# Reality in mathematics assessment

## An analysis of task-reality concordance in Finnish and Swedish national assessments

Torulf Palm & Lars Burman

There are a number of possible purposes that can be served by the inclusion of applied tasks in school mathematics. However, the attainment of several of these purposes is likely to be dependent on the applied tasks closely emulating real life task situations. The study reported in this paper is an analysis of this relation between school tasks and problems encountered in real life beyond school that require the application of mathematics for their solutions. The analysis was made on national assessments in Finland and Sweden. The assessment tasks were analyzed in relation to a number of aspects of real life task situations that are a part of a framework for considerations of the quality of real life simulations in school mathematics. There were a substantial proportion of applied tasks in the assessments. The results of the analysis display a large difference between the aspects, in the proportion of tasks that simulate them with reasonable fidelity. There was also a marked difference between the assessments for the different courses in the two countries, in the proportions of tasks simulating the aspects well.

## 1 Introduction

A substantial part of the tasks students encounter in school mathematics is contextualized in such a way that the situation described in the task, the figurative context (Clarke & Helme, 1998), includes elements of some task situation in real life beyond school. Specifically this is the case in national assessments in Finland and Sweden, which is part of the results of the study reported in this paper. However, there is no consensus on the role, and therefore the necessary features, of these tasks.

**Torulf Palm**, *Umeå University*
**Lars Burman**, *Åbo Akademi University,*

In one view, such tasks are not principally seen as having strong intrinsic relations to task situations in real life beyond school. The figurative contexts may involve only objects belonging to the real world, or they may also involve fictitious objects. The whole task situation described in the school task may be in high concordance with some real life situation beyond school or it may not. In this view, such concordance is not of prime priority. The objects involved in the figurative contexts are mainly seen as support for thinking about mathematical concepts and relations (Toom, 1999). The purpose of these tasks "is to convey a *mathematical meaning*, that is the use of suitable concrete objects to represent or reify abstract mathematical notions. (...) They are allegories, *mental manipulatives* or *reifications* which pave children's way to abstractions" (Toom, 1999, p. 37). With this emphasis there is no need for contextualized tasks to have features equivalent to tasks encountered in life beyond school. In fact, the simulation of contextual features of task situations in real life outside school, others than those absolutely needed for a solution to the tasks, is not even desirable. In this view, "The real world is full of waste, redundancy, confusion and boredom, all of which should be excluded from the mathematics classroom" (Toom, 1999, p. 38).

The proponents of an another view share a perspective in which the concordance between school tasks with an out-of-school figurative context and task situations encountered in real life beyond school is emphasized (e.g. Boaler, 1994; Greer, 1993; Nesher, 1980; Reusser, 1988; Verschaffel & De Corte, 1997). Tasks with such concordance could serve a number of purposes in addition to functioning as support for abstract mathematical thinking. They could facilitate an experience of school mathematics as useful and powerful for solving meaningful task situations in real life beyond school. That students obtain such a picture of school mathematics is a goal in itself, specifically emphasized in the Finnish curriculum (Utbildningsstyrelsen, 1994, p. 76, 79). In addition, experiencing school mathematics as useful for out-of-school activities could be a boost for many students' motivation for learning mathematics.

The use of tasks with out-of-school figurative contexts could also provide the students with opportunities to engage in activities requiring not only pure mathematical competencies, but also the specific skills involved in the process of mathematical modeling. Such skills exceed the skills needed in solving pure mathematical problems, not directly involving non-mathematical objects, in that they include the integration of knowledge of the real world and knowledge of pure mathematics. They include the interpretation of task situations in the real world, the creation of real models, the mathematization of such models into mathematical models, the monitoring of the work against both the mathematical and the real world, and the interpretation of mathematical results in relation to the original real situations. Furthermore, by means of these tasks students could be engaged in task-solving situations close to those existing in real life outside of school. Since the sum of the parts does not equal the whole, engagement in the whole process of solving such applied tasks is an important ingredient in education.

An important part in education is played by assessment. Since the efficient application of subject matter knowledge in situations encountered in life outside school is a central goal of education, permeating, for example, both the Finnish curriculum (Utbildningsstyrelsen, 1994) and the Swedish curriculum (Swedish Ministry of Education and Science, 1994), it is of course of utmost importance that such competence is captured in assessment results. This may be accomplished in different ways, but with the use of tasks with high 'task-reality concordance' inferences from the observed performance to performance in task situations in life beyond school are more direct than with the use of other kinds of tasks. Such relatively direct inferences can have a positive affect on, not only the evidential aspect of validity, but also on the consequential aspect. That is, since assessments signal what is important the inclusion of such tasks could accentuate the important competence of applying subject matter knowledge in situations beyond school. Such consequences are of special importance in external assessments such as high-stake summative national examinations. There are strong indications that teachers are influenced by these external assessments and adjust their priorities and aim their efforts in a direction that is in line with the assessments. Evidence of a "pervasive influence of externally-mandated high-stakes assessment on both classroom assessment and instructional practice" is presented by Barnes, Clarke & Stephens (2000). In addition, viewing assessments as good learning opportunities, the inclusion of tasks serving the purposes outlined in this introduction could even more directly contribute to the attainment of these purposes.

Thus, compared to the requirements of tasks serving mainly as support for thinking about mathematical concepts and relations, the requirements on tasks also aimed at serving the additional purposes outlined above are more comprehensive. These purposes require that the tasks emulate task situations encountered in real life beyond school. But, the constitution of some of these tasks has come under attack by proponents of the perspective of task-reality concordance for not meeting these demands. As exemplified in the following the tasks are described as being pseudo-realistic, non-engaging and not serving these outlined purposes very well.

According to Boaler (1994) and Nesher (1980) students have not formed the belief that their skills in mathematics are useful in situations in real life out of school, and Gerofsky (1996) and Sowder (1989) claim that students do not like word problems. Boaler (1994) suggests the pseudo-real contexts in many tasks in school mathematics as a contributing factor to this view and disinterest in mathematics. Furthermore, these tasks are said to be promoting superficial solution strategies, such as the 'key word strategy' and the assumption that 'multiplication always makes bigger', instead of heuristic and metacognitive strategies suitable for attacking tasks encountered in real life situations (Greer, 1992; Reusser, 1988). Indeed, the use of word problems in their current shape would seem to foster an attitude that solving such tasks does not include considerations of the real world (Reusser, 1988; Verschaffel & De Corte, 1997, p. 82, 88). The frequent focus on incorrect assumptions and strategies not based on mathematical properties of the task situation at hand, even when such strategies do not work, is also said to indicate a weak understanding of arithmetic operations as models of real situations (Greer, 1992). Moreover, Boaler (1993, 1994) and Cooper & Dunne (2000) criticize the constitution of some of these tasks for requiring partly participation in the task. Requirements of using some common sense and knowledge of the real world, while at the same time having to leave other such knowledge out is reported to be biased to social class (Cooper & Dunne, 2000) and gender (Boaler, 1994). It would be obvious that inferences from observed performance on such tasks to performance in task situations in real life beyond school are not very direct, and that the inclusion of these tasks in assessments does not send a signal of the important goal of applying subject matter knowledge in situations beyond school.

Hence, there are several applicable purposes of tasks with figurative contexts including 'real' objects, and the choice of purposes influences the constitution of the developed tasks. However, there is considerable potential in school tasks that are in close concordance with task situations in real life beyond school. In addition, there are claims that the use of tasks with an out-of-school figurative context, and with a weak concordance with corresponding real life task situations, may have negative impact on learning, self-conception, view and interest in mathematics, and on the validity of interpretations of assessment results (e.g., Boaler, 1994). This makes it important to thoroughly investigate the features of 'applied' tasks used in mathematics education.

The study reported in this paper is an analysis of the national assessments in mathematics for grades 10–12 in Finland and Sweden. The aim of the study was to describe in what way and to what extent the 'applied' tasks in the assessments are in concordance with task situations in real life beyond school. First the assessment systems in the two countries are described. In the following section the method and the tool used for the analysis is presented. Section 4 consists of the results of the analysis, and in the last section the results are discussed.

