

Uncommon vocabulary in mathematical tasks in relation to demand of reading ability and solution frequency

ANNELI DYRVOLD, EWA BERGQVIST
AND MAGNUS ÖSTERHOLM

This study reports on the relation between commonness of the vocabulary used in mathematics tasks and aspects of students' reading and solving of the tasks. The vocabulary in PISA tasks is analyzed according to how common the words are in a mathematical and an everyday context. The study examines correlations between different aspects of task difficulty and the presence of different types of uncommon vocabulary. The results show that the amount of words that are uncommon in both contexts are most important in relation to the reading and solving of the tasks. These words are not connected to the solution frequency of the task but to the demand of reading ability when solving the task.

When using tests to assess students' knowledge, one aspect of validity is to measure the intended latent variable, in our case mathematical competence, and nothing else. One thing that might disturb the validity of the assessment is the language, since there is always a possibility that we measure not only the students' mathematical competence but also their reading ability. In some tests, for example PISA (Program for International Student Assessment) (OECD, 2006), the explicit aim is to use as simple language as possible, to measure mathematics ability (or rather mathematical literacy in this particular case) and not reading ability. Still, reading and writing are essential in order to express, think about, and do mathematics, since there are particular words, symbols, phrases, and grammatical structures that mathematics cannot exist without. To include a language component in mathematical knowledge is also supported by the aspect of communication being included as a part of mathematical proficiency in several frameworks (e.g. NCTM, 2000; Niss & Højgaard, 2011).

Anneli Dyrvold, *Umeå University*
Ewa Bergqvist, *Umeå University*
Magnus Österholm, *Umeå University*

Dyrvold, A., Bergqvist, E. & Österholm, M. (2015). Uncommon vocabulary in mathematical tasks in relation to demand of reading ability and solution frequency. *Nordic Studies in Mathematics Education*, 20(1), 5–31.

Within the complex relationship between aspects of reading and aspects of mathematics, there are many unanswered questions. In this paper we focus specifically on connections between vocabulary in mathematical tasks and aspects of reading ability. A pilot study has been conducted (Bergqvist, Dyrvold, & Österholm, 2012) and experiences from that study have been used to refine the method for the present study. Both the present study and the pilot study use Swedish PISA tasks and the results for Swedish students as data.

Background

There are several reasons why it is of interest to know more about the relation between linguistic features of a mathematical task text and some aspect of the solving of the task. One reason is the general validity of the test: to assess mathematical competence and nothing else. Another reason is to shed light on whether some subgroup of students is disadvantaged by the presence of some linguistic aspect. Furthermore, it is important to gain information about which linguistic aspects that are, or could be, demanding for students. Besides aspects of difficulty and validity of tests, the relation between linguistic aspects of a task text and the solving of the task can inform us on the complex relation between language and mathematics.

Based on our focus on relations between linguistic aspects of mathematical tasks, in particular vocabulary, and aspects of the solving of such tasks, we first discuss empirical studies similar to our own. Thereafter, we discuss methodological and theoretical issues in such studies, and in particular we outline the theoretical perspectives used as basis for our study.

Linguistic aspects of tasks and solution frequency

In earlier research, the presence of various linguistic features in tasks and how such features may relate to the solution frequency, has been studied using different methods. Several linguistic features that could be expected to cause difficulty have empirically shown to be correlated with solution frequency. One example is a study focusing on solution frequency of mathematical tasks that are linguistically simplified by, for example, shortening of long nominals, changing passive verb forms to active, and clarifying relational and conditional clauses. These simplified tasks were solved to a higher frequency than the original tasks, especially by students that were less proficient in the language of the test (Abedi & Lord, 2001). A similar relation is found for science tasks in TIMSS (Trends in International Mathematics and Science Study) where the tasks that are solved to a lower frequency by students less proficient in the language

are characterized by nominalizations, passive voice, logical connectives, and many qualifiers (Dempster & Reddy, 2007). However, much is still not known concerning the relation between linguistic aspects and solution frequency, since earlier research is conflicting and the total picture is incomplete. For example, there are results showing that correlations between linguistic features of a task text and solution frequency of the task may occur for different features in different grades: A study investigating fifteen different linguistic features revealed that in grade ten only the presence of mathematics vocabulary in the tasks was significantly related to solution frequency, but in grade four also, for example, the presence of ambiguous words and passive voice had a statistically significant relation to solution frequency (Shaftel, Belton-Kocher, Glasnapp, & Poggio, 2006). We do not know where this difference stems from, but it indicates that it is wise not to draw too wide ranging conclusions from correlations between some linguistic aspect and solution frequency that occur for one particular category of students.

Vocabulary in tasks is one linguistic feature that has been studied in relation to solution frequency. The relation between reading and vocabulary has been studied through readability formulas or other methods with focus on different aspects of words that are presumed, or suspected, to be related to task or text difficulty. Aspects of words focused on in such studies are for example number of syllables, number of letters, and word difficulty/familiarity based on lists or expert judgements (e.g. Dempster & Reddy, 2007; Homan, Hewitt, & Linder, 1994; Shaftel et al., 2006; Shorrocks-Taylor & Hargreaves, 2000; Stahl, 2003). For mathematics tasks, one study found a negative correlation between the number of words that are classified by experts as ambiguous, as well as unusual or difficult mathematics vocabulary, and the solution frequency of the tasks (Shaftel et al., 2006). For science tasks from TIMSS, a study found that tasks that are solved to a lower frequency by students less proficient in the language of the test contain more words with multiple meaning (Dempster & Reddy, 2007). One aspect of vocabulary in tasks is how the amount of uncommon words in tasks may correlate to solution frequency, something that is part of the focus in the current study. A relation between the amount of uncommon words and solution frequency is revealed in a study where tasks in which uncommon words had been replaced, were solved to a higher frequency than the original tasks (Abedi & Lord, 2001). However, when comparing TIMSS science tasks that were solved to different degrees, unfamiliar words were not present to a significantly higher extent in the more difficult tasks (Dempster & Reddy, 2007). This somewhat contradictory picture indicates that the relation between uncommon or unfamiliar words and solution frequency is important to investigate further.

Methodological issues

Many studies use statistical methods to investigate the relation between different aspects of language in tasks and the students' performance on the tasks, but we argue that these methods often have serious limitations.

