Gender Aspects of Sense Making In Word Problem Solving

Torulf Palm
Umeå Mathematics Education Research Centre (UMERC),
Department of Mathematics, Technology, and Science Education, Umeå University, Sweden
torulf.palm@math.umu.se

Peter Nyström
Umeå Mathematics Education Research Centre (UMERC),
Department of Educational Measurement, Umeå University, Sweden
peter.nystrom@edmeas.umu.se

Abstract
This study focuses on gender differences in the extent to which students take real-world considerations into account when working with word problems in mathematics. Previous studies have found that students have a tendency to neglect an appropriate use of real-world knowledge in their word-problem solving, which leads to solutions that are inconsistent with the ‘real’ situations described in the tasks. Research has also shown that the authenticity of the word problems can influence students’ activation of their knowledge of the ‘real’ situations described in the tasks, as well as their use of this real-world knowledge in order to provide ‘realistic’ solutions. The study reported here investigates whether there are gender differences in the students’ activation and use of real-world knowledge when working with word problems in mathematics. In addition, it investigates whether task authenticity influences boys and girls differently with respect to these real-world considerations. The results show that even though some of the tasks used in the study affected boys and girls differently, across all tasks and students no evidence of gender differences with respect to real-world considerations were found.

Keywords: Authenticity, gender, mathematical modelling, sense making, word problems

1. Introduction

1.1. Previous research

1.1.1. Sense making in students’ work with word problems

The study presented in this paper is about word problems. However, the term has been given different meanings in different publications and sometimes its intended meaning is not clear. In this paper the term “word problems” is used in accordance with the following description provided by Verschaffel, Greer & De Corte, (2000, p. ix): “textual descriptions of situations assumed to be comprehensible to the reader, within which mathematical questions can be contextualized”. Verschaffel et al. note that word problems “provide, in convenient form, a possible link between abstractions of pure mathematics and its application to real-world phenomena” (p. ix). This is a broad definition that includes pure mathematical tasks “dressed up” in a real-world context that for their solution merely require that the students’ “undress” these tasks and solve them. It also includes tasks requiring that students’ be involved in the full mathematical modelling cycle.

During the last three decades a large number of studies have investigated students’ work with word problems of the kind 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 (for a comprehensive overview of these studies see Verschaffel, Greer & De Corte, 2000; Verschaffel, Greer, Van Dooren & Mukhopadhyay, 2009 in press). An example of such a word problem is the following: “John’s best time to run 100 m is 17 sec. How long will it take him to run 1 km?” (Greer, 1993). Clearly, a solution that does not consider the fatigue of the runner will not be consistent with the ‘real’ situation described in the task. The collective outcome of these studies is that 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.

The reasons for the ‘unrealistic’ answers have been investigated and identified in a few studies. Based on the interpretations of the interviews and the written responses of 161 fifth-grade students’ Palm (2008) concluded that for these students’ there were two main, interconnected, reasons for providing solutions that were inconsistent with the situations described in the tasks. These reasons are similar to the ones identified in the interview with 15 students’ by Hidalgo (1997, in Verschaffel et al., 2000, p. 26) and a study on undergraduate students’ by Inoue (2005). The first reason was the frequent use of solution strategies that can be characterized as superficial. These solution strategies do not involve a careful analysis of the task situations but tend to focus on the numbers given in the task. Thus, the strategies include neither a thorough reflection of the applicability of the mathematical models used for the solutions nor an evaluation of the answers in relation to the ‘real’ situations described in the tasks. The second 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. These beliefs do not include the requirement that school mathematics and real life outside school must be consistent. On the contrary, they 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 (Palm, 2008).

The possibilities of influencing students’ work with word problems through modifications of the tasks and test conditions have also been investigated. Studies show that the tendency to provide ‘unrealistic’ solutions to word problems seems to be strong and cannot be easily overcome by hints (Greer, 1997; Reusser & Stebler 1997a; Verschaffel, De Corte & Lasure, 1999; Yoshida, Verschaffel & De Corte, 1997). The effect of different working conditions on the students’ behaviour 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’ modelling behaviour. Examples of such changes are arrangements for the students’ to make actual telephone calls when the task is to order the required number of buses for a school trip (practically arranged by using two additional rooms, two telephones, and an unknown person answering at the other end) (DeFranco & Curcio, 1997), providing students’ with concrete materials such as planks, a saw, and a meter-stick (Reusser & Stebler, 1997b), and having the students’ work with tasks in a social science class setting instead of in the setting of a mathematics lesson (Säljö & Wyndhamn, 1993).

