Educ Stud Math (2008) 67:37–58
DOI 10.1007/s10649-007-9083-3

Impact of authenticity on sense making in word problem solving

Torulf Palm

Published online: 29 June 2007

Abstract The study presented in this paper seeks to investigate the impact of authenticity on the students' disposition to make necessary real world considerations in their word problem solving. The aim is also to gather information about the extent to which different reasons for the students' behaviors are responsible for not providing solutions that are consistent with the 'real' situations described in the word problems. The study includes both written solutions to word problems and interview data from 161 5th graders. The results show an impact of authenticity on both the presence of 'real life' considerations in the solution process and on the proportion of written solutions that were really affected by these considerations. The students' frequent use of superficial solution strategies and their beliefs about mathematical word problem solving were found to be the main reasons for providing solutions that are inconsistent with the situations described in the word problems.

Keywords Authenticity · Beliefs · Real life · Sense making · Word problems

1 Introduction

1.1 Background

In a large number of studies students have been confronted with word problems such that if solved in a stereotypical way, without much attention paid to the realities of the 'real' situations described in the tasks, the solutions would not match, and in some cases even be absurd in relation to these 'real' situations. The collective outcome of these studies is that

T. Palm (✉)
Department of Mathematics, Technology and Science Education,
Umeå University, 901 87 Umeå, Sweden
e-mail: torulf.palm@educ.umu.se

when confronted with such tasks elementary and secondary school students from different parts of the world often do provide solutions that are inconsistent with the 'real' situations described in the tasks. The general conclusion is that students have a tendency not to make proper use of their real-world knowledge and to suspend the requirement that their solutions must make sense in relation to the 'real' situations. This tendency also has been found to be shared by pre-service teachers (for a comprehensive overview of these studies see Verschaffel et al. 2000).

The tendency to provide such 'unrealistic' solutions seems to be strong and not easily overcome by hints (Greer 1997; Reusser and Stebler 1997a; Verschaffel et al. 1999; Yoshida et al. 1997). The effect of different working conditions on the students' behavior when solving this kind of problematic word problem has also been addressed in experimental studies. These studies show that substantial changes in the task solving conditions may influence students' modeling behavior. Examples of such changes are arrangements for the students to make actual telephone calls to order buses (practically arranged by using two additional rooms, two telephones, and an unknown person answering in the other end) (DeFranco and Curcio 1997), providing students with concrete materials such as planks, a saw, and a meterstick (Reusser and Stebler 1997b), and having the students working with tasks in a social science class setting instead of in the setting of a mathematics lesson (Säljö and Wyndhamn 1993).

Several reasons for the 'unrealistic' answers have been suggested in the literature. One suggested explanation is that students' solution strategies comprise mindless calculations and do not include considerations of the real life aspects of the situations described in the tasks. Another explanation that has been put forth is the students' beliefs about word problem solving. They adapt to the socio-cultural norms of schooling and consciously decide not to consider the realities of the 'real' situations described in the tasks. The seemingly 'unrealistic' answers may also stem from interpretations of the real situations that does not coincide with the interpretations of the researchers. It may also be that the students do not always possess the real world knowledge that is necessary to make the required 'realistic' considerations.

However, studies providing empirical evidence for the suggested explanations are rare. An interview with 15 children in a study by Hidalgo (cited in Verschaffel et al. 2000), gives some support to the idea that the students' 'unrealistic' responses may partly be due to their solution strategies, which keep them from noticing that their solutions are not consistent with the 'real' situations described in the tasks. The study also provides some support for the explanation that students' beliefs, such as that all mathematical tasks have a single numerical solution, may hinder them from providing appropriate answers. Some of the students in the study did not seem to have the necessary real-world knowledge. Inoue (2005) set out to systematically investigate which of these explanations that account for undergraduate students 'unrealistic' responses to similar tasks (except the possession of the required real life knowledge). The study found that the students' solutions could be interpreted with these explanations and that each of the explanations was responsible for some of the 'unrealistic' solutions provided by the students.

In another line of research Cooper and Dunne (2000) observed that student solutions that *do* include realistic considerations sometimes include the task developers' intended specific realistic considerations, but sometimes they include general realistic considerations, which not always render full credit to assessment tasks. They also showed that a major reason for this difference in the type of realistic considerations made is a different interpretation of what the tasks demand and that the likeliness of each type of interpretation is dependent of social class.

1.2 Aims

The above studies give ample evidence of students' tendencies to provide 'unrealistic' solutions to the word problems administered in the studies. However, the tasks used in this line of research, for example in the study by Verschaffel et al. (1994), and the replication studies by others, are often themselves 'unrealistic' in the sense that important aspects of the 'real' situations described in the tasks are not well emulated in the word problems (this is further described in Section 2.2).

Thus, the students have been faced with 'unrealistic' mathematical school tasks and have in many cases provided 'unrealistic' responses to them. Such an observation calls for an investigation of the extent to which the 'realism' of the mathematical tasks the students are set to solve influences the 'realism' of the solutions the students provide. There is no consensus in the mathematics education community of which term to use for this 'realism' of a task or concordance between a school task and a real life task situation. Different terms have been used to label tasks that in some way emulate real life task situations (e.g. authentic tasks, realistic tasks, real life tasks), and in addition many different meanings have been attached to each one of them (for a review of different meanings attached to these terms see Palm 2002). In this paper the concordance between on the one hand a mathematical school task that includes a description of a 'real' out-of-school situation and on the other hand the actual real life situation will be called the authenticity of the school task and its more precise meaning will be described in the next section.

The effect of some task features on the students' realistic considerations was also the interest of Cooper and Harries (2002, 2003). They put specific attention on the students that *did* include some realistic considerations to a 'realistic' task (Cooper and Dunne 2000, p. 84) defined a task as 'realistic' "if it contains either persons or non-mathematical objects from 'everyday' settings"). They investigated these students' capacity and/or willingness to introduce an *extended* range of realistic considerations to the task by providing the students with four different responses (by fictitious other students) to the task and asking them if they were willing to accept these as correct answers. Based on interviews they concluded that a significant minority of the students was able and willing to extend their realistic considerations and offer support for some of the answers that were provided in the task. Based on this research they also suggested that "given suitable 'realistic' problems, many children may be more willing and able to introduce realistic responses in a testing context than earlier research might lead us to expect" (Cooper and Harries 2002, p. 1).

The main aim of this study is to investigate the impact of task authenticity (as defined in the next section) on students' use of real world knowledge in their solutions to word problems. Specifically, the study investigates the hypothesis that the students would introduce 'realistic' responses more often if the word problems were more authentic. For the results to be most useful for teachers it was also decided to impose the restriction that a higher degree of task authenticity should not require substantial changes of the students' task solving conditions but has to be accomplished within the frames of the practicalities of normal classroom procedures, even though this limits the extent to which a task can be made authentic. Specifically, the higher degree of authenticity is to be accomplished solely by a modification of the task text. A second aim is to gather information about the main reasons for not providing solutions that are consistent with the situations described in the tasks.