It is common to use statistical analysis to describe to what extent different linguistic properties of a task correlate with the difficulty of the task, but such analysis does not inform us on why. One concrete (but simplified) example of the limitations of this type of analysis is the following. Assume that the task variables *number of long words* and *solution frequency* correlate, showing that tasks with more long words are generally more difficult than tasks with fewer long words. One reason could be that students have problems reading long words and therefore have problems understanding the task. Another reason could be that cognitively advanced mathematical words often are longer. In this case, the students could have difficulty grasping the (mathematical) meaning of these words and therefore perform worse. However, that students perform worse when they solve more cognitively challenging mathematical tasks is completely reasonable and should be expected. Our conclusion is therefore that based on a statistical relationship between a variable describing a linguistic property of a task and students' performance on the task, we cannot conclude that this linguistic property is only related to reading ability.

When linguistic properties of tasks *do* relate to reading ability it is important to reflect upon which linguistic properties may be considered relevant or irrelevant for a mathematical task. When a linguistic feature is related to a mathematical communication competence, it could be argued that it is not only reasonable to include such a feature in mathematical tasks, but also important to include it if tasks are supposed to assess all competences that represent mathematical proficiency (see e.g. NCTM, 2000; Niss & Højgaard, 2011; OECD, 2006). It is not easy to separate between linguistic features that are related to mathematical competence and those that are not, but it is desirable for several reasons. First, information about unnecessary linguistic complexity (i.e. linguistic features not related to mathematical communication competence) is valuable in order to avoid such features in tests and thereby increase the validity of tests. Second, information about relevant linguistic features is interesting from a learning perspective, in order to create tasks that address all aspects of mathematical competence. One first step towards knowledge about these different types of linguistic features of mathematical tasks could be to not only rely on relationships between text features and solution frequency, but to utilize also students' results on a test of reading comprehension. Several studies have done so and have

used different statistical methods, primarily correlations and regressions in different ways. However, these types of methods have been examined by Österholm and Bergqvist (2012a) and the results show that methods found in other studies all have problems with aspects of validity or reliability. Based on those results, Österholm and Bergqvist (2012a) suggest an approach using a principal component analysis, which is shown to have good properties regarding both validity and reliability. This method, which creates a measure of a task's *demand of reading ability*, is used also in the current study, and is described in more detail in later sections.

A previous study, where the aforementioned principal component analysis was used to analyze PISA tasks, revealed significant correlations between both linguistic features and task type properties (e.g. if it is a multiple choice question) and the demand of reading ability (Österholm & Bergqvist, 2012b). The linguistic features that correlated with demand of reading ability were the proportion of long words (more than six letters) in the tasks and the information density (number of nouns divided by the number of verbs) of the tasks. However, the proportion of words in the tasks that were among the most common in two corpora did not correlate with demand of reading ability (Österholm & Bergqvist, 2012b). In conclusion, earlier research on common and uncommon words in tasks has come to different types of conclusions: Commonness of the vocabulary in mathematical tasks is not connected to the tasks' demand of reading ability (Österholm & Bergqvist, 2012b) and there is a connection between the amount of uncommon words and solution frequency in mathematical tasks (Abedi & Lord, 2001) but not in science tasks (Dempster & Reddy, 2007).

Theoretical issues

Many studies highlight different linguistic challenges for students in mathematics education. Some studies broadly characterize several different types of challenges (e.g. Austin & Howson, 1979; O'Halloran, 2008; Pimm, 1989; Schleppegrell, 2007) while other focus on specific issues, such as children's interpretations of lexical ambiguity (Durkin & Shire, 1991) or reading comprehension of symbols in mathematical texts (Österholm, 2006). Mathematical register is a notion that captures the types of specific linguistic properties of mathematical practice that are different from other practices (e.g. everyday language) concerning words and structures in the use of language (Halliday, 1978).

In her research review, Schleppegrell (2007, p. 141) summarizes some of the key linguistic features of the mathematical register, including the use of multiple semiotic systems and the use of specific grammatical

patterns, such as technical vocabulary and dense noun phrases. In the present study, we focus on vocabulary in relation to the mathematical register, concerning variations of vocabulary regarding how common words are in a mathematical context and in an everyday context. In particular, we are interested in math-specific words (words common in mathematics but uncommon in an everyday context), which are sometimes labeled as technical vocabulary (Schleppegrell, 2007). In research literature, these types of words are very often described as an essential part of what is special about the mathematical language (Österholm & Bergqvist, 2013). We are also interested in words that are not part of the mathematics register, since the use of such words in mathematical tasks could cause a decreasing validity of the task by demanding an irrelevant aspect of reading ability to fully comprehend the task.

In our study, we examine relations between the vocabulary of mathematical tasks and aspects of reading and solving of the tasks. However, we do not view reading and solving as two separate steps in the process when students are given mathematical tasks, in the way that for example Hegarty, Mayer, and Monk (1995) do. Instead, we view reading and solving as integrated, in particular because a separation between reading and solving tends to create a separation between reading and mathematics. This perspective is discussed in more detail elsewhere (see Bergqvist & Österholm, 2010). For the present study, when we discuss issues of solving mathematical tasks we include the whole process from when students are given a mathematical task to when they have finished working with the task. This process includes different types of activities, such as reading and calculating, and also includes the use of different types of abilities, such as reading ability and mathematical ability. It is questionable if these two abilities can be viewed as two (separate) things, since they both include many aspects of the ability in question, such as phonological, syntactical, and semantical aspects of reading (see e.g. Nation, 2005) and different competences of mathematics (see e.g. NCTM, 2000; Niss & Højgaard, 2011). However, for the present study we rely on a (simplified) model of students' reading ability and mathematical ability as two, partially overlapping, different types of abilities (see figure 1).

In the present study, we focus on issues of reading ability, that is, areas B and C in figure 1. Area B illustrates the part of reading ability that is relevant, and potentially specific, for mathematics. Area C illustrates the part of reading ability that is not part of mathematical ability. If a student must utilize the ability in area C to solve a task, this can be seen as a sign of lacking validity since this type of ability is not part of mathematical ability. Each arrow in figure 1 symbolizes how much of the variation of a task's solution frequency can be explained by a certain type of ability

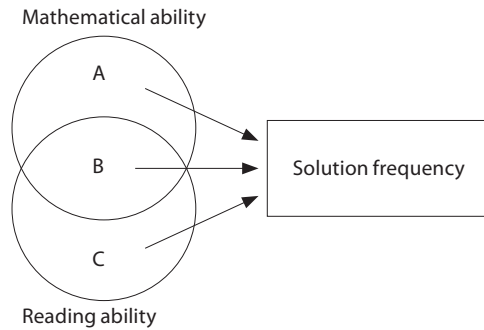


Figure 1. Schematic illustration of relations between abilities and of the meaning of demand of reading ability (the arrow from area C).

