Difficult to read or difficult to solve?

The role of natural language and other semiotic resources in mathematics tasks

Anneli Dyrvold

To my mother

Abstract

When students solve mathematics tasks, the tasks are commonly given as written text, usually consisting of natural language, mathematical notation and different types of images. This is one reason why reading and interpreting such texts are important parts of being mathematically proficient, at least within the school context. The ability utilized when dealing with aspects of mathematical text is denoted in this thesis as *a mathematical reading ability*; this ability is useful when reading mathematical language, for example, in task text. There is, however, a lack of knowledge of what characterizes this mathematical language, what students need to learn regarding the mathematical language, and exactly which mathematical language that tests should preferably assess. Therefore, the purpose of this thesis is to contribute to the knowledge of aspects of difficulty related to textual features in mathematics tasks. In particular, one aim is to distinguish between a difficulty that has to do with a mathematical ability and another that has not. Different types of text analyses are utilized to capture textural features that might be demanding for the students when reading and solving mathematics tasks. Aspects regarding vocabulary are investigated both in a literature review and in a study where corpora are used to analyse word commonness. Other textual analyses focus on textual features that concern mathematical notation and images, besides natural language. Statistical methods are used to analyse potential relations between the textual features of interest and both task difficulty and task demand on reading ability. The results from the research review are sparse regarding difficult vocabulary, since few of the reviewed studies analyses word aspects separately. Several of the analysed textual features are related to aspects of difficulty. The results show that tasks with more words that are uncommon both in a mathematical context and in an everyday context, may favour students with good reading ability rather than students with good mathematical ability. Another textual feature that is likely to be demanding for students, is if the task texts contains many meaning relations, for example, when several words refer to the same or similar object. These results have implications for the school practice both regarding textual features that are important from an educational perspective and regarding the construction of tests. The research does also contribute to an understanding of what characterizes a mathematical language.

Table of Contents

Acknowledgements/Författarens tack

Jag vill börja med att rikta ett stort tack till mina handledare Ewa Bergqvist och Magnus Österholm, som genom sitt sätt att handleda gjort min forskarutbildning till en förstklassig utbildning. Tack för alla rättframma och djupgående handledningssamtal där huvudfokus varit att föra forskningen framåt och att jag ska lära. Tack också för ert *genuina* engagemang i min forskarutbildning och i den forskning jag genomfört. Jag vill även tacka min tredje handledare. Tack Johan Lithner, för dina värdefulla råd och för kritisk läsning av mina texter. Tack också för att du som föreståndare för *Umeå Forskningscentrum för Matematikdidaktik* (UFM) i olika typer av beslut lägger stor vikt vid att doktorandens perspektiv ska beaktas. Det är stort. Och klokt.

Mitt första möte med den matematikdidaktiska forskarmiljön i Umeå var vid det årliga forskarmötet inom UFM (retreaten) som jag besökte någon månad innan jag påbörjade mina doktorandstudier. Min omedelbara känsla när jag tog del i de diskussioner som fördes i gruppen var att "jag har kommit hem". Det låter möjligen klyschigt, men får så göra eftersom det verkligen är sant. Därför, tack till alla ni som tillsammans bidar till att göra UFM till det det är. Tack också till er licentiander, doktorander och forskare inom UFM som kritiskt diskuterat min forskning vid seminarier.

En gedigen läsning av min framväxande avhandling bidrog också Andreas Ryve med. Tack, Andreas för en givande 90%-opponering och för konstruktiv feedback på kappans upplägg och på artiklarna.

Den forskarskola jag deltagit i, *Ämnesspråk i matematiska och naturvetenskapliga praktiker* har jag också att tacka för mycket, bland annat för möjligheten att genomföra den forskning som presenteras i avhandlingen. För detta tackar jag även forskarskolans finansiär, Vetenskapsrådet. Jag vill särskilt tacka forskarskolans vetenskapliga ledare Caroline Liberg och koordinator Åsa af Geijerstam. Tack för det arbete ni lagt ner med att göra forskarskolan till en bra miljö för oss doktorander att utvecklas i. Tack också till alla er andra, som inom forskarskolans ramar läst och kommenterat texter och forskningsplaner. En ovärderlig del i forskarskolan har varit de fyra kollegorna (numera doktorerna): Ida Bergvall, Judy Ribeck, Marie Ståhl och Tomas Persson. Jag är så tacksam att jag fått dela denna tid med er. Det har varit en ynnest att ha vänner som brottas med liknande svårigheter och som verkligen förstår. Med er har jag kunnat dela allt. Tack också Judy, för att du ställt upp som jourhavande lingvist.

Two researchers have welcomed me to visit them at their Universities, something that I am very grateful for. Thank you, Kay O'Halloran for welcoming me to your research group at Curtin University. I much appreciated to discuss multisemiotics with you, at an early stage of in my PhD studies.

Thank you also, for arranging for me to visit the Multimodal Analysis Lab in Singapore. Thank you, Candia Morgan for welcoming me to your research group, at the Institute of Education in London. Thank you, for valuable discussions regarding my research. I am also grateful for your thorough reading of one of my drafts, and the following discussion in Uppsala, early in my PhD studies.

I have received very useful and constructive feedback on the English in this thesis. Thank you, Cris Edmonds-Wathen for volunteering to help with the language when I needed it the most. Thank you also Jill James for thorough commenting on the language in the thesis.

Till kollegorna på Institutionen för Matematik och matematisk statistik riktar jag ett generellt tack för hjälp och stöd i smått och stort. Ett specifikt tack vill jag dock rikta till vännen Mathias Norqvist som jag delat doktorandvardagen med. Utan dina goda råd om tråkiga datorprogram hade jag förspillt en massa onödig energi på datortjafs (tack för hjälp med framsidan!). Det mest värdefulla du bidragit med är dock ett sunt förhållningssätt. Tack Mathias, för givande samtal om livet och vad som är viktigt – egentligen.

Jag vill också tacka Anki Jakobsson. Tack för att du trodde på mig och hjälpte mig på traven när jag insåg att jag ville doktorera.

Till sist. Jag är stort tack skyldig maken Einar som alltid accepterat att jag prioriterat jobbet. Att jag råkar vara gift med den snällaste och bland de mest energiska män jag känner har varit avgörande för att jag orkat när det varit som tuffast. Tack kära du för markservice. Jag vill också tacka sönerna som genom sitt varande gör vår familj till min bästa kraftkälla. Tack för att jag fått tanka energi av er genom knyckta kramar. Att disputera och åstadkomma en avhandling är att betrakta som någon typ av prestation. Dock, Pontus och Viktor, vill jag att ni ska veta att avhandlingen är en petitess på det stora hela. Det mest fantastiska jag åstadkommit och någonsin kommer att åstadkomma i mitt liv är ni – mina älskade!

Bonässund, september 2016

List of papers

1. Österholm, M., Bergqvist, E., & Dyrvold, A. (preprint). The study of difficult vocabulary in mathematics tasks: a framework and a literature review.

2. 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), 101-128.

3. Dyrvold, A. (2016). The role of semiotic resources when reading and solving mathematics tasks. *Nordic Studies in Mathematics Education, 21*(3), 51-72.

4. Dyrvold, A. (preprint). Relations between various semiotic resources in mathematics tasks – a possible source of students' difficulties.

The published papers are reproduced with permission of the relevant publisher.

The first two studies are conducted together with Bergqvist and Österholm (Paper 1-2). All three of us have participated in the design, the implementation, and the writing of the studies. In Paper 1, I am the first author since I have done a bit more of the writing. In Paper 2, Österholm is the first author since he contributed most to the design of the study. During the research process, however, the three of us have shared the work equally and therefore my contribution is almost a third.

1. Introduction

1.1 Setting the scene

Questions about in what sense mathematics tasks are demanding to read and solve are important from several perspectives. In this thesis, two types of difficulties, in relation to textual features of task texts, are addressed: unwanted difficulties and reasonable difficulties. The difference between those perspectives has to do with what features of the task text are perceived as part of the mathematics. In the current thesis, the language of mathematics is seen as one important part of mathematics. This mathematical language in written text is constituted of several semiotic resources and is therefore multisemiotic. Four different semiotic resources are considered in this thesis, namely: natural language (words and letters), two types of images, and mathematical notation (the mathematics symbolic language).

The multisemiotic language is an important part of mathematics; for example, the diagrams and symbols used have been essential for the development of the mathematics (see e.g., O'Halloran, 2005). One particular property of mathematics is many of the objects are not accessible as physical objects; for example, the derivative is not a 'thing' but it can well be represented symbolically as a mathematical object. Therefore it is only through these symbolic representations that we have access to the abstract mathematical objects (see e.g., Duval, 2006; Moreno-Armella & Sriraman, 2010), and in that sense the mathematics language is essential. A discursive perspective on language such as Sfard's (2008) can also illustrate the essential role different representations have in mathematics. Simplified, with a discursive perspective the different representations are seen not as various representations of the same object but a collection of realizations that together construct the discursive object. Still another aspect of the intrinsic role that the different semiotic resources have within mathematics is the importance of relations that exists between those semiotic resources (see e.g., Duval, 2006). It has also been argued that students' use of several different semiotic resources is important for their development of a deeper understanding of the mathematics (Ainsworth, Bibby, & Wood, 1997).