1.3 Authenticity

To study the concordance between a school task and a real life task situation it is useful to have a definition of what constitutes such a relationship. That is, in more precise terms what we mean

with a school task that well emulates a real life task situation. In this paper the term *authentic* is used for this relationship and the term is used in relation to the out-of-school situation that is described in the task, the *task context*. The underlying idea lies in the following quotation by Fitzpatrick and Morrison (1971, p. 239): "if a performance measure is to be interpreted as relevant to 'real life' performance, it must be taken under conditions representative of the stimuli and responses that occur in real life." This means that (with the use of the term authentic in the sense described above) for a school task with an out-of-school task context to be authentic it must represent some task situation in real life, and important aspects of that situation must be simulated to some reasonable degree. The more specific operational framework (Palm 2002, 2006) includes a number of such aspects of real life task situations which simulations have been argued to be important for the possibilities of the students engaging in the mathematical activities attributed to the simulated real life situation.

A school task can, of course, never completely simulate an out-of-school task situation. Nevertheless, sometimes the school situation can be organized and the assignment formulated in such a way that many of the aspects of a real life task situation may be simulated fairly well following that the students' task solving can take place under conditions fairly close to those in the simulated situation. Other times, for example in large-scale high stakes testing, the conditions under which the task solving takes place put severe restrictions on the possibilities to simulate many of the aspects with high fidelity. Thus, such circumstances restrict the possibilities of developing very authentic tasks. However, under both of these circumstances tasks may be developed that simulate more or fewer of the aspects with more or less fidelity. Regardless of the existing restrictions of the school situation at hand the framework for authenticity used in this study may be helpful both in distinguishing between tasks in terms of their authenticity and for developing tasks aiming at highest possible authenticity under the existing circumstances.

One of the drawbacks of choosing tasks from the literature as a starting point for task development, which is done in this study, is that it is not possible to simulate all of the aspects with reasonable fidelity by only modifying the task text. However, some of them can be simulated. The following is a condensed and short description of some of the aspects in the framework, which are used in the discussion and development of the word problems in this study. For a more comprehensive description and argumentation of the framework, see Palm (2002, 2006).

Event This aspect refers to the event described in the task. In a simulation of a real life task situation it is a prerequisite that the event described in the school task have taken place or have a fair chance of taking place.

Question This aspect refers to the concordance between the assignment given in the school task and in a corresponding out-of-school situation. The question in the school task being one that actually might be posed in the real life event described is a prerequisite for a corresponding real life task situation to exist and therefore also for the whole simulation enterprise.

Purpose in the task context The appropriateness of the answer to a task, and thus the necessary considerations to be made, sometimes depends on the purpose of finding the answer. In other tasks the whole solution method is dependent on the purpose. Thus, the purpose of the task solving in the task context needs to be as clear to the students in the school situation as it would be in a corresponding real life situation. This experienced clarity might arise as a result of an explicit declaration of the purpose in the task, or might be experienced as implicitly clear from the task context.

Language use This aspect refers to the terminology, sentence structure, and amount of text used in the presentation of the task situation. In a simulation of this aspect, with a reasonable degree of fidelity, the school task does not, for example, include difficult terms that hinder the students in their task solving if the corresponding difficulties do not occur in the simulated out-of-school situation.

Information/data This aspect refers to the information (including values, models and given conditions) on which the solution to a problem can be based. The aspect is divided in the following three subaspects:

- *Existence of information/data*. If this aspect is simulated with high fidelity then the same kind of information accessible in the simulated real life situation is also accessible in the school situation. Differences in accessible information can arise if information that would have been known in the simulated task situation is not given in the school situation, or if additional important information is added to the school task. Lack of information in school tasks occurs when, for example, numerical values are withheld from the students, but also when the description of the situation that is simulated is so short of contextual features that the students do not get a clear overall picture of the situation. Differences can also arise if the information given in the school task have been substantially simplified or made more difficult than in the simulated situation.
- *Realism of information/data*. In a simulation of this aspect, with a reasonable degree of fidelity, numbers and values given are realistic in the sense of identical or very close to the corresponding numbers and values in the simulated situation.
- *Specificity of information/data*. In a simulation of this aspect, with some reasonable degree of fidelity, the information given is specific and not general. The task text describes a specific situation in which the subjects, objects, and places in the task context are specific.

1.4 Research questions

Given the use of the term authenticity in this paper, the aims described earlier can now be stated more precisely as the following two research questions:

Q1: Do the 5th grade students, in the Swedish city used for the present study, provide a larger proportion of solutions that are consistent with the 'real' situations described in the word problems if the tasks simulate some of the aspects of real life task situations, described above, with higher fidelity? That is, do they provide more 'realistic' solutions if the tasks are more authentic?

Q2: What are the main reasons why the 5th grade students do not provide solutions that are consistent with the 'real' situations described in the word problems?

2 Materials and method

2.1 Procedure and participants

The participants in the study were 161 students from eight fifth grade classes (11-year olds) in a middle-size city in Sweden. The eight classes were a random selection from the 33 fifth grade classes in the city. Most of the schools in the city include students from different

socio-economic backgrounds but the majority of the students in the city belong to the middle class. There were no indications that this random selection of schools was different from the population of all 33 schools in this respect. The selection included about the same number of boys and girls. Based on the interviews with the teachers a common learning activity for the students had been the work with exercises in the textbook. The students in all classes had dealt with a lot of traditional word problems but had not had much experience with tasks in which substantial engagement in the task context is essential for successful solutions to the tasks.

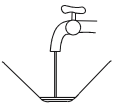
Each student in the sample was given one of two versions of a written test. One test version consisted mainly of word problems from studies in the literature (e.g. Verschaffel et al. 1994) and the other version consisted of *more authentic* variants of these word problems. The word problems taken from the literature will be called *less authentic* task variants. An equal number of tests of each version were randomly assigned to the students in each class. Since the relative order of the tasks could affect the results on each task, two different orders of presentation of the tasks were used for each version. After the test had been carried out each student was interviewed to gather further information about the responses. In the analysis, both, written and oral information was used to compare the responses to the word problems from the literature with the responses to the more authentic task variants.

Before the tests were handed out, the students were told that the reason for the test was to learn about how students work with mathematical tasks. The administration of the tests was made by the teacher as a part of a normal mathematics lesson. The students were instructed to treat the task solving as an ordinary classroom activity and to write down their solutions and answers as usual. In addition, they were asked to write comments in a commentary area if they did not understand a task or if they thought it was strange in some way. They were instructed not to ask any questions once they had started solving the task. Otherwise the instructions were kept to a minimum. The teacher was instructed not to answer any questions but to direct uncertain students to the commentary area. Each student had a calculator available at the desk, although most of them had limited experience in using it.