However, the mere concordance between the school word problems and the corresponding out-of-school task situations described in the tasks (task authenticity) can affect both the students’ engagement in the task context and the extent to which they require that their solutions make sense in the ‘real-life’ situations described in the tasks. Palm (2008) showed that an increase in task authenticity, even when accomplished solely by a modification of the task text, increased students’ tendencies to use their real world knowledge effectively in the solutions to word problems (see the section on Authenticity below for a more detailed description of the meaning of the concept of task authenticity). The students’ confronted with more authentic versions of word problems more often reflected over the problematic feature of the ‘real’ situations described in the tasks and more often provided solutions that were consistent with the realities of the ‘real’ situations.

Thus, this line of research has identified students’ tendencies to neglect an appropriate use of real-world knowledge and also reasons for this phenomenon. The phenomenon is important in itself but carries even more significance since the main reasons for providing ‘unrealistic’ answers to the word problems are solution strategies that do not involve considerations of the properties of the mathematical objects involved and a view of mathematics as disconnected to the world outside school. As a consequence, the measures that can be taken to influence this tendency, students’ beliefs and ways to deal with mathematical word problems are of great relevance. The study presented in this paper will investigate further the influence of task authenticity with a focus on potential gender differences.
1.1.2. Word problems and gender

In addition to the general issue of students’ problem-solving behaviour in relation to word problems, research has focused on differences between categories of students’ in this respect. Research with gender perspectives on word problems, contextualization, and realistic considerations in solving tasks in mathematics has identified differences between male and female students’ and also tried to understand these differences. Cooper & Dunne (2000) conclude that there is a range of research on problem-solving claiming that gender is associated with “different typical approaches to tackling problems, and some of this work at least is directly relevant to the issue of responses to ‘realistic’ mathematics questions” (p. 35).

Cooper & Dunne (2000) focused on the differences between contextual and non-contextual tasks in mathematics. They studied the performance of 10-11 year-old students’ and found that differences between students’ performance on items with an out-of-school context (‘realistic’) and items involving only mathematical objects (‘esoteric’) were related to social class, but also to gender. The ratio of percent correct on ‘realistic’ items to percent correct on ‘esoteric’ items was higher for boys’ than for girls’, indicating that relative to their performance on ‘esoteric’ items, boys’ performed better than girls’ on ‘realistic’ items.

Similar results were reported by Gipps & Murphy (1994) who summarized studies of examination results for older students’ in the UK (age 16) indicating that contextualized questions gave a slight advantage to boys’, whereas questions which favoured girls’ tended to be non-contextualized. In a more recent review of research on physics education Murphy & Whitelegg (2006b) conclude that girls’ are more likely than boys’ to give value to and “notice” the social context in which tasks are posed in defining a problem.

In addition to differences attributed to the degree of contextualization, researchers have explored the impact of differences in the kind of contextualization found in mathematics and related areas, primarily differences between (stereotypical) male and female contexts. Zohar & Gershikov (2007) conclude that a common trend emerges from the literature, repeatedly showing that the context of tasks in mathematics affects girls’ more than boys’. The empirical part of their study concerned the way 523 children from kindergarten (age 5) to grade six (age 10) interacted with tasks set in contexts that were stereotypically male, stereotypically female, or neutral. In the study girls’ and boys’ performed equally well on tasks set in a neutral context but girls’ performed significantly less well than boys’ in stereotypically boys’ contexts. In addition, some stereotypically girls’ contexts had a negative effect on girls’ mathematical performance. The way girls’ were affected by gendered contexts varied partly with their age.

Murphy & Whitelegg (2006a) describe how girls’ can be negatively affected by a male context by pointing to research supporting that some girls’ may not, even if they can do the tasks, respond to a data-handling question (in science or mathematics) when the data are about cars or about workshop parts, i.e. with a content that can be categorized as ‘masculine’. Similarly some boys’ avoid questions that emphasize ‘feminine’ content. They conclude that what the task is about can create a barrier to students’ access and engagement in the task, regardless if the students’ can do the actual task or not.

An example of how stereotypically female contexts can affect girls’ negatively is described by Boaler (1994). In a small-scale study involving 50 students’ from two schools, she exposed students’ to word-problems with a stereotypical gender characteristic. Another characteristic of some of these tasks was that they seemed to invite some students’ to engage deeply in the context and use more knowledge of this real-world situation than was intended, and such solutions were not considered appropriate (also called ‘over-involvement’). The results showed that girls’ did less well in the ‘fashion task’, which was intended to be set in a more feminine context, than on a similar problem set in an abstract context or the stereotypical masculine context of football. In one of the schools in the study the students’ were discouraged from taking account of real-world variables in school problems. Boaler’s interpretation is that boys’ were more successful in employing this strategy and were more able to focus only upon the numbers in the task, which was what was expected of them. As many as two-thirds of the girls’ in her study used their common sense as well as their mathematical knowledge and were penalised for doing so.

