-items, *not* because the students neglected these illustrations or because these illustrations did not evoke realistic considerations during the solution process, but because their beliefs about word problems ultimately drove them towards the routine NR. To explore this second hypothetical explanation, we asked students at the end of the test session, first, to indicate how confident they were about the correctness of their answers to each previously solved P-item and, second, to explain these confidence scores. So, if this second explanation holds true, students should show weak confidence in the correctness of their NRs to P-items (*Hypothesis 2a*). Moreover, these doubts about the correctness of their NRs to the P-items should be stronger in the RI-condition than in the other two conditions because of the availability of the representational illustrations (*Hypothesis 2b*). Furthermore, a substantial amount of students' explanations for their hesitations about the correctness of their NRs to the P-items should refer to realistic considerations these students had generated during their solution process without including them in their final answer (*Hypothesis 2c*), and this amount should again be larger in the RI-condition than in the other two conditions (*Hypothesis 2d*).

Method

Participants

Thirty students (27 females and 3 males) with an age between 18 and 24 years (mean age 20.83) took part in the experiment. Every student had a normal or corrected to normal vision (5 students by contact lenses and 2 students by glasses). None of the participants reported any colour vision deficiencies or any other eye dysfunction. Because students

from different study fields could voluntarily subscribe to participate in the study, there was some variety in participants' educational background. One-third were studying to become primary school teachers, one-third were studying psychology and one-third came from other study programs such as social sciences, economics, or arts. The students were randomly assigned in the three conditions, meaning that each condition contained ten students. At the beginning of the experiment students signed a form for their consent to voluntarily participate in the study and with the information that they could stop the experiment anytime or could refuse to participate. They received six Euros for their participation in the experiment.

Material and procedure

We presented eight P-items from the study of Verschaffel et al. (1994) to each student individually. Eight additional S-items (i.e. word problems for which the application of an arithmetical operation gave a straightforward answer to the word problem) were included in the test as buffer items. Depending on the condition, the word problems were presented together with a representational illustration (RI-condition), without an illustration (NI-condition), or with a decorative illustration (DI-condition). To be able to track participants' eye movements, the word problems were presented on a computer screen and they had to say their answer out loud. After solving the problems on the computer, participants received a paper questionnaire with the same 16 problems, that asked to indicate for each problem whether they had hesitated about their solution, and if so, why they had doubted the correctness of their answer.

The illustrations that accompanied the word problems were either representational (RI-condition) or decorative (DI-condition). The representational illustrations were the same as the ones that were used in the research of Dewolf et al. (2014). These illustrations were directly connected to the word problem; they depicted the (problematic) situation without containing essential information to solve the problem. The decorative illustrations, which were adapted from Martin (2011), had no connection whatsoever with the real-world situation that is described in the word problem; they depicted regular (mathematics) class situations such as a student holding a book, a teacher in front of a blackboard, etc. In the NI-condition were no illustrations. The illustrations in the DI-condition and RI-condition were in colour. Table 1 gives an overview of all eight P-items and their decorative and representational illustrations.

In the first part of the experiment, students were seated in front of the computer screen. After communicating the instructions and signing the informed consent, the eye movement device, the Eyelink II, was installed and calibrated using a nine-point calibration procedure. A chin rest was used to reduce head movements and to maintain the distance between the eyes and the screen (approximately 60 cm). The word problems appeared one by one on the screen. The text was presented at the left side (text area) and the illustration at the right side (illustration area). In the NI-condition, the illustration area was blank. Before the presentation of each word problem there was a fixation point in the middle of the screen. Students had to press a button when they knew their answer to the problem and had to say the answer out loud (since using the eye tracking device the students were not able to write their answers down). By pressing the button again they could move to the following problem. Students were asked to press a button before saying their answer out loud, for two reasons. First, it allowed us to make a clear distinction between the eye movements when silently reading the question and thinking about the answer, on the one hand, and those while verbally stating the answer, on the other hand. Second, it avoided motion artefacts in

Table 1 The eight P-items with the representational and decorative illustrations

Item		Representational illustrations	Decorative illustrations
P1	Carl has 5 friends and George has 6 friends. Carl and George decide to give a party together. They invite all their friends. All friends are present. How many friends are at the party?		
P2	Steve would like to make a bookshelf. He has bought 4 planks of 2.5 m each. How many planks of 1 meter can he get out of these planks?		
P3	Martin pours a full bucket of water at 40° and a full bucket of water at 80° in a container. How many degrees is the water in this container?		
P4	250 soldiers must be bused to their training site. How many busses are needed if you know that each bus can hold 20 soldiers?		

Table 1 continued

	Item	Representational illustrations	Decorative illustrations
P5	John's best time to run 100 m is 17 s. How long will it take to run 1 km?		
P6	Bruce and Alice go to the same school. Bruce lives at a distance of 17 km from the school and Alice at 8 km. How far do Bruce and Alice live from each other?		
P9	The distance between two poles is 12 meter. How many pieces of rope of 1.5 meter long should one tie together to connect both poles?		
P10	A conic vase is being filled with water. After 10 s the water is 4 cm high. How high will the water be after 30 s?		

* In the present study we maintained the item numbers from the research of Verschaffel et al. (1994)

the recorded eye movement data due to saying out loud the answer in the relevant recordings. To practice pressing the button and saying the answer out loud, students received six practice items before the actual experiment. After this short practice session the instructions were repeated and the actual experiment started. In these instructions the students were told that there was no time limit and that they would receive no feedback during the experiment.

In the second part of the experiment, which started after the last word problem was solved and the eye movement device was turned off, students received a paper-and-pencil questionnaire with the same 16 word problems. In this questionnaire students were asked to indicate, for each item, to what extent they had hesitated about their answer [by responding: (a) not at all, (b) a little bit, (c) quite a bit, and (d) a lot], and if so why. Depending on the condition, the items were again presented with their representational or decorative illustration or without an illustration.