2.2 The tests

Both test versions consisted of seven tasks. Table 1 provides a list of all the word problems with their sources (some of the tasks taken from the literature have been slightly modified to fit this study). The first task (WP 1) was a straightforward less authentic word problem, which was identical in both versions of the test. This task was included for the purpose of giving the students a comfortable start in their work. The other six tasks in each test version, *the target tasks*, were the tasks of real interest. In one of the two test versions all of the target tasks were *less authentic word problems* (WP 2-WP 7). Five of these were near replications of tasks used in the studies by, for example, Reusser and Stebler (1997a), Verschaffel et al. (1994), and Yoshida et al. (1997), and one was a new task (WP 5). In the other test version the target tasks were *more authentic variants* of the tasks in the first test version (WPA 2–7). These tasks were modified versions of the less authentic word problems and are considered to better simulate some of the aspects of real life task situations outlined in the used framework for authenticity. The extent to which the tasks could be made authentic was limited by the decision to take the tasks from the literature as a starting point for the development of the more authentic task variants, instead of starting with real instances of real life task situations, and to restrict the changes to modifications of the task text. However, the first of these restrictions allowed a close connection to the

Table 1 The word problems in the study

WP 1	You are buying candy in a candy store. The candy costs 12.50 Kr, and you give the store assistant 20 Kr. How much money should you get back?
WP 2	Grandfather gives his 4 grandchildren a box containing 18 balloons, which they share equally. How many balloons does each grandchild get? (Davis 1989)
WP 3	360 students shall go by bus on a school trip. Each bus can hold 48 students. How many buses are required? (Carpenter et al. 1983)
WP 4	Anton has bought 4 planks of 2.5 m each. How many planks of 1 m can he saw out of these planks? (Kaelen 1992, in Verschaffel et al. 1994)
WP 5	Elin is planning to ride horses each day for 4 days. Each day she has 45 minutes of free time to do this. How many 10-minute rides does she have the time to do during these days?
WP 6	Martin's best time to run 100 m is 10.00 sec. How long will it take him to run 10 000 m (=1 Swedish mile)? (Greer 1993)
WP 7	The bowl in the picture is being filled from a tap, which flows at a constant rate. After 1 minute the depth of the water is 2 cm. The height of the bowl is 20 cm. How long time does it take until the bowl is full? (Greer 1993) 
WPA 2	You have a birthday party, and if you count yourself you are 4 children at the party. You receive a bag with 18 balloons from your grandfather. For no one to be sad he wants you to share them equally so that each of you get the same number of balloons. How many balloons shall you give to each child?
WPA 3	All students in the school will go on a school trip together on the 15th of May. You and the other organizing students have decided that everyone will go by bus, and that you will order the buses. You have seen in the student rosters that there are 360 students in the school. Your teacher said that you can order the buses from Swebus, and that each bus can hold 48 students. Fill in the note below, which you are going to send to Swebus to order the buses SWEBUS – Busorder Your name:…….. School:……. Date of the trip:…… Number of buses to order:……. Other requirements:…… ……

WPA 4 You are building a cabin and as walls you want to use planks that are 1 m long. You are at the moment short of thirteen 1-meter planks. A friend says that she has found 4 planks, each 2.5 m long. You are wondering if that is enough to finish the walls. How many 1-meter planks can you saw out of the planks she found?

WPA 5 You are going to a camp for 4 days, but you also want to ride horses. Your dad sees in the camp brochure that you have 45 minutes free time each day, and that horses can be rented for tours on a path in the woods that takes 10 minutes. To know how much money you will bring you must calculate how many tours you have time to ride. How many 10-minute tours do you have the time to do during these days?

WPA 6 There is an athletics competition on TV. You and a friend watch when the fastest man in the world, Maurice Green, wins the 100 m race in 10.00 sec. The next race you watch is 10 000 m, which is won by Haile Gebrselassie in 26 min. and 5 sec. What do you answer when your friend asks you: How long time do you think it would take Maurice Green to run 10 000 meters (= 1 Swedish mile)?

WPA 7 It's a crisis! The pipe that leads water to the toilet has broken. You find the bowl in the picture and put it under the pipe. The water flows at the same speed the whole time. You want to know if the water will overflow the bowl while you get help. You see that after 1 minute the water level in the bowl has risen to 2 cm from the bottom, and you measure the height of the bowl to 20 cm. How long time will it take till the bowl is full?

literature on sense making in word problem solving and the aspects from the framework of authenticity that were chosen for this study was possible to consider at least to some extent. The changes made in the WP-tasks were based on considerations of these aspects. Thus, a difference in authenticity between the WP-tasks and the WPA-tasks was attained.

In the less authentic variants of all of the tasks the description of the real life task situation is minimal. The students are given very little information about the circumstances of the situation, including the purpose of the question. In addition to changes in specific tasks all of the more authentic variants of the tasks include a more thorough description of the task context and the purpose of solving the task in the simulated situation. Also, the students are included in the described situation. However, at the same time, to limit the possible problems for students with reading difficulties, attempts have been made to avoid too long descriptions of the task context. That is, attempts have been made to increase the fidelity in the simulations of the aspects Existence and Specificity of information without to much decrease in the fidelity of the simulations of the aspect Language use.

The target tasks represent three different types of tasks included in the three above-mentioned studies. The following is a more in-depth discussion of the tasks, and it focuses on the aspects that are considered to be simulated with different fidelity in the different task variants. All of the target tasks in the test will be included in the quantitative analysis. However, for reasons of space only one of the tasks in each of the three categories will be focused in the qualitative analysis of the tasks and the students' responses to them.

2.2.1 Category 1 word problems

A successful solution to WP 2–3 and WPA 2–3 requires an interpretation of a remainder, and this type of tasks is in this study called *Category 1 word problems*. In the bus tasks the answer to the calculation 360/48 has to be rounded up to achieve an appropriate answer to the tasks. For an answer to be consistent with the realities of the situation described in the task it has to be a whole number of buses since buses do not function well in halves. Another characteristic of the Category 1 word problems that differentiates them from the other two word problem categories is that when the students write their 'response sentences' (e.g. "They will need... buses") which most of them do, they will get a hint to activate their real-world knowledge about the task context when the answer to the calculation is 7.5. A hint of the same magnitude will not occur when the students write their response sentences to the category 2 and category 3 word problems (Verschaffel et al. 1994).

The bus tasks The aspect of most importance in this task may be the purpose in the task context. In the less authentic task variant it is not known if the question is asked so that the solution could be used directly to order the buses, or if the purpose is to use the answer as a part of a basis for the discussion of such an order. In the more authentic task variant the purpose of the question in the task context is made clearer through requiring the students to fill in an ordering sheet.

2.2.2 Category 2 word problems

WP 4–5 and WPA 4–5 represent *Category 2 word problems*. The solution to these tasks involves recognizing that all of the quantity of interest cannot be used. To make proper use of the knowledge of the situation described in the plank tasks it must be noticed that 0.5 m of each 2.5 m plank will remain after the sawing has been done. Therefore, only eight 1-m planks can be sawed out of the four 2.5-m planks.