Boaler concluded that her results can be understood in terms of the assertion of Murphy (1990) and others that boys’ are less likely to be affected by context, and hypothesizes that girls’ were
unaffected by male contexts in her study because they were less interested in these contexts and thereby able to distance themselves from them to a greater extent.

In addition to interest, several mechanisms have been suggested as explanations to observed gender differences in relation to word problems. Research has shown that the context of a task plays a critical role in determining what task pupils perceive and what response they consider appropriate. How contexts are defined by assessors and interpreted by pupils does require more thorough consideration. (Gipps & Murphy, 1994) (p. 226)

After reading a word problem, students’ can construct different perceptions of what they are expected to do, and the way the task is designed can be crucial to that process. Murphy & Whitelegg (2006b) argue that research has shown that differences between what girls’ and boys’ have learned is relevant, and valuably affects the problems they perceive. They refer to national surveys as well as case-studies in classrooms presenting evidence that girls’, more than boys’, tend to value the context of science tasks and take this in consideration when constructing meaning in the task. Based on research on design and technology, Murphy & Whitelegg (2006b) also argue that situations always have more information than people notice, and in the process of selecting information people take note of what has salience for them. The selection process is determined by familiarity and familiarity is determined by experience and our purposes for engaging in activities. Gender is relevant here since children’s gendered socialisation influences what becomes familiar to them.

The idea of socially constructed gendered domains and gendered beliefs are potentially powerful explanatory factors that can be useful for understanding the observed gender differences. Gendered domains have been used to denote the different real-world situations which are perceived as typical male or female contexts (Browne & Ross, 1991). The construction of gender imposes classifications of contexts as either male or female, and this classification affects students’ perceived competence; male students’ perceive that they are more competent in what they interpret as male situations and female students’ perceive that they are more competent in what they interpret as female situations. Murphy & Whitelegg (2006a) suggest that these gendered domains might be the cause of differences in students’ perception of barriers in word problems.

The concept of gendered domains has also been used to capture beliefs about different school subjects, and research has shown that students’ view mathematics as a male domain in some respects, but as female and gender-neutral domain in other respects (see e.g. Forgasz, Leder, & Gardner, 1999). Brandell, Leder, & Nyström (2007) found that one of the most typical gender-related beliefs about mathematics in Sweden was that girls’, more than boys’, expressed the importance of understanding mathematics. Boaler found similar views in her study of teaching and its impact on student learning in two secondary schools in England (Boaler, 1997, 2002). In the school she describes as traditional (‘Amber hill’), the most prominent gender difference was related to girls’ emphasis on understanding mathematics contrasted by boys’ view of mathematics as a competition or a game. She concludes that:

As a result of a number of different data sources, I became convinced that it was this desire to understand, rather than any difference in understanding, that differentiated some of the girls’ from the boys’ (Boaler, 2002, p. 140).

The girls’ at Amber hill also had preferences towards working in their own pace, working in groups, and working on open-ended projects, and Boaler argues that these preferences were caused by girls’ feeling able to understand their work in these modes of teaching.

Gilligan’s (1982) notion of ‘separate’ and ‘connected’ knowledge, and the suggestion from researchers that women and men have differential preferences for ways of knowing and subsequent ways of working, is another conceptualisation of gendered domains that has been suggested as an explanation for the differences between how girls’ and boys’ react to word problems (Boaler, 1997). This theory hypothesizes that women tend to value ‘connected’ knowledge (involving intuition, creativity and experience), whilst men tend to value ‘separate’ knowledge (characterized by logic, rigour and abstraction). (See also Marrs & Benton (2008) and Schommer-Aikins & Easter (2006)).

1.2. Aims

As described above there is a substantial body of research showing 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. There are also many studies that have investigated gender differences in students’ mathematical task solving. But the research literature seems to lack studies focusing on gender differences regarding students’ tendencies to neglect an appropriate use of real-world knowledge, when such a use is required for providing solutions that are consistent with the ‘real’ situations described in the tasks. Following the research direction summarized in Verschaffel et al. (2000) the study presented here focuses on tasks that do require the use of real-world knowledge, and to which students’ have been shown to neglect such a use (in contrast to the studies by Cooper & Dunne (2000) and Boaler (1994) which have investigated students’ responses to tasks where students’ are not intended to use too much real-world knowledge of the task context). The study is intended to contribute to the literature on gender and word problem solving by investigating (1) potential gender differences in students’ tendencies to neglect an appropriate use of their real-world knowledge in word problem solving and (2) by investigating the effect of task authenticity on this knowledge use. These aims can be stated more precisely as the following two research questions:

Q1: Do the 5th-grade girls’, in the Swedish city used for the present study, more often than the boys’ activate their real-world knowledge of the ‘real’ situations described in the word problems, and do they more often let this knowledge affect their written responses?