Apparatus and software

The Eyelink II system (SR Research Osgood, ON, Canada) uses the combined corneal reflection and pupil mode if possible (resulting in an approximate 0.5° of visual angle accuracy of the measurement of the gaze position, well below the size of the text and image area of the stimuli). For participants where the corneal reflection was not visible for all positions on the screen, the 'pupil only mode' was used instead. Gaze position was recorded at 250 Hz.

Data-analysis

A proper test of our hypotheses required that we first had to replicate the findings of Dewolf et al. (2014). So we first of all analysed students' responses on the P-items to see if they also solved them unrealistically, and if there was again no effect of the illustrations on the number of RRs on the P-items. The recorded responses were coded by one researcher in the same way as in the study of Verschaffel et al. (1994) and of Dewolf et al. (2014). Realistic reactions (RRs) and non-realistic reactions (NRs) were distinguished. The intra-rater reliability was calculated for 20 % of the participants. There was an almost perfect agreement (K = 0.906).

Eye movements were used to test *Hypothesis 1a* and *Hypothesis 1b*. The eye movements were measured from the onset of the presentation of the word problem until the moment the participant pressed the button to give his response. Only the movements of the left eye were analysed with a view to determine their eye fixations. Fixations are periods in which the eyes are relatively still and visual input enters the eyes, allowing detailed processing of that visual input (Duchowski 2007; Rayner 1998). Because of the sampling rate of 250 Hz, fixations were measured with a temporal accuracy of 4 ms. All fixations were included in the analysis.

To test *Hypothesis 2a* and *2b*, we analysed students responses to the question how much they had hesitated about their answer (not at all, a little bit, quite a bit, or a lot), for all P-items on which they had given a NR.

Furthermore, students' explanations about why they had hesitated about the correctness of their NR to a P-item were used to test *Hypothesis 2c* and *2d*. These explanations were coded in three different categories: 'No Explanation', when students did not give any reason for their hesitation, 'Realistic Explanation', when they referred to the realistically problematic nature of the problem as the reason for their hesitation, or 'Other Explanation',

when they gave another kind of explanation, for example a computational difficulty (due to the size or nature of the numbers) or uncertainty about what arithmetic operation(s) to perform on the given numbers. The inter-rater reliability was calculated for 20 % of the participants. There was an almost perfect agreement between the two coders (K = 0.915).

Students' responses on the P-items, their eye movements and their answers on the questionnaire were analysed with a logistic regression analysis in SPSS, whereas their hesitation scores were analysed with an ANOVA.

Results

Students' responses

The percentage of students' RRs on the P-items was only 27.9 %. This percentage is somewhat higher than in Verschaffel et al.'s (1994) study with upper elementary school children (17.0 %) but lower than in Verschaffel et al.'s study (1997) with pre-service teachers (48.0 %). So, like the participants in all previous studies with P-items, higher education students involved in this experiment had a strong tendency to neglect their real-world knowledge when solving P-items.

Second, differences in percentage RRs on P-items between the three conditions were analysed with a logistic regression analysis, and, more specifically, by means of a Generalized Estimating Equations analysis. The analysis revealed no significant differences between the three conditions for the P-items, $Wald\ X^2(2,240) = 4.38$, $p = 0.112$ ($B_1 = -0.521$ and $B_2 = -0.521$), even though the percentages of RRs in the RI-condition (31.3 %) and the DI-condition (31.3 %) tended to be higher than in the NI-condition (21.3 %). This indicated that, as for elementary school pupils (Dewolf et al. 2014), providing representational illustrations also had no beneficial effect in higher education students on the number of their RRs on P-items.

Students' eye movements

To test *Hypothesis 1a* and *1b*, we analysed students' eye fixations on the illustrations accompanying the P-items. More specifically, we looked at the percentage and raw number of fixations on the illustration area for the P-items in the three different conditions. The data show that students barely looked at these illustrations next to P-items: only 0.9 % or 91 fixations out of a total of 10,684 fixations were in that area, with no fixations on that area in the NI-condition. The extremely small percentages of fixations on the illustrations accompanying the P-items in the RI-condition and the DI-condition support *Hypothesis 1a* stating that learners scarcely attend to illustrations accompanying P-items. Still, as predicted in *Hypotheses 1b*, the percentage of fixations on the illustrations of the P-items was significantly larger for the RI-condition than for the DI-condition, $Wald\ X^2(1,7158) = 29.80$, $p < 0.001$ ($B_1 = 1.460$), with respectively 2.0 % or 74 from the 3,626 fixations and 0.5 % or 17 from the 3,532 fixations in the illustration area.

Although the number and percentage of fixations on the illustration area were extremely small, one could argue that one single fixation may be already sufficient to process an illustration. Therefore, we decided to look not only at the distribution of the fixations in the illustration area of the P-items among the different conditions, but also at the percentages and raw number of cases wherein an illustration was fixated at least once for the minimum amount of time needed for the processing of (pictorially presented) scenes. Indeed,

previous research has shown that viewers can get the gist of a scene very early in the looking process, even from a single glance (Biederman et al. 1982; Fei-Fei et al. 2007; Potter 1975, 1976). More specifically, Rayner et al. (2009) showed that viewers needed to glance a scene for at least 150 ms to process it somehow. Therefore, we decided to work with this very strict minimum of 150 ms. To calculate how many illustrations were at least minimally processed, the longest fixation time on each illustration for each P-item and each student (for the RI-condition and the DI-condition) was identified. Overall, only 23.1 % of these illustrations of P-items elicited at least one fixation longer than 150 ms (or 37 fixations of the 160). The finding that in only about 1/4 of the cases, the illustration was fixated for at least 150 ms provided additional support for *Hypothesis 1a*, as it implies that in about 3/4 of the cases the illustration was not processed at all. Confirming *Hypothesis 1b*, a logistic regression analysis on the data of the P-items showed that there were significantly more illustrations that were processed for a minimum of 150 ms in the RI-condition than in the DI-condition, *Wald* $X^2(1,160) = 5.37$, $p = 0.020$ ($B_1 = -1.271$), respectively 27 fixations of the 80 fixations or 33.8 % and 10 fixations of the 80 fixations or 12.5 %, providing additional evidence that representational illustrations of P-items attracted more attention than decorative ones.