## 2 The National assessment systems

The Swedish upper secondary school (grades 10–12) is not compulsory, but nevertheless 98 % of the students in compulsory school are pursuing their studies in grades 10–12. The upper secondary school includes both vocational and more theoretical programs. School mathematics in the Swedish upper secondary school is divided in five consecutive courses, A–E. Course A is studied by all students and is often the only mathematics course taken by the students following a vocational program. Students studying a program with aesthetic or linguistic emphasis are also required to take course B, and students studying with social science emphasis were at the time of the study required to extend their mathematics studies to include course C. For students following the natural sciences program course D is required, but many students also take course E.

The Swedish tests analyzed in this study, given in 1999-2000, were carried out in a system in which the tests were not compulsory but were taken by most students. There were national tests for all courses and the tasks required short answers or more extended responses. One of the tasks in each test required about an hour's work (25 % of the test time). The students received a grade on the test, determined by the cut scores, but the course grade was based on both the test result and on other performances during the course. The teacher decided on the relative weight that each performance should have on the course grade, as well as on the course grade itself. The course grades were one way to attain university entrance.

Also in Finland most students carry on their studies after their first nine years in compulsory school. Some of them go on to the vocational upper secondary school, and some proceed in the more theoretical upper secondary school. The proportion of the students attending each education form varies over the country, but at an average more than half of the cohort carries on their studies in the more theoretical upper secondary school. In this school form the students take either the *short course* (similar to the Swedish courses A-C) or the *long course* (similar to the Swedish courses A-E). Approximately two thirds of the students take the short course, and one third takes the long course. The short course is mostly taken by students planning to continue studying with a social science or mercantile emphasis, while students aiming at university studies in mathematics, natural sciences, medicine, or other programs in which mathematics is needed take the long course.

The Finnish national examinations are compulsory for the students in the more theoretical upper secondary school and the students receive a test grade, which depends on the scores on the test. In addition, the students receive a course grade based on the performances during the whole course. The teacher alone is responsible for the course grade. Both grades are then used for entrance to the university. The mathematics part of the national examination consists of 15 tasks and the students are to choose 10 of them. An extended response is required on all tasks. Students attending the vocational upper secondary school do not in general take the national examinations.

## 3 Method

### 3.1 *Procedure and analyzed tests*

The analysis of the assessment tasks was done by comparing the 'applied' tasks included in the assessments with the real life task situations indicated by their figurative contexts. For such an analysis of the tests one requires an operational definition of mathematical school tasks that are in concordance with the reality indicated by their figurative contexts. For this purpose such a framework has been developed (Palm, 2001b), in which the terminology of simulation is employed. In this framework the founding idea lies in the following quotation of 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 for an assessment task to be in concordance with a corresponding out-of-school task situation it must represent some task situation in real life and important aspects of that situation must be simulated to some reasonable degree. The framework includes a number of such aspects of real life task situations that are essential in that they clearly can have an impact on the extent to which the students may engage in the mathematical activities attributed to the simulated situation. That is, the simulation of the aspects is essential for the possibilities of the mathematics required for solving the assessment task to be the same as the mathematics required in the simulated situation.

The framework includes 18 aspects of real life task situations. To attain a lucid description of the assessment tasks in relation to reality it was decided to decrease the number of aspects included in the analysis. The selection of aspects was made by judgments of their relative importance for attaining the goals set out in the introduction of this paper, and of their usefulness in the development of high-stakes and large-scale tests. Because of the restraints caused by the tests being large-scale and of high-stakes the aspects *Mode of the task presentation*, *Discussion opportunities*, *Consultation and collaboration*, and *Consequences of solution success/failure* were considered to have limited possibilities of being simulated to a reasonable degree. Therefore these aspects were not chosen to be included in the study. The aspect *Time constraints* was not included in the analysis since the available time on specific tasks is dependent on each individual's time usage on other tasks. Finally, the aspects *Purpose in the social context* and *Plausible strategies* were not included in the study due to restricted knowledge of the different and specific learning environments in different schools. For a description of these categories, see (Palm, 2001b). The remaining 11 aspects were used in the analysis and are described later in this section.

To make a description of the tasks in the tests in relation to these aspects, classification categories can be attached to each aspect. In this study, to all but one aspect, two classification categories were defined in relation to each aspect. The two categories correspond to whether the tasks were judged as simulating the specific aspect to a reasonable degree (Category 1) or not (Category 2). Three classification categories were attached to the aspect *Specificity of information/data*, since in relation to this aspect three different task features, partly serving different purposes, were considered discernible (see the description of this aspect in section 3.3). In relation to Specificity of information/data the tasks are classified in Category 2 when this aspect is considered as partly simulated and in Category 3 when it is judged as not being reasonably simulated. The degree of fidelity in the simulation of each aspect required for a classification in Category 1 (and also for Category 2 in the case of Specificity of information/data) is outlined in the description of the classification categories below.

The tasks were also classified in relation to the mathematical content required for their solutions. The content of the mathematics courses in the two countries was divided in 10 categories. The division was made so that the categories would essentially be consistent with the division of the subject matter made in the syllabi in both countries.

A short summary of the selected aspects, classification categories, and mathematical content categories are described in section 3.3, and illustrated by examples from the set of tasks in section 3.2.

Each task was not classified in relation to all of the selected aspects. In the classification procedure the aspects were considered in the order of their descriptions in section 3.3. If the classification on one aspect was such that the task was considered as not having a corresponding real life out-of-school task situation, in which the mathematical competence needed for the task solving was judged to be similar to the competence needed for solving the assessment task, then no more aspects were considered in relation to this task. That is, it was not considered to be reasonable to further classify a task that was judged as not emulating an aspect with such fidelity that a corresponding real life task situation requiring the same mathematics could be identified. Such a termination of the classification procedure only occurred on the aspects *Event*, *Question*, *Existing information* and *Availability of external tools*, since it was only low fidelity simulations of these aspects that caused fundamentally different mathematical requirements between the assessment tasks and corresponding situations beyond school.

For example, the information given in Example 5 in section 3.2 is distinctly different from the information that would have been known in a corresponding real life out-of-school situation. Because of the difference in the available information, the mathematics required for a solution to Example 5 is very different from the mathematics that would be used in a corresponding out-of-school situation. In such a real life situation the values of the mean, the standard deviation, and the probability would not have been known and therefore not used in the task solving. Thus, there does not seem to exist a corresponding out-of-school situation in which the mathematics required would be the same as in Example 5. Therefore, it is not considered reasonable to consider the aspects following in the list. For example, it is not considered useful to judge whether the same external tools, such as a computer or a calculator, would be available since different external tools would be needed to different extent in the assessment task and in the out-of-school situation. It should be noted however, that there are tasks that are classified in Category 2 in relation to the aspects Existence of Information and Availability of external tools and yet also classified in relation to the other aspects. In these cases it has been judged that despite the low-fidelity simulations of these aspects it is possible to solve the assessment task in a similar way as would have been reasonable to do in a corresponding out-of-school situation.

The low-fidelity simulations of the other aspects did not produce situations in which it was impossible to find a corresponding out-of-school situation requiring similar mathematics as in the assessment tasks. For example, the temperature fluctuation in Example 6 in section 3.2 was judged as non-realistic and the task was classified in Category 2 in relation to the aspect *Realism of information/data*. Nevertheless, similar mathematics, as in situations with more realistic temperature fluctuations, is required so it is considered reasonable to classify this task in relation to other aspects as well.

The Swedish tests selected for the analysis were the two latest tests given in each course in 1999-2000. The Finnish tests chosen were the four tests for the short course and the four tests for the long course given during the same time period.