The plank tasks It is not likely that anyone has already bought four 2.5-m planks without knowing how many planks of the desired length could be obtained from them. The development of a more authentic variant of the task therefore included finding an event in which the question would make sense. For the question to be relevant and a clear purpose to be apparent the situation in the more authentic task variant involves the idea of enough planks and therefore includes the number of planks that is needed. This information would in a less authentic word problem often be regarded as superfluous information since it is not needed for the calculation of 2×4. The number of missing planks is provided with letters for the purpose of not making it salient information. Nevertheless, this extra information might affect the students' solution strategies, but whatever effect the extra information may have it is considered to be a part of the simulated real life situation (see Results Section 3.3 for possible effects).

2.2.3 Category 3 word problems

WP 6–7 and WPA 6–7 are called *Category 3 word problems*. Successful solutions to these tasks have to include estimations, and the use of a linear model based only on the numbers in the tasks is not a good description of the 'real' situations described in the tasks. Successful solutions to the running tasks, WP 6 and WPA 6, have to take into account that a runner can not keep the same speed for 10,000 m as he can in 100 m.

The running tasks In these tasks the question was of main concern. Together the information and the question presented in the less authentic variant of the task imply that the person asking (which in this task is clear to be the task developer and not someone in the task context) knows that the person who is supposed to answer the question does not know anything about Martin other than the data given in the task. In such situations the question in real life situations would not be "What *will* his time be?" but "What do you *think* his time would be?". The word 'think' more clearly opens up the possibilities of not knowing or not being able to specify an exact time. The use of the formulation "will be" together with the information provided in the task is a combination that insinuates, in a way that a question in life beyond school would not, that there is one specific correct answer to the task.

Thus, the question in the more authentic variant is formulated "How long do you *think*...." The purpose of the question is not explicitly formulated, but from the description of the situation the possibilities to interpret this purpose is similar in the real situation and in the school situation. The distance chosen for both variants is 10,000 m (compared to 1,000 m in the original task variant in Greer 1993). One reason for this was that it would be more likely that most students would know that this is a distance too long for the runner to be able to maintain his 100 m speed. The other reason was that in the World Championships in athletics there is no competition for 1,000 m but they do compete in 10,000 m.

2.3 The interviews

The interviews were conducted with one student at a time starting with the first student immediately after all of the students had finished the test. The purpose of the interview was explained to be that the interviewer, being a researcher, wanted to learn from them about students' mathematical task solving (in the end of the interview this was also taken as the purpose for the request that the students wait to talk to each other about the interview until the last student had been interviewed). One of the reasons for giving this, true, explanation was to support a relaxed interview situation in which the students would have as little incentive as possible to say something that was not true but that would make their task solving look good.

The aims of the interviews were to investigate (1) if realistic considerations of the task contexts had been made in their task solving, but not put on paper; (2) why these considerations did not affect the written solution (in the cases they did not); (3) whether the students really had the real world knowledge required for solving the tasks successfully; and (4) why realistic considerations had not been made (in the cases it had not).

The interviews were structured and based on the questions and requests below. The word problems and the students' responses to them were put on the table in front of the students and were used as the basis for the interviews. The first two questions were formulated to detect real world considerations made during the task solving with minimal influence from the interviewer. If no such considerations were detected the third question, specifically pointing to a task specific phenomenon, was posed. This question was included in the

interview since some students may not connect such considerations with the generally formulated first two questions. Then the students were asked why they did not use their real-world knowledge (if they had made such considerations but not used them in their written solution). Finally, to find out if the students really had the required real-world knowledge the fifth question was asked. The answers to the questions were discussed with the students with a special focus on validating their answers through the discussion. This was judged to be especially important if the students claimed to have made realistic considerations, but not until confronted with the third of the questions below. The students' answers to all of the questions and the discussion around them were, together with their written responses to the word problems, used to answer the question of why realistic considerations were not made (when they were not). The interview questions were the following:

1. Describe your thinking on which you based your solution.
2. Did you think of anything else, and did you make any other considerations, during the solution process?
3. A question about whether the students had made relevant realistic considerations during the solution process, explicitly pointing to a task specific phenomenon. For example: did you think about whether the runner could keep the same speed during the whole race?
4. Why did you not pursue this line of thinking in your solution to the task? (if the student had made realistic considerations but not used them in the solution to the task)
5. A task specific question aimed at finding out whether the students had the real world knowledge required for solving the task successfully. Concerning the running task this question was: Do you think the runner can keep the same speed during the whole race?

2.4 Data analysis

The students' work was analyzed on the basis of three different sources: the written solutions given in the solution area, the written comments given in the comment area, and the discussion in the interviews. To find an answer to research question 1 the solutions were first categorized in a number of *Base categories* developed before the analysis. This categorization was then summarized by merging some of these categories into a few *Summary categories* used in the statistical analysis. The different categories are described below in Sections 2.4.1 and 2.4.2. The information about the reasons for providing 'unrealistic' answers, research question 2, were not categorized in preformulated categories. For the purpose of being as open as possible for different reasons it was decided to define such categories during the interpretation of the gathered information.

2.4.1 Base categories

The solutions given in the solution area were, with considerations also taken to the information obtained in the interviews, categorized in ten categories (see below), some of them adopted from Verschaffel et al. (1994).

- Expected answer (EA). Answers that are inconsistent with the realities of the 'real' situation described in the task and which are the results of a straightforward solution method. The label for this category stems from the hypotheses that the students would demonstrate a strong tendency to exclude considerations of the real world in their word problem solving, and therefore solve these tasks in a

straightforward way without considering the realities of the 'real' situation described in the task. In accordance with the categorization in, for example, (Verschaffel et al. 1994) an answer to the bus tasks (WP3/WPA3 in Table 1) given as a fraction of buses is categorized as an expected answer (EA).
- Technical error (TE): Solutions based on similar solution strategies as the EA answers but that involves technical errors in the computations or other similar mistakes.
- Realistic answer (RA): Answers that follow from a solution process involving appropriate use of knowledge about the 'real' situation described in the task.
- No answer (NA): This category was used when no answer was provided.
- Other answers (OA): Answers that did not fit into one of the other categories.

In addition to the above five categories the solutions could also be categorized in the following categories:

- EA(T), TE(T), OA(T), and NA(T): The T stands for "thought" and denotes that the interviews indicated that the students had made realistic considerations in the solution process although they did not pursue this line of thought in the written response to the task. An example of a response to one of the running tasks categorized as EA(T) is the written answer 1,000 s from a student who in the interview said that he "did not think he (Martin) was going to make it, but the task did not say anything about that. So I only used what it said. Otherwise it is not possible to solve it".
- NA(a): This last category consists of responses such that the students did not provide a written solution, but did so because they had made a realistic consideration and could not find a solution that was consistent with the realities of the 'real' situation described in the task. The "a" denotes that the (lack of a) written solution is affected by the realistic considerations made. The information backing this classification comes from the interview, but the category is different from the category denoted by "NA(T)." The NA(T) category also describes solution processes including realistic considerations, but in that category these considerations were not the reason for the choice of not providing a written solution or answer. An example of an NA(a) classification is the non-written response to one of the running tasks in combination with the interview statement "I did not know how to calculate because he was going to get tired."