Responses on the questionnaire

To test *Hypotheses 2a* and *2b* about students' weak confidence in the correctness of their NRs to P-items, particularly in the RI-condition, students' confidence scores were analyzed. For an overview of the distribution of the four levels of hesitation in general and per condition for all P-items where students had given a NR (i.e., in 72.1 % of all cases) see Table 2.

In line with *Hypothesis 2a*, students expressed quite some doubt about the correctness of their NRs to the P-items, as only in 38.2 % of all cases they expressed no hesitation at all. To put these percentages in perspective, it is interesting to compare them to the percentage of "no hesitation at all" about the correctness of their answer to the eight S-items, which was 69.2 %.

To test *Hypothesis 2b*, we checked if the confidence in the correctness of the NRs on the P-items was smaller in the RI-condition than in the other two conditions. Each category of hesitation was assigned with a numerical value to have a measure of hesitation which allowed us to compare means with a univariate analysis of variance (ANOVA). The values 1–4 were assigned successively to the categories (a) not at all, (b) a little bit, (c) quite a bit and (d) a lot. An ANOVA was performed with hesitation as the dependent variable and condition as independent variable. The analysis revealed that there were no differences between conditions, $F(2,170) = 0.885$, $p = 0.415$, $M_{RI\text{-condition}} = 2.15$ and $SD_{RI\text{-condition}} = 0.15$, $M_{DI\text{-condition}} = 2.02$ and $SD_{DI\text{-condition}} = 0.15$, and $M_{NI\text{-condition}} = 2.29$ and $SD_{NI\text{-condition}} = 0.14$. So, contrary to our *Hypothesis 2b*, students were not less confident about their NRs on P-items in the RI-condition than in the DI-condition or the NI-condition.

We now turn to the test of *Hypotheses 2c and 2d*, which, respectively claim that students' expected low confidence in the correctness of their NRs on the P-items would be mainly caused by the interference of realistic considerations during the solution process (*Hypothesis 2c*), especially when confronted with representational illustrations (*Hypothesis 2d*). To test these two hypotheses, we analyzed the explanations for the hesitations about the correctness of the NRs to the P-items of all students from the three conditions who had responded the question with "a little bit", "quite a bit", or "a lot". As shown in Table 3, when students gave some explanation for their doubts about the correctness of their NR to

Table 2 Percentages of hesitation about the correctness of the NRs for the P-items per condition and in general for Experiment 1

	RI-condition	DI-condition	NI-condition	Total
a. Not at all	34.5	45.5	34.9	38.2
b. A little bit	27.3	20.0	20.6	22.5
c. Quite a bit	27.3	21.8	25.4	24.9
d. A lot	10.9	12.7	19.0	14.5

a P-item, it was, in line with *Hypothesis 2c*, mostly based on realistic considerations (44.9 %).

When focusing on the effect of condition on the nature of the explanations for students' hesitations about their NRs on P-items (*Hypothesis 2d*), students indeed tended to refer more to realistic considerations in the RI-condition (61.1 %) than in the two other conditions (DI-condition 30.0 % and NI-condition 41.5 %), but the logistic regression revealed that this effect was not significant, *Wald* $X^2(2,107) = 5.36$, $p = 0.069$ ($B_1 = -0.797$ and $B_2 = 0.502$).

Conclusion and discussion

The results of Experiment 1 showed that students fixated the illustrations next to P-items very rarely, as only a negligibly small percentage of the fixations were on the illustration area. Or, in terms of number of at least minimally processed illustrations, only about 1/4 of the illustrations accompanying the P-items were processed. So *Hypothesis 1a* was confirmed. But when students did look at the illustration next to a P-item, they looked significantly more to representational than to decorative illustrations, confirming *Hypothesis 1b*. Furthermore, students who had responded to a P-item in a non-realistic way were generally quite hesitant about the correctness of their NR, as claimed in *Hypothesis 2a*. On the other hand, *Hypothesis 2b* had to be rejected because students' doubts about their answer on P-items were not stronger in the RI-condition than in the two other conditions (even though students had looked significantly more at the illustrations in the RI-condition, see Hypothesis 1b). Finally, in support of *Hypothesis 2c*, students who had responded to P-items with a NR gave a lot of explanations for their hesitation about the correctness of their answer that reflected realistic considerations, and the data for the three conditions were in line with *Hypothesis 2d* that the RI-condition would yield the largest number of such explanations, although this latter effect did not reach significance. Overall, these confidence scores for the NRs on the P-items and their accompanying explanations showed that students, particularly those from the RI-condition, did notice frequently the realistic modelling complexity during their solution of a P-item but nevertheless responded with a NR, probably because of their beliefs about mathematical word problems and how to solve them in the mathematics class.