All assessment tasks with figurative contexts including elements of some task situation in real life beyond school were included in the study. The tasks in the tests for both countries sometimes consisted of a number of subtasks (e.g. 1a, 1b, and 1c). Each of these subtasks was analyzed and classified separately. In addition, in the assessments more than one question was sometimes posed within one task or subtask. These different questions were treated as different subtasks and therefore they were also classified separately. The classification was carried out by both authors separately. A majority of the classifications were the same for both authors but when different classifications had been made a discussion took place and decisions were made in consensus.

It is not always easy to determine the most appropriate classification category for each task. Regarding for example the aspect Event, in some tasks there is not really an event presented, but merely a question or an assignment. In some other tasks the event is so poorly contextualized that it is difficult to see to which real situation the event might correspond. In other tasks the situation is so general that almost any situation could fit in, and in some tasks it is just not easy to know if such an event could occur. Nevertheless, choices have been made and the descriptions of the aspects and the classification categories attached to them, together with the examples referred to, are guides to the choices made.

### 3.2 *A set of tasks*

This section consists of 12 tasks that are used in section 3.3 to illustrate each aspect and classification category. A summary of the classifications of these tasks is provided in the end of section 3.3. Some of the tasks are from the national assessments, but since most of the Swedish national tests are under secrecy, and it was not considered suitable to emphasize tasks from the tests of one country, it was decided that tasks would be taken from other sources as well. The source of each task is indicated within brackets.

**Example 1** (Björk, Brolin, & Ekstig, 1994a, authors' translation)
An urn contains 12 red marbles and an unknown number of white marbles. Every time you pick a marble you note the color and put it back in the urn. The experiment is repeated many times. As a mean you pick 40 red marbles on 100 picks.
How many white marbles would there be in the urn?

**Example 2** (Danielsson, Gabrielsson, & Löfstrand, 1995, authors' translation)
A cow is tethered at the middle of one side of a quadratic meadow, the length of the sides being 10.00 m, as shown in the figure.
How long should the rope be for the cow to be able to graze half of the area of the meadow?

**Example 3** (Danielsson et al., 1995, authors' translation)
Carina and Per are standing under a lamppost with a height of 4.0 m. Carina is 1.60 m and Per is 2.00 m tall. They start walking away from the lamppost at a speed of 0.80 m/s.
a) At what speed does Carina's shadow increase?
b) At what speed does Per's shadow increase?

**Example 4** (Palm, 2001b)
A student has just moved to an apartment of his own. The account for this month consists of the following: rent 3545 SEK, telephone 452.50 SEK, electricity 627.50 SEK, and TV/radio 453 SEK.
How much is the account this month?

**Example 5** (Finnish national exam item, 1996, authors' translation)
A person leaves home for work by car every morning at the same time, and arrives at the parking area nearby his work at a point of time that is normally distributed, with the mean 8.50 a.m. and standard deviation 5 minutes. In the parking area the person finds a parking space with a probability of 65 %, and from there the walk to his work takes 5 minutes. If all parking spaces are occupied the person can drive to another parking area 5 minutes away by car. In this parking area there are always parking spaces available. From there the walk to his work takes 10 minutes.
With what probability does the person arrive at work after 9.00 a.m.?

**Example 6** (Developed for this study)
During a week the following measures of the temperature were taken at 12.00 p.m. each day. Calculate the mean temperature at 12.00 p.m. that week.

| Day | Mon | Tue | Wed | Thu | Fri | Sat | Sun |
|---|---|---|---|---|---|---|---|
| Temperature °C | -8 | 8 | -9 | 9 | -10 | 10 | -10 |

**Example 7** (A modified version of an example in Björk, Brolin & Ekstig (1994b))
The cost and the income per day for a company to produce one of its products are:

$$K(x) = x^3 - 6x^2 + 13x + 15\ , \qquad 0 < x < 9$$

and

$$I(x) = 28x$$

where $K$ is the cost in SEK, $I$ is the income in SEK, and $x$ is the number of thousand produced units.
Determine the maximum profit per day.
(The profit function $P(x) = I(x) - K(x)$)

**Example 8** (Swedish national test item, 1996)

| *Chocolate ball was for 17 000* | |
|---|---|
| Probably the world's largest chocolate ball was produced by Åhlens' bakery in Umeå. It was exhibited at Rådhus square in June 1988. The record ball was produced to celebrate the 350th anniversary of the city of Umeå. It consisted of 135 kg of butter, 180 kg of sugar, 162 kg of oatmeal, 22.5 kg of cocoa, 2.7 kg of vanilla, and 2.7 kg of mocha. During the day, 17 000 chocolate lovers got a taste. | The world's longest Swiss roll, 2 053 meters long, was produced by the bakery of Konsum in connection with the People's Forum of Umeå in 1989. The cake was decorated with butter cream. It was sold in 25-cm pieces. The proceeds went to a forest project in Kenya. It took 10 people 400 hours to make the Swiss roll. |

The text above was placed on a milk carton from the dairy "Norrmejerier". Use the information to answer the following questions.

b) With the ingredients, each dm3 of the finished chocolate ball weighed 1.0 kg.

What was the volume of the chocolate ball?

c) The Swiss roll is cylinder shaped. A dissection straight through the cake produces a circular shape with a diameter of 7 cm.

Which one had the largest volume, the Swiss roll or the chocolate ball?

(items a and d do not add anything to the discussion in this paper)

**Example 9** (Palm, 2001b)
You work at the local Bank. A client needs a loan of 1 200 000 SEK to buy a house in the city. The client wants to know if he can afford it. You know that you can offer him a monthly interest of 0,65 %. The client wants to have the loan for as long as possible, which is 40 years and to be paid with the same amount every month.
What will the client's monthly cost for the loan be?

**Example 10** (Palm, 2001b)
Your mother is going to buy balloons for your birthday party. You have invited 6 friends to your party and she wants you to have two balloons each.
How many balloons does your mother have to buy?

**Example 11** (From National Pilot Mathematics Test Summer 1992, Band 1–4, paper 1, in (Cooper, 1992)).
This is the sign in a lift at an office block:

| This lift can carry up to 14 people |
| --- |

In the morning rush, 269 people want to go up in this lift.
How many times must it go up?

**Example 12** (Swedish national test item, 1997)
Anna is going to make blackcurrant jam. Her recipe contains the following ingredients:

1 kg blackcurrant
750 g sugar
2.5 dl water

When Anna has picked and topped the berries she establishes their weight to be 800 g.
How much sugar and water must she use to make the jam?

## 3.3 *Tool of the analysis*

The following is a description of the selected aspects, the classification categories attached to them, and the categories of mathematical content (for a more comprehensive description and argumentation of the aspects, see Palm, 2001b). The classification categories are illustrated by relating them to the set of tasks in section 3.2. An overview of the classification of these tasks is provided in Table 1 at the end of this section.

*Aspect 1. Event*. This aspect refers to the event in the figurative context. A prerequisite for concordance between a school task and a task situation in real life beyond school is that the event described in the school task must have happened or might happen in real life beyond school.

Tasks in which the event in the figurative context is one that could be encountered in real life outside school are categorized in Category 1. Example 2 is considered to be such a task. The event in this task is about a cow that is tethered in a meadow. It is true that this is not the usual condition for a cow, but sometimes it does happen. Examples 3–12 are also judged as belonging to this category.

Tasks in which the figurative context is about an imaginary event not belonging to this world (e.g. an event describing Martians invading earth) and events including objects from this world only, but which still are fictitious events, are classified as belonging to Category 2. Example 1 is an example of such an event. People do no not normally pick marbles from an urn, noting their colors, putting them back in the urn, and repeating that experiment many times. There might exist similar events, for example bingo, but it is the particular event described in the task that is classified.

*Aspect 2. Question*. This aspect refers to the question or assignment given in the figurative context. A task does not always include an explicitly posed question or assignment. However, in analogy with the aspect Event, if a school task includes a question or an assignment, then it must be one that has been posed or might be posed in a real life out-of-school task situation for the school task to be concordant with such a situation.