The comments given in the comment area were classified by a "+" sign or a "–" sign. If a comment indicated that realistic considerations had been made, whether or not used in the written solution, the solution was classified with a "+" in addition to the categorization of the response in the solution area. If there were no comments given, or if the given comments did not indicate that realistic considerations had been made, the solution was classified with a "–" in addition to the categorization of the response in the solution area.

2.4.2 Summary categories

The categorization was summarized in three ways. In the first summary the students' responses were summarized in the two categories Realistic written reactions (RR) or Non-realistic written reactions (NR). This summary is made to compare the students' responses on the 'less authentic' task variants in this study with the responses from the students in the previous studies by Reusser and Stebler (1997a), Verschaffel et al. (1994), and Yoshida et al. (1997). Since the proportion of realistic responses in theses studies are based solely on

the students' written work only the written responses, and not explanations made in the interviews, were used as a source of information for the interpretation of the solutions and comments. Thus, the RR-category basically consists of the responses RA− and all categories categorized with a "+" sign.

In the second summary the responses were summarized in the categories Realistic practice (RP) or Non-realistic practice (Non-RP). The former of these categories consists of the responses based on realistic considerations of the 'real' situations described in the tasks. This category consists of the responses categorized as RA−, RA+, NA(a)− or NA(a)+. The latter category consists of all other responses. The purpose of this summary is to provide an answer to Research question 1 (Q1). However, when a student did not come far enough in a solution to really have to make the crucial task specific realistic consideration this solution was not included in the statistics of RP and non-RP (and RC and non-RC, see below). Such solutions can be exemplified by students who did not provide a written answer on a task since they did not know how to go about the task (categorized as NA−) and a student who 'solved' the plank task (WP 4 in Table 1) by calculating 2.5×13 (categorized as OA−).

In the third summary the students' responses were summarized in the categories Realistic considerations (RC) or Non-realistic considerations (Non-RC). The former of these categories consists of the responses in which realistic considerations had been made in some way during the solution process, whether or not they affected the written solution. Thus, this category consists of all responses classified as RA−, NA(a)−, and all responses classified with a "+" or a "T." The latter category consists of all other responses. The purpose of this summary is to find out if the *activation* of relevant knowledge of the 'real' situations is also influenced by the authenticity of the tasks (Q1), but also to provide information that is helpful in answering Research question 2 (Q2).

3 Results

The Results section starts with the comparison of the Swedish students' responses to the 'less authentic' task variants with the responses provided by the students from other countries in some of the earlier studies. This comparison provides some background information about these Swedish students' tendencies to provide 'unrealistic' solutions. Then, in Section 3.1, the results about the first research question are presented. These results include a comparison of the proportions of 'unrealistic' solutions given to the 'less authentic' task variants with the proportions of 'unrealistic' solutions given to the 'more authentic' task variants. Section 3.2 includes a description of the reasons identified for the 'unrealistic' solutions (research question 2). Section 3.3 treats the reasons for the impact of authenticity on the proportions of unrealistic solutions. These reasons are described in terms of the reasons for providing 'unrealistic' solutions and are therefore presented after Section 3.2.

The results of the comparison of the task solving by the Swedish students in this study and the task solving by the students from Switzerland, Belgium and Japan reported by Reusser and Stebler (1997a), Verschaffel et al. (1994) and Yoshida et al. (1997) respectively is displayed in Table 2 below. In the table the proportion of RR on tasks similar to WP2, WP3, WP4, WP6 and WP7 are given. The table shows similar proportions of 'realistic written reactions' (RR) for the students in this study and the students in the other three studies. Thus, the results indicate that the students included in this study share the same tendencies to provide 'unrealistic' solutions as their counterparts in the other three countries and add this study to the bulk of research displaying this tendency in students from different countries around the world.

Table 2 The proportion (in percentages) of solutions categorized as RR in this study and in the studies by Reusser and Stebler (1997a), Verschaffel et al. (1994) and Yoshida et al. (1997)

	WP 2	WP 3	WP 4	WP 6	WP 7
Reusser & Stebler	75	49	14	5	0
Verschaffel et al.	59	49	14	3	4
Yoshida et al.	52	62	0	7	4
Palm	67	63	24	5	2

The table does not include WP 5 since the earlier three studies did not include a similar task.

3.1 Research question 1 (Q1) The impact of authenticity on the proportion of 'unrealistic' solutions

The results of the study show that an increased task authenticity, *even when it has to be accomplished solely by a modification of the task text*, can increase students' tendencies to effectively use their real world knowledge in the solutions to word problems. The students who were faced with the more authentic task variants provided written solutions that were consistent with the realities of the 'real' situations in the tasks (solutions categorized as Realistic practice (RP)) in 51 % of the solutions, while the corresponding proportion for the students faced with the less authentic task variants was 33%. The difference was statistically significant according to the Mann–Whitney test ($p<<0.05$). (As can be seen in the note under Table 3, below, a total of 51 solutions out of 966 were not included in the RP statistics. The exclusion of these solutions influenced the proportions of RP by two units of percent for both the less authentic and the more authentic tasks).

Focusing on separate tasks it can be seen in Table 3 that the proportion of responses categorized as RP was higher on the more authentic task variant than on the less authentic task variant on all tasks. All of these differences were statistically significant ($p<<0.05$ according to the chi-square test including Yates' correction) except for the balloon tasks (WP 2/WPA 2) and the bowl tasks (WP 7/WPA 7).

Authenticity also affected the students' tendency to activate their knowledge of the 'real' situations described in the tasks, whether or not it affected their written solutions (solutions categorized as Realistic considerations (RC)). In 57% of the solutions to the more authentic task variants the students had reflected over the problematic feature of the 'real' situation in the tasks, and the corresponding proportion for the less authentic task variants was 43%. The difference was statistically significant according to the Mann–Whitney test ($p<<0.05$). Significant differences between the proportions of responses categorized as RC occur on all tasks except the balloon tasks and the bowl tasks (see Table 3).

3.2 Research question 2 (Q2) The reasons for providing 'unrealistic' solutions

Based on the interpretations of the interviews and the written responses it appears that there were two main reasons for providing solutions that were inconsistent with the situations described in the tasks. One of these reasons was the frequent use of solution strategies that can be characterized as superficial. This is also consistent with the indications from the interview with 15 students by Hidalgo (cited in Verschaffel et al. 2000, p. 26) and the study on undergraduate students by Inoue (2005). These solution strategies do not involve a careful analysis of the task situations but a focus on the numbers given in the task. Thus, the strategies do not include a thorough reflection of the applicability of the mathematical models used for the solutions and not an evaluation of the answers in relation to the 'real' situations described in the tasks.