Since different semiotic resources have a crucial role in mathematics, students' abilities to use and interpret these resources are essential. Therefore, knowledge of difficulties related to how semiotic resources are used in written text is important for us to develop mathematics education. Another perspective on why natural language and other semiotic resources in mathematics are important to focus on is that mathematics high stake tests often are assessed through tasks represented as written text. Mathematics tests should aim at assessing mathematical ability, nothing else, and therefore questions

about whether potentially difficult textual features are part of what the assessment aim at are highly relevant.

In summary, many features of mathematics language are essential both in the learning of mathematics and from an assessment perspective. Therefore the study of difficulties related to reading and solving mathematics tasks can contribute to knowledge useful both from a teaching perspective and in relation to test construction.

1.2 Purpose and research questions

The purpose of this thesis is to *contribute to the knowledge of aspects of difficulty related to textual features in mathematics tasks*. In addition to the interest in the broad concept of difficulty, a particular emphasis has been made to distinguish between what is difficult from a mathematics point of view and difficulties that have to do with other aspects than the mathematics. In particular, analyses of whether a non-mathematics specific reading ability is applicable in the solution of tasks are performed, something that is referred to as the tasks' demand on reading ability.

Throughout this thesis, the broader concept of text is used to include images, mathematical notation, and natural language (see e.g., Björkvall, 2010). This means that 'textual' will also refer to other textual features than natural language. Three research questions are formulated in relation to the purpose of the research.

RQ 1) Are there any particular textual features in mathematics tasks that are related to task difficulty, and if so, how?

RQ 2) Are there any particular textual features in mathematics tasks that are related to task demand on reading ability, and if so, how?

RQ 3) Regarding textual features that in any way are related to how difficult the tasks are to read or solve—is a particular textual feature's difficulty a mathematics specific difficulty or not?

The concept 'difficulty' is important in this thesis; this importance is reflected in both the purpose and the three research questions. However, it is crucial to note there are different types of difficulties addressed. Difficulty in RQ 1 regards a more general difficulty; that is the combination of difficulties affecting students' problem solving. This type of difficulty in the studies is analysed using student's scores on mathematics tasks. RQ 2 regards demand on reading ability; this demand does not address mathematics. This implies the difficulty addressed in RQ 2 is a difficulty unwanted in mathematics assessments. RQ 3 regards the difference between the two types of difficulties addressed in RQ 1 and RQ 2. Therefore, the difficulty in RQ 3 is referred to as mathematics specific. The purpose regards all these types of difficulties,

and thus, *aspects of difficulty* is used to signal the inclusion of different types of difficulty. Figure 1 in section 2.1.2 illustrates how these aspects of difficulty are related to mathematical ability and reading ability.

Altogether four studies (see List of papers on page V) are conducted to fulfil the purpose of this thesis and to answer the research questions. Common to all four of the studies is they regard whether some textual feature is potentially difficult. Therefore the results of all four studies contribute to answering RQ 1. Study 2-4 also regard whether some textual feature have the potential to cause a non-mathematics specific demand on reading ability and therefore those studies contribute to answering RQ 2. The last question, RQ 3 is answered based on results concerning task difficulty in relation to the tasks' demand on reading ability for each particular textual feature. Therefore, only the studies within which both these difficulty aspects are investigated contribute to answering RQ 3, namely Study 2-4.

One important notion in relation to the purpose of the thesis is *demand on reading ability* (DRA), which is explained in section 3.3. Essential is that DRA represents a non-mathematics specific reading ability. Such a reading ability is exactly what a reading test should assess but not in a mathematics test. Accordingly, RQ 2 deals with textual features that we do not want to assess in a mathematics test.

There is an inconsistency in how DRA is referred to in the thesis due to review comments on an article. In Study 1-2 DRA is called demand *of* reading ability and in the coat and Study 3-4 it is called demand *on* reading ability; the meaning is however the same (see also section 3.3).

2. Background

The overarching theme of this thesis is the language in mathematics tasks; both the purpose and the three research questions of the thesis concern the textual features of mathematical tasks. Therefore, the first section of the background focuses on *mathematics and language*. Written mathematics, which is studied in this thesis, is often referred to as multisemiotic, since it consists of a combination of different semiotic resources, such as natural language (words and letters), mathematical notation, and different types of images. The second section concerns the *multisemiotic mathematical language*, especially in mathematics tasks. Finally, since the thesis examines mathematics tasks, the last section of the Background focuses on *assessment* in mathematics, especially regarding validity.

2.1 Mathematics and language

The relation between language and mathematics has been previously studied with different purposes and based on different understandings of the role that language has within mathematics. The purpose of this thesis is to contribute to the knowledge of aspects of difficulty related to textual features in mathematics tasks. The relation between mathematics and language is important to this purpose because when assessing mathematics, it is essential to distinguish whether text and language is perceived as a means, or part of the mathematics. In the following sections, the relation between mathematics and language is discussed from several perspectives. The first section concerns the role language has within mathematics and the second section presents empirical results on the relation between language ability and mathematical ability.

The term language describes different types of systems used in communication within various contexts. Oxford English dictionary defines language as "The system of spoken or written communication used by a particular country, people, community, etc., typically consisting of words used within a regular grammatical and syntactic structure" (language, OED online). Language, as a system, can take place through other means than print, for example by sound when we speak. The term language can be used to refer to gestures and other wordless communication (e.g., Radford, Edwards, & Arzarello, 2009). This means research on the mathematical language includes a variety of research. This thesis focuses on a sample of the written mathematical language used in task text, including images and mathematical notation. The aspect of mathematical language investigated is therefore referred to as different textual features.

2.1.1 The relation between mathematics and language

There are opposing standpoints when discussing whether mathematics *is* a language or a *means* to communicate pure mathematics. These two standpoints exist, but in reality there is a spectrum between them creating an intermediate third standpoint.

Arguments for the first standpoint—to see mathematics as a language—are given by several researchers (e.g., Schweiger, 1992; Usiskin, 1996; Wakefield, 2000). Usiskin (1996) presents several arguments for perceiving mathematics as a language; mathematics is constructed as natural language with expressions that are sentence like, it has syntax and symbols that function as verbs, and learning mathematics is like learning a second language.

The second standpoint is to perceive language as a tool useful to communicate mathematics. This perspective is sometimes revealed in research articles where mathematical language is referred to as simply the means to express the mathematics. For example, Sato, Rabinowitz, Gallagher, and Huang (2010) describes a linguistic modification of task text as "intended to increase student access to tested content by minimizing the language load associated with the text in a test item" (Sato et al., 2010, p. 16). A similar way to talk about language in mathematics is found in Tindal's (2014) study where reading and writing are referred to as "access skills". The perception of language as a tool is also found in a study by Adu-Gyamfi, Bossé and Faulconer (2010); the study refer to reading and writing as *tools* to articulate mathematical understanding. This naming of reading and writing as tools must not imply a view on mathematics as 'free from language', but it signals a separation between language and mathematics.

The third and intermediate standpoint is: mathematics *has* a language. Examples of this mathematical language is not only defined by technical vocabulary and grammatical patterns, such as dense noun phrases (Schleppegrell, 2007), but also by the use of multiple interacting semiotic resources (O'Halloran, 2008). The third standpoint about mathematics having a language is compatible with the concept of mathematical literacy and of mathematical ability consisting of several mathematical competencies. Historically, mathematics education have been tightly bound to the discipline mathematics, but more recently there has been an emerging shift to a view of mathematics as a configuration of literacies (Cobb, 2004). Cobb (2004) and de Lange (2003) argue that there are different forms of mathematical literacy, but in the PISA framework (OECD, 2013) mathematical literacy is referred to as *one* capacity students can have. Mathematical literacy is in the PISA framework defined as "an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts" (OECD, 2013, p. 25). This definition illustrates that language ability plays an important role in what it

means to be mathematically literate; in this context the words "formulate" and "interpret" imply language ability is needed.

The mathematical literacy concept is tightly bound to a perception of mathematical ability consisting of several competencies (see e.g., de Lange, 2003). Mathematical competencies (or corresponding phenomena) are defined in several competence frameworks (e.g., Kilpatrick, Swafford, & Findell, 2001; Niss & Højgaard, 2011); based on those definitions it is also apparent that mathematics has its own language: a language that must be mastered as part of a mathematical ability. Two apparent examples are the communicative and reasoning competencies. In the KOM framework, the communicative competency is described as "being able to communicate in, with, and about mathematics" (Niss & Højgaard, 2011, p. 67). Communicating *in* and *with* mathematics implies the communication takes place within a mathematics language. What characterises mathematical reasoning – according to Kilpatrick, Swafford, and Findell (2001)–is a pathway of statements and arguments. The NCTM standards (NCTM, 2000) include in the reasoning competence category: the ability to develop and evaluate conjectures and arguments in mathematics. Both constructing and evaluating conjectures and arguments demand abilities that are founded in language ability.