Category 1 can be described as consisting of tasks in which the questions are such that they have been asked, or might be asked, in the event described in the tasks. The answers to the questions are of practical value or of interest for others than just the people very interested in mathematics. Examples 4–12 illustrate such tasks.

Category 2, on the other hand, can be described as consisting of tasks in which the question or assignment given is judged not to have been asked, and neither would be asked, in the event described in the task. Tasks belonging to this category are of different kinds. In some tasks the choice of category is not a problem. The tasks were never intended to represent a real task situation, and this is evident from the questions posed. Example 2 is considered as an example of such tasks. In other tasks the questions could be asked in the described event, but they do not belong to task situations in which the answers to the questions are of any practical use. In such cases a distinction is made. Tasks with questions that are judged to be of interest just for the sake of mathematics itself are classified in Category 2. The same goes for questions people generally do not go about thinking of, but which might appeal to people very interested in mathematics. Example 3 is judged as such a task. In contrast, tasks in which the question is of interest for others than those interested in mathematics are classified in Category 1.

*Aspect 3. Purpose in the figurative context*. There is always a more or less explicit purpose of solving task situations in real life. This aspect refers to those purposes. The decisive issue in categorizing the tasks in relation to this aspect is the students' possibilities, when working with the assessment task, of experiencing the same clarity of the purpose of the real life task solving as in the corresponding real life situation.

Category 1 consists of tasks in which the purpose of the question in the figurative context is judged as being as clear in the assessment situation as it would have been in the simulated 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 figurative context. Examples 5, 9, 10 and 12 are judged to be tasks in which the purpose of the real life task solving is experienced as clear for most students.

Category 2 consists of tasks in which the purpose of the task solving in the simulated situation is judged as unclear. The figurative contexts could be generally described, not pointing to a specific situation, resulting in many possible situations and purposes of the task solving. In other tasks the situation described in the task is more specific but still open for more than one purpose. In example 4 it is not possible to know if the purpose in the figurative context is to fill in a giro form or to make a rough estimate of the expenses of the month. The mathematics used is likely to be different. The first interpretation of the purpose would need some mathematics appropriate for attaining an answer with high accuracy, and the other interpretation would lead to the possibility of a solution based on a rough mental estimate. Examples 6, 7, 8 and 11 would also be classified as belonging to Category 2, even if the interpretation of these tasks would not be as problematic. Tasks in which no purpose of the question can be found are not classified in relation to this aspect since such tasks are considered as not having a question that would be posed in relation to the described event. Therefore, the classifications of those tasks are terminated after the classification of the aspect Question.

*Aspect 4. Existence of information/data.* This aspect refers to the existence of directly available or obtainable information, on which the solution to a problem can be based. If this aspect is simulated with reasonable fidelity then the same kind of information accessible in the simulated situation is also accessible in the school situation. Discrepancies in such information between the school situation and the simulated situation often lead to differences between the mathematical activities performed in the two situations.

The decisive factor in the classification in relation to this aspect is the relevant and important information on which to base the solution. If the information/data that is important for the solution in the simulated situation coincide with the accessible information/data in the assessment situation, then the task is classified in Category 1. Examples 4 and 6–12 are judged to have concordant information/data and would have been classified in Category 1.

If the information/data that is important for the solution in the simulated situation is not the same as the accessible information/data in the assessment situation, then the task is classified in Category 2. Also, if this information is accessible only by applying other competencies that are different from those required in the simulated situation, and for many students this is a major obstacle in the task solving, then the task is classified in Category 2. Such differences in accessible information can arise if information that would have been known in the simulated task situation is withheld from the students in the assessment situation, or if additional important information is added to the assessment task. Differences can also arise if the information given in the assessment task have been substantially simplified or made more difficult than in the simulated situation. Example 5 is a task that would have been classified in Category 2. In this example the students have available numerical values of mean, standard deviation and probability, values on which they are supposed to base the solution. In the simulated situation these numerical values would not be available and the task would have been solved in a completely different way. The task solving in the two corresponding situations have very little in common, and the assessment task and the corresponding simulated task situation are therefore not in high concordance. Other factors belonging to this aspect, such as contextual richness and if the information has to be retrieved or is explicitly given in the task, do not influence the classification in this particular study.

*Aspect 5. Realism of information/data.* This aspect is *about* the existing information. In a simulation of this aspect, with a reasonable degree of fidelity, numbers and values given are realistic in the sense of being identical to or very close to the corresponding numbers and values in the simulated situation. Such tasks are classified in Category 1. Examples 4 and 7–12 would be classified in this category. A task that does not fit this description is classified in Category 2. The large fluctuation of temperatures in example 6 would not be considered as realistic and therefore this task would be classified in Category 2.

*Aspect 6. Specificity of information/data*. This aspect is also about the existing information. In a simulation of this aspect with high fidelity the information given is specific and not general. The text of the task describes a specific situation in which the subjects, objects, and places in the figurative context are specific. Such simulations may help to provide evidence of real life situations in which school mathematics is useful. These tasks are classified in Category 1 in relation to this aspect. Examples 8–10 and 12 would be classified in this category. In tasks classified as belonging to Category 2 the situations described in the tasks may not be specific, but at a minimum the objects that are the foci of mathematical treatment are specific. With a familiar figurative context this makes it possible for students to monitor their work and check their results in relation to the real situation described in the task. Examples 4, 6 and 11 are considered to be such tasks. If the situation in the figurative context is a general situation in which the subjects and objects are not specified the tasks are classified in Category 3. Example 7 would be classified in this category.

*Aspect 7. Language use in the task presentation*. This aspect refers to the terminology, sentence structure, amount of text etc. used in the presentation of the task. Tasks that are linguistically similar to the corresponding simulated situations are categorized in Category 1. Examples 4, 6, 7 and 9–12 are considered to be such tasks. Tasks are categorized in Category 2 if the terminology, sentence structure or amount of text is judged to affect a more than an insignificant proportion of students in such a way that the possibility to use the same mathematics in the school tasks and in the simulated situations is greatly impaired. The experience with example 8 is that more than just a few students had difficulties with the term "dissection", which hindered them in their task solving. Since this problem would not have occurred in the out-of-school situation this task would be classified in Category 2.

*Aspect 8. Availability of external tools*. This aspect refers to the external tools (e.g. calculators, computers and maps) available in the task situation. Generally, the available external tools have to be the same in the school situation and in the simulated real life situation for the competencies important for solving the task to be the same in the two situations.

If there is a match in the availability of external tools, important for the solution of a task, in the assessment situation and the simulated real life situation, the task is classified in Category 1. Examples 4, 6–8 and 10–12 are considered to be such tasks. If there is a discrepancy between these tools in the two corresponding situations the assessment task is classified in Category 2. Example 9 is such a task. In the simulated situation the clerk would use his computer with specially designed software for these occasions. The students on the other hand, would most likely not have access to a computer, or at least not the software, when solving the assessment task. The consequence is that totally different competencies are needed in solving the assessment task.

*Aspect 9. Guidance.* This aspect refers to help provided to the task solver by means of explicit or implicit hints on, for example, solution methods and types of answers required. The tasks are divided in the different categories depending on whether there is a match in the guidance provided in the assessment situation and the guidance provided in the simulated real life situation. Category 1 consists of tasks that provide the same guidance as is provided in a corresponding out-of-school task situation. Examples 4, 6, 8 and 10–12 are considered to be such tasks. Category 2 consists of tasks that do not match in the guidance given between the two corresponding situations. Example 7 would be classified in this category.

*Aspect 10. Availability of solution strategies.* This aspect refers to the task solvers' available solution strategies. The tasks are categorized in different categories depending on whether the available solution strategies of most students match those of the type of persons solving the tasks in the simulated situations. A task in which the students' available solution strategies allow them to solve the task in the same way as the taken character in the simulated situation would have done, is categorized in Category 1. Other tasks are categorized in Category 2. Examples 4, 6–8 and 11–12 would be categorized in Category 1. If example 10 were given to 6-year-olds it would be classified in Category 2, since they may well solve it but they do not have skills in multiplication available, which is the way their mothers probably would solve it.