(area A, B or C). What we in the previous section have labeled as a task's demand of reading ability corresponds to the arrow from area C.

We want to examine if and how this type of demand of reading ability is related to variations in task properties regarding different types of uncommon vocabulary. We hypothesize that reading ability of technical vocabulary could be located in area B, while reading ability of vocabulary uncommon in a mathematical context could be located in area C. That is, we expect that the amount of technical vocabulary in a task text would not correlate with demand of reading ability, while the amount of vocabulary uncommon in a mathematical context could correlate with demand of reading ability.

Purpose of the study

The purpose of this study is to increase the understanding of aspects of reading and solving mathematical tasks and how those aspects are related to the commonness of the vocabulary used in tasks. In particular, we are interested in the relation between *different types of uncommon vocabulary* and *different aspects of difficulty* for mathematical tasks. As mentioned in the background, we focus primarily on two types of uncommon vocabulary: technical vocabulary, since it is a key feature of the mathematical register, and vocabulary that lies outside of the mathematical register, since the presence of such words could be related to lacking validity of the tasks. Regarding aspects of task difficulty, as mentioned in the background, we are particularly interested in *demand of reading ability*, but we also consider the more common variable *solution frequency*. To reach the purpose we focus on the following research question:

1. What are the characteristics of a possible connection between different aspects of task difficulty and the presence of different types of uncommon vocabulary?

To this we add a second research question that is relevant primarily as a methodological question. It enables us to compare different types of analyses and draw conclusions depending on possible differences between the results. Two aspects of the tasks are considered: leading text and unique words. The *leading text* of a task is an introductory text that is common for several sub-tasks (often a, b, c and so on). A methodological question is whether the leading text should be included or not in the analysis of each sub-task. Similarly, when examining the vocabulary in a task, some words are repeated several times in the task. The methodological question is whether recurring words should be included in the analysis once (i.e. counting only unique words) or once for every occurrence (i.e. counting all words). The second research question is therefore:

2. How does the existence of any connection between different aspects of task difficulty and the presence of different types of uncommon vocabulary depend on whether the leading text in tasks is included or not in the analyses, and whether unique or all words are analyzed?

Method

The method essentially consists of three steps. First, values for the variables *demand of reading ability* (DRA) and *solution frequency* are calculated for each mathematical task. Second, variables that describe the number and proportion of different types of uncommon vocabulary are calculated. Third, the correlations between the variables from the first and second step are analyzed. Each step is presented in more detail below.

The tasks analyzed in the study are Swedish PISA tasks from 2003 and 2006. PISA is chosen since it includes tasks measuring both mathematical and reading literacy and many students' performances on these tasks.

Demand of reading ability and solution frequency

The first variable concerning task difficulty is *demand of reading ability* (DRA). To measure a mathematics task's DRA, a *principal component analysis* (PCA) is used. This method is presented and discussed in more detail in a previous paper (Österholm & Bergqvist, 2012a). A PCA converts a set of observations of variables into a set of new variables, called principal components. The components are constructed in such a way

that the first principal component explains as much of the variation in the data as possible, and each subsequent component explains as much of the remaining variation as possible. In this study, all Swedish students' results on all PISA mathematical literacy tasks and reading literacy tasks from 2003 and 2006 are entered into the PCA. We use Promax as the method for rotation (an oblique rotation, since we expect the components to correlate) and only the first two components, which are expected to correspond to the two abilities of mathematics and reading, are extracted. From this analysis, each mathematics task receives a loading value for each of the two components. The loading value on the reading component is taken as a measure of the demand of reading ability since it can be interpreted as a measure of the genuine effect of reading ability when the effect of mathematical ability has been excluded (see Tabachnick & Fidell, 2006). Among the 84 PISA mathematics tasks, the PCA resulted in a positive loading value on the reading component for 63 of the tasks. Only these 63 tasks are used in the analyses in this study since we focus on a presence of a *demand of reading ability* in tasks, and tasks with a negative loading value are seen as qualitatively different types of tasks.

The second variable concerning task difficulty is *solution frequency*, which is calculated by dividing the total score for all students who attempted to solve a PISA task with the highest possible total score for those students on that task.

Different types of uncommon vocabulary

The second step of the method is to create variables that describe the number and proportion of different types of uncommon vocabulary. This study primarily focuses on two types of uncommon vocabulary: technical vocabulary and vocabulary that lies outside of the mathematical register. Here we define technical vocabulary as words common in a mathematical context but uncommon in an everyday context. The second type of vocabulary is defined here as words uncommon in a mathematical context. For each word in each Swedish PISA task, we therefore determine if the word is common or uncommon in both a mathematical and an everyday context, which gives four categories of words (see table 1).

The reason we calculate both number and proportion of the different types of words is connected to aspects of reliability described in detail in the section Correlation analyses. Thus, for every task, eight variables are created that describe the *number* and *proportion* of words in each of the four categories. This procedure is described in detail in the following subsections.

Categories of words

To get an unbiased measure of how common the words used in PISA are in different contexts we use two corpora. Therefore, what in this article are referred to as the mathematical and the everyday context, are in practice two different corpora. A corpus is "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (Sinclair, 2005, p. 23). The corpus we use to represent the written everyday language that is familiar to the test takers is available through Språkbanken¹ (The Swedish Language Bank) at the University of Gothenburg. Our corpus is composed of 81 novels² (7.2 million words), newspapers³ (237.4 million words) and blog texts⁴ (344.8 million words). In the analysis, for each word we take the relative frequencies obtained from the three different corpora and calculate the mean value, which means that each corpus has equal impact on the frequencies we use for the words. The corpus used to represent the written mathematical language that is familiar to the test takers consists of two mathematics textbooks intended for year 8 students (the same age group as the students that take the PISA tests; about 100,000 words), which are part of the OrdiL project (Lindberg & Kokkinakis, 2007). With these two corpora as references, we categorize every word in the tasks as common or uncommon, which gives us four categories of words (see table 1). The category Ueveryday corresponds to technical vocabulary, and both categories UU and Umath (middle column in table 1) correspond to vocabulary that lies outside of the mathematical register. Since the categories UU and Umath are qualitatively different, we perform the analyses with these two categories separately. Also, even though we are mostly interested in the three categories Ueveryday, UU, and Umath (based on our rationale for the study), since our method results in four categories we will perform the same analysis on all four, for comparison.