Q2: Concerning the activation and use of real-world knowledge in their task solving: Do the girls’ benefit more than the boys’ from more authentic word problems?

1.3. Authenticity

There is no consensus in the mathematics community of which term to use for a 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 and real life tasks), and in addition many different meanings have been attached to each one of them. (The issue of the meaning of authenticity of application tasks is addressed in for example Galbraith, 2007, and Niss, Blum & Galbraith, 2007, p. 19. For a review of different meanings attached to each one of these terms see Palm, 2002). In this paper the term authenticity is used to denote the concordance between a school task and a real life task situation, and the term is used in relation to the out-of-school situation that is described in the task, the task context. Thus, authenticity concerns contextualized tasks and an authentic school task is one that well emulates a real life task situation. The underlying idea for the more precise description of such tasks lies in the following quotation by Fitzpatrick & 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, 2009 in press) includes a number of such aspects of real life task situations which simulations have been argued to be important for the students’ possibilities to engage 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 are simulated fairly well, enabling students’ to solve the task 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 these 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. 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, 2009 in press).
Gender Aspects of Sense Making In Word Problem Solving

**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 has taken place or has 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 statement describing 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 occur 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.

2. Method

2.1. Procedure and participants

The method is similar to the one used and described in Palm (2008). 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, and all of the selected classes accepted to participate in the study. The selection process was carried out using a list of all fifth-grade classes in the city, from which eight classes was randomly selected. 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. The sample included about the same number of boys’ and girls’. Based on interviews with the teachers it appears that the predominant learning activity in the mathematics classroom is individual work with exercises in the textbook. The students’ in all classes had been given 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 found in the literature (e.g., Verschaffel, De Corte & Lasure, 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 tests were administrated 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 (WP 2-WP7), 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. Five of these were near replications of tasks used in the studies by, for example, Reusser & 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. 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 authenticity framework. 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 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 two versions of these tasks was achieved.

<table>
<thead>
<tr>
<th>Table 1. The word problems in the study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Less authentic version</strong></td>
</tr>
<tr>
<td>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?</td>
</tr>
</tbody>
</table>
360 students’ shall go by bus on a school trip. Each bus can hold 48 students’. How many buses are required? (Carpenter, Lindquist, Matthews, & Silver, 1983)

All students’ in the school shall the 15:th of May go on a school trip together. You and the other organizing students’ have decided that everyone shall go by bus, and that you shall order the buses. You have seen in the student namelists 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: | | |

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)

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?

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?

You are going on a camp for 4 days, but do also want to ride. Your dad sees in the camp papers that you have 45 minutes free time each day, and that horses can be rent for tours round a path in the woods that takes 10 minutes. To know how much money you shall 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?

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)

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 on the time 10.00 sec. The next race you watch is 10 000 m, which is won by Haile Gebrselassie on the time 26 min. and 5 sec. What do you answer when your friend asks you: What time do you think it would take Maurice Green to run 10 000 meters (= 1 Swedish mile)?

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)

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?

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. To limit the possible problems for students’ with reading difficulties, attempts have been made to avoid unnecessarily 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 too much
decrease in the fidelity of the simulations of the aspect Language use. But even so, one unavoidable characteristic of the more authentic tasks is that they contain longer texts, and thus may be more demanding to read. However, the relationship between text length and readability in mathematics tasks is complex and longer texts may sometimes be easier to read (see e.g. Nyström, 2008).

The target tasks represent three different types of tasks included in the three above-mentioned studies. The following is a discussion of each task category, focusing on the aspects that are considered to be simulated with different fidelity in the different task variants. One of the tasks in each category is used as an example.

2.2.1. Category 1 word problems

A successful solution to WP 2-3 requires an interpretation of a remainder, and this type of task 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 it has to be a whole number of buses since buses do not function well in halves. In these tasks the most important aspect of authenticity 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 order sheet.

2.2.2. Category 2 word problems

WP 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-meter planks can be sawn out of the four 2.5-meter planks. The less authentic version of this task pictures that someone has already bought four 2.5-meter planks without knowing how many planks of desired length could be obtained from them, which is not a very likely scenario. 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.

2.2.3. Category 3 word problems

WP 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 task have to take into account that a runner can not keep the same speed for 10 000 m as he can in 100 m, and in this task the question is 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 implies, 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.