The different perspectives on the role that language has within mathematics consist of not only the two extreme standpoints, but one intermediate perspective and can therefore be divided into three standpoints.

1. Mathematics is a language.
2. Mathematics has its own language, and that language is part of the mathematics itself.
3. Mathematics is a science that exists independent from human activity. Language is only a means by which humans communicate mathematics.

The standpoint taken in this thesis is the second one. It relates to the theoretical model presented in the next section (2.1.2); if mathematics has its own language, the existence of a specific mathematical reading ability is also reasonable.

In summary, different perceptions of the relation between language and mathematics exist; some perceptions where language is seen as part of mathematics and others where mathematics is seen as separated from language. In definitions of mathematical literacy and mathematical competencies, the presence of some aspect of language ability is evident. Therefore, if the description of what mathematics is were to be adjusted to what it means to be mathematically proficient, language ability should be a part of the description.

2.1.2 The relation between mathematical ability and reading ability

If the standpoint is taken that mathematics has its own language and that language is part of the mathematics itself, the question whether mathematical ability is related to reading ability follows naturally. Therefore, another topic relevant when examining textual features in mathematical tasks is the potential relation between mathematical ability and reading ability. The relation is important to the purpose of the thesis since a reading ability is used when mathematics tasks are read and solved. This section first presents examples of results showing that there is an empirical relation between mathematical ability and reading ability. In the next section a theoretical model of this relation is presented.

Empirical relation between the abilities

Independent of whether the school subject is mathematics, science, or social science, language ability is of importance for the students to benefit from teaching (e.g., Schleppegrell, 2004). Therefore, the relation found by, for one example, Grimm (2008) between early reading comprehension skills and later achievements in mathematics, is expected. There is however convincing evidence for a relation between the two abilities of mathematics and reading.

Many studies analyse the relation between reading ability and mathematical ability using different test scores as data. For example, Edge and Friedberg (1984) found significant correlations between scores on a language test in English and both an algebra test and grades on the first algebra course. The results give evidence for a relation between language abilities and mathematical ability, both as quantified teacher judgements and as test scores. Another large study by Chen (2010), with 21,000 students from 1,300 schools, reveals that reading scores explain as much as 44-54% of the variance in mathematics scores; this shows there is a strong relationship between reading ability and mathematical ability.

The relation between reading ability and mathematical ability is also studied by Hickendorff (2013). An analysis on results from 2,000 students in grade 1-3 reveals a strong relation between mathematical ability and reading ability for all three school years. In both Hickendorff's (2013) and Chen's (2010) studies, a decrease in how strongly language ability is related to mathematical ability, from earlier to later grades, can be observed.

A study in which students' results from an international reading literacy study (PIRLS), evaluated in relation to the same students' results on TIMSS science and mathematics, revealed strong correlations between results for all three subjects (Caponera, Sestito, & Russo, 2016). The importance of reading literacy on the results of the science and mathematics tests is also evident of the crucial role reading has on the correlation between mathematics and

science results. In the same study, Caponera et al. compared the correlation between mathematics and science results with and without reading literacy as a control variable; the study's shared variance decreased from 79% to 31% when reading literacy was controlled. Altogether, the four studies above reveal a consistent relation between reading ability and mathematical ability.

The relation between reading and mathematical ability has also been shown through studies that focus on genetics. For example Kovas, Haworth, Harlaar, Petrill, Dale, and Plomin (2007) found that for 10-year-olds, the same genes influence poor reading and poor mathematics performance. Another study, where the results for 4,000 pairs of 12-year-old twins were analysed using a group twin correlation, revealed a high correlation between the two abilities on a mathematics and a reading test (Haworth et al., 2009). For the whole sample the genetic correlation between mathematics and reading was 0.58 (Haworth et al., 2009). A high genetic correlation is an indication of a high degree of overlap in the genes that influence the respective abilities, in this case mathematics and reading.

In summary, the empirical evidence for the relation between mathematics and reading is convincing and there is no need to question whether the two abilities are related. Still, it is yet unclear how the relation can be explained and what implications the relation has for school practice.

Theoretical model of the relation between abilities

In this thesis the standpoint taken is that mathematics has its own language and that language is part of the mathematics itself (see section 2.1.1). Having taken this standpoint also affects how a mathematical ability is perceived, namely as an ability that includes a kind of mathematics specific reading ability. The relation between a mathematics specific reading ability and the abilities of reading and mathematics can well be represented through a diagram (Figure 1). The overlap between mathematical ability and reading ability in the diagram represents a mathematics specific reading ability. This mathematical reading ability can be thought of as the mathematical ability that is used, for example, in mathematical reasoning and oral and written mathematical communication. Field 1 in Figure 1 represents a mathematical ability that does not have to do with reading ability, and field 3 represent a reading ability that is not part of mathematical ability. This model is explained in Article 2, only a brief summary is presented here.

The schematic illustration is adequate to illustrate the proposed mathematical reading ability, but the figure also illustrates three different abilities that can contribute to the test result on a mathematics test. A mathematics test should test mathematical ability, noting else, and therefore the arrow pointing from the third field in Figure 1 represents an unwanted factor in test results. If a mathematics test assesses this reading ability it is a construct

irrelevant factor in the test (see also section 2.3 about construct irrelevant factors). Only the reading ability that is part of the mathematical ability (field 2) and the mathematical ability that does not have to do with reading (field 1) should be assessed in a mathematics test.
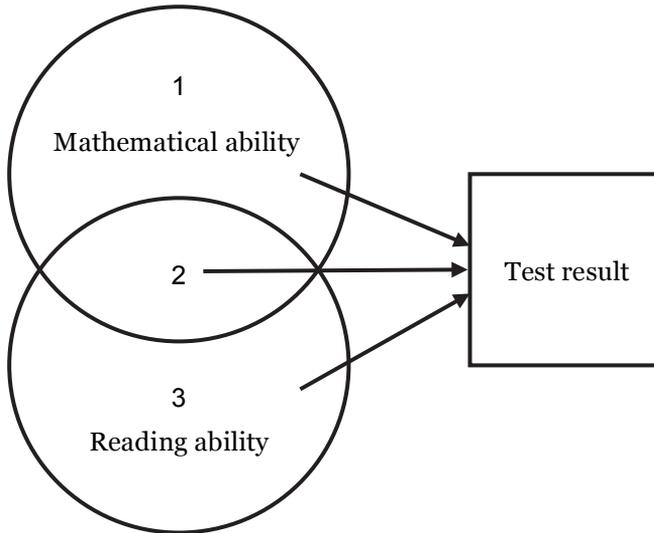
Figure 1: *Illustration of relations between and within abilities in relation to test result* (modified after Dyrvold, Bergqvist, & Österholm, 2015).

This model of the relation between reading ability and mathematical ability is very useful in a discussion about the validity of assessments in mathematics; more precisely, this model is useful in relation to aspects of mathematical language and language ability. One important criterion for a valid assessment is the desired construct (i.e., mathematical ability) is assessed and nothing else. This is illustrated by the first two arrows in Figure 1. The existence of field 2 shown in the model also illustrates the standpoint taken in the thesis, namely that 'mathematics has its own language, and that language is part of the mathematics itself' (see 2.1.1).

The previously presented empirical relation between mathematical ability and reading ability is not visualised in the model since it is theoretical, however, the empirical relation is explained to some extent by the existence of field 2.

Note that the mathematical reading ability (field 2) can also represent the ability to read different diagrams and mathematical notations, not just words. The ability to read other semiotic resources beyond natural language is important within mathematics since the mathematical language to a large extent is multisemiotic.

2.2 The multisemiotic mathematical language

Mathematics text (e.g., in test tasks) often consists of a combination of different semiotic resources; this includes natural language, mathematical notation, and different types of images. Texts with this combination of semiotic resources can be described as *multisemiotic*. This thesis is about the multi-semiotic mathematics text and therefore some research within the area of multisemiotics is presented here. The textual features studied in relation to the three research questions concern both separate semiotic resources and the interactions between them; these relationships are reflected in the four subsections presented here.

2.2.1 Mathematics tasks as multisemiotic texts

Some distinguishing features of multisemiotic texts have been described by, for example, Kress and van Leeuwen (Kress, 2010; Kress & van Leeuwen, 2006). In Kress' description of texts with different semiotic resources, the concept affordance is central. Different semiotic resources are characterized by different affordances, for example, natural language has the affordance of sequence, something that cannot be said about an image (Kress, 2007, 2010). Kress exemplifies how different semiotic resources have different means to convey a message, for example, colour saturation in an image can express emphasis or prominence (Kress, 2007). Images also have the means of spatial organisation, for example, when the relative location of parts in an exploded-view drawing is visualised or when time is represented as distance in a diagram (Kress & van Leeuwen, 2006). This difference between different semiotic resources has also been emphasized by, O'Halloran (2005), according to whom it is impossible to say the same thing with different semiotic resources in mathematics. Altogether, according to research a multisemiotic text enables the author not just to say things differently than with natural language alone, but to say different things.