Table 1. *Categorization of the words in the mathematics tasks, according to their frequencies in the corpora*

	Uncommon in mathematics corpus	Common in mathematics corpus
Uncommon in everyday corpus	Words uncommon in both corpora (UU)	Words common in mathematics, uncommon in everyday corpus (Ueveryday)
Common in everyday corpus	Words uncommon in mathematics, common in everyday corpus (Umath)	Words common in both corpora (CC)

When dealing with the specialized vocabulary of mathematics, two types of words are often noted: "words that exist only in mathematics" and "borrowed/familiar words that are used in a special sense or manner" (Österholm & Bergqvist, 2013, p.10). The second kind is not examined in the present study since our method cannot distinguish between when the same word is used in different manners.

Words included in the analysis

This study focuses on the vocabulary in tasks, that is, the words included in task texts. It is not easy to give a strict, and also functional and relevant, definition of the notion of word. Here we rely on a more colloquial type of delimitation of what a word is: a combination of letters that is a linguistic element that can be found in a dictionary, including compounds using a hyphen (e.g. "two-thirds"). Therefore, the following types of (combinations of) characters are not included in the analysis: numbers or other symbols, Roman numerals, combinations of letters denoting variables, designations of objects, abbreviations, and response options (e.g. "x", "A", or "CDE"). According to our delimitation of what a word is, different mixtures of upper and lower case letters (e.g. "woRd", "word", and "WORD") should be seen as the same word. However, due to technical limitations in the search procedure, words with unusual mixtures of upper and lower case letters (e.g. "woRd") are treated as separate words, but "Word", "word", and "WORD" are treated as the same word.

Categorization of words and creation of variables

The analysis of word frequency in each corpus gives us two values for every word, used to label each word as common or uncommon in the two corpora respectively (see table 1). How common words are refers in this study to commonness relative to how common the other words used in the PISA tasks are. A *relative* limit is chosen since no real limit for when a word is uncommon exists and since we are interested in differences between the words used. The limit is set at the median frequency for all the analyzed words, one median for each corpus. Words with higher frequency than the median are categorized as common in that corpus and words that occur to a lower frequency as uncommon. Words with the same frequency as the median are categorized in the same way as the category that has the least number of words when the words with the same frequency as the median are not included. Instead of using the median as a limit and dividing the words in common and uncommon it would be possible to use, for example, three categories to get a more fine grained classification. Words with frequencies close to the median could also have been excluded to obtain a clearer

difference between the groups. Still, we use the median as a limit and two categories since the number of words analyzed are limited and the advantage of more categories or excluded words is outweighed by having more words in the analysis. According to the two limits, one in each corpus, we sort the words into *four categories* (see table 1). These four categories of words are used to define eight different variables for every task: *number* and *proportion* of words in each of the categories UU, Ueveryday, Umath, and CC. An example of the analysis is found in appendix A.

Correlation analyses

The final step of the method is to perform correlation analyses between the two variables regarding task difficulty (*demand of reading ability* and *solution frequency*) and the eight variables regarding type of vocabulary (*number* and *proportion* of each of the categories UU, Ueveryday, Umath, and CC). We use two-tailed non-parametric partial correlations (Spearman R coefficient) with a significance level of .05.

In the correlation analyses we do not want to measure a potential effect of the total number of words in tasks. This risk is evident since the total number of words in tasks in some cases correlates with the number of words in a specific category, with the proportion of words in a specific category, with solution frequency, and with demand of reading ability (see table B1 in appendix B for correlation coefficients). Therefore, we use partial correlations in all analyses, where the total number of words in tasks is controlled for, in order not to measure any indirect effect of the total number of words in tasks. When the total number of words in tasks is controlled for, variables describing *number* and *proportion* of words can be said to measure the same thing. Therefore, significant correlations are seen as reliable if they occur for both these types of variables and this is the reason why we calculate both number and proportion of the different types of words.

Four different analyses

Most of the methodological choices made in this study are based on the purpose and the first research question. Some choices are based more on what is possible, for example, the availability of different types of corpora. In two cases, however, exactly two versions of the method are possible, and neither is more clearly connected to the purpose. In these cases we choose to do the analysis in both possible ways, to answer the second research question: How does the existence of any connection

between different aspects of task difficulty and the presence of different types of uncommon vocabulary depend on whether the leading text in tasks is included or not in the analyses, and whether unique or all words are analyzed? We therefore perform the analyses both with and without the leading text of the tasks. Also, when counting words in different categories for a task, we first count all words (i.e. a repeated word is counted as many times as it is repeated) and then only unique words (i.e. a repeated word is counted only once). Altogether, two versions for each of two choices give us four different analyses that each include the correlations between two variables of task difficulty and eight variables regarding type of vocabulary.

Validity and reliability

A previous methodological study (Österholm & Bergqvist, 2012a) shows that there is good validity and reliability when using a principal component analysis (PCA) to measure PISA mathematics tasks' demand of reading ability (DRA). Regarding validity, the use of a PCA created an anticipated clear division into two components, where almost all reading tasks were placed in one component and most of the mathematics tasks were placed in the other component. When it comes to reliability of this method, the characterization of PISA tasks regarding their DRA proved to be very consistent when comparing analyses based on students' results from PISA 2003 with analyses based on students' results from PISA 2006.

Since creating a corpus for use in research is a very complex task, we rely on existing corpora that are not ideal but we argue that they are representative enough for our study. The everyday corpus represents the everyday language of an average citizen, and the *written* language fifteen year olds meet can be assumed to be reasonably close to the texts chosen in our everyday corpus. In particular, we include three different types of texts: newspapers, blog-text, and novels. The mathematical corpus is composed of two mathematical textbooks, which represent the written mathematical language that is familiar to the test takers.