*Aspect 11. Solution requirements.* This aspect refers to explicit or implicit requirements on the solution to a task. Such requirements may, in the assessment situation, be expressed in the task text or in the judgment on the solution. The tasks are divided in the different categories depending on whether the solution requirements in the assessment situation are considered to be the same as in the simulated situation. Tasks in which the requirements are considered to be the same are classified in Category 1, and tasks not fulfilling this condition are classified in Category 2. Examples 4, 6–8, 10 and 12 are considered to belong to Category 1. Since the marking scheme does not give any credit to solutions in which realistic considerations have been made (e.g., that people on the lower floors might get tired and take the stairs and that people with wheelchairs might take more space) example 11 would be classified in Category 2.

As stated earlier the tasks were also classified in relation to the mathematical content required for their solutions. The following is a description of the categories of mathematical content used in the study.

*Categories of mathematical content*

| | |
|---|---|
| Algebra | Algebraic expressions, linear and quadratic equations and functions, linear simultaneous equations, linear and quadratic inequalities |
| Arithmetic 1 | Addition, subtraction, multiplication, division, fractions, percent, number sense |
| Arithmetic 2 | Logarithms, powers, number series, polynomial equations (degree 3 and higher) |
| Derivatives 1 | Derivatives of polynomials and exponential functions |
| Derivatives 2 | Derivatives of product, quotient, and trigonometric functions |
| Differential equations | Differential equations |
| Geometry | Theorem of Pythagoras, angle, area (not requiring trigonometry), volume (not requiring integration) |
| Integration | Includes determining volume by the use of integration |
| Statistics and probability | Measures of mean and dispersion, normal distribution, statistical surveys, probability |
| Trigonometry | Trigonometry |

Table 1. *Classification of the tasks in section 3.1*

| Tasks (Example No) | Mathematical content | Event | Question | Purpose | Existence of information | Realism | Specificity (3 categories) | Language use | Availability of external tools | Guidance | Availability of—strategies | Solution requirements |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Statistics | 2 | | | | | | | | | | |
| 2 | Geometry | 1 | 2 | | | | | | | | | |
| 3 | Derivatives 1 | 1 | 2 | | | | | | | | | |
| 4 | Arithmetic 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 |
| 5 | Statistics | 1 | 1 | 1 | 2 | | | | | | | |
| 6 | Arithmetic 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 |
| 7 | Derivatives 1 | 1 | 1 | 2 | 1 | 1 | 3 | 1 | 1 | 2 | 1 | 1 |
| 8 | Geometry | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 |
| 9 | Arithmetic 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | | |
| 10 | Arithmetic 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 |
| 11 | Arithmetic 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 |
| 12 | Arithmetic 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*Note.* A task is classified in Category 1 in relation to an aspect if it is judged to simulate the aspect with reasonable fidelity, and in Category 2 or 3 otherwise.

## 4 Results

The tests in the study were made up of 486 subtasks distributed over the tests as displayed in Table 2. In total just under 50 % of the subtasks were considered to be applied tasks in the sense of the figurative context including elements other than pure mathematical objects. The proportions of applied tasks differed significantly between the tests for the different courses within each country, but were relatively similar for corresponding courses in the two countries. The proportion of applied tasks seems to be dependent on the level of the mathematics tested. The proportions of applied tasks in the tests for the most advanced courses were under 30 %. The proportions of applied tasks in the tests for the Swedish B and C courses was 43 %, and the proportions in the tests for the Finnish Short course and the Swedish Course A were about 70 % (see Table 2).

Table 2. *The number of applied tasks in the assessments*

| | Finland | | | Sweden | | | | Both countries |
|---|---|---|---|---|---|---|---|---|
| | Short course | Long course | Total | Course A | Courses B-C | Courses D-E | Total | Total |
| Total number of tasks | 107 | 93 | 200 | 75 | 104 | 107 | 286 | 486 |
| Number of applied tasks | 74 | 27 | 101 | 54 | 45 | 28 | 127 | 228 |
| Proportion of applied tasks (%) | 69 | 29 | 51 | 72 | 43 | 26 | 44 | 47 |

The results of the classification of the applied test items are displayed in Tables 3–6 below. Tables 3 and 4 show that the proportion of tasks simulating the different aspects in the framework to a reasonable degree varied substantially between the tests for the different courses. For example, in most items the event described was judged to be an event from real life beyond school, but the proportions of such items still varied from 78 % in the tests for the Finnish short course to 100 % in the tests for the Finnish long course.

Table 3. *The proportion of applied tasks in the Finnish assessments that were classified in Category 1 in relation to each aspect*

| | Short course | | Long course | | Total | |
|---|---|---|---|---|---|---|
| | $n^a$ | Category 1 (%) | $n^a$ | Category 1 (%) | $n^a$ | Category 1 (%) |
| Event | 74 | 78 | 27 | 100 | 101 | 84 |
| Question | 58 | 53 | 27 | 37 | 85 | 48 |
| Purpose | 31 | 13 | 10 | 10 | 41 | 12 |
| Existence of information | 31 | 39 | 10 | 10 | 41 | 32 |
| Realism of information | 19 | 100 | 1 | 100 | 20 | 100 |
| Specificity of information | 19 | 16 | 1 | 0 | 20 | 15 |
| Language use | 19 | 79 | 1 | 100 | 20 | 80 |
| Availability of external tools | 19 | 84 | 1 | 0 | 20 | 80 |
| Guidance | 17 | 94 | 0 | - | 17 | 94 |
| Availability of solution strategies | 17 | 100 | 0 | - | 17 | 100 |
| Solution requirements | 17 | 94 | 0 | - | 17 | 94 |

*Note.* [a] *n* denotes the number of tasks that were classified in relation to each aspect.

Table 4. *The proportion of applied tasks in the Swedish assessments that were classified in Category 1 in relation to each aspect*

| | Course A | | Courses B-C | | Corses D-E | | Total | |
|---|---|---|---|---|---|---|---|---|
| | $n^a$ | Category 1 (%) | $n^a$ | Category 1 (%) | $n^a$ | Category 1 (%) | $n^a$ | Category 1 (%) |
| Event | 54 | 91 | 45 | 96 | 28 | 89 | 127 | 92 |
| Question | 49 | 59 | 43 | 72 | 25 | 40 | 117 | 60 |
| Purpose | 28 | 14 | 31 | 35 | 10 | 20 | 69 | 25 |
| Existence of information | 28 | 32 | 31 | 71 | 10 | 0 | 69 | 45 |
| Realism of information | 12 | 83 | 23 | 100 | 0 | - | 35 | 94 |
| Specificity of information | 12 | 58 | 23 | 43 | 0 | - | 35 | 49 |
| Language use | 12 | 92 | 23 | 100 | 0 | - | 35 | 97 |
| Availability of external tools | 12 | 83 | 23 | 65 | 0 | - | 35 | 71 |
| Guidance | 10 | 80 | 15 | 87 | 0 | - | 25 | 84 |
| Availability of solution strategies | 10 | 100 | 15 | 100 | 0 | - | 25 | 100 |
| Solution requirements | 10 | 100 | 15 | 100 | 0 | - | 25 | 100 |

*Note.* [a] *n* denotes the number of tasks that were classified in relation to each aspect.

The proportion of tasks classified in Category 1 in relation to the aspect Question was much lower than for the aspect Event. The proportions of tasks including a question that would possibly be posed in the real life event described in the task varied from 37 % in the tests for the Finnish long course to 72 % in the Swedish tests for the B and C courses.

The proportion of tasks demonstrating concordance with a corresponding real life situation in the existence of information on which to base the solution varied substantially between the different course tests. In the Finnish tests for the long course one (10 %) of the ten classified tasks was classified in Category 1, and in the Swedish tests for the most advanced courses none of the ten classified tasks was classified in Category 1. In contrast, 22 (71 %) of the 31 classified tasks in the Swedish tests for the B and C courses were classified in Category 1 in relation to the aspect Existence of information/data.