Multisemiotic texts differ from texts with only natural language in many ways. Something that can be both an asset and a source for difficulties; when multisemiotic texts are read, there exists the possibility (or necessity) to interpret the different semiotic resources together. Based on Duval's (2006) analyses of comprehension of mathematics texts with different semiotic resources, it can be concluded that reading such texts is different from reading texts consisting only of natural language. There are several differences between texts with only natural language and multisemiotic texts, but possibly one of the most prominent differences is multisemiotic texts are not necessarily read linearly since relations between the different semiotic resources may direct the reader back and forth (Unsworth & Cléirigh, 2009). Kress goes as far as to call the reader the "designer" of a multisemiotic text, since the reading of multisemiotic texts follows the reader's interest, engagement

and attention, not the order of the elements (Kress, 2010). Reading pure mathematical expressions also differs from reading natural language because pure mathematical expressions often are grouped in clusters with another logic than sentences, for example, left-right sides of equalities (see also, Kirshner, 1989). There are also empirical results showing a relation between how students read a mathematics task and how difficult they perceive the task to be (Beitlich, Lehner, Strohmaier, & Reiss, 2016). Additionally, empirical results show the experienced difficulty in comprehension of the texts is influenced by the type of relations between images and natural language (Unsworth & Chan, 2008).

One aspect of both the complexity and benefit of several semiotic resources in texts is objects are represented in different ways within the different resources. Both Kress (2010) and Duval (2006) address this issue. An object can be represented more than once within the same semiotic resource (e.g., in two sentences), or it can be represented in different semiotic resources (e.g., in words and as an image). Both Kress and Duval argue that it is more complex to understand how two representations of the same object relate to each other when the two representations use different semiotic resources than when they use the same semiotic resource. When two different semiotic resources represent the same object, the change between the different semiotic resources is called a conversion (Duval, 2006). Duval describes conversions as alterations of representations that entail a change in semiotic resource without changing the object that is being denoted. Based on earlier empirical studies on students' work with mathematical tasks, Duval describes a systematic variation in performance related to conversions between different semiotic resources (Duval, 2006). Duval argues that this difficulty stems from a cognitive conflict, and that this conflict can lead to students seeing two representations of the same mathematical object as two different objects. For this not to happen, the students need to dissociate the cognitive object from its representation.

In summary, multisemiotic texts differ from texts with only natural language in many ways. For multisemiotic task text one apparent difference is that linear reading cannot be assumed. There are several affordances associated with the mathematics multisemiotic, but also challenges. Different aspects of how the mathematical multisemiotic text might be demanding have also been investigated empirically. A summary of that research is given in the following sections.

2.2.2 Different semiotic resources in mathematics tasks
Besides natural language, two other semiotic resources are often used in mathematics tasks, namely images and mathematical notation. These semiotic resources are essential to mathematics texts and have been a focal point in some earlier research. For example, previous studies have focused on the

demand students face when different aspects of images and mathematical notation are present and on the role a semiotic resource has in problem solving. The following short summary gives a few examples to present a snapshot of the type of research conducted regarding images and mathematical notation in mathematics tasks.

Images in mathematics texts have been investigated from several perspectives in earlier research. There are, for example, studies that focus on the features of particular mathematical images. One example is Alshwaikh's (2011) framework for different types of geometrical diagrams and their role in construction of mathematical meaning. Another study that also focuses on diagrams in geometry is Dimmel and Herbst's (2015) study that examines how geometry diagrams differ from each other and what visual resources are used in those diagrams. Dimmel and Herbst analyse 2,300 diagrams in mathematics textbooks to capture every variance in their scheme of how the features of geometry diagrams can vary. These two studies contribute to the developing knowledge of the role of images in mathematics.

Different types of images in mathematics text have also been studied in relation to students' problem solving and in aspects of task difficulty. A study focusing on the role of images in problem solving revealed that pictorial images were not helpful, whereas organisational, representational, and informational images were (Elia & Philippou, 2004). Other studies are more concerned with whether particular types of images are helpful in learning or if different types of images are difficult. Chen and Herbst (2013) found that more challenging diagrams are advantageous when it comes to the development of students' mathematical reasoning. When the diagrams are constrained in what they reveal (i.e., fewer of the elements in the image are drawn or labelled), students are more prone to make reasoned conjectures about the diagrams (Chen & Herbst, 2013). De Kirby and Saxe (2014) also study diagrams in mathematics, but they focus on the students' interpretation of the diagrams. In analyses of classroom observations, they noted students did not manage to see through the diagrams to the represented mathematical object. In this study, the students interpreted a point in a diagram as the visible small dot instead of a mathematical point that has no size. A complementary experimental study revealed students were helped in the interpretation of the diagrams if they had access to mathematical definitions that distinguished between the diagram and the idealized mathematical object (de Kirby & Saxe, 2014). De Kirby and Saxe's study is an illustrative example of the cognitive conflict described by Duval (2006). De Kirby and Saxe's description of the difficulty of 'seeing through' the diagram would, in Duval's terminology, be described as a difficulty with the dissociation of the cognitive object from its representation.

When it comes to research about mathematical notation, there are both a variety of studies in which symbols that are investigated and which issues

mathematical notation that the studies concern. One particular symbol that has been focused on in many studies is the equal sign. For example, difficulties and misunderstandings related to the equal sign have been extensively studied (e.g., Baiduri, 2015; Kieran, 1981; Li, Ding, Capraro, & Capraro, 2008). There is also research that focus particularly on numbers, for example, Rousselle and Noël's study about the understanding of numbers in primary school (Rousselle & Noël, 2007). They found that, for children with mathematical disabilities, the critical aspect when working with tasks on magnitude was the Arabic digits.

After primary school, however, the use of letters as mathematical notation can be more demanding than the Arabic digits. The use of mathematical notation in learning algebra is investigated in a study by Susac, Bubic, Vrbanc, and Planinic (2014) The conclusions drawn from an analysis of students' solutions are that younger students (age 13-15) tend to use concrete strategies, such as inserting numbers in the equations, whereas older students (age 16-17) use more abstract strategies. Also, younger students are both slower and less accurate in solving equations with letters than with numbers (Susac et al., 2014). These results show the difference in complexity between the different notations is visible both in students choice of notation and in solution speed.

Difficulties students experience when solving mathematics tasks with mathematical notation have also been investigated with different methods. Several studies compare results with or without mathematical notation; it can be concluded based on the studies that the presence of mathematical notation is difficult compared to natural language, but not compared to schematic images. Driver and Powell (2015) analyse the difference between second grade students' scores on mathematics tasks with and without mathematical notation. Both students with and without mathematics difficulties scored significantly lower on the tasks with mathematical notation (Driver & Powell, 2015). Difficulties related to the reading of mathematical notation have also been discovered for older students. A study comparing reading comprehension of the same mathematical content in text with or without mathematical symbols showed the existence of mathematical notation made the text more demanding (Österholm, 2006).

Koedinger and Nathan (2004) also found that tasks with mathematical notation are difficult for students. In their study, they compared students' results on algebra story problems with solutions on mathematically equivalent equations. In addition to the difference in performance between the two task types, Koedinger and Nathan found that the students use different strategies depending on what form the task was presented–whether it was presented as a story or as an equation. There is also research that shows the relation between students' performance on tasks with mathematical notation to their performance on tasks with schematic images. Lin, Wilson, and

Cheng's (2013), as well as Yang and Huang's (2004) research shows the type of semiotic resource present in a task is related to the students' performance, demonstrating that students perform better on tasks where multiple choice answers are given as mathematical notation compared to when answers are given as schematic images.

In summary, there exists a variety of research that concerns aspects of images or mathematical notation. Results from several studies indicate the aspects of difficulty are related to both when the mathematical notation is presented in different ways and that the types of images used in tasks have an impact on the learning and success of solving mathematics tasks.

2.2.3 Translations and relations between different semiotic resources in mathematics tasks

The notion translation is sometimes used in mathematics education research to refer to the construction of a new representation of an object in a semiotic resource (e.g., to represent y=x as a graph). It is also used to denote the act of connecting instances in task text when reading, both within and between different semiotic resources. Common for both types of translations is that they concern the connections between different representations of the same mathematical object. Research about both types of translations is relevant in relation to the multisemiotic analyses; in the thesis some empirical results about translations are therefore presented. Relations that do not concern the translations have also been studied empirically, for example, cohesive relations, that are semantic relations that link instances in the text together. A few studies about other types of relations are also presented in the end of this section.

Several studies focus on translations between different semiotic resources and the results reveal that translations tend to be difficult. For example, Chahine (2011) shows students who practice translations between different semiotic resources perform significantly better on mathematics tests than students in a control group. Moreover, the students become more flexible in how they use different representations to understand important concepts. Chahine (2011) argues different representations helped the students to shift from procedural strategies in problem solving, to reasoning strategies (Chahine, 2011). Earlier research also reveals a lack of flexibility in the use of several semiotic resources seems to be one reason behind low achievement rate. Moreover, the ability to translate between semiotic resources in problem solving seem to be related to a good problem solving ability (Delice & Sevimli, 2010).