One problem with the mathematics corpus is that it is small. Therefore, there is a pronounced flooring effect, which means that we cannot distinguish between those words that do not appear in the corpus since all have zero frequency. However, the flooring effect does not directly affect the classification of words as common or uncommon since we classify according to the median frequency, which is larger than zero. Furthermore, there is consistency in word frequency between the two mathematics books ($r=.79$ in table 2), indicating that there is

homogeneity in the mathematics corpus. The similarities between the mathematics corpus and the everyday corpus is smaller ($r < .69$) than the similarities within each corpus ($r = .79$ for mathematics corpus and $r > .94$ for everyday corpus). Thus, despite its size and the fact that the two mathematics books cover slightly different content, the mathematics corpus represents one distinct register.

Table 2. *Correlations (two-tailed, non-parametric) between frequencies of the analyzed words (N=1336) obtained from different corpora (all significant at level .001)*

Measure	1	2	3	4	5
1. Math book 1	–				
2. Math book 2	.786	–			
3. Newspapers	.634	.689	–		
4. Novels	.624	.672	.943	–	
5. Blogs	.645	.688	.961	.946	–

Results

In this article we study the relation between different types of uncommon vocabulary and different aspects of difficulty for mathematical tasks. In this section we present four different analyses of correlations between eight task variables regarding type of vocabulary (number and proportion of words from the categories UU, Ueveryday, Umath, and CC) and two variables of task difficulty (*demand of reading ability* and *solution frequency*). Note that *technical vocabulary* is represented by the variable Ueveryday and that the categories UU and Umath together correspond to *vocabulary that lies outside of the mathematical register* (see table 1 for all categories of words).

All correlations presented here are partial correlations, where total number of words is controlled for, even though they are mostly referred to just as *correlations*. Results from all analyses are found in table B2 in appendix B. There are no significant correlations between any of the variables representing number and proportion of words from different categories and the variable *solution frequency*, and therefore we only use the significant correlations with demand of reading ability (DRA) to present and compare the results from the four analyses, see table 3.

The proportion of words that are uncommon in the everyday corpus and common in the mathematics corpus (Ueveryday, technical vocabulary), significantly correlates with DRA in one of the analyses, with a

Table 3. Significant partial correlations between both number and proportion of words in different categories and demand of reading ability in all four variants of analyses ($N=63$)

	With leading text	Without leading text
Unique words	Number of UU (.39**) Proportion of UU (.35**) Proportion of Ueveryday (-.26*)	Number of UU (.33**) Proportion of UU (.30*)
All words	Number of UU (.37**) Proportion of UU (.32*)	Number of UU (.39**) Proportion of UU (.36**) Number of CC (-.28*) Proportion of CC (-.35**)

Note. * $p < .05$. ** $p < .01$. *** $p < .001$

negative correlation coefficient. We see this correlation as coincidental, since the correlation only occurs for proportion of words in the category, and not for number of words (see section Correlation analyses). Reliability is found in the correlations between DRA and the number and proportion of words in the category words uncommon in both corpora (UU), where a correlation is also found in all four analyses. The correlation between the category of words that are common in both corpora (CC) and DRA is the only reliable correlation that does not occur in all four variants of the analysis. This correlation is only significant in the analysis without leading text and with all words included in the analysis. The fact that this correlation only occurs in one of the analyses indicates that the difference does not derive from either the difference between the analyses with or without leading text, or with all or unique words alone. Instead, it occurs according to some type of interaction, which is difficult to interpret, between a design of method where all words in the task text are analyzed but the leading text is excluded from the analysis. Thus, there are no unambiguous differences in the results between analyses with or without leading text or between analyses with all or unique words.

Conclusions

Based on our analyses, we can answer our research questions accordingly. The first question concerns the characteristics of a possible connection between aspects of task difficulty (*demand of reading ability* and *solution frequency*) and the presence of different types of uncommon vocabulary. The results show that there is a connection between demand of reading ability (DRA) and commonness of vocabulary, but no connection between

solution frequency and commonness of vocabulary. This conclusion is based on the fact that there is no correlation between amount of words in any category and solution frequency, but reliability in correlations between one category of uncommon words (UU) and DRA. Furthermore, the result shows only coincidental correlations between amount of technical vocabulary, that is, words that are uncommon in the everyday corpus and common in the mathematics corpus (Ueveryday), and aspects of task difficulty. Based on the above, we therefore draw the conclusion that the connection between DRA and the presence of different types of uncommon vocabulary is mainly created by the types of words that are *globally* uncommon (i.e. uncommon in both a mathematical context and an everyday context).

The second research question concerns how the existence of any connection between aspects of task difficulty and the presence of different types of uncommon vocabulary depends on whether the leading text in tasks is included or not in the analyses, and whether unique or all words are analyzed. The results show that overall there are no differences depending on these methodological choices, since for all four methodological variants of the analyses, almost all correlations in our analyses are of similar type (either significant or nonsignificant).

Discussion

No category of words correlates with solution frequency of the tasks. This lack of correlation means that none of the categories of words appear to a significantly higher (or lower) amount in tasks solved to a lower frequency. We note that earlier research has shown a relation between solution frequency and uncommon words (Abedi & Lord, 2001) and also between solution frequency and unusual or difficult mathematics vocabulary (Shaftel et al., 2006). However, in the latter study, expert judgments with the following coding instructions are used: "unusual or difficult but specific mathematics vocabulary words" (Shaftel et al., 2006, p. 126), that is, there is no single categorization of words based on commonness. Abedi and Lord (2001, p. 221) on the other hand, focused on "unfamiliar or infrequent" vocabulary that was not mathematics vocabulary. In their study, the potentially difficult linguistic features analyzed were not analyzed separately. The relation between solution frequency and uncommon vocabulary is actually a relation between solution frequency and linguistically modified tasks, where uncommon vocabulary is one of several features that were modified (Abedi & Lord, 2001). In summary, we see that the differences between the results of those two studies and the current study can be due to differences in method since the

relations between solution frequency and some type of unfamiliarity of vocabulary are relations of quite different type in all three studies.