Although these findings shed some light on why representational illustrations do not help learners in solving P-items in a realistic way, one should be careful in interpreting them. One limitation of this experiment was that we could not exclude that the illustrations that were not *fixated* but were nevertheless *processed*. In other words, it is still possible that students who did not fixate a representational illustration, saw it 'in the corners of their eyes' and thus may still have processed it somehow. In this respect, we refer to a study by Thorpe et al. (2001), who examined if people could categorize natural photographs

Table 3 Percentages of students' explanation for their hesitation about the correctness of their NRs for the P-items per condition and in total for Experiment 1

	RI-condition	DI-condition	NI-condition	Total
No explanation	16.7	26.7	17.1	19.6
Other explanation	22.2	43.3	41.5	35.5
Realistic explanation	61.1	30.0	41.5	44.9

presented in the peripheral retina. The authors demonstrated that participants were able to answer to the question 'did the photograph contain an animal or not?' purely with peripheral vision. Therefore, our conclusions about the small number of cases in which students actually looked at the (representational) illustrations of P-items and thus could have profited from processing them might be questioned. Another limitation of the present investigation is that, since the illustrations were looked at so rarely in general, there was perhaps a too small amount of data available for testing the hypotheses concerning the hesitations that were caused by representational illustrations, as compared to purely decorative and no illustrations.

To address both limitations, we set up a second experiment in which we experimentally forced students to attend to the (representational) illustrations accompanying P-items.

Experiment 2

Research question and overall design

Departing from one of the main findings of Experiment 1, namely that students scarcely look at the representational illustrations accompanying P-items, we set up a second experiment in which students in all conditions could not but look at the illustrations that accompanied the word problems. This was done by presenting the illustration already for 5 s before confronting students with the actual word problems. In a first condition (DI-condition) all illustrations were decorative, whereas in a second condition (RI-condition) they were representational. We also added a third condition (RIW-condition), which also involved representational illustrations, but they were accompanied by an extra warning that these illustrations were helpful to imagine the problem situation and to solve the problem.[2] We predicted that, due to the expected impact of the *processed* representational illustrations on students' problem representations, they would respond more realistically to P-items when these items were presented together with representational illustrations (RI-condition) than with decorative illustrations (DI-condition), and that they would produce even more realistic reactions when the representational illustrations were given together with an additional warning (RIW-condition) (*Hypothesis 1*). However, since we still expected a significant number of NRs to the P-items (because of the second explanation based on students' beliefs, for which Experiment 1 also provided empirical evidence), we further hypothesised, as in Experiment 1, that students who still would respond with a NR

---

[2] It must be noted that this warning was of a different nature than the one used in the study of Dewolf et al. (2014) and in other previous studies with P-items (Verschaffel et al. 2000). In those studies learners received a general alert that the test could involve some non-standard word problems that may require a different kind of response, whereas in the present experiment the warning aimed at convincing participants of the usefulness of the accompanying illustrations for solving the word problems.

would show weak confidence in the correctness of their NRs to the P-items (*Hypothesis 2a*), and that these doubts would be the weakest in the DI-condition, stronger in the RI-condition, and the strongest in the RIW-condition (*Hypothesis 2b*). Furthermore, when learners had doubts about the correctness of their NR on a P-item, they should attribute their weak confidence to the realistically problematic nature of the items. Therefore, a substantial number of students' explanations should refer to realistic considerations (*Hypothesis 2c*) and this number should be larger in the RI condition than in the DI-condition, and even larger in the RIW-condition (*Hypothesis 2d*).

Method

Participants

The participants were 142 first year university students from educational sciences (135 females and 7 males), between the ages of 17 and 24 years (mean age 18.6, SD = 0.96). They volunteered in participating in the study and received a course credit in return. The students were randomly assigned to three conditions; 47 students (43 females and 4 males) in the condition with decorative illustrations (DI-condition), 47 students (47 females and 0 males) in the condition with representational illustrations (RI-condition), and 48 students (45 females and 3 males) in the condition with representational illustrations and an additional warning (RIW-condition).

Material and procedure

As in Experiment 1, students were administered eight P-items from the study of Verschaffel et al. (1994), as well as eight S-items that served as buffer items. There were also two additional S-items that served as practice items at the beginning of the experiment. Depending on the condition, students received the word problems together with decorative illustrations (DI-condition), representational illustrations (RI-condition), or representational illustrations and an additional warning (RIW-condition). After solving these items, students received, in the second part, the same questionnaire as in Experiment 1, that asked to indicate for each item, how much they had hesitated about their answer and if so why. At the end of this questionnaire, we also asked, for purely exploratory reasons, whether the illustrations had helped to solve the word problems.

The 16 items and the two practice items were presented individually to the students who were seated in front of a computer. The researcher shortly explained that the experiment consisted of two parts and that the instructions could be found on the instruction sheet in front of each computer. On this sheet it was explained how the computer-based test should be started. After a repetition of the instructions on the screen, a fixation cross appeared for 1 s, and was followed by an illustration (decorative or representational). The illustration was presented in the middle of the screen for 5 s. Students could not skip the illustration, so they had to look at it. After 5 s the corresponding word problem appeared on the left side of the screen, and the illustration (which was initially presented in the middle of the screen) moved to the right side next to the problem. Students were asked to solve each item, and to write down their answer on the sheets in front of them. This allowed them to extensively write down their calculations, and additional explanations on how they handled the problem. After writing down their answer, students could go to the next problem by pressing the "enter" key.

The decorative illustrations (DI-condition) and representational illustrations (RI-condition and RIW-condition) were the same as in Experiment 1 (see supra Table 1).

In the RIW-condition, a warning was added to encourage students to process the representational illustrations in detail and to make active and productive use of these illustrations to imagine the situation and solve the problems. The warning, stating that the illustrations could help to imagine the situation and to solve the word problem, was part of the general instructions, and was repeated on top of the computer screen for each item.

When a student was ready with the first part of the experiment, (s)he gave a sign to the researcher, who collected the answer sheet and gave the paper-and-pencil questionnaire, which was exactly the same as in Experiment 1, except for the final question whether the illustrations had helped them to solve the problems. Students had to answer that question by indicating (a) (almost) never, (b) sometimes, (c) often, (d) (almost) always and if so explain how it at helped.