Potential difficulties related to translations between different semiotic resources have also been studied with a focus on different student ability levels. The results reveal that students with different ability levels process the

translations between mathematical notation and graphical forms in different ways and that high ability students are more capable of correctly performing the translations. The high ability students are more flexible in the choice of process to use in the translations (Bossé, Adu-Gyamfi, & Chandler, 2014). Several other studies address questions of difficulties related to students' translations between different semiotic resources (e.g., Capraro & Joffrion, 2006; Janvier, 1987; Lech, Post, & Behr, 1987). The research about translations focus on different aspects, but it can be summarized that aspects of difficulty can be attributed to the need to translate between different semiotic resources when solving mathematics tasks.

Other results demonstrating the importance of being able to work with and relate between different semiotic resources can be found, for example, in a study by Bagni (2006). Based on analyses of two classroom experiments, the conclusion was that the ability to distinguish between and coordinate the meaning presented through different semiotic resources is important for students' knowledge attainment. Bagni's results are not surprising, but important. The study concerns set theory but it is reasonable to believe that the results can be generalized to other areas in mathematics as well.

The importance of how the solver connects meanings from different instances in the text is also a part of the results in Hegarty, Mayer, and Monk's study (1995). Hegarty et al. conduct two experiments to test whether there is any difference between the strategies used by successful and unsuccessful problem solvers. An eye fixation analysis reveals that successful problem solvers construct a model that describes the problem and base their solution on that model. In contrast, unsuccessful problem solvers base their solution on numbers and keywords. The study also reveals that the integration of meaning of different instances in the text is crucial for the construction of the successful problem solvers' model (Hegarty et al., 1995). Hegarty et al. do not focus particularly on relations between different semiotic resources, but the integration of meanings from the task text involves integration of different semiotic resources.

There are also results indicating that textual features that urge the reader to relate instances in text (to cycle) can be difficult. For example, Turner Blum and Niss (2009) focus particularly on relations between different semiotic resources within the text. Their analysis reveals a relation between task difficulty and task features that urge the solver to cycle in the text. Qualitative methods have also been used to study potential difficulties that students experience when connecting different semiotic resources during problem solving. For example, Moon, Brenner, Jacob and Okamoto's (2013) analysis of students' work reveals that the students had difficulties in making connections between mathematical notation and graphs. The researchers conclude that one reason behind these difficulties is the students lack under-

standing of so called big ideas[1] related to the use of the represented mathematics. This lack of understanding disturbed their ability to make connections between the semiotic resources. Similarly, Acartürk, Taboada, & Habel (2013) focus on relations between different semiotic resources in the text, but analyse potential differences in difficulty depending on features of the relations in text. Their analysis of different types of cohesion (how explicit reference) between natural language and images reveals how the type of cohesive relation used influences the reading. The results reveal differences in both eye movement parameters and retention of the material, depending on type of reference used. The texts used in the analysis were mainly science texts, but the results are still relevant in relation to mathematics task text since the same types of references are used between instances in mathematics tasks.

The research presented here does, in different ways, concern relations between different semiotic resources in text, something very much related to cohesion. The concept cohesion has to do with particular types of relations in text, and cohesion has been studied both in natural language and in multisemiotic texts. The concept cohesion and a few studies about cohesion in multisemiotic text are presented in the next section.

2.2.4 Cohesion in multisemiotic texts

The purpose, as well as the three research questions posed in this thesis, focus on textual features in mathematics tasks. Some of these are relatively straightforward, such as how common different words are, but some are more complex, such as cohesion. Since cohesion is not a commonly used concept in mathematics education, the concept is defined and exemplified in this section. An explanation is also given about exactly which parts of Hasan's (1989) framework for cohesion are used in this thesis, what cohesion in multisemiotic texts is, and how it differs from cohesion in natural language.

Cohesion is an essential feature of what we call text. A text with only natural language that has no cohesion is purely a list of words (Hasan, 1989); accordingly, a multisemiotic text with no cohesion would be a collection of natural language, mathematical notation, and images where no message can be derived from some interaction between or within the different parts of the text. The word cohesion refers to "the action or condition of cohering; (…) sticking together" (cohesion, OED online). This act of sticking together reflects also what cohesion in text means. Cohesion in text is defined by particular meaning relations that relate multiple parts in the text together, for example, when a pronoun refer to a noun (it – triangle) the meaning relation between the words makes the text stick together (e.g., Hasan, 1989).

---

[1] E.g., the Cartesian connection: "a point is on the graph of the line L if and only if its coordinates satisfy the equation of L" (Moschkovich, Schoenfeld, & Arcavi, 1993, p. 73)

Cohesion is a concept commonly used within linguistics (about natural language), but in later years the concept is also used concerning meaning relations in multisemiotic text. The concept cohesion is by Halliday and Matthiessen (2004) described as a function of a text. It is a comprehensive term that includes a collection of semantic relations in text. Those semantic relations occur in a text when the interpretation of some elements in the text depends on the interpretation of other elements (Halliday & Hasan, 1976). This interdependence between elements in the text consist of meaning relation and thus cohesion is defined by "relations of meaning that exist within the text, and that defines it as text" (Halliday & Hasan, 1976, p. 4). Those relations are called cohesive ties.

The only types of cohesive ties that are identified in this thesis (Study 4) are focused on in this section: ellipsis, reference, and lexical cohesion at the word level (not between whole phrases). *Ellipsis* is when a word is excluded since it can be derived by the context, for example, when the word "box" is avoided at the end of the sentence "I take this box and the other". *Reference* is when a word explicitly refers to another word. For example, in the sentence "Lena draws one line and then she draws another line parallel to the first line", 'Lena' and 'she' are cohesively tied through reference; 'she' refers to 'Lena'. Both ellipsis and reference are grammatical cohesive ties, since they are achieved through grammatical constructions. *Lexical cohesion* is achieved through the choice of lexical items, that is, single words, parts of words, or a chain of words. In the example above regarding "Lena", there are cohesive ties between the three instances of the word 'line' (repetition). Two other types of lexical cohesive ties are hyponymy (i.e., words within the same class of objects such as quadrilateral and square) and synonymy (i.e., using a synonym to a word), (see e.g., Halliday & Hasan, 1976; Hasan, 1989). There are several other types of cohesion and those that are used in the analysis are explained in Article 4.

The research on cohesion in natural language is substantial and many studies have shown relations between cohesion and readability of a text, as well as the relations between cohesion and reading comprehension (see e.g., the research review by Irwin, 1986). Similar relations have also been found in more recent research. A study by Bayraktar (2011) reveals a relationship between the ability to recognize lexical cohesive relations (ties) in text and comprehension scores. The same study also shows that lexical cohesive ties are more difficult to identify when they occur across paragraphs than within the same paragraph. Research has also shown that subject specific words are overrepresented among words that are cohesively tied compared to all words in a particular text (Teich & Fankhauser, 2005).

Cohesion has also been studied in multisemiotic texts, but since the research area is still developing, there are different conceptualisations of multisemiotic cohesion (how parts of different semiotic resources in a text are

integrated to a coherent whole). One essential feature of multisemiotic cohesion, emphasised by both Liu and O'Halloran (2009) and Royce (2007), is cohesion between different semiotic resources enables a meaning expansion, something that is aided by the combination of semiotic resources with different affordances. Such a meaning expansion has also been argued for elsewhere (see e.g., Lemke, 1998; Lemke, 2002; Unsworth & Cléirigh, 2009). Lemke (2002) explains that this expansion of meaning, when several semiotic resources are used, is caused by the selection of all possible combinations of types of capacities from the different semiotic resources. Liu and O'Halloran (2009) argue that intersemiotic cohesion is more than a mere linkage between two semiotic resources since the intersemiotic meaning relations create *integration* of, for example, words and images. Without this integration, the multisemiotic text would only consist of the co-occurrence of the different semiotic resources. The meanings created by one semiotic resource can alter the meaning revealed by another semiotic resource and therefore the set of possible meanings are multiplied when several semiotic resources are used together (Lemke, 1998).

The differences between different semiotic resources also lead to differences in how the cohesive ties are realised. All types of cohesive ties in multisemiotic texts are different from their counterparts within natural language, this relates to the affordances of the different semiotic resources. For two particular cohesive ties, this difference is substantial. Therefore, in the thesis the cohesive ties between different semiotic resources are given other names than the corresponding cohesive ties between words. This is done when the cohesive ties in the multisemiotic text is similar to synonymy (i.e., two words refer to the *same* object) and repetition (the *same* word is repeated). Similar cohesive ties occur between different semiotic resources, but because of the differences between the semiotic resources the *same* object is not represented in the occurrence. The concept *intersemiotic correspondence* replaces the cohesive ties synonymy and repetition in natural language, since different semiotic resources cannot offer the same meaning. For example between the word "graph" and a graph in a diagram there is a meaning relation similar to synonymy or repetition, but the image offers a very different meaning than the word and therefore it is more accurate to talk about an intersemiotic correspondence between the two representations of 'graph' (see Jones, 2007).