Our empirical results concerning connections between different types of uncommon vocabulary and demand of reading ability are in line with our theoretical assumptions. In particular, there is no reliable correlation between the amount of technical vocabulary and demand of reading ability, which supports the assumption that the reading of technical vocabulary is part of a mathematical reading ability. Furthermore, there is a reliable correlation between the amount of vocabulary that is uncommon in a mathematical context and demand of reading ability, which supports the assumption that the reading of this type of vocabulary is part of an irrelevant type of reading ability, that is, the part of reading ability that is not part of mathematical ability. However, it is not all types of vocabulary uncommon in a mathematical context that are connected to demand of reading ability. It is the category of *globally uncommon vocabulary* (UU) that is correlated to demand of reading ability in all four analyses (see table 4), which means that for tasks with a high amount of those words, reading ability is needed, or at least can be utilized, when solving the task. We note the difference between this result and a previous study (Österholm & Bergqvist, 2012b) where no relation was found between the proportion of common words and demand of reading ability. However, there are differences in methodology between the previous and the present study: In the previous study, where no significant correlations were found, one type of corpus at a time was utilized in the analyses, while in the present study, the interaction between commonness in different corpora is used. Similarly as in the previous study, in the present study there are no significant correlations for words uncommon only in one corpus, but in the present study there are significant correlations for words that are uncommon in both corpora (mathematics and everyday).

Presence of globally uncommon vocabulary (UU) significantly correlates with demand of reading ability but not with solution frequency. This result may be interpreted as a presence of a linguistic factor (the uncommon words) that causes a higher demand of reading ability in tasks, but at the same time does not affect the solution frequency. That is, the amount of globally uncommon words (UU) in a task seems not to be a crucial property in relationship to students' ability to solve the task, but that this property of the task is primarily connected to what type of ability that is possible to utilize when solving the task. Furthermore, it is important to remember that our measure of demand of reading ability consists of an aspect of reading ability that is not overlapping with mathematical ability. Therefore, we can conclude that tasks with more globally uncommon words have lower validity since a mathematically irrelevant

type of reading ability can be utilized to a higher degree by students when solving these tasks. More generally, the difference between the results concerning demand of reading ability and solution frequency is also a reminder of the theoretical and methodological complexity in studies that investigate relations between aspects of difficulty of mathematical tasks and some linguistic feature (cf. Österholm & Bergqvist, 2012a).

We have concluded that there is a correlation between the amount of certain types of words in mathematical task texts and the demand of a type of reading ability that is not part of mathematical ability. As in all studies of correlations, it is not possible to directly draw conclusions about causality, even though some results can be interpreted as such (e.g. as in our discussion above). Therefore, in relation to the present study, other types of studies are also needed concerning the phenomenon in question, to more directly examine aspects of causality. For example, qualitative studies focusing on the process of solving mathematical tasks could be utilized to examine if and how different types of words affect this process. However, if the results from the present study are assumed to reflect a causality, a direct implication of the results is a recommendation to avoid globally uncommon vocabulary (UU) in mathematical tasks since these types of words create a demand of reading ability that is not part of mathematical ability, that is, the use of these types of words decreases the validity of mathematical tasks.

Besides this more practical implication based on the results of the present study, there are also two important methodological implications. First, the main result focuses on globally uncommon words, which have been possible to examine through the simultaneous use of two different corpora (i.e. by examining commonness of words in two different contexts). That is, studies about uncommon words need to specify in which context the words are uncommon, and also that it seems essential to examine words that are uncommon in two relevant contexts. Second, the main result focuses on demand of reading ability, which has allowed us to more directly examine aspects of decreasing validity of mathematical tasks, since it is measuring the part of reading ability that is not part of mathematical ability.

Previous empirical studies have most often only examined uncommon words more broadly, either by relying on manual coding (e.g. Shaftel et al., 2006) or with reference only to a type of everyday context (e.g. Helwig et al., 1999), and/or have most often only utilized solution frequency as a measure of task difficulty (e.g. Abedi & Lord, 2001). Our point here is that solution frequency is a blunt instrument since it is a measure of all kinds of knowledge and abilities that students utilize to solve the task. Commonness of vocabulary measured either by manual coding or by

utilizing only one corpus is also a blunt instrument. In particular, the category of technical vocabulary includes words that are uncommon in an everyday context but more common in a mathematical context and also since the results in the present study highlight the category of globally uncommon words as important.

Finally, certain limitations of the present study need to be mentioned. Our method using computerized analyses has its benefits, in particular that large amount of data can be analyzed very quickly in a reliable manner. However, there are also technical limitations that can cause a decrease in validity. For example, in our analyses we could not take into account lexical ambiguity (i.e. words spelled the same way but with different meanings). Therefore, there is a need for similar types of studies utilizing other methods of analyses, either manual analyses or by using more advanced computer software.

Another type of limitation is that the data from PISA utilized in this study focuses on certain aspects of reading ability and mathematical ability through the specific tasks and the specific setting (in particular, mathematical word problems within a standardized test situation). Therefore, there is a need for similar types of studies utilizing the same, similar, or very different types of tasks or settings to draw any more general types of conclusions concerning relations between different types of uncommon vocabulary and aspects of difficulty for mathematical tasks. However, in such studies it is essential to examine commonness in both an everyday and a mathematical context and also to examine not only solution frequency in order to be able to draw conclusions about the level of validity of the tasks.

References

- Abedi, J. & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14 (3), 219–234.
- Austin, J. L. & Howson, A. G. (1979). Language and mathematical education. *Educational Studies in Mathematics*, 10, 161–197.
- Bergqvist, E., Dyrvold, A. & Österholm, M. (2012). Relating vocabulary in mathematical tasks to aspects of reading and solving. In C. Bergsten, E. Jablonka & M. Raman (Eds.), *Evaluation and comparison of mathematical achievement: dimensions and perspectives. Proceedings of Madif 8* (pp. 61–70). Linköping: SMDF. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-51411>