Table 3 Results of the categorization of the students' work with the word problems

Category	WP 2	WPA 2	WP 3	WPA 3	WP 4	WPA 4	WP 5	WPA 5	WP 6	WPA 6	WP 7	WPA 7
Realistic practice (RP)	72[a]	84	75[b]	95[c]	30[d]	58[e]	19[f]	49[g]	0	22	2	3
Realistic considerations (RC)	94[a]	90	84[b]	97[c]	32[d]	58[e]	21[f]	49[g]	24	42	9	11
Realistic reactions (RR)	67	78	74	90	24	32	16	43	5	22	2	3

The classification is based on their written work and the interviews. The numbers in the table are percentages.
The data are based on the responses and considerations made by 161 students from eight fifth-grade classes. Eighty-two students solved the less authentic variants of the word problems and 79 students solved the more authentic variants. Students not confronted with the situation of having to make the crucial realistic consideration in a specific word problem were not considered in the proportion of RPs and RCs.

[a] The proportion is based on 79 students.

[b] The proportion is based on 76 students.

[c] The proportion is based on 77 students.

[d] The proportion is based on 76 students.

[e] The proportion is based on 66 students.

[f] The proportion is based on 68 students.

[g] The proportion is based on 72 students.

A representative example of an 'unrealistic' answer to the bus tasks interpreted to have been provided because of a superficial solution strategy is the written answer 7.5 buses combined with the following answer to the third interview question: "I did not think about buses, it was just 7.5." Similar responses to the plank tasks and the running tasks were interpreted in the same way. Another example of the lack of significant engagement in the running task context are the students that answered with 1,000 s and in the interview said that they, when discussing the solution in the interview, were not certain whether the person in the running tasks had run the 100 m as fast as he could (following that maintaining this speed may be possible).

The other main reason for the 'unrealistic' solutions that appeared in the study was the students' beliefs about school mathematics task solving in general and word problem solving in particular. This is also consistent with the indications from the interview with the 15 students by Hidalgo (cited in Verschaffel et al. 2000 pp. 26–27). It is also consistent with the results from the study on undergraduate students by Inoue (2005). These beliefs do not include the requirement that school mathematics and real life outside school must be consistent. On the contrary, they do include the ideas that all tasks have a solution, that the solution is attainable for the students, and that the answer is a single number.

Most of the 'unrealistic' responses (to the three focused tasks) that were judged to have been provided due to the students' beliefs were given to the running tasks. These students said in the interview that they had thought about the fatigue of the runner but still provided a solution based on a constant 100 m pace since if they had based their solution on the realistic assumption that the runner would not be able to maintain the same speed, then the task would be unsolvable (the example provided to exemplify an EA(T) response in Section 2.4.1 is a representative example of such responses). To a follow-up question about if they were sure that their solution would be considered correct some students said "no" but that they did not think they had any other possibilities. Some other students said that they were confident that their solution was correct. As one student put it: "they don't want to know how long time it will take, they want to know if I can calculate 10×100." Some students justified their solution by pointing to earlier experience. They claimed that in word problems one should consider exactly what is written in the task and nothing else. If they were supposed to make other considerations it would be stated in the task. In addition, a couple of students claimed that there were tasks in their textbooks in which unrealistic assumptions, such as a 100 m pace for 10,000 m, were stated as a precondition in the task. This experience was also shared by some of the teachers.

The students' beliefs about mathematics and mathematical word problem solving on the one hand and their solution strategies on the other hand are not independent of each other. In fact, the conscious or unconscious choice of strategy is affected by the students' beliefs about mathematical task solving. However, in this analysis of the reasons for providing 'unrealistic' solutions the category "beliefs" only comprises the beliefs that are not directly responsible for the students' choice of general solution strategy. They include beliefs about the appropriateness of solutions and answers to mathematical school tasks in general and to word problems in particular. They also include beliefs about the relation between mathematical word problem solving in school and the use of mathematics in real life outside school. The beliefs in this category are mostly in play when the students' have made some calculations and are considering the appropriateness of the potential answer or solution, or when they are about to make some calculations and are considering the appropriateness of these calculations as a solution method. Examples of beliefs included in this category have been given above in this subsection. Other beliefs, for example beliefs about what mathematical task solving is all about, that may be one of the causes for the students' choice of general solution strategy are not included in this category. These beliefs, and other causes for the students' superficial solution strategies, are mostly not directly

identified in this study. Instead the focus is on their more directly observable outcome – the students' solution strategies (exemplified above). This means that when a student uses a general solution strategy that is characterised by its lack of a sufficiently thorough analysis of the problem situation (no matter why this strategy is chosen) and this strategy is judged to be the cause for the student giving an 'unrealistic' answer then the reason for the 'unrealistic' answer is categorised as the use of a superficial solution strategy. With these meanings of the two described categories of reasons it has been possible to sort the reasons to one of the categories (or to one of the additional categories described below).

In addition to the two most common reasons for not providing written answers consistent with the realities of the task context, some students did not provide realistic answers due to the lack of the necessary real world knowledge and because of communication problems. The former seems to have been a hindrance only in the running tasks and not for more than a few students. The latter reason occurred for solutions to the bus tasks. A few students giving the answer 7.5 buses said that they meant 7 buses with students on every seat and 1 bus that was only half filled with students.

3.3 Research question 1 (Q1) revisited. Reasons for the impact of authenticity on the proportion of 'unrealistic' solutions

The reasons for the higher proportions of 'realistic' answers to the more authentic task variants seem to involve both the students' solution strategies and beliefs. The students confronted with the more authentic word problems seem to have engaged more often in thorough task analyses than the other students. A higher proportion of the students working with the more authentic versions observed the problematic situation with half buses, leftover plank parts and keeping maximum speed for a long time (see the proportions of Realistic considerations (RC) in Table 3). One part of the difference in the plank tasks consists of the 14 students who gave the answer 10 planks to the more authentic task variant but described in the interview that they were thinking about gluing or nailing the planks together (three of the students that worked with the less authentic task variant provided this combination of answer and explanation). That is, they really analyzed and significantly engaged in the situation described in the task. However, Table 3 shows that there is a difference between the three word problem categories concerning the effect that a tendency to use superficial solutions strategies has on the proportion of 'unrealistic' responses. The proportion of students that did not notice the task specific realistic feature to consider was much higher in the Category 2 and Category 3 word problems (such as the plank tasks and the running tasks respectively) than in the Category 1 word problems such as the bus tasks (see the proportions of RC in Table 3). The part of the 'real' situation that needs to be considered for a successful solution appears to be less visible in these tasks.