Data analysis

We coded and analysed students' written responses to the P-items as well as their indications of how much they had hesitated about their answers, and why in the same way as in Experiment 1. The inter-rater reliability was calculated for 20 % of the participants. There was an almost perfect agreement for the codes RR or NR (K = 0.946) and for pupils' explanations for their hesitation (K = 0.962). We furthermore analysed students' answers on the exploratory question whether and how the illustrations had helped them.

Results

Students' responses on the word problems

First, we analysed the responses of the students on the P-items to test our first hypothesis stating that because the illustrations are now inevitably processed, learners would respond more realistically to P-items in the two conditions with representational illustrations than in the condition with decorative illustrations, with most RRs in the RIW-condition. In 7 out of 1,136 cases students had pressed the button to go to the following P-item too soon and consequently did not solve these items. These cases were excluded from the analysis, so we analysed the data for 1,129 solutions of P-items.

The overall percentage of RRs on the P-items, across all three conditions, was 53.0 %. This percentage was higher than in Experiment 1, in which only 27.9 % of the reactions of the students were realistic, but more or less the same as in the study of Verschaffel et al. (1997) conducted with pre-service teachers. When looking at the data for each condition separately, the scores on the P-items were quite similar: 53.1 % RRs in the DI-condition, 52.7 % in the RI-condition and 53.1 % in the RIW-condition. A logistic regression analysis (i.e., a Generalized Estimating Equations) with the response (RR = 1 and NR = 0) as dependent variable and condition as independent variable revealed that these differences between conditions were not significant, $Wald\ X^2(2,1129) = 0.01$, $p = 0.995$ ($B_1 = 0.003$ and $B_2 = 0.018$). So, in contrast to *Hypothesis 1*, even though the experimental procedure had forced students to attend to the illustrations, there was still no positive effect of the representational illustrations on the realistic nature of students' responses to the P-items, even not when accompanied with an additional warning.

Responses on the questionnaire

To test *Hypothesis 2a* and *2b* about students' weak confidence in the correctness of their NRs to P-items, we analysed their confidence scores and their accompanying explanations. As mentioned above, some students had skipped an item due to hitting the button too quickly. So the data on the questionnaire for these items were again not included in the analysis. Table 4 gives an overview of the distribution of the four levels of hesitation per condition and in general, for all P-items being answered with a NR (i.e., 47.0 % of all responses to P-items).

Contrary to *Hypothesis 2a*, students did not hesitate so much about their NRs to the P-items. According to their responses to the questionnaire, students indicated that they hesitated at least a bit in only 34.3 % of the cases wherein a P-item was answered non-realistically. So students were quite confident about the correctness of their NRs on the P-items. This is in contrast with Experiment 1, where the percentage of such hesitations was 61.9 %. So, possibly as a result of the change in experimental arrangement, students from Experiment 2 who noticed a realistic modelling difficulty were more inclined to incorporate it in their answer than students from Experiment 1, leading to comparatively more RRs on the P-items but, in turn, less hesitation about the correctness of their remaining NRs to these items. However, a comparison of this percentage of 34.3 % of hesitation for the P-items in Experiment 2 with the percentage of hesitation for the eight S-items shows that there was still considerably more doubt for the former than for the latter (i.e., only 10.8 % hesitation for S-items).

Despite the fewer cases where students had doubts about their NRs on the P-items, we tested whether the doubt was weakest in the DI-condition, stronger in the RI-condition and the strongest in the RIW-condition *Hypothesis 2b*. As in Experiment 1, the differences in hesitation per condition were analysed by means of an ANOVA. To each hesitation category a numerical value from 1 to 4 was assigned and an ANOVA was conducted with hesitation as dependent variable and condition as independent variables. The findings show that there again (as in Experiment 1) were no differences between conditions, $F(2,528) = 1.360$, $p = 0.258$, $M_{DI\text{-}condition} = 1.71$ and $SD_{DI\text{-}condition} = 0.07$, $M_{RI\text{-}condition} = 1.59$ and $SD_{RI\text{-}condition} = 0.07$, and $M_{RIW\text{-}condition} = 1.55$ and $SD_{RIW\text{-}condition} = 0.07$. So the students were in all three conditions equally hesitant about their NRs on P-items.

To test *Hypothesis 2c* and *2d*, namely that a significant number of students' explanations about their hesitation for their NRs on P-items should reflect realistic considerations, and that this number would be the least in the DI-condition, larger in the RI-condition and the largest in the RIW-condition, we performed a qualitative analysis of students' responses to the question why they had hesitated. All the P-items for which students had given a NR and had indicated at least some hesitation were analysed. As shown in Table 5, and in line with *Hypothesis 2c* and the results of Experiment 1, most explanations (59.9 %) belonged again to the category of realistic explanations.

However, there was no significant effect of condition, *Wald* $X^2(2,182) = 2.23$, $p = 0.328$ ($B_1 = 0.174$ and $B_2 = 0.693$). The number of realistic explanations was the same in all three conditions, so *Hypothesis 2d* was rejected.