The lowest proportion of test items in concordance with a corresponding real life situation in relation to a single aspect, taken as an aggregate of proportions on all tests, was found for the aspect Purpose. Twenty percent of the tasks were classified in Category 1 in relation to this aspect, and the proportions on the tests for the different courses ranged from 10 % in the Finnish tests for the long course to 35 % for the Swedish tests for the courses B and C.

The remaining aspects will only be commented on in relation to the lower courses in each country. No tasks in the Swedish D or E course tests and only one task in the tests for the Finnish long course were classified in relation to these aspects. In the other tasks the simulations of the aspects Event, Question or Existence of information/data were judged to be of such low degree that there was not considered to exist corresponding real life task situations in which the mathematics required for the task solving would be similar to the mathematics required in the assessment tasks.

In relation to the aspects Realism of information/data, Language use in the task presentation, Guidance, Availability of solution strategies, and Solution requirements the proportions of tasks classified in Category 1 were fairly high in all of the tests for the lower courses. This was not the case for the aspect Specificity of information/data. About half of the Swedish tasks and 16 % of the tasks in the tests for the Finnish short course were considered to include a description of a task situation that was specific, and were classified in Category 1. However, in relation to this aspect there were three classification categories, and all but one of the items that were not classified in Category 1 were classified in Category 2. The proportions of tasks with concordance in the availability of relevant tools varied from 65 % in the Swedish tests for the B and C courses to 84 % in the Finnish tests for the short course.

The distribution of the applied tasks over different mathematical content will be commented only briefly and not displayed in a table. The distribution followed the distribution over different courses without any significant exceptions. That is, Statistics, Algebra, Arithmetic 1 and Geometry, contents that belong to the less advanced courses, were required in many applied tasks, while Derivatives 2, Integration and Differential equations, contents belonging to the most advanced courses, are contents that were required in relatively few applied tasks. It is to be noted, however, that while the Finnish long course includes most of the mathematical content categories, the Swedish D and E courses only include Trigonometry in addition to these three content categories.

For most aspects, the proportion of applied tasks classified in Category 1 was similar for all the mathematical content categories and the proportion of category 1 classifications depending on content will not be displayed in a table either. Some exceptions existed though. Only 70 % of the applied Arithmetic 2 tasks, but most of the applied tasks in the other content categories, were considered to simulate the aspect Event with reasonable fidelity. In contrast, nearly all of the applied Arithmetic 2 items classified in relation to the aspect Question were judged as simulating this aspect reasonably well. In relation to this aspect the proportion of Geometry items classified in Category 1 stood out as comparably low. The proportion of Category 1 classifications for the geometry items was 32 % while the mean proportion for all content categories was 55%. Substantial differences occurred in relation to the aspect Existence of information/data. Only a few applied tasks requiring algebra or mathematical content belonging to the most advanced courses were judged to simulate this aspect with a reasonable fidelity. In contrast, a majority of the applied items requiring the use of the mathematical content included in the lower courses were classified in Category 1 in relation to this aspect.

Tables 5 and 6 show the frequencies and proportions of applied assessment items that, in relation to the four aspects on which the classification of each item could come to an end (see section 3.1), were judged to simulate a specific aspect, as well as the preceding aspects, to such a degree that the tasks were considered to require the same mathematics for their solutions as would be required in corresponding real life situations. Just over one third of the applied items in the tests for the most advanced courses in the two countries (37 % in the Finnish tests and 36 % in the Swedish tests) were considered to both describe an event that might occur in real life beyond school and include a question that might be posed in that event. The tables also show that none of the items in these tests were judged to simulate all of these aspects to such a degree that one could recognize corresponding real life task situations that require the same mathematics. In fact, only one of the items in these tests "made it through" the classification of the aspect Existence of information.

In the tests for the lowest course in each country about half of the items were considered to both describe an event that might occur in real life beyond school and to include a question that might be posed in that event. About 25 % of the tasks also simulated the aspect Existence of information with relatively high fidelity, and about 20 % of the tasks in these tests were judged to simulate all of these aspects to such a degree that one could recognize corresponding real life task situations that require the same mathematics.

For the Swedish B and C course tests, 69 % of the applied tasks were considered to both describe an event that might occur in real life beyond school and to include a question that might be posed in that event. The proportion of tasks that were considered to simulate the aspect of Existence of information was 51 %, and 33 % of the tasks were judged to simulate all of these aspects to such a degree that one could recognize corresponding real life task situations that require the same mathematics.

Table 5. *The frequencies and proportions of applied assessment items that, in relation to the four aspects on which the classification of each item could come to an end, were judged to simulate a specific aspect, as well as the preceding aspects, to such a degree that the tasks were considered to require the same mathematics for their solutions as would be required in corresponding real life situations*

| | Finland Short course | | Finland Long course | | Finland Total | |
|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *n* | % |
| Total number of tasks | 74 | 100 | 27 | 100 | 101 | 100 |
| Event | 58 | 78 | 27 | 100 | 85 | 84 |
| Question | 31 | 42 | 10 | 37 | 41 | 41 |
| Existence of information | 19 | 26 | 1 | 4 | 20 | 20 |
| Availability of external tools | 17 | 23 | 0 | 0 | 17 | 17 |

Table 6. *The frequencies and proportions of applied assessment items that, in relation to the four aspects on which the classification of each item could come to an end, were judged to simulate a specific aspect, as well as the preceding aspects, to such a degree that the tasks were considered to require the same mathematics for their solutions as would be required in corresponding real life task situations*

| | Sweden Course A | | Sweden Courses B-C | | Sweden Corses D-E | | Sweden Total | |
|---|---|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *n* | % | *n* | % |
| Total number of tasks | 54 | 100 | 45 | 100 | 28 | 100 | 127 | 100 |
| Event | 49 | 91 | 43 | 96 | 25 | 89 | 117 | 92 |
| Question | 28 | 52 | 31 | 69 | 10 | 36 | 69 | 54 |
| Existence of information | 12 | 22 | 23 | 51 | 0 | 0 | 35 | 28 |
| Availability of external tools | 10 | 19 | 15 | 33 | 0 | 0 | 25 | 20 |

## 5 Discussion

Assessments need to be validated in different aspects. If several of the purposes discussed in the introduction are going to be served and "if mathematics education is going to be realistic, problems will have to be sought that respect assumptions about life outside school" (Nunes, Schliemann, & Carraher, 1993, p. 148). The tool of analysis, partly used in this paper, may function as an instrument for validation of the links between school tasks and real life task situations beyond school. For example, it may, as in this study, be used to analyze both in what ways and to what extent the applied tasks in national assessments are concordant (or not concordant) with out-of-school task situations. Furthermore, success in taking on such a challenge of providing students with an experience of a strong connection between school mathematics and real life beyond school is facilitated by having available a framework for reflecting on the issue. The fidelity of the simulation of the aspects may be analyzed and conclusions of the analysis may then be used for different decisions such as task revision.

However, the fidelity of the simulation of these aspects is not to be seen as a general measure of the quality of items, assessments or use of assessment scores. Such evaluations need to be much more comprehensive and also consider other quality constructs, such as reliability and different aspects of validity.

Furthermore, there can be several purposes for including tasks comprising non-mathematical objects in their figurative contexts in mathematics education, and not all of these purposes require emulation of certain aspects of a real life task situation to be attained. Neither does the simulation of these aspects tell the whole story of the links between school mathematics and real life beyond school. For example, it does not say anything about which mathematical competencies are important and required in life beyond school. However, it does tell part of the story about the relation between school tasks and tasks encountered in real life beyond school. In addition, considering that high concordance between these two sets of tasks could facilitate the attainment of several important goals, this link between school mathematics and reality beyond school is of significant value.