In summary, cohesion is a function of a text that is essential for making the text a coherent whole. Cohesion, in previous research, has been extensively studied in natural language, but there is also a developing research field focusing on cohesion in multisemiotic texts.

2.3 Assessment in mathematics

The research presented in this thesis contributes to the knowledge of aspects of difficulty related to textual features in mathematics tasks. Several types of difficulty is included in the expression 'aspects of difficulty', something that is explained in section 1.2. The results do not only concern which textual features are potentially difficult, but which textual features are, or should be the desired construct in assessments. This thesis, assumes the existence of a mathematical *reading* ability that is essential to a mathematical ability, and should therefore be assessed as part of the mathematical ability (see section 2.1.2). The concept *construct validity* is very useful in relation to questions about what a test should reasonably assess; it is therefore described in relation to mathematics tests in the first of the following two sub-sections. In the second sub-section a few research examples are given about how language accommodations of mathematics tests are carried out. The second sub-section is relevant in relation to the thesis since both questions about validity and language in mathematics texts, which are central in the thesis, are very relevant in relation to research about language accommodations in mathematics tests.

Assessment and construct validity

The goal of a mathematics test is to assess mathematics and nothing else; this often is referred to as the validity of the assessment. In this thesis it is assumed there is a particular mathematics language and that mathematical ability includes a mathematical *reading* ability. Therefore, a valid assessment of mathematics can, and perhaps must, include mathematical language aspects, such as technical vocabulary, particular grammatical structures, and several semiotic resources. The validity of an assessment can be threatened in different ways, but since the current thesis concerns language and mathematics, this section focuses on aspects of validity pertaining to the language and text.

The concept of validity is relevant in relation both to conclusions drawn from the results and to consequences from the conclusions. The aspect relevant in this thesis concerns the conclusions drawn regarding the construct being assessed, often referred to as *construct* validity. The concept construct validity has evolved over time, but one comprehensive definition is as the extent to which the interpretation of the scores from a particular measurement are related to the certain construct the scores are supposed to reflect (see e.g., Linn, 2014). The construct can be an ability, an attribute, or a proficiency that is clearly defined (see e.g., Brown, 1996). If the construct validity is high, it is the desired construct that is being assessed. The theory of construct validity is very comprehensive and construct validity is sometimes used as an overarching concept that includes (up to) six sub aspects (e.g.,

Messick, 1995).

Within the theory of validity and construct validity, construct irrelevant factors are important; these construct irrelevant factors are not part of the desired construct (Messick, 1995). The most important construct irrelevant factor for this thesis is construct irrelevant difficulty. One example of construct irrelevant difficulty, that applies very well to mathematics tests, is "the intrusion of undue reading comprehension requirements in a test of subject matter knowledge" (Messick, 1995, p. 742). What Messick refer to as undue reading comprehension requirements is represented, in the theoretical model of abilities (2.1.2), as a part of the reading ability that is not included in the mathematical ability.

Textual features that have the potential to be difficult are analysed in this thesis; one particular aim has been to distinguish between textual features that are construct relevant and construct irrelevant (e.g., by using statistical methods as described in section 3.3.1). In other words, to distinguish between textual features that can be seen as part of the mathematics and textual features that cannot be seen as such. The differences between these two types of textual features are also important to consider when assessments are accommodated to particular student groups; this is discussed in the next section.

Language accommodations of mathematics assessments
In relation to questions about construct validity of mathematics tests and in relation to the assumed mathematics reading ability (section 2.1.2), a particular area of research is relevant, namely studies of language accommodations of mathematics tests. This area is relevant, since one crucial aspect is to distinguish between language that is relevant in assessing part of a mathematical ability from language that is not. This separation, which of course is not an exact one, is addressed in research question three in this thesis: *Regarding textual features that in any way are related to how difficult the tasks are to read or solve—is the particular textual feature's difficulty a mathematics specific difficulty or not?* A few examples of studies focusing on language accommodations of mathematics tests are presented in this section because the question about which language is relevant to assess is central to the studies.

Different types of language accommodations have been tested in several studies, but the usefulness of the studied accommodations has also been questioned. For example, a substantial amount of research has been conducted with an explicit aim to investigate how aspects of language in task text may disfavour students less proficient in the natural language of the test; these students are often second language learners (SLL). For example, many of the studies included in the research review (Study 1) investigate differ-

ences between the performance of SLL and non SLL (e.g., Choi et al., 2013; Lee & Randall, 2011). A method often used is to rewrite mathematics tasks so that one version is simplified in regards to several challenging linguistic features (e.g., Abedi, 2000; Johnson & Monroe, 2004). Not so surprisingly, the result is often the simplified tasks are solved correctly to a higher extent than the original tasks, by students less proficient in the language of the test. These results can be used to accommodate tests for second language learners. For example, in a study by Lee and Randall (2011), it is explicitly stated that the results have implications for test construction and for text accommodations. However, one crucial question in relation to such accommodations is, if a mathematical language ability should be assessed; this question is not always addressed. In Lee and Randall's (2011) study, one aspect investigated is vocabulary use. The study shows a difference in performance depending on whether the words 'equivalent' and 'lowest terms' are used or not. It is important to note the authors do not comment on the relevance of these words in a mathematics assessment (see Lee & Randall, 2011). However, based on the standpoint that there is a mathematical language (see section 2.1.2), these words are probably relevant to include in a mathematics assessment and should not be changed solely because they are difficult.

Other studies reveal some differences in performance on mathematics tests when read-aloud accommodations are provided for students who are less proficient in the language used in the test, as shown in the studies of Bolt and Thurlow (2007) and, Helwig, Rozek-Tedesco, Tindal, Heath, and Almond (1999). Both studies reveal particular demanding language aspects that the students can overcome if the task is read aloud. There are, however, additional results that call in question read-aloud accommodations, since they can alter what is being assessed. A study by Bolt and Ysseldyke (2006) reveals that the amount of tasks that are easier to solve (not just understood correctly) when a read aloud accommodation is used is substantial. Around a quarter of the tasks were easier to solve when the students received the read-aloud accommodation. As suspected, the read aloud accommodation was more appropriate on a mathematics test than on a reading and language arts test, but there is still a risk of high contamination from the accommodation on the mathematics test as well. When the read aloud accommodation was used for one group, the measurement was not comparable between the different student groups.

The reason why accommodations are used is to avoid construct irrelevant variance in the assessments. This is important, but it is hard to in a good way implement the accommodations because it is not just to take away potentially difficult language. Research on accommodations of tests with the purpose of reducing construct irrelevant variance is important because tests are used to draw conclusions about students' abilities in mathematics. An obvious potential source for construct irrelevant variance is lack of accommodations

for lower language ability students. Through a meta analysis, Kieffer, Lesaux, Rivera and Francis (2009) evaluate the effectiveness of different accommodations for students with lower language ability. Kieffer et al. focus on students that do not have English as their first language. Seven different types of accommodations are used in the reviewed studies: simplified language, dictionaries or glossaries, bilingual dictionaries or glossaries, tests in the native language, dual language test booklets, dual language questions for some passages, and extra time. Based on the review, the authors suggest that accommodations are not particularly efficient to reduce the disadvantage for the students with limited language proficiency. Only *one* of the tested accommodations resulted in a significant improvement on the SLL performance. Providing a dictionary or glossary served to reduce the gap between SLL and non SLL students, but only a small reduction in the performance gap was identified. Based on the findings, the authors suggest that accommodations are not the solution to promoting academic skills of SLL students. One way to overcome the difficulties related to language can be to improved by how academic language is taught; this is suggested by Kieffer et al. (2009). They argue that intensive instruction on the language that is used in the content area is needed. Knowledge of the mathematical language and how to teach it is important. Unfortunately, many researchers make statements regarding what characterizes the mathematical language, even when the empirical ground for the statement is weak (Österholm & Bergqvist, 2013). Fortunately, the research field is emerging and the collection of empirical results regarding the mathematical language is growing. For example, corpus analyses on textbooks reveal that in comparison with language in natural science the mathematical language has more subject specific words, more longer words, and more sentences with a typical structure (see Ribeck, 2015). Still, instruction on mathematical language would be easier to implement if more research on the characteristics of the mathematical language was available.

3. Methods and methodological considerations

The method section is structured in three subsections related to the research questions in different ways. The research questions addressed in this thesis are:

RQ 1) Are there any particular textual features in mathematics tasks that are related to task difficulty, and if so, how?,

RQ 2) Are there any particular textual features in mathematics tasks that are related to task demand on reading ability, and if so, how?, and

RQ 3) Regarding textual features that in any way are related to how difficult the tasks are to read or solve - is the particular textual feature's difficulty a mathematics specific difficulty or not?

The method section begins with a presentation of the data (section 3.1), then explains the three different methods used to analyse the task text (section 3.2). Research questions 2 and 3 both concern *reading* of mathematics tasks and the statistical method used to obtain a measure for the tasks' demand on reading ability is presented in the last section (3.3). All three questions are about potential relations between the textual features and aspects of difficulty. Several types of difficulty is included in the expression 'aspects of difficulty', something that is explained in section 1.2. The two main aspects of difficulty, a difficulty that includes everything that may affect the success on solving and a difficulty that have to do with undue reading demand, are either the measure based on solution frequency or the measure for a non-mathematics specific demand on reading ability (DRA). The statistical methods used to analyse such relations are commonly used methods and are therefore only shortly summarised in the beginning of section 3.3.