Additional explanatory question about the illustrations

Finally, we asked the students at the end of the questionnaire to what extent the illustrations had helped them to solve the word problems. Two students from the DI-condition did not fill in this final question, resulting in a data set of 140 students. In general, 84.3 % of

Table 4 Percentages of hesitation about the correctness of the NRs for the P-items per condition and in general for Experiment 2

	DI-condition	RI-condition	RIW-condition	Total
a. Not at all	61.9	67.0	68.2	65.7
b. A little bit	11.4	11.9	13.4	12.2
c. Quite a bit	21.0	16.5	14.0	17.1
d. A lot	5.7	4.5	4.5	4.9

the students indicated that the illustrations (almost) never helped, 13.6 % indicated that the illustrations helped sometimes, 1.4 % that the illustrations often helped and one student (0.7 %) that the illustrations helped (almost) always. The comparison between the three conditions revealed that there were no significant differences for the response alternatives "(almost) never", $Wald\ X^2(2,140) = 4.46$, $p = 0.108$ ($B_1 = -1.030$ and $B_2 = 0.424$) and "sometimes", $Wald\ X^2(2,140) = 2.52$, $p = 0.284$ ($B_1 = 1.030$ and $B_2 = -0.025$), whereas for "often" and "(almost) always" the scarcity of the responses did not allow to perform a statistical analysis. So, neither the nature of the illustrations nor the presence of an additional warning about the usefulness of the illustrations (in the RIW-condition) had an effect on students' feeling of usefulness of the illustrations to solve the word problems.

Conclusion and discussion

Hypothesis 1 stated that, when the illustrations are actually processed, learners would respond more realistically to P-items when these items were presented together with representational illustrations than with decorative illustrations, and even more realistically when the representational illustrations were complemented with an additional warning. Our findings showed that the number of RRs on the P-items did not differ between the three conditions. So even when students could not but attend to the illustrations and even when they were informed about of the usefulness of the illustrations for solving the items, no positive effect on the realism of their reactions was found. Thus *Hypothesis 1* was rejected. *Hypothesis 2a* stated that students who would respond with a NR to the P-items would show weak confidence in the correctness of their answers. We did not find strong support for this hypothesis either, as in 1/3 of all cases, students reportedly hesitated about their NR to a P-item, which was considerably less than in Experiment 1. *Hypothesis 2b*, which stated that the doubts about the correctness of a NR answer to a P-item would be weakest in the DI-condition and strongest in the RIW-condition, was rejected too. *Hypothesis 2c* about students' explanations was confirmed. In those cases wherein students hesitated about their NR to P-items, they did so most frequently because of realistic considerations. However, there were no significant differences between conditions with respect to the number of such realistic considerations, so *Hypothesis 2d* was also rejected.

General conclusion and discussion

Previous research has shown that learners—both elementary school pupils and students in higher education—demonstrate a strong tendency to respond to realistically problematic word problems (P-items) without seriously taking into account realistic considerations

Table 5 Percentages of students' explanation for their hesitation about the correctness of their NRs for the P-items per condition and in total for Experiment 2

	DI-condition	RI-condition	RIW-condition	Total
No explanation	13.4	8.6	10.5	11.0
Other explanation	23.9	41.4	22.8	29.1
Realistic explanation	62.7	50.0	66.7	59.9

(Verschaffel et al. 2000), and that attempts to increase the number of realistic reactions (RRs) to those problems by providing representational illustrations of the problem situation have little or no effect (Dewolf et al. 2014). The present study examined why these representational illustrations did not help learners to answer P-items more realistically. To this end we conducted two subsequent experiments. In Experiment 1, higher education students received eight P-items and eight S-items together with either representational illustrations (RI-condition), purely decorative illustrations (DI-condition), or without illustrations (NI-condition). Their eye movements were recorded during the test. Afterwards, they received a questionnaire in which they were asked how much they had hesitated about their answer for each word problem, and if so why. In Experiment 2 students received the same P-items and S-items, but this time they were confronted with the illustration not only during the solution process but already before they saw the word problem, to make sure that they looked at the illustration. In the first condition these illustrations were decorative (DI-condition), in the second representational (RI-condition) and in the third representational with an additional warning (RIW-condition). Afterwards students received the same questionnaire as in Experiment 1, complemented by the general question whether the illustrations had helped to solve the word problems.

The results of Experiment 1 lead to the conclusion that representational illustrations did not help to solve P-items realistically, because students simply did not look at these illustrations. Experiment 1 also revealed that in many cases, students showed some awareness about the realistic modelling complexity involved in a P-item and therefore were in doubt about how to respond to it, but nevertheless ultimately decided to give a non-realistic reaction (NR). However, there was no evidence that the presence of a representational illustration had a positive impact on the frequency of these realistically inspired doubts about the correctness of their NR.

Experiment 2 revealed that forcing students to attend to these illustrations and motivating them to actually use the illustrations in their solution yielded no positive effect on the number of RRs to P-items. Furthermore, probably as a consequence of the somewhat higher number of RRs to P-items than in Experiment 1, students in Experiment 2 were somewhat less hesitant about the correctness of their NRs. However, there still was a substantial amount of hesitations, the vast majority of which were due to realistic considerations. Still, as in Experiment 1, there was no significant impact of condition either on the strength of these hesitations or on their underlying reason. Finally, the vast majority of the students (84.3 %) indicated that the illustrations (almost) never helped them to solve the word problems. Only 13.6 % reported that the illustrations were at least sometimes helpful. There were again no significant differences between conditions.

Altogether, the results of Experiment 1 and 2 raise the question why the representational illustrations being used in these two experiments (as well as in the study of Dewolf et al. 2014) did not help to respond the P-items more realistically. In what follows we will discuss several possible and complementary explanations for these findings from two

perspectives. First, in terms of the phase of the word problem solving process where the potential positive impact of the illustration could have occurred (=the *when* perspective). Second, in terms of the factors that may account for the fact that the illustration failed to realize its expected impact in the various phases of the solution process (=the *why* perspective).