The high proportion of applied tasks in the tests included in this study could serve several of the purposes outlined in the introduction, such as providing good opportunities for students to experience strong links between school mathematics and real life beyond school. This is especially true for the tests used in the lower courses, which display a higher proportion of applied tasks than do the tests for the higher courses (the proportions of 69 % and 72 % applied tasks in the tests for the Finnish short course and the Swedish A-course respectively must be regarded as considerable proportions). This difference in the proportion of applied tasks between the tests for the less advanced courses and the tests for the more advanced courses seems to mirror the emphasis put on real life applications in the curricula and syllabi in both countries (Swedish Ministry of Education and Science, 1994; Swedish National Agency for Education, 1996; Utbildningsstyrelsen, 1994).

For several of the outlined purposes to be well served though, it is also probably required that a substantial proportion of these tasks really have close similarities with task situations in life beyond school on important aspects of such situations. The results in this study, which show that a significant proportion of the tasks are not classified in category 1 in relation to several of the aspects, are in agreement with claims in the research literature that many tasks in school mathematics with an out-of-school figurative context are pseudo-realistic and not in accordance with the reality indicated in the tasks. Therefore, it may be that despite the large proportion of applied tasks the tests in the study do not facilitate the attainment of several of the outlined purposes. On the other hand, the results show that in a significant minority of the tasks most of the aspects are simulated with a reasonable degree of fidelity.

However, the framework used in this study allows a more thorough analysis of the relationship between the assessment tasks and situations beyond school than just a general judgment of the realism of the tasks. The classification of the tasks in relation to the different aspects in the framework also reveals in what ways the tasks are concordant with real life beyond school or not.

The results of the classification vary among the tests, but for some of the courses the proportion of the applied tasks classified in Category 1 in relation to the aspect Question was high (up to 72 % for the Swedish B and C courses). This may be a good basis for supporting the students' formation of the belief that school mathematics is useful in life beyond school and for fostering some of the specific skills involved in the mathematical modeling process. In some other tests, less than half of the applied items (down to 37 % in the tests for the Finnish long course) include a question that would have been posed in the real event described in the task. Such proportions of applied items including a question that is relevant in the real event described in the task may not be enough to make the students form the belief that school mathematics is useful and powerful for solving problems in life beyond school. Indeed, it would be quite natural for students to be skeptical to the real life usefulness of school mathematics when the mathematics education community seems to have troubles coming up with examples in which school mathematics is useful in life beyond school. In addition, such proportions of questions that are relevant for the real event might not be high enough for the students who have real life usefulness as a facilitator for motivation.

Furthermore, a low proportion of tasks that include a relevant real life question may not facilitate the learning of some of the specific skills involved in the mathematical modeling process. On the contrary, if the irrelevance of the question is obvious to the students this may instead facilitate the forming of the belief some students have (e.g. Palm, 2002) that solving applied tasks is a game, with rules not necessarily consistent with the rules of real life problem solving. It may be one of the factors that lead students to the belief that the purpose of the task solving is not to find out anything about the real situations described in the tasks but to show that some mathematical tools are mastered, a belief held by some students (Palm, 2002). In this context it may not be natural to develop a habit of making careful validations of models and solutions, which are important competencies in mathematical modeling and which seem to be characteristics that are absent in many students' word problem solving endeavors (Greer, 1993; Palm, 2002; Reusser & Stebler, 1997; Silver, Shapiro, & Deutsch, 1993; Verschaffel, De Corte, & Lasure, 1994; Yoshida, Verschaffel, & De Corte, 1997).

In relation to the aspect Existence of information, only in the tests for the Swedish B and C courses much more than a third of the applied tasks were classified in Category 1 (71 %). In the tests for the most advanced courses the included information, upon which the solutions was to be based, was in concordance with the information available in a corresponding real life situation for only one of the twenty items classified in relation to this aspect. This caused different mathematical skills to be required for the solutions of the assessment tasks and for the solutions of the problems in the corresponding real life situations. If this is obvious for the students such low proportions of category 1 classifications, as in the tests for the more advanced courses, are not likely to facilitate the development of the belief that school mathematics is useful in the world beyond school, the motivation of the students (those that call for real-life usefulness), or the learning of some of the skills that are specific to mathematical modeling. In addition, the purpose of making relatively direct inferences to the competence of solving problems in certain situations in real life beyond school would not be well served.

The classification in relation to the aspects Purpose and Specificity shows that many tasks do not provide a well specified real life task situation with a clear purpose of the task solving in the given real situation. These low proportions will probably not facilitate the development of the belief that there are many out-of-school situations in which school mathematics is useful. In addition, when the purpose of solving the problems in the figurative context is not clear, the incentive to validate answers against the real situation is likely to be lower than when the tasks include a clear purpose of the task solving. These aspects display the largest differences in Category 1 proportions between the tests in the two countries. The tests for the Swedish A, B and C courses displayed three times as many Category 1 classifications in relation to the aspect Specificity as did the tests for the Finnish short course. In relation to the aspect Purpose, the Swedish tests included twice as many Category 1 classifications as did the Finnish tests. It should be noted, though, that regarding Specificity only one of all of the items classified in relation to this aspect was classified in Category 3. The reasons that these aspects rendered the lowest proportions of category 1 classifications cannot be determined from this study. Possible reasons may be a focus on including the numbers needed for the computations and in addition wanting to keep the task text down, both to save space and not to bring reading difficulties into the solution process. At the same time this aspects either are not considered important or one has not reflected over them.

There were also a number of aspects on which most of the items were in concordance with a corresponding real life situation. Realism of information/data, Availability of external tools, Availability of solution strategies, and Solution requirements were such aspects. These high proportions on Realism and Solution requirements provide the students with possibilities to validate answers against available data and their real world knowledge. They also make it possible for the students to experience that their solutions are given credit for similar features as would have been given credit in the real situations beyond school. This would enhance the validation part of the mathematical modeling process.

To summarize the results of the classification in this study, the tests, and especially the tests for the less advanced courses, included a significant proportion of applied tasks. Depending on the course, the proportion of tasks that were considered to both describe an event that might occur in real life beyond school and to include a question that might be posed in that event was between 36 % and 69 % in the tests. None of the tasks from the tests for the most advanced courses, and 19 % to 33 % of the tasks in the tests for the less advanced courses, were considered to simulate all of the aspects Event, Question, Existence of information, and Availability of external tools to such a degree that one could recognize corresponding real life task situations that require the same mathematics. In relation to the aspect Purpose the proportion of Category 1 classifications was quite low, and in relation to most of the other aspects it was quite high. In a comparison between the tests in the two countries some of the Swedish tests include substantially higher proportions of Category 1 classifications in relation to some of the aspects (see, for example, the proportion of category 1 classifications in relation to the aspect Specificity of information for the tests for both the Swedish A course and the Swedish B and C courses), while the proportions are similar for the rest of the aspects. Similar courses in the two countries also often displayed similar proportions of category 1 classifications.

The tests for the most advanced courses differed from the other course tests. For none of the included applied tasks there was considered to exist a corresponding real life situation beyond school requiring the same mathematics. There may be several reasons for this, possibly acting together. One of the reasons may be that the test developers' interpretations of the curricula and syllabi imply that there should be a greater emphasis on task-reality concordance for the less advanced courses. That is, the differences in task-reality concordance may reflect differences in the views of the role (see the introduction) that the applied tasks should play. For example, one opinion may be that there is a greater need for high-fidelity simulations of the aspects in the tasks included in the tests for the less advanced courses, due to a greater need for the students doing only these tests to experience the out-of-school usefulness of school mathematics.

The differences in task-reality concordance between the tests for the more advanced courses and the less advanced courses may also mirror that there are relatively many applications in ordinary life that only require the more basic school mathematics and that do not demand special equipment. The more complex mathematical content included in the higher courses may make it more difficult to find applications that lend themselves to high fidelity simulations of the aspects used in this study. This may be particularly so when time and equipment constraints, as in assessment situations, are taken into account.