- Bergqvist, E. & Österholm, M. (2010). A theoretical model of the connection between the process of reading and the process of solving mathematical tasks. In C. Bergsten, E. Jablonka & T. Wedege (Eds.), *Mathematics and mathematics education: Cultural and social dimensions. Proceedings of MADIF 7* (pp. 47–57). Linköping: SMDF. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:um u:diva-31890>
- Dempster, E. R. & Reddy, V. (2007). Item readability and science achievement in TIMSS 2003 in South Africa. *Science Education*, 91, 906–925.
- Durkin, K. & Shire, B. (1991). Primary school children's interpretations of lexical ambiguity in mathematical descriptions. *Journal of Research in Reading*, 14(1), 46–55.
- Halliday, M. A. K. (1978). *Language as social semiotic: the social interpretation of language and meaning*. London: Edward Arnold.
- Hegarty, M., Mayer, R. E. & Monk, C. A. (1995). Comprehension of arithmetic word problems: a comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, 87(1), 18–32.
- Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B. & Almond, P. J. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *The Journal of Educational Research*, 93(2), 113–125.
- Homan, S., Hewitt, M. & Linder, J. (1994). The development and validation of a formula for measuring single-sentence test item readability. *Journal of Educational Measurement*, 31(4), 349–358.
- Lindberg, I. & Kokkinakis, S. J. (Eds.). (2007). *OrdiL – en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år*. University of Gothenburg. Retrieved from <http://hdl.handle.net/2077/20503>
- Nation, K. (2005). Children's reading comprehension difficulties. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: a handbook* (pp. 248–265). Malden: Blackwell Publishing.
- NCTM. (2000). *Principles and standards for school mathematics*. Reston: National Council of Teachers of Mathematics.
- Niss, M. & Højgaard, T. (Eds.). (2011). *Competencies and mathematical learning: Ideas and inspiration for the development of mathematics teaching and learning in Denmark*. Roskilde University. Retrieved from http://milne.ruc.dk/ImfufaTekster/pdf/485web_b.pdf
- OECD. (2006). *Assessing scientific, reading and mathematical literacy: a framework for PISA 2006*. Paris: Author.
- O'Halloran, K. (2008). *Mathematical discourse: language, symbolism and visual images*. London: Continuum.
- Pimm, D. (1989). *Speaking mathematically: communication in mathematics classrooms* (paperback edition). London: Routledge.
- Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: a research review. *Reading & Writing Quarterly*, 23(2), 139–159.

- Shaftel, J., Belton-Kocher, E., Glasnapp, D. & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment*, 11(2), 105–126.
- Shorrocks-Taylor, D. & Hargreaves, M. (2000). Measuring the language demands of mathematics tests: the case of the statutory tests for 11-year-olds in England and Wales. *Assessment in Education*, 7(1), 39–60.
- Sinclair, J. (2005). Corpus and text – basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: a guide to good practice* (pp. 5–24). Oxford: Oxbow books.
- Stahl, S. A. (2003). Vocabulary and readability: how knowing word meanings affects comprehension. *Topics in Language Disorder*, 23(3), 241–247.
- Tabachnick, B. G. & Fidell, L. S. (2006). *Using multivariate statistics* (Vol. 5 rev. ed.). Boston: Allyn and Bacon.
- Österholm, M. (2006). Characterizing reading comprehension of mathematical texts. *Educational Studies in Mathematics*, 63, 325–346.
- Österholm, M. & Bergqvist, E. (2012a). Methodological issues when studying the relationship between reading and solving mathematical tasks. *Nordic Studies in Mathematics Education*, 17(1), 5–30.
- Österholm, M. & Bergqvist, E. (2012b). What mathematical task properties can cause an unnecessary demand of reading ability? In G. H. Gunnarsdóttir, F. Hreinsdóttir, G. Pálsdóttir, M. Hannula, M. Hannula-Sormunen, et al. (Eds.), *Proceedings of Norma 11, the sixth Nordic Conference on Mathematics Education* (pp. 661–670). Reykjavík: University of Iceland Press. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-39699>
- Österholm, M. & Bergqvist, E. (2013). What is so special about mathematical texts? Analyses of common claims in research literature and of properties of textbooks. *ZDM – The International Journal on Mathematics Education*, 45(5), 751–763.

Notes

- 1 <http://spraakbanken.gu.se>
- 2 58 of the novels from the 1990s and the other 23 novels are the novels that were published at Norstedts Agency in year 1999.
- 3 Göteborgs-Posten from year 1994 and 2001–2011.
- 4 The most popular blogs in Sweden according to top lists on www.bloggportalen.se

Appendix A

Example task

The task Coloured Candies is a released task from PISA. It is displayed to explain the method used. The task text is English here, but in the example of the analysis, the actual Swedish words used in the sentence in the Swedish test are shown. In our analysis, every word in the task is analysed, also the names of colours in the diagram. Nothing is analysed from the response alternatives here since they only contain numbers, signs and single letters that are not actual words. The word "Robert's" would not be analysed since it contains an apostrophe, but genitive is written without apostrophe in Swedish, so also "Roberts" is analysed. The task title, Coloured candy, is not analysed, neither is the word "Question".

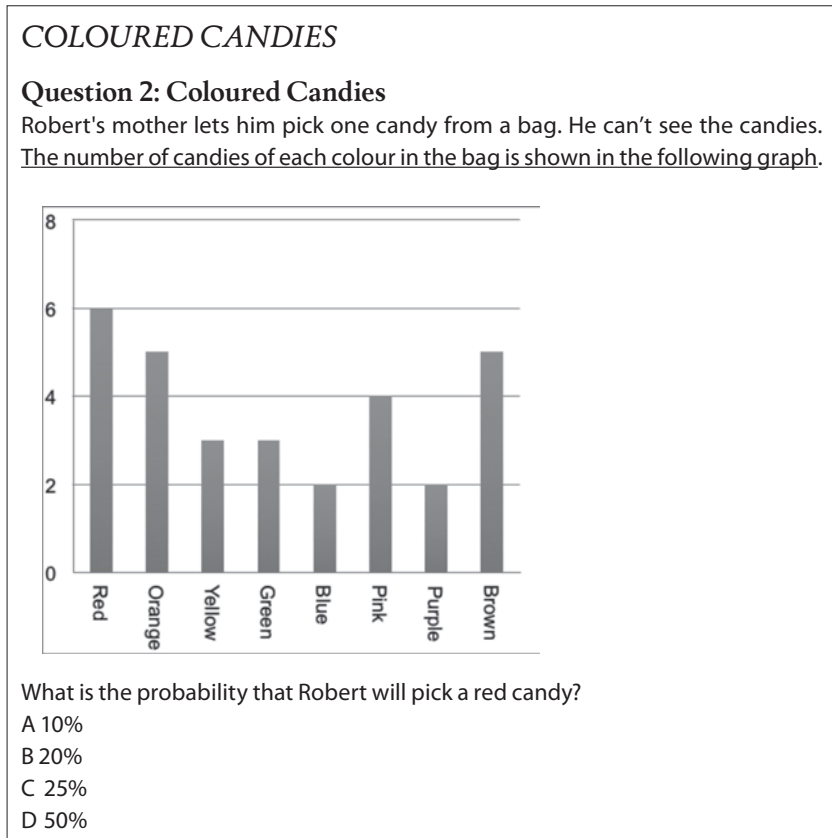


Figure A1. Example task from PISA, task number M467Q01 (The task can be retrieved from <http://www.oecd.org/pisa/38709418.pdf>).