With respect to the first perspective, it seems that the illustrations failed to bear their intended effect at three different moments of the participants' solution processes. First, as revealed by the eye-movements of Experiment 1 and as suggested by the questionnaire data of both experiments, many participants did not look at all at the illustrations or looked at them only very superficially (i.e., in basically the same way as they did at the decorative illustrations). Arguably, if an illustration is not or hardly processed, it cannot bear its expected positive effect. Second, some participants may have looked at the illustrations and processed them, but still not have built a more elaborated mental model including the realistic modelling complexity underlying the P-item, which could have led to a more appropriate mathematical model and ultimately a RR. Third, some participants may have built a more elaborated mental model including the realistic modelling complexity, and, consequently, may have considered to construct a more appropriate mathematical model and a more realistic solution to the problem, but nevertheless ultimately have chosen for the NR. Unfortunately, the design of our experiments does not allow us to determine how many NRs originated at each of these three phases of participants' solution processes.

With respect to the second perspective, four factors may account for the illustrations' failure to realize their expected positive impact on the realism of the participants' answers. A first factor relates to the nature of the illustrations themselves. In retrospect, it seems that the representational illustrations used in the present study as well as in the study of Dewolf et al. (2014) did not depict in a sufficiently salient way the realistic modelling complexity involved in the P-items. All our illustrations were designed in such a way that they evoked the global scene described in the problem, with a view to trigger the problem solver's real-world knowledge about that problem, without, however, pointing directly or explicitly to the specific realistic modelling complexity. For instance, for the rope item, we deliberately did not make a drawing zooming in on the hands of a boy who is actually knotting together two pieces of rope, whereby it is clearly shown that a significant amount of rope is needed just for the knotting. Given the aims and scope of the present study, we did not want to change the nature of the illustrations (compared to the study of Dewolf et al. 2014). However, given the disappointing results of both the study of Dewolf et al. (2014) and the present study, it would be interesting to set up a new investigation that would compare the effect of the illustrations used in Dewolf et al. (2014) and in the present study with illustrations that depict the problematic nature of the P-items in a more salient way (along the lines suggested above for the rope item). Actually, we are planning a study in which an extra element is added to the illustrations that makes the realistic modelling complexity more apparent.

A second factor relates to the nature of the word problems being used in these studies, and, more particularly, to the seemingly simple and straightforward nature of the problems in the eyes of their solvers (even though these P-items were, when given some deeper thought, quite non-trivial and complex mathematical modelling tasks). As a consequence of this task characteristic, participants may have glanced at the word problem and immediately perceived it as a simple S-item, without carefully reading and analyzing the problem and consequently discovering the mathematical modelling complexity. This immediate perception of the problem as a simple S-item may also have prevented the participants from processing the accompanying illustration (as one would expect to do

when being confronted with a word problem that one does perceive as a non-trivial and complex mathematical modeling task). It would be therefore interesting to further investigate whether there is a relation between solvers' initial perception and interpretation of a word problem (as a simple S-item or as a challenging mathematical modeling task) and their subsequent looking behaviour towards the accompanying illustration.

A third factor relates to the precise relationship between the illustrations and the texts of the word problems. In both experiments, we presented the illustrations next to the texts of the word problems rather than trying to completely integrate text and picture, and in the second experiment, the text was only added to the illustration after a period of 5 s wherein only the illustration was presented. This implies that the presentation of text and picture was not in line with the split-attention effect principle (Schnotz 2005). So, both in terms of space and time, the illustrations were presented under conditions that were suboptimal from an instructional design point of view. This experimental design of presenting illustrations next the P-items, allowed us to see if the illustrations accompanying the text were looked at after all, and, if so, to what extent. Moreover, from a practical perspective, the way in which we presented these illustrated word problems corresponds with how illustrations are added to word problems in regular mathematics textbooks and tests. Although, in some textbooks and tests that are based on a more authentic or realistic approach to mathematics education, alternative forms of combining textual and pictorial elements may also be used. In the vast majority of cases when a word problem is accompanied by a picture, constructors of textbook and tests will present them separately, as in the present study. So, both for reasons of internal and external validity, and since the goal of the present study was to unravel and better understand (by means of eye-tracking) why the representational illustrations added to word problems in the study of Dewolf et al. (2014) had not worked as expected, we decided not to integrate text and illustrations. Still, we acknowledge that in doing so, the potential impact of the split-attention effect may have been underestimated, and that it may be a possible explanation for why there was no effect of the representational illustrations. Therefore, we are planning a new study wherein we will compare the impact of various ways of combining text and illustrations of P-items on the realistic nature of learners' solution processes and outcomes, with a view to see if the number of RRs on P-items can indeed be improved by alternative text-illustration combinations that minimize the split-attention effect, instead of the mere juxtaposition of text and illustration as in the present study and the study of Dewolf et al. (2014).

Fourth, and finally, the representational illustrations may not have worked because of participants' metacognitive beliefs about solving school *word problems*. While some partial and indirect evidence for this third explanation was found in the questionnaire data the nature of that questionnaire did not yield rich and reliable findings about these beliefs and how they actually affected students' solutions of the P-items. Additionally, learners also may have certain beliefs about the informative or trustworthy value of illustrations accompanying mathematical text in general and word problems in particular, which also may explain the absence of an effect of the illustrations. So, we are planning future research in which we investigate these beliefs about word problems and their illustrations in a more systematic and fine-grained way, by means of individual interviews wherein students are explicitly asked about these beliefs and when and how students process the word problems and their accompanying illustrations. Alternatively, researchers could also try to alter these beliefs, for instance, by telling participants that alternative ways of reacting to the word problems are allowed or that illustrations are not merely decorative but are of significant help to solve the items. Clearly, some efforts in this direction have

already been done, both in the study of Dewolf et al. (2014) and in the present study, but apparently these efforts were too weak.