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 aim of the interviews that pertain to the specific study presented in this paper was to investigate if realistic considerations of the task contexts had been made in their task solving, but not put on paper. 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 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 categorized in two different ways on the basis of the written solutions, the written comments given in the commentary area, and the discussion in the interviews. A total of 51 solutions out of 966 were not possible to categorize using the criteria described below because the students’ had not come far enough in a solution to really have to make the crucial task-specific realistic consideration. The exclusion of these solutions only influenced the proportions of RP (see below) by two units of percent for both the less authentic and the more authentic tasks. The categorization was made by one person (the first author), and the validity of the categorization was supported by high transparency through clear descriptions of the classification categories and ample examples of classifications of common student solutions. For a more detailed description of the categorization see Palm (2008).

The first kind of categorization explores whether or not the students’ provided solutions that are consistent with the ‘real’ situations described in the word problems. Students’ work with the tasks was categorized as Realistic practice (RP) or Non-realistic practice (Non-RP). The RP-category consists of all answers that follow from a solution process involving an appropriate use of the students' knowledge about these ‘real’ situations. This includes written answers that are consistent with the situation described in the task, such as the answer 8 buses and not 7.5 buses. The RP category also includes responses where students’ did not write any answer but showed in the interview that 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. An example of such a solution 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 Non-RP-category consists of all responses not categorized as RP. This includes 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 without considering the realities of
the situations described in the tasks. Examples of such answers are answers to the bus tasks given as a fraction of buses, the answer 10 planks to the plank tasks, and an answer to the running tasks based on the possibilities that the runner can keep a 100 m pace for 10 000 m.

In the second kind of categorization students’ work was analyzed in a slightly different way. Here the focus was on whether or not students’ showed an activation of relevant knowledge of the ‘real’ situations, regardless whether or not this activation affected the students’ written responses. The data were categorized into Realistic considerations (RC) or Non-realistic considerations (Non-RC). The \textit{RC-category} 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, and the \textit{Non-RC-category} consists of all other responses. The RC-category is related to the previously described RP-category as it consists of all responses categorized as RP, but also those where the interviews or the students’ written comments in the commentary area 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 such a response to one of the running tasks is the written answer 1000 s from a student who in the interview said that “I 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”. This response was categorized as RC but not RP, since the realistic consideration made did not affect the student’s written solution.

3. Results

The design of the study gave each student six opportunities to make realistic considerations (RC), and to let these considerations affect their written solutions (Realistic Practice, RP). The proportions of student responses categorized as RC and RP are shown in Table 1.

<table>
<thead>
<tr>
<th>Task variant</th>
<th>Girls’ RC</th>
<th>Boys’ RC</th>
<th>Girls’ RP</th>
<th>Boys’ RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less authentic</td>
<td>34</td>
<td>31</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>More authentic</td>
<td>53</td>
<td>46</td>
<td>57</td>
<td>50</td>
</tr>
<tr>
<td>All tasks</td>
<td>43</td>
<td>38</td>
<td>49</td>
<td>46</td>
</tr>
</tbody>
</table>

As can be seen from Table 1, the girls’ participating in the study fairly consistently more often than the boys’ activated their real-world knowledge of the ‘real’ situations described in the tasks as well as let this knowledge affect their written responses. This holds for both the less authentic and the more authentic task variants. However the differences are small, and a test of statistical inference regarding the frequencies of RC and RP does not support a conclusion that the observed differences are a representation of differences in the population (p > 0.1, Mann-Whitney). In our study we find no conclusive evidence that girls’ and boys’ differ in how frequently they make necessary realistic considerations or letting such considerations affect their written solutions to mathematical word problems.

Furthermore, Table 1 shows that the proportions of RC and RP are higher on the authentic variants of the tasks for both boys’ and girls’. The differences in the distributions of the number of RC and RP have been tested for statistical significance (Mann-Whitney) and the results support the conclusion that a higher task authenticity has a positive effect on the frequency of RC and RP for both boys’ and girls’ (p<0.05).

As also can be seen from Table 1 the increases in RP and RC from the less authentic tasks to the more authentic tasks are larger for the girls’ than for the boys’, but these differences are only marginal and not statistically significant (see below). The increases are also shown in Figure 1 (RP) and Figure 2 (RC). The diagram in Figure 1 graphs the proportions of Realistic Practice (RP) and illustrates the interaction effect of gender and task type. There is a small difference in slope, indicating that, regarding RP, girls’ are slightly more affected by the change in task authenticity than the boys’. The same holds for the proportion of RC, which is illustrated in Figure 2. The statistical significance
of these interaction effects can be studied with an analysis of variance. Even though the data cannot be expected to be normally distributed this method is fairly robust towards such deviations due to relatively large samples.

For both RP and RC, the ANOVA shows that the interaction effects are non-significant. (p = 0.538 for RP and p = 0.344 for RC). Thus, the interaction effects indicated by Figure 1 and 2 cannot be shown to be statistically significant, and therefore cannot be assumed to represent a similar difference in the whole population.