The underlined sentence is used to show how variables for commonness are created. In Table A1, the frequencies for the Swedish words' occurrences are displayed (translated in parenthesis). The division of words below or above the median frequency leads to the division of the words in the four categories (Table A2). The amount of words in each of those four categories gives values for the variables of different types of commonness.

Table A1. *Frequencies for the words in the example sentence in both corpora*

Frequencies for words in the mathematics corpus			Relative frequencies* for words in the everyday corpus	
Word	(in english)	Frequency	Word	Frequency
godisbitar	(candies)	0	godisbitar	0.65
		(median)	diagram	4.84
visas	(is shown)	6		(median)
färg	(colour)	21	visas	78.32
diagram	(graph)	22	följande	151.32
följande	(the following)	60	färg	217.53
finns	(there is)	112	varje	1318.93
varje	(every)	155	många	2791.97
många	(many)	603	finns	3438.93
det	(it)	926	hur	5095.99
hur	(how)	1652	det	58805.72
i	(in)	1983	i	60054.70

Note. * Occurrences per one million words.

Table A2. *The words in the example sentence divided into the four categories*

Uncommon in mathematics	Common in mathematics
Uncommon in everyday	godisbitar
Common in everyday	diagram
	visas, följande, färg, finns, varje, många, det, hur, i

When the words are divided into the four categories, variables describing number of or proportion of words in each category are created. To simplify, we see the example sentence as one task here.

Table A3. *Values for the variables in the analysis example*

Variable name and explanation in parenthesis	Values for the sentence "as a task"	
	Number of words in each category	Proportion of words in each category*
UU (uncommon in both corpora)	1	.091
Ueveryday (uncommon only in the everyday corpus)	1	.091
Umath (uncommon only in the mathematics corpus)	0	0
CC (common in both corpora)	9	.818

Note. * Since we only analyze one sentence, we get the proportion by dividing with 11 here.

Appendix B

All Correlations

Table B1 displays correlations between categories of words and total number of words in tasks. The decision to use partial correlations where total number of words is controlled for is based on the results in Table B1. Table B2 displays all partial correlations between categories of words and difficulty as well as demand of reading ability (DRA) in all variants of analyses. That is, only the significant correlations in Table B2 are presented in the result section. All correlation analyses are two-tailed non-parametric correlations with Spearman R coefficient.

Table B1. *Correlations between number of/proportion of words in the four categories and total number of words in tasks in four variants of analyses (N=63)*

Version of analysis	First variable	Number of				Proportion of			
		UU	Umath	Ueveryday	CC	UU	Umath	Ueveryday	CC
Unique words, no leading text	Total no. words	.83***	.68***	.67***	.97***	.22	.36**	.18	-.38**
Unique words, with leading text	Total no. words	.76***	.59***	.49***	.96***	.17	.12	.03	-.26*
All words, no leading text	Total no. words	.83***	.69***	.67***	.97***	.20	.32*	.14	-.32*
All words, with leading text	Total no. words	.77***	.57***	.48***	.96***	.11	.06	-.05	-.02

Note. The categories of words are: words uncommon in both the everyday and the mathematical corpus (UU), words common in both the everyday and the mathematical corpus (CC), words uncommon in the everyday corpus and common in the mathematical corpus (Ueveryday) and words uncommon in the mathematical corpus and common in the everyday corpus (Umath) (see also table 1). * $p < .05$. ** $p < .01$. *** $p < .001$

Table B2. Results from all four partial correlation analyses (controlled for total number of words) between demand of reading ability (DRA) and the categories of words and between difficulty and the categories of words (N=63)

Version of analysis	Task variable	Number of				Proportion of			
		UU	Umath	Uevery day	CC	UU	Umath	Uevery day	CC
Unique words, no leading text	DRA	.334**	-.014	-.147	-.217	.297*	-.011	-.158	-.242
Unique words, with leading text		.391**	.051	-.229	-.123	.353**	.056	-.260*	-.203
All words, no leading text		.388**	-.025	-.135	-.283*	.357**	-.027	-.134	-.346**
All words, with leading text		.369**	.016	-.200	-.151	.323*	.004	-.237	-.242
Unique words, no leading text	Diffi-culty	-.010	.070	.053	-.038	.015	.079	.034	-.067
Unique words, with leading text		-.033	.196	.038	.014	.011	.134	-.001	-.100
All words, no leading text		-.158	.034	.050	.119	-.116	.048	.003	.094
All words, with leading text		-.108	.151	.023	.089	-.038	.121	-.031	-.016

Note. * $p < .05$. ** $p < .01$. *** $p < .001$

Anneli Dyrvold

Anneli Dyrvold is a PhD student in mathematics education at the Department of Mathematics and Mathematical Statistics at Umeå University. She is one of the students in the doctoral programme: The language of schooling in mathematical and science practices (<http://forskning.edu.uu.se/langmathscience/>). She is also a member of Umeå Mathematics Education Research Centre (UMERC). Her research interest is the affordances, and distinguishing characteristics, of the language of mathematics and what students need to learn in order to master it.

anneli.dyrvold@umu.se

Ewa Bergqvist

Ewa Bergqvist is an assistant professor in mathematics education at the Department of Science and Mathematics Education at Umeå University. She is a member of Umeå Mathematics Education Research Centre (UMERC) and teaches mathematics education for pre-service mathematics teachers. Her research focuses mainly on aspects of reasoning and language in upper secondary and university level mathematics.

ewa.bergqvist@umu.se

Magnus Österholm

Magnus Österholm is a docent (associate professor) in mathematics education and works at the Department of Science and Mathematics Education at Umeå University. He is also a member of Umeå Mathematics Education Research Centre (UMERC). His research interests deal primarily with mathematics education at the upper secondary and university levels, where cognitive and metacognitive perspectives are of special interest, together with studying language and communication in the learning and teaching of mathematics.

magnus.osterholm@umu.se