Besides the suggestions for further research raised above, a final perspective for further research could be to explore the efficacy of more direct instructional interventions aimed at increasing learners' RRs to (illustrated) P-items. For example explaining to learners that word problems sometimes require the inclusion of realistic considerations, putting them in a situation wherein they can actually experience that looking at illustrations may help to better understand and solve a word problem. Teaching them, e.g. through worked-out examples, how to represent and solve P-items, instructing them how to make use of textual and pictorial information to construct an integrated model of the problem situation, making learners aware of their beliefs about word problems and illustrations and how these beliefs may negatively affect the solution process, etc. Although some of these more direct instructional interventions have to some extent already been explored in intervention studies (for an overview see Verschaffel et al. 2009), or in the RIW-condition in experiment 2 of the present study, it would be interesting to investigate them more systematically in future research.

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, it could be argued that the representational illustrations should have helped participants to create an elaborated mental model of the problem situation of the P-items (=situation model), which leads to a more appropriate mathematical model and ultimately, to a realistic answer. This was clearly *not* what we found in our two studies and what was found by Dewolf et al. (2014). 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 (in our case: the mathematics class setting, or prior experiences with such settings for higher education students) or 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 may be a valuable challenge for future research.

At an even more general theoretical level, the ineffectiveness of our experimental manipulations can be viewed as a case of what instructional designers have called non-compliant learner behaviour (Elen 2013; Goodyear 2000). With this term they refer to the well-documented phenomenon that learners do not use, or inadequately use, the supportive tool provided by the designer (Elen and Clarebout 2006, 2007). Students in our two studies were non-compliant, as they were neglecting the representational illustrations and the warning that were intended to help them to solve the P-items more realistically. Relying on Perkins (1985), Elen (2013) proposes three possible general explanations for students' non-compliancy. First, learners may not make use of the learning opportunity because the opportunity was not actually present. Applied to our studies, the representational illustrations that we perceived as a helpful tool to solve the P-items more realistically, may in fact be ineffective to help students to solve the items more realistically due to some intrinsic shortcomings as mentioned above. Second, learners have to be knowledgeable about the learning opportunity, that is they have to be informed about the tool's usefulness. Perhaps students in our studies perceived the representational illustrations as useless (e.g., because of their previous experiences with the mainly purely decorative illustrations

accompanying word problems in typical mathematics textbooks) and consequently did not perceive them as helpful for solving the P-items. Finally, learners have to be motivated to take the opportunity offered by the tool. Students in our studies may not have been motivated to pay attention to and make active use of the illustrations, because they lacked the necessary internal or external motivation to perform well on the task.

References

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*, 143–177.

Blum, W., & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects: State, trends, and issues in mathematics instruction. *Educational Studies in Mathematics, 22*, 37–68.

Burkhardt, H. (1994). Mathematical applications in school curriculum. In T. Husén & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (2nd ed., pp. 3621–3624). Oxford: Pergamon Press.

Caldwell, L. (1995) Contextual considerations in the solution of children's multiplication and division word problems. Master's thesis, Belfast, Queen's University, Northern Ireland.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review, 14*, 5–26. doi:10.1023/A:1013176309260.

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. doi:10.1080/00220973.2012.745468.

Duchowski, A. (2007). *Eye tracking methodology. Theory and practice*. Retrieved from http://www.springerlink.com/.

Elen, J. (2013, August). "Instructional disobedience": Challenging instructional design research. *Plenary lecture given at the 15th European conference for research on learning and instruction,* Munich.

Elen, J., & Clarebout, G. (2006). The use of instructional interventions: Lean learning environments as a solution for a design problem. In R. E. Clark & J. Elen (Eds.), *Handling complexity in learning environments: Research and theory* (pp. 185–200). Oxford: Elsevier.

Elen, J., & Clarebout, G. (2007). Supporting Learners: Increasing complexity? *Computers in Human Behavior, 23*, 1163–1166.

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: University College.

Eyelink II [Apparatus and software]. (2004). Mississauga: SR Research.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*, 1–29.

Gerofsky, S. G. (1999). The word problems as genre in mathematics education. Unpublished doctoral dissertation or master's thesis, Simon Fraser University.

Goodyear, P. (2000). Environments for lifelong learning. Ergonomics, architecture and educational design. In J. M. Spector & T. M. Anders (Eds.), *Integrated and holistic perspectives on learning, instruction and technology* (pp. 1–18). Dordrecht: Kluwer Academic Publishers.

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

Hannus, M., & Hyönä, J. (1999). *Utilization of illustrations during learning of science textbook passages among low- and high-ability children, 24*, 95–123.

Inoue, I. (2005). The realistic reasons behind unrealistic solutions: The role of interpretive activity in word problem solving. *Learning and Instruction, 15*, 69–83.

Lave, J. (1992). Word problems: A microcosm of theories of learning. In P. Light & G. Butterworth (Eds.), *Context and cognition: Ways of learning and knowing* (pp. 74–92). New York: Harvester Wheatsheaf.

Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science, 41*, 811–831.

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology, 30*, 195–232.

Martin, P. (2011). Free kids clip art by Phillip Martin. Retrieved from http://www.phillipmartin.info/clipart/homepage.htm.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*, 1–19.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43–52.

Paivio, A. (1986). *Mental representations. A dual coding approach*. New York: Oxford University Press.

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

Perkins, D. (1985). The fingertip effect: How information-processing technology shapes thinking. *Educational Researcher, 14*, 11–17.

Potter, M. C. (1975). Meaning in visual search. *Science, 187*, 965–966.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 509–522.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science, 20*, 6–10.

Renkl, A. (1999, August). The gap between school and everyday knowledge in mathematics. *Paper presented at the Eighth European Conference for Research on Learning and Instruction,* Sweden: Göteborg.

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.

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49–69). New York: Cambridge University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*, 141–156.

Schoenfeld, A. H. (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, NJ: Erlbaum.

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience, 14*, 869–876.

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.

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

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

Verschaffel, L., Greer, B., Van Dooren, W., & Mukhopadhyay, S. (2009). *Words and worlds: Modelling verbal descriptions of situations*. Rotterdam: Sense Publishers.

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.