3.1 Data
All three research questions are about aspects of difficulty in mathematics task text; different types of data are used to answer those research questions. The first type is research articles: these are used in a structured literature review (Study 1). The second type of data is mathematics tasks, which are analysed to reveal potentially difficulties related to the task text (Study 2-4). This section focuses mainly on the second type, that is, the type of tasks used in the analyses. Some differences between the tasks are explained and then Table 1 describes how different types of tasks are used in the analyses conducted to answer the research questions. Mathematics tasks from two different tests, Programme for International Student Assessment (PISA) (OECD,

2013) and from the Swedish national test in mathematics for grade 9 (Skolverket, 2013), hereafter referred to just as SweNT, are analysed. The two tests are taken by 15-year-old students, and, for the SweNT, students that will turn 15 during the particular school year are included. The two tests differ from each other from several perspectives, and this difference gives breadth to the data. Properties of the two tests and some notable differences between them are described here. After the description of the properties, an overview is given of which data is used in order to answer each research question.

PISA mathematics and the Swedish National Test in Mathematics
PISA tasks are chosen for one particular reason, namely PISA includes both a reading and mathematics test. It is therefore possible to use the same student's result on both reading and mathematics tasks; this in turn makes it possible to use the results on the reading tasks to draw conclusions about the mathematics tasks' demand on reading ability. The statistical method used to do this is presented in section 3.3.1. PISA tasks are also chosen as data because it is possible to compute solution frequencies based on a large number of students' results for those tasks. The SweNT is considered a good complement because more tasks were available than for PISA and the SweNT is a test constructed based on the Swedish curriculum.

The main reason for the inclusion of two test samples is that it gives a greater breadth in the data. The PISA framework is based on the concept of mathematical literacy, whereas the SweNT is composed to assess the competences described in the Swedish curriculum. Despite this difference, the language of mathematics is important within both those constructs (the PISA framework and the Swedish curriculum), making the two tests good choices as data in research focusing on textual features in mathematics. Both tests assess mathematical competencies and there is good agreement between them concerning which competencies are assessed.

Some differences between the tests are also evident. For example, the focus on mathematical literacy in PISA is apparent through the focus on four particular context categories; this is not as explicitly addressed in SweNT (Skolverket, 2015). Other apparent differences between the two tests are that multiple choice questions are more common in PISA and, whereas very short tasks are more common in SweNT (e.g., "Calculate 4·0,75+0,5=_"). Those differences in the two samples are seen as beneficial since the inclusion of those two types of tasks gives a larger breadth to the data. There is also a difference in the test situation, that may affect the results on the two tests, namely that the tests have different roles for the students. The results on the SweNT are important for the students as individuals since the SweNT has a large impact on the grading of the students in their mathematics course. The

same cannot be said about the PISA test. In that sense, SweNT may be advantageous as data. The SweNT is desirable because the solution frequencies are based on solutions where the students have made an effort. Otherwise, there is a larger risk of influence from other factors than mathematical ability on the results.

Another benefit of using two different test samples is that a comparison between results of the analyses can be used as a reliability measure, to some extent. Differences between the results of the two samples can be reasonable according to differences between the tasks, and at the same time if the results are similar despite those differences, it is likely to be a sign of reliability in the results.

The use of the two test types in relation to the different research questions
All three research questions are about textual features in task text, therefore mathematics tasks are used as data. Research questions two and three both concern the demand on reading ability. Since a measure for the tasks' demand on reading ability was possible to obtain only for PISA mathematics, those tasks were used to answer research question two and three. Task difficulty is addressed in research question three and research question one. Measures for task difficulty are obtained based on solution frequencies that are available for both PISA mathematics and SweNT tasks. Accordingly, both PISA tasks and SweNT tasks are used as data in the analyses conducted to answer research question one and three.

The data analysed in relation to each research question is presented in Table 1. Recall that two different measures are used to represent how demanding the tasks are to read and solve: "difficulty", which is based solely on the tasks' solution frequency, and "DRA", which explains the tasks' demand on reading ability (DRA). Since it was not possible to obtain useful values for DRA on every PISA task in the samples, fewer tasks were used in the analyses regarding DRA.

Table 1: *Type of dependent variable and data used in the analyses conducted to answer each research question and in which study the data is used.*

Research question	Dependent variable	Tasks from PISA	Tasks from SweNT	Study
1 & 3	difficulty	84 tasks from 2003 & 2006		2
1 & 3	difficulty	133 tasks from 2003 & 2012	354 or [a] 364 tasks from 2004 - 2013	3-4
2 & 3	DRA	63 tasks from 2003 & 2006		2
2 & 3	DRA	105 tasks from 2003 & 2012		3-4

[a]In the SweNT sample in Study 4 one task from each test year was excluded and therefore one sample was smaller than in Study 3.

Measures of DRA and difficulty are based on results from around 1,500 students for each PISA task and around 2,000 students for each SweNT task. Solution frequencies are freely available for both PISA tasks (OECD homepage: http://www.oecd.org/pisa/pisaproducts/) and SweNT tasks (The Swedish National Agency for Education: http://www.skolverket.se). For some years, the solution frequencies for SweNT were not available when the studies were conducted. For those years access to the solution frequencies were provided by Katarina Kjellström at the Swedish National Agency for Education.

Some of the SweNT tests are freely available at the homepage for the PRIM group who are responsible for the SweNT in mathematics for grade 9 (http://www.su.se/primgruppen/). For both the SweNT tests that are not freely available and the PISA tasks used, access has been applied for and granted through The Swedish National Agency for Education.

3.2 Analyses of task text

The purpose of this thesis is to contribute to the knowledge of aspects of difficulty related to textual features in mathematics tasks. One step to fulfil this purpose is completing different types of analyses of mathematics task text. This section is about the different analyses of text and a few crucial methodological considerations made in relation to each analysis. The three research questions all concern different textual features of the task text. Two main types of textual features have been analysed, namely aspects of vocabulary and features dealing with multisemiotics. These two types of textual features are chosen to encompass several important features of the written mathematics text. However, the main focus is on multisemiotics since that is a central feature of mathematical language (see section 2.2).

Several aspects of vocabulary are focused on in the literature review (Study 1) but are not considered here since no textual analyses were conducted. One textual feature analysed is word commonness. The commonness of vocabulary is chosen since there are convincing results showing a relation between a person's vocabulary and reading comprehension (Qian & Schedl, 2004). It is reasonable to assume that a person with a large vocabulary knows more uncommon words whereas a person with small vocabulary does not. Therefore word commonness is an important textual feature to investigate. Word commonness is also related to particular categories of words. For example, there is a particular category of vocabulary a student needs to grasp to be mathematically proficient; this set of vocabulary can be categorised based on word commonness. Different categories of uncommon vocabulary are therefore relevant to focus on. For example, words that are common within the mathematical context can be perceived as relevant in mathematics tasks.

Corpora (i.e., large structured sets of texts) are used to as a reference to capture how common words are in particular contexts. This process is described in section 3.2.1. For the analyses that focus on multisemiotics, text is categorised in four different semiotic resources. Those four semiotic resources are defined in section 3.2.2. The analyses of multisemiotics concern both the presence of semiotic resources, which is described in section 3.2.2, and aspects of multisemiotic cohesion, which is described in section 3.2.3.

3.2.1 The use of corpora in text analysis

In one analysis (Study 2) the textual feature focused on is word commonness. A particular emphasis is laid on the distinction between words that are relevant in relation to a mathematical reading ability and words that are not relevant; this distinction is made based on word commonness (Figure 2).

Different categories of words can be defined in relation to how common the words are in two different contexts, the everyday context and the mathematical context. If every word in a task is categorised as either common or uncommon in these two contexts the result is four categories of words. The four categories of words are explained and exemplified in relation to Figure 2, after an explanation of how the corpora are used in the analysis. The four categories of words–either common or uncommon in the two contexts–are either part of the mathematics register (technical words) or not. It is only the words that are part of the mathematics register that are relevant to test in a mathematics assessment. This distinction is also illustrated in the theoretical model (section 2.1.2) through the division of the reading ability into two parts, where only one part is included in mathematical ability.

Two different corpora are used as reference when judging if a word is uncommon or not. Each corpora represent one particular context, the everyday context and the mathematical context. A corpus is a collection of texts or spoken language that are used to study and describe a language (McEnery, Xiao, & Tono, 2006). Corpora are used since they provide reliable measures of word commonness in relation to the particular contexts of interest. In this thesis corpora are used as references for the everyday context and the mathematical context. By using one corpus representing the mathematical register and one corpus representing the everyday language, each word could be reliably categorized as common or uncommon within those particular contexts.