The authenticity of the tasks also seems to have affected the participating students' willingness to disregard their classroom *didactical contracts* (Brousseau 1997) in the interpretation of the meaning of the task context and the rules for the task solving. A comparison of the proportions of Realistic practice (RP) and Realistic considerations (RC) in Table 3 shows that in addition to the decisive part of making the relevant real world observation, the students' beliefs about mathematical word problem solving played a significant role in the running tasks, but not in the Category 2 word problems such as the plank tasks. The reason for this seems to be the nature of the running tasks. If considerations are taken of the inability of a runner to maintain a 100 m maximum pace for 10,000 m, then the answer to the word problem will not be a single number (or an exact time) that is possible to calculate only from the information given in the task and without making estimations. The

plausibility of such solutions did not seem to be included in many of the students' beliefs about word problem solving: To the less authentic task variant of the running task none of the 20 students who noticed the problem with keeping the 100 m speed gave a written solution in accordance with this consideration. But 17 of the 33 students who noticed this problem in the more authentic task variant did. Eight of these students said in the interview that they refrained from giving an answer to the task since they could not accept a solution that would be inconsistent with the real life running situation described, and at the same time they said that they did not believe they were allowed to make an estimation of the runner's 10,000 m time. They claimed that because of fatigue of the runner it is not possible to know what time the runner would have only on the basis of the given data. In addition, one other student wrote as his answer that the time is not possible to know and another student wrote that "it would take much longer time," explaining in the interview that because of fatigue it was not possible to know the exact time. Of the remaining seven students, four of them did not make any calculations and estimated Greens time on the basis of Gebrselassies time and whom they thought would win. The three remaining students based their answers on calculations. Two of them calculated the time it would take Green to run the 10,000 m with his 100 m pace and then added a couple of minutes because he would get tired. The third student calculated Gebrselassie's 100 m mean time, added a few seconds and then calculated a 10,000 m time based on this new 100 m mean time. A few of these estimations would not be considered realistic in the perspective of a person knowledgeable of such running realities, but they were estimations based on these students' perceptions of reality and therefore the solutions were classified as Realistic practice. Very few of the students that noticed the problematic feature of half buses in the more authentic task variant of the bus task provided 'unrealistic' answers to it. The students solving the less authentic task variant more often did so.

The two variants of the plank task and the two variants of the running task differed in a way the other four pairs of tasks did not. The two more authentic variants of these tasks included numbers that were not included in the less authentic task variants. In the plank task this number (thirteen) is of no use in an appropriate solution to the task. However, it may have contributed to the more thorough analysis of, and the meaning assigned to, the task context but this further information may also have made the task analysis more difficult for some students. Indeed, compared to the less authentic task variant, seven more students refrained from giving an answer (categorized as NA−) or providing a solution of the kind 2.5×13 (categorized as OA−) to the more authentic task variant (in the other Category 2 task, the riding task, it was the other way around). In the more authentic running task the provided reference time could be seen by the students as a time to use for a rough estimation of the reasonableness of their answer or even as a base for their solution. However, only five of the students used this information in their solution. Four of them guessed Green's 10,000 m time based on the reference time and one student used the reference time in his calculations of a reasonable 100 m mean time for Maurice Green.

Some brief comments about the three tasks not focused in the paper will conclude the Results section. Both the balloon tasks and the riding tasks showed similar response patterns as the other tasks belonging to the same category of word problems, the bus tasks and the plank tasks respectively (see Table 3). The two Category 3 tasks, though, differed in the proportions of RP and RC. A possible reason for this seems to be the beliefs about pictures in word problems, which the students revealed in their work with the bowl tasks. The picture was considered to be subordinated to the text in a task and the text was in this case understood as providing sufficient information. Thus, the students did not pay much attention to the picture of the bowl and therefore did not detect their 'unrealistic' responses. The increase in authenticity was not enough to change this understanding of the task.

Another possible reason for the low proportions of RC and RP could be that the students had no experience or mental structure for using the picture to think non-linearly. However, this explanation does not fit well with a follow-up study that was carried out in which the form of the bowl did not influence the proportions of RP and RC, but that showed that when measures of the height and radii were provided in one of the figures used in the study (consisting of two cylindrical parts with the same height but with different radii and the smaller part being on the top of the bigger part) the proportions of solutions classified as RP and RC were much higher than without these measures in the same figure. This shows that (1) it was necessary that another feature of the figure was changed for the students to analyze it more carefully and (2) the students could use (at least) that figure for thinking about a non-linear model for the solution. Further support for the first explanation (the students' beliefs about pictures) are students saying in the interview that they did not put much attention to the picture, they did not do so because they thought the text gave them sufficient information, the purpose of the picture being there was decoration or for clarification that the water fell into a bowl, and that they had the knowledge that the depth of the water would not rise with constant speed (for a fuller discussion of the responses to the balloon tasks, the riding tasks and the bowl tasks, see Palm 2002).

4 Discussion

This study provides evidence about the reasons that students provide solutions that are not consistent with the realities of the 'real' situations described in the word problems used in the investigation. The study shows that not all of the 'unrealistic' answers provided by the students in the study stem from total 'suspension of sense-making.' In many of the solution attempts ending up in 'unrealistic' answers the students really had made use of their personal understanding of the task situation, although their understanding, which is consistent with their experience of the socio-cultural context of the school situation, is different from the understanding of the situation that many researchers and teachers have. Nevertheless, the frequently used superficial solution strategies that do not include a thorough analysis of the situation, or the appropriateness of the mathematical model used to model it, are not efficient for solving non-routine tasks, regardless of whether the tasks are word problems with an out-of-school task context or pure mathematical problems. Furthermore, since the reasoning that leads to the chosen mathematical calculations does not involve much consideration of the properties of the mathematical concepts that are involved in the reasoning, these strategies are not efficient for the development of conceptual understanding of these concepts. On the contrary, these strategies share important properties with the superficial solution strategies described by Lithner (2000) as a major reason for undergraduate students' difficulties with mathematics. In addition, when the intention with word problems is to practice real life task solving, then dismissing important conditions of real life, which were consciously made by students due to their beliefs about word problem solving, is not an efficient practice since it misses out important aspects of task solving beyond school. Thus, this study shows that the underlying reasons for the participating students' 'unrealistic' responses are serious (and may so be for the students in many of the other studies using similar tasks, see Verschaffel et al. 2000). It seems that the didactical contracts predominant in many of these students' classrooms do not foster adequate beliefs about mathematical task solving, and the idea that a thorough task analysis is not necessary for mathematical task solving is a major obstacle both to the learning and to the application of mathematics.

Several authors (e.g. Maier 1991) have suggested that important factors of the learning environment that are responsible for students' 'non-realistic' behavior to word problems are the task features that make the tasks being perceived as having little in common with those faced in life and seen as "school problems, coated with a thin veneer of 'real world' associations" (Maier 1991). There is, however, a need for empirical evidence for the claims about the impact of authenticity (in the sense of the term used in this paper) on students' responses to such word problems. This study contributes such evidence. The study shows how the characteristics of the tasks, even when they only differ as a result of a modification of the description of the task context, can affect the students' engagement in the task context as well as the willingness to disregard the rules of their didactical contracts. Considering the beliefs and solution strategies of the students' participating in this study, and the insights gained from the study about the effect that authenticity can have on the meaning and the analysis the students make of word problems, the results of the study support proposals that authenticity is a factor to be seriously considered in the development and choice of word problems to be used in the classroom. In addition, the described task dependency of the requirements on the students' beliefs and solution strategies (as was seen in the comparison of the responses to the three different categories of word problems) would be useful when developing learning tasks that fit a particular purpose of supporting students development of their beliefs and/or solution strategies. The characteristics of the tasks used for helping students learn mathematics should be clearly connected to the desired outcome of the learning activity. Given the students' focus on superficial solution strategies, shown in this paper and elsewhere (e.g. Lithner 2000), and the many descriptions of learning environments consisting mainly of stereotyped tasks (e.g. Reusser and Stebler 1997a; Verschaffel et al. 2000) an improvement of this purposeful choice of learning tasks would be of significant importance.