**Figure 1**  Proportions of Realistic Practice (RP) for the less authentic and the more authentic set of tasks, for girls’ and boys’.

**Figure 2**  Proportions of Realistic Considerations (RC) for the less authentic and the more authentic set of tasks, for girls’ and boys’.

In addition to the analysis of an aggregate of the students’ solutions to all of the tasks we have also investigated gender differences concerning the occurrence of Realistic Considerations (RC) and Realistic Practice (RP) on individual tasks. The proportions of student responses categorized as RC and RP on each task variant are shown in Table 2.

**Table 2**  Percentage of student responses categorized as Realistic considerations (RC) and Realistic Practice (RP) on each task variant
It is obvious from Table 2 that potential gender differences cannot be detected on the Bowl tasks, the Bus tasks and possibly the Balloon tasks due to either dominance or a lack of Realistic Considerations and Realistic Practice. In order to analyse the differences between girls’ and boys’, with respect to the occurrence of Realistic Practice and Realistic Considerations, the number of students’ who provided a response that was categorized as RP and Non-RP (and RC and Non-RC) were cross-tabulated with gender for each version of the tasks. Table 3 shows an example of such a crosstabulation concerning RP for the more authentic version of the Running task.

<table>
<thead>
<tr>
<th>Task</th>
<th>Gender</th>
<th>More authentic</th>
<th>Less authentic</th>
<th>Frequency of students’</th>
<th>showing RP</th>
<th>not showing RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balloon</td>
<td>Girls’</td>
<td>70</td>
<td>89</td>
<td>Girls’</td>
<td>25 (71%)</td>
<td>10 (29%)</td>
</tr>
<tr>
<td>(WP2)</td>
<td>Boys’</td>
<td>74</td>
<td>80</td>
<td>Boys’</td>
<td>21 (48%)</td>
<td>23 (52%)</td>
</tr>
<tr>
<td>Bus</td>
<td>Girls’</td>
<td>83</td>
<td>100</td>
<td>Girls’</td>
<td>35</td>
<td>38</td>
</tr>
<tr>
<td>(WP3)</td>
<td>Boys’</td>
<td>80</td>
<td>96</td>
<td>Boys’</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Plank</td>
<td>Girls’</td>
<td>35</td>
<td>64</td>
<td>Girls’</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>(WP4)</td>
<td>Boys’</td>
<td>26</td>
<td>52</td>
<td>Boys’</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Ride</td>
<td>Girls’</td>
<td>20</td>
<td>62</td>
<td>Girls’</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>(WP5)</td>
<td>Boys’</td>
<td>18</td>
<td>38</td>
<td>Boys’</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>Running</td>
<td>Girls’</td>
<td>0</td>
<td>14</td>
<td>Girls’</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>(WP6)</td>
<td>Boys’</td>
<td>0</td>
<td>30</td>
<td>Boys’</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Bowl</td>
<td>Girls’</td>
<td>0</td>
<td>3</td>
<td>Girls’</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>(WP7)</td>
<td>Boys’</td>
<td>4</td>
<td>2</td>
<td>Boys’</td>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>

The statistical significance of the differences in distribution between girls’ and boys’ on each task version was examined with Fischer’s exact test. The results of this significance analysis are presented in Table 4, which with respect to the studied population intends to answer the question: Are realistic practice and realistic considerations more frequent for one sex compared to the other, for each version of the tasks? The occurrence of realistic considerations (RC) on the Running task shows significant differences between boys’ and girls’ in this analysis. In addition, Table 4 signals almost significant gender differences on the more authentic version of the Ride task. Realistic considerations are more common for boys’ on the Running task and more common for girls’ on the Ride task.

<table>
<thead>
<tr>
<th>Task</th>
<th>RP Less authentic</th>
<th>More authentic</th>
<th>RC Less authentic</th>
<th>More authentic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balloon</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Bus</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Plank</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Ride</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Running</td>
<td>No</td>
<td>No</td>
<td>Yes*</td>
<td>Yes*</td>
</tr>
<tr>
<td>Bowl</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

* \( p < 0.05 \) (Fischer’s exact test)

1 \( p = 0.057 \)

2 No realistic considerations shown by any student
In order to study the possibility of differential effects of changing the tasks from less authentic to more authentic versions of the tasks another kind of cross-tabs was created. For boys’ and girls’ separately the frequencies of students’ solutions categorized as RP and Non-RP (and RC and Non-RC) were cross-tabulated with the task version (less authentic and more authentic). An example of this analysis is shown in Table 5, which presents the frequency of RP for girls’ on the less authentic and the more authentic versions of the Running task.