It is, of course, not possible to determine a limit to the proportion of tasks that have to be judged as simulations with reasonable fidelity on these aspects for the purposes outlined in the introduction to be fairly well served. But tests including a high proportion of applied items in high concordance with a real life situation would probably facilitate the attainment of these purposes. In addition, a test analysis such as this could be useful for reflections on the issue and for test development. For example, concerning the national assessments analyzed in this paper, one might reflect on the possibilities that it would be worthwhile to see if some of the aspects could be simulated with even higher fidelity and in a higher proportion of the applied tasks. It may be possible to do so without too much loss in other respects, such as a lower degree of construct validity due to the increase in reading comprehension difficulties that may arise if more information has to be added to some of the tasks.

The results of analyses like the one presented in this paper are not independent of the classifiers, and since the tasks were classified by only two persons generalizations to other possible classifiers are limited. The many factors influencing this discrepancy between the raters' classifications include the restricted descriptions of the simulated situations in the tasks, and the lack of clarified purposes of the task solving in the described real life situations. Such factors make it difficult to decide on a classification (which also makes it difficult for the students to experience the simulated situation as real and important), which impedes rater agreement.

However, the fact that most classifications were the same for both authors and that consensus decisions were easily attained suggest that the main features of the results are likely to be similar for many other researchers, especially since the authors come from different countries with different school cultures and sociocultures. The main features are also consistent with the results in an analysis of the tests for the Finnish short course and the Swedish A, B and C courses 1998–1999 (Palm, 2001a).

Another factor that influences the discrepancy between raters' classifications is the different life experiences and views of the classifiers. The classifiers in this study may be seen as representatives of the teaching and task development community. That is, people from the adult community that provides the tasks the students are to work with. Since this community contributes significantly to the students' learning environment, studies of task features from this perspective are important.

However, the purposes of using applied tasks that rely on the students' experiencing the usefulness of school mathematics in life beyond school are dependent on the conceptions the students themselves have of the tasks. That is, the tasks have to be high fidelity simulations of meaningful real life situations in the eyes of the students. Therefore, it could be useful to investigate further how the students themselves see the tasks they are set to solve. The results of such a study may be different from analyses made by researchers and teachers.

## *References*

Barnes, M., Clarke, D., & Stephens, M. (2000). Assessment: the engine of systemic curricular reform? *Journal of Curriculum Studies*, 32(5), 623-650.

Björk, L.–E., Brolin, H., & Ekstig, K. (1994a). *Matematik 2000, Kurs B*. Stockholm: Bokförlaget Natur och Kultur.

Björk, L.–E., Brolin, H., & Ekstig K. (1994b). *Matematik 2000, Kurs* C. Stockholm: Bokförlaget Natur och Kultur.

Boaler, J. (1993). The role of contexts in the mathematics classroom: Do they make mathematics more "real"? *For the Learning of Mathematics*, 13(2), 12-17.

Boaler, J. (1994). When do girls prefer football to fashion? An analysis of female underachievement in relation to "realistic" mathematics contexts. *British Educational Research Journal*, 20(5), 551-664.

Clarke, D. J., & Helme, S. (1998). Context as construction. In O. Björkqvist (Ed.), *Mathematics teaching from a constructivist point of view* (pp. 129-147). Vasa, Finland: Faculty of Education, Åbo Akademi University.

Cooper, B. (1992). Testing national curriculum mathematics: Some critical comments on the treatment of 'real' contexts for mathematics. *Curriculum Journal*, 3(3), 231-244.

Cooper, B., & Dunne, M. (2000). *Social class, sex and problem-solving.* Buckingham, Philadelphia: Open University Press.

Danielsson, R., Gabrielsson, G., & Löfstrand, B. (1995). *Räkna till max: Grundbok och övningsbok kurs E*. Malmö, Sweden: Gleerups Forlag.

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.

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

Greer, B. (1992). Multiplication and division as models of situations. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 276-295). New York: McMillan.

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

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

Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics.* Cambridge University Press.

Palm, T. (2001a). *School mathematics and reality: An analysis of the relation between national assessment tasks and task situations in life beyond school* (Research reports, No 2, in Mathematics Education). Department of Mathematics, Umeå university, Sweden.

Palm, T. (2001b). *Word problems as simulations of real world task situations: A proposed framework* (Research reports, No 3, in Mathematics Education). Department of Mathematics, Umeå university, Sweden.

Palm, T. (2002). *Impact of authenticity on sense making in word problem solving* (Research reports, No 2, in Mathematics Education). Department of Mathematics, Umeå university, Sweden.

Reusser, K. (1988). Problem solving beyond the logic of things: Contextual effects on understanding and solving word problems. *Instructional Science,* 17, 309-338.

Reusser, K., & Stebler, R. (1997). Every word problem has a solution – The social rationality of mathematical modeling in schools. *Learning and Instruction*, 7(4), 309-327.

Silver, E. A., Shapiro, L. J., & Deutsch, A. (1993). Sense making and the solution of division problems involving remainders: An examination of middle school students' solution processes and their interpretations of solutions. *Journal for Research in Mathematics Education*, 24(2), 117-135.

Sowder, L. (1989). Searching for affect in the solution of story problems in mathematics. In D. B. McLeod & V. M. Adams (Eds.), *Affect and mathematical problem solving: A new perspective*. New York: Springer-Verlag.

Swedish Ministry of Education and Science. (1994). *1994 Curriculum for the non-compulsory school system*. Stockholm: Swedish Ministry of Education and Science.

Swedish National Agency for Education. (1996). *Study programme: Information for upper secondary schools and upper secondary adult education programme: Natural sciences programme*. Stockholm: Swedish National Agency for Education.

Toom, A. (1999). Word problems: Applications or mental manipulatives. *For the Learning of Mathematics*, 19(1), 36-38.

Utbildningsstyrelsen. (1994). *Grunderna för gymnasiets läroplan*. Helsingfors: Utbildningsstyrelsen.

Verschaffel, L., & De Corte, E. (1997). Word problems: A vehicle for promoting authentic mathematical understanding and problem solving in the primary school. In P. Bryant & T. Nunes (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 69-97). Howe: Psychology Press.

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.

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.

## *Acknowledgements*

The authors wish to thank Hans Wallin and Johan Lithner for their most valuable comments on an earlier version of this paper.

### Torulf Palm

Torulf Palm is a researcher in mathematics education and assessment at the Department of Mathematics and the Department of Educational Measurement at Umeå University, Sweden. His special research interests are assessment, mathematical reasoning, mathematical modeling and the authenticity of word problems.

### Lars Burman

Lars Burman is a lecturer in mathematics education at the Department of Teacher Education at Åbo Akademi University, Vasa, Finland. His special research interests are problem solving and assessment in mathematics education.


# Sammanfattning


Det finns ett antal möjliga syften med att inkludera tillämpade uppgifter i skolmatematiken. Starka argument finns dock för att en icke obetydlig andel av dessa uppgifter behöver vara verklighetsnära för att flera av dessa syften ska kunna uppnås. Denna artikel innehåller en analys av relationen mellan tillämpade matematikuppgifter i skolsystemet och problemsituationer utanför skolan där matematiken kan vara tillämplig. Analysen är baserad på studentexamensprov från Finland och nationella kursprov från Sverige. Som analysverktyg användes ett ramverk som innehåller ett antal aspekter av verkliga situationer som bedömts vara väsentliga att beakta vid simulering av verkliga situationer utanför skolan. Provuppgifterna analyserades och klassificerades i olika kategorier i förhållande till respektive aspekt beroende på hur väl aspekten bedömts vara simulerad.. Analysen visar att ungefär hälften av alla uppgifter i proven kan anses vara tillämpade, och att det finns en betydande skillnad mellan aspekterna vad gäller andelen uppgifter som i hög grad bedömdes simulera dem. En tydlig skillnad finns också, i båda länder, mellan proven för olika kurser beträffande andelen uppgifter som i hög grad simulerar respektive aspekt.