It cannot be concluded from this study which of the changes in the tasks that contributed the most to the higher proportion of 'realistic' solutions to the more authentic task variants. However, informal discussions with the students in the study indicated that the formulation of the question in the more authentic running task was important in opening up for a more thorough task analysis since it was not suggesting, as the question in the less authentic variant, that a specific time based on a linear model of the situation was expected. Furthermore, in all of the less authentic tasks the students are given very little information about the circumstances of the situation, including the purpose of the question. This leads to an unspecified and somewhat unclear situation. The students are also assigned the role of a passive outsider, not belonging to the real life situation in any way. It may be that when given so little information about the situation in which the question is posed, this may be indicating to the students that the situation described in the task is not of major concern for the task solving. This may lead to the interpretation that the important context of the task solving is the social context of a school situation. Thus, the students comply with those implicit rules of the didactical contract that have been experienced to be valid for school word problem solving. A thorough analysis of the task contexts, considering the mathematical objects and their relation to the reality indicated in the tasks, may not be a part of these rules.

The conclusions in this study are based on interview data and the students' written solutions and comments. Generally, interpretations of most kinds of data are impaired with some amount of uncertainty. In this study there is the risk that the students' descriptions of the reasoning carried out during the task solving included thoughts that arose as a result of the interview questions. For example, the students' may (1) not be able to distinguish between the thoughts that came up during these two occasions and (2) want to appear as high performing students and therefore add new thinking to the description of the reasoning

that took place during the task solving. Verschaffel and De Corte (1997) reported that during a classroom discussion (following the task solving and a group discussion) some of the students in their study invented far-fetched contextual conditions in which their answer would still hold. Some of the interview procedures in this study were chosen in order to minimise the risk of misinterpretations of the interview data due to the two risk factors mentioned above. To decrease the risk of point (1) the students were made aware of this risk and it was emphasised that the students should try to remember the thinking and reasoning carried out during the task solving and separate that from the thoughts that came up in the interview. When considering point (2) in the planning of the interview procedure it was assumed that the risk of reconstructions such as the one reported in Verschaffel and De Corte (1997) would be more substantial if the students' incentive for making them is more substantial. For reasons of achieving or maintaining high status it may, for example, be important for the students that their classmates perceive that they have produced an appropriate answer. To minimize such situations the interviews were done individually, and it was explicitly pointed out that the purpose of the interview was not to judge the quality of the student's work but to help the researcher in understanding how students think when solving mathematical school tasks. In addition, the interview situation took place in a relaxed environment in order not to put unnecessary pressure on the students.

Restricting the data collection to written information would also have imposed interpretation difficulties. Such a method would miss out useful and necessary information, which is one of the methodological limitations that Verschaffel and colleagues attributed to their 1994 study (Verschaffel et al. 1994). Relying only on written data may underestimate the number of solution processes in which there was activation of real-world knowledge since some of these reflections may not have been written down. Indeed, the results of the present study show that the use of only written data would have produced such an underestimation. Thus, one of the strengths with the methodology used in this study is the use of several sources of information.

Additional research about the impact of authenticity on students' word problem solving could be useful in that it might provide further insights about students' selection of solution strategies. It could also be of value to compare the results of this study with the outcome of a study including tasks that simulate more of the aspects in the framework with high fidelity. Such a study could explore the hypothesis that further increase in authenticity would result in larger proportions of 'realistic' solutions and more thorough reflections of the applicability of the mathematical models used for the solutions. Such more authentic tasks could more easily be developed if, for example, the constraints of increasing the task authenticity of existing word problems solely by modifying the task text were abolished. Another possibility could be to start the task development from real examples from daily life, even if this would make it more difficult to compare the results of that study with the results of earlier studies.

Acknowledgements The author wishes to thank Hans Wallin and Johan Lithner for their valuable comments on an earlier version of this paper.

References

Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Dordrecht: Kluwer.

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

Cooper, B., & Dunne, M. (2000). *Assessing children's mathematical knowledge: Social class, sex and problem-solving*. Buckingham: Open University Press.
Cooper, B., & Harries, T. (2002). Children's responses to contrasting 'realistic' mathematics problems: Just how realistic are children ready to be? *Educational Studies in Mathematics, 49*, 1–23.
Cooper, B., & Harries, T. (2003). Children's use of realistic considerations in problem solving: Some English evidence. *Journal of Mathematical Behavior, 22*, 451–465.
Davis, R. B. (1989). The culture of mathematics and the culture of schools. *Journal of Mathematical Behavior, 8*, 143–160.
DeFranco, T. C., & Curcio, F. R. (1997). A division problem with remainder embedded across two contexts: Children's solutions in restrictive vs. real-world settings. *Focus on Learning Problems in Mathematics, 19*(2), 58–72.
Fitzpatrick, R., & Morrison, E. J. (1971). Performance and product evaluation. In R. L. Thorndike (Ed.), *Educational measurement,* (2nd ed.) (pp. 237–270). Washington, DC: American Council on Education.
Greer, B. (1993). The modeling perspective on wor(l)d problems. *Journal of Mathematical Behavior, 12*, 239–250.
Greer, B. (1997). Modelling reality in mathematics classrooms: The case of word problems. *Learning and Instruction, 7*, 293–397.
Inoue, N. (2005). The realistic reasons behind unrealistic solutions: The role of interpretive activity in word problem solving. *Learning and Instruction, 15*, 69–83.
Lithner, J. (2000). Mathematical reasoning in task solving. *Educational Studies in Mathematics, 41*, 165–190.
Maier, E. (1991). Folk mathematics. In M. Harris (Ed.), *School, mathematics and work*. Basingstoke: Falmer Press.
Palm, T. (2002). *The realism of mathematical school tasks – Features and consequences*, Doctoral dissertation, Department of Mathematics, Umeå University, Sweden.
Palm, T. (2006). Word problems as simulations of real-world situations: A proposed framework. *For the Learning of Mathematics, 26*(1), 42–47.
Reusser, K., & Stebler, R. (1997a). Every word problem has a solution – The social rationality of mathematical modeling in schools. *Learning and Instruction, 7*(4), 309–327.
Reusser, K., & Stebler, R. (1997b). Realistic mathematical modeling through the solving of performance tasks. In Seventh European Conference on Learning and Instruction, Athens, Greece.
Säljö, R., & Wyndhamn, J. (1993). Solving everyday problems in the formal setting: An empirical study of the school as context for thought. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 327–342). Cambridge: 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*(5), 577–601.
Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 22*(4), 273–294.
Verschaffel, L., De Corte, E., & Lasure, S. (1999). Children's conceptions about the role of real-world knowledge in mathematical modeling of school word problems. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 175–189). Oxford: Elsevier.
Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, The Netherlands: Swets & Zeitlinger.
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*(4), 329–338.