<table>
<thead>
<tr>
<th>Task version</th>
<th>Frequency of girls’ showing RP or not on the two versions of the running task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency of girls’ showing RP</td>
</tr>
<tr>
<td>Less authentic</td>
<td>37 (100 %)</td>
</tr>
<tr>
<td>More authentic</td>
<td>30 (86 %)</td>
</tr>
<tr>
<td>Total</td>
<td>67 (93 %)</td>
</tr>
</tbody>
</table>

Fischer’s exact test: p = 0.023

The statistical significance of the differences in distribution between less and more authentic tasks was examined with Fischer’s exact test. The results from this significance analysis are presented in Table 6. The results presented in Table 6 are intended to answer the following question in relation to the investigated population of students’: Are realistic considerations and realistic practice more frequent for one version of each task compared to the other, for girls’ and/or boys’? For the Ride task, girls’ show significantly more realistic practice and realistic considerations on the more authentic version of the task compared to the less authentic version. No such significant differences are detected among boys’. This indicates that the only task exhibiting differential effects of the change from the less authentic to the more authentic versions is the Ride task. On this task girls’ benefit more than boys’ from a higher task authenticity.

<table>
<thead>
<tr>
<th>Task</th>
<th>Girls’</th>
<th>Boys’</th>
<th>Total</th>
<th>Girls’</th>
<th>Boys’</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balloon</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Bus</td>
<td>Yes*</td>
<td>Yes*</td>
<td>Yes**</td>
<td>No*</td>
<td>No</td>
<td>Yes**</td>
</tr>
<tr>
<td>Plank</td>
<td>Yes*</td>
<td>Yes*</td>
<td>Yes**</td>
<td>No*</td>
<td>Yes*</td>
<td>Yes**</td>
</tr>
<tr>
<td>Ride</td>
<td>Yes**</td>
<td>No</td>
<td>Yes**</td>
<td>Yes**</td>
<td>No</td>
<td>Yes**</td>
</tr>
<tr>
<td>Running</td>
<td>Yes*</td>
<td>Yes**</td>
<td>Yes**</td>
<td>No</td>
<td>No</td>
<td>Yes*</td>
</tr>
<tr>
<td>Bowl</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

* p<0.05 (Fischer’s exact test)  
** p<0.005 (Fischer’s exact test)  
1 p = 0.051  
2 p = 0.055

4. Discussion

The aim of this study was to further investigate the documented tendency that elementary and secondary school students’ all over the world do not make necessary real-world considerations in their word problem solving (Verschaffel et al., 2000; Verschaffel et al., 2009 in press). More specifically this phenomenon was studied from a gender perspective in order to better understand whether girls’ and boys’ behave differently in this respect.

The first research question asked if there are gender differences in the students’ tendencies to neglect an appropriate use of real-world knowledge. The results show that although the girls’ in the study sample more often provided ‘realistic’ solutions than the boys’ this difference was not large and not statistically significant. This holds for both the less authentic and the more authentic word
problems. In addition, there were no statistically significant differences between boys’ and girls’ in the proportion of tasks for which they activated their knowledge of the ‘real’ situations described in the tasks, whether or not it affected their written solutions. The results indicating that girls’ did not notice important aspects of the contexts more than the boys’ may appear somewhat surprising, since the literature on gender differences in task solving indicates that girls’ have a stronger desire than boys’ to understand their mathematics (Boaler, 2002, p. 140; Brandell et al., 2007; Forgasz et al., 1999) and are more likely to take the context into account (Murphy & Whitelegg, 2006b). Such tendencies could have led the girls’ to have noticed the problematic features of the word problems to a larger extent than the boys’, and thus have activated their knowledge of the real-world situations described in the tasks whether or not it affected their solutions to the tasks. This however was not the case to any significant extent so the superficial solution strategies found to be responsible for students’ not thoroughly analyzing the task contexts (e.g. Palm, 2008) seem to have been used by boys’ and girls’ to a similar extent. Also the reluctance to act on those problematic situations that really were detected, which has been indicated to be due to students’ beliefs about word problem solving (e.g. Palm, 2008), was shared to a similar extent by boys’ and girls’. Thus, the need for actions to develop these task-solving approaches and beliefs apply similarly to both genders.