An assumption is made that there is a larger risk that words are more difficult if they occur to a lower frequency in the chosen corpora. Such an assumption is very reasonable since the relation between word frequency and word difficulty has been proven statistically (Breland, 1996). The textual feature analysed in the task text is the amount (number or fraction) of words in different categories of the task text. Figure 2 illustrates how every word in

the task text is sorted into one of four categories, depending on the word's frequency in each of the two corpora. The dotted lines represent the median frequency of the analysed words in the particular corpora.
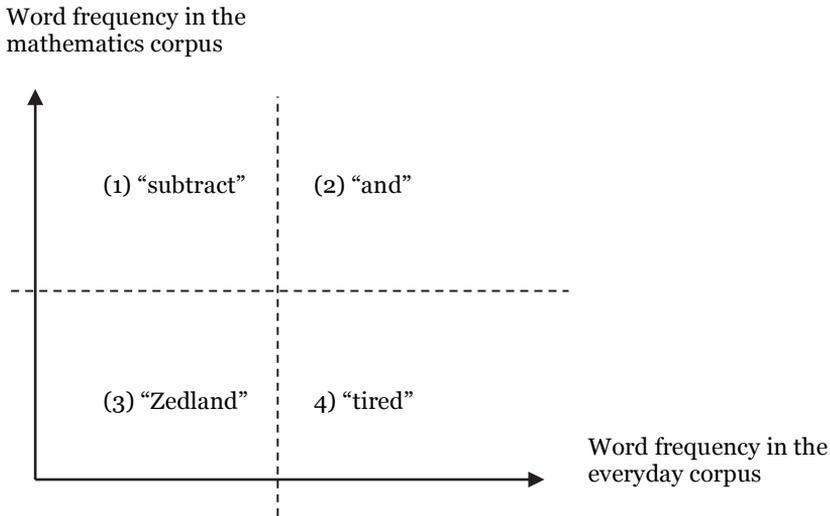
Word frequency in the
mathematics corpus

Figure 2: *Illustration of how the analysis results in four different categories of words depending on their frequency in two corpora.*

The words in the first category (1) that are common in the mathematics corpus but uncommon in the everyday corpus is referred to as a technical vocabulary. The second category (2) of words is common in both corpora. Category (3) and (4) are both uncommon in the mathematics corpus but category (3) is also uncommon in the everyday corpus, while (4) is not. The four words in Figure 2 exemplify the four word types, for example "Zedland", a word used in the PISA test, is uncommon in both corpora. Category (3), generally common words, is not of interest in the analyses since the aim is to reveal potential difficulties. Among the other three categories, only one word belongs to the mathematics register (1); this is because mathematics vocabulary must be common in the mathematical context and the other category that is common in the mathematical context is not specific for mathematics. When words from this particular category (1) are read it is assumed that a mathematical reading ability is utilized and it is also reasonable to assume difficulties related to reading such words are part of what a mathematics test should assess. This potential difficulty that is reasonable to assess is illustrated in Figure 1 in section 2.1.2 by the arrow from the field representing a mathematical reading ability to "test result".

Crucial in this analysis of word commonness is that a relative measure for commonness is used. Corpora are used as reference, but since no absolute limit between when a word is common or uncommon exists, the distinction

is made based on all words in the analysed tasks. For example, all words in the PISA tasks that occur more often in the mathematics corpus than the word with median frequency is categorised as words common in the mathematics corpus. This choice means the results concern commonness in relation to the particular corpora used and in relation to the vocabulary in the PISA test.

3.2.2 Four different semiotic resources in task text

To facilitate an analysis of the semiotic feature of a task text in relation to aspects of difficulty, four different semiotic resources are defined, namely *natural language, mathematical notation, schematic images,* and *pictorial images*. The crucial part of the analysis is the definition of the four semiotic resources since the definition affects the categorisation of the text and what is possible to catch in the analyses. This section will therefore focus on how the four semiotic resources are defined and why it is reasonable to divide a text in these four types. The choice of semiotic resources is made based on earlier research and practical issues, and is done in relation to which type of conclusions are possible to draw depending on how the text is categorised.

Natural language is defined as language in the form of sentences, phrases, single words, or even single (Latin) letters. Natural language is analysed as one semiotic resource based on the fundamental feature of natural language that distinguishes it from both mathematical notation and images. Natural language consists of words composed of single letters, letters that correspond to particular speech sounds. Experienced readers do not read letter by letter, but even when word recognition is used in reading, the single letters and syllables are crucial (e.g., Balota, Yap, & Cortese, 2006). Therefore, reading natural language is very different from reading for example mathematical notation, such as $\div$.

Mathematical notation is defined as symbols that are used following special conventions in mathematics. Based on Pimm's (1987) categorisation of mathematical symbols, symbols from four categories are coded as mathematical notation: logograms (e.g., π and $\div$), pictograms (e.g., $||$ and $\angle$), letters (e.g., μ and β), and punctuation marks (e.g., ! and]). There is one exception from Pimm's categories; Latin letters used as mathematical notation are categorised as natural language, not as mathematical notation. This exclusion is made since Latin letters differ from other mathematical notation because each letter corresponds to a particular speech sound used in everyday language. This distinction is also done by for example Drouhard and Teppo (2004) who argues that linguistic signs like x are part of a natural language, unlike other mathematical notation. Therefore they mean that letters such as x and y should be studied with a linguistic framework whereas other mathematical notation should not.

Mathematical notation is categorized as a separate semiotic resource compared to images. This distinction is partly based on the notion that mathematical notation is considered essential within the written mathematical language and partly on differences in what message images and mathematical notation can convey. Images differ from mathematical notation both in that they have the affordance of spatiality and that a central feature of images is they reflect or depict something. Mathematical notation on the other hand is more abstract. A single symbol can represent something very exact and comprehensive (e.g., the symbol $\sum$). A particular category of mathematical notation, pictograms, have similarities with images, since they can be seen as simplified images that are reduced in size. Pictograms that are mathematical notation are however different form small images since only the stylized forms of pictograms that are often used in mathematics (e.g., $\angle$) are categorised as mathematical notation.

The category images is divided in two different categories: pictorial images and schematic images. Those categories are not self-evident. For example Shorrocks-Taylor and Hargreaves (1999) defines three types of image categories, but those categories are based on the role the image has in a task, for example if it is illustrative. In the current research, the aim is to draw conclusions about particular features of the image, not the role that the images take in the tasks; therefore Shorrocks-Taylor and Hargreaves' categories are not considered adequate. Instead, the categories of semiotic resources are made only based on form.

Schematic images are images where the relations between parts in the image are emphasized and *pictorial images* are images where the likeliness with the object depicted is emphasized. The division of images in mathematics text into those two categories has been used in earlier research (e.g., Blatto-Vallee, Gaustad, Porter, & Fonzi, 2007; Hegarty & Kozhevnikov, 1999; Martiniello, 2009). An argument shown through results from research for the two types of images, schematic and pictorial, is that the two types of images are connected to two types of visualizers. One type of visualizers prefers spatial/schematic imagery in problem solving and another group prefers visual/pictorial imagery. The spatial/schematic visualizers are better at dealing with abstract spatial relations and process images analytically, part by part, whereas the other group of visualizers process images holistically, as a single perceptual unit (Kozhevnikov, Hegarty, & Mayer, 2002; Kozhevnikov, Kosslyn, & Shephard, 2005).

This division of images into the two categories, schematic images and pictorial images, leads to a division where all mathematical diagrams, such as tables, graphs and 'naked' geometrical figures, are categorised as schematic. Schematic images are also images on which the emphasis for the layout is on the represented parts and the relations within the image, for example an underground map.

In summary, the categorisation of a task text in these four different semiotic resources enables a clear distinction between the different semiotic resources differing substantially in the means used to convey the message. The four categories facilitate the analysis of two different textual features focused on in this thesis, namely i) which semiotic resources that are present in task text, and ii) whether meaning relations occur between different semiotic resources or not. To be more exact, in the first analysis (i), the four different semiotic resources facilitate analyses considering several features regarding presence: presence of each semiotic resource, co-occurrence of particular semiotic resources, and number of different semiotic resources. In the second analysis (ii), the categorisation of the text into different semiotic resources is crucial for a clear distinction of whether a meaning relation exists between different semiotic resources or not. This particular analysis of meaning relations is presented in the next section.

3.2.3 Analysis of cohesion in multisemiotic text

All three research questions focus on textual features, and one such textual feature investigated is *meaning relations* in multisemiotic text. This particular analysis is based on a framework of cohesion. Cohesion in natural language and in multisemiotic texts is defined in section 2.2.4 as particular meaning relations between or within sentences in the text (Halliday & Matthiessen, 2004).

This section describes which parts of Hasan's (1989) conceptual framework of cohesion in natural language are used and how. Only parts of the framework are used because when multisemiotics are added, each aspect, or part of the framework, creates several additional dimensions resulting in the analysed categories becoming too numerous with only a small sample of cases in each. Therefore a decision was made to not include these parts. Cohesion can be studied at word, phrase, and sentence level; the analysis may also consider objects, circumstances, or processes. The analysis in this thesis focuses on abstract or concrete *objects* studied at *word level*. For mathematical notation