The second research question asked if girls’ benefit more than boys’ from more authentic word problems. The results indicate that benefits from more authentic word problems, in the sense of a higher proportion of realistic solutions, apply to a similar extent to both boys’ and girls’. The similarity in benefit holds for the activation of necessary real-world knowledge as well as for the use of this knowledge in the task solving. This means that the students’ solutions to the more authentic tasks included a more thorough analysis of the situation and the appropriateness of the mathematical model used to model the situation. Such solution strategies are more efficient than superficial solution strategies for solving non-routine tasks, regardless whether the tasks are word problems with an out-of-school task context or pure mathematical problems. Since proficiency in solving non-routine tasks is a significant learning goal it is important that students’ are provided with tasks that elicit this kind of reasoning. In addition, since this kind of reasoning to a larger extent than superficial solution strategies involve considerations of the properties of the mathematical concepts that are involved in the reasoning these strategies are also more efficient for the development of conceptual understanding. Furthermore, that both boys’ and girls’, when having detected the problematic features of the tasks, also used their real-world knowledge to a larger extent in their solutions to the more authentic task variants means that these tasks did not to the same extent elicit conscious dismissals of important conditions of real life due to students’ beliefs about word problem solving. When the intention of using word problems is to practice real-life task solving, then not dismissing such conditions is a more efficient practice since it does not to the same extent miss out important aspects of task solving beyond school. Thus, the positive effect of using more authentic tasks holds for both boys’ and girls’.

The research questions were also examined from the perspective of individual tasks. The results from that part of the study indicate that the lack of large gender differences pertains to most tasks. Of the four tasks to which the whole group provided more ‘realistic’ answers to the more authentic task variant the boys’ as well as the girls’ benefitted significantly. And on the two tasks to which no significant differences between the task variants were found for the whole group, no significant differences were found for either girls’ or boys’ (for a discussion of why the bowl task did not render any differences between the two task variants see Palm (2008)). However, a comparison of the two Category 2 word problems, which have a similar structure, reveals some interesting clues. The authenticity of both tasks influenced the students’ but while the boys’ proportion of real-world knowledge activation were about twice as high on the more authentic variant than on the less authentic variant of both tasks the girls’ corresponding proportion were less than twice as high on the more authentic version of the plank task but over three times as high on the riding task. This may have to do with the context of the tasks. The plank task can be seen as having a more ‘male’ context and the riding task as having a more ‘female’ context, and this difference in type of context may have affected the girls’ more than the boys’. It seems that, especially for the girls’, the riding context was experienced as reasonably interesting or familiar and some more authenticity was what a large number of girls’ needed to engage in the context to a sufficient extent to be able to detect the problematic nature of the task. This may also be compared with the running tasks. On these athletics-competition-on-TV tasks, which can be regarded as including a more ‘male’ context, the boys’ did better on both
task variants. Thus, in line with the literature the girls’ did not engage as much as the boys’ in this more ‘male’ context (Murphy & Whitelegg, 2006a), and the authenticity of the task was not sufficient for the same number of girls’ as boys’ to engage in the context.

Thus, the study indicates that the gender differences due to task context that have been found in achievement on more common tasks (Zohar & Gershkov, 2007) may have its analogue on tasks where it is required to engage sufficiently well in a context to detect when the answer is inconsistent with the ‘real life’ situation described in the task. In both cases it seems that a contributing reason for these differences is differential engagement in the task context depending on its gender type. This emphasizes that if one wants to obtain the positive aspects of more authentic tasks described above for both boys’ and girls’ it would be important to use word problems with many different kinds of contexts to reduce the risk of one gender not gaining as much as the other from the word problems being more authentic.

A methodological limitation of this study is the number of participating students’. 161 students’ is not a small sample but when divided into four subgroups (boy/girl doing less authentic or more authentic task variants) then the number of students’ in each subgroup is more modest. This means that fairly large differences in proportions may not always be statistically significant. This is not a big problem in the comparison between boys’ and girls’ on an aggregate on all of the tasks since these differences were quite small and small differences seldom have any important educational significance regardless of they are statistically significant or not. But for some of the individual tasks the differences were larger and a larger sample of students’ would have been desirable for judging the extent to which these differences would extend to the population from which the sample was drawn. Another issue concerns the generalizability of the results of the study. To be able to generalize results obtained from any study to other students’ and conditions, there must be trustworthy arguments for why this is possible. In this study the participating students’ came from a random sample of fifth-grade classes in the Swedish city from which the sample was taken. Thus, the results of the study are argued to be generalizable to this population. However, the students’ in this sample have been shown to share the tendency to neglect an appropriate use of real-world knowledge in their word problem solving with the student samples from many other studies around the world (Palm, 2008), and this increases the likelihood that these results will hold in many other populations as well. But some phenomena are culturally bounded and it may be that differences between how boys’ and girls’ perceive mathematical word problems differ between cultures. Thus, based on these two methodological notes replications of this study would be needed before making conclusions about the extent to which they may generalize to other populations.

5. Acknowledgements

This study was made possible through a research grant from the Swedish Riksbankens Jubileumsfond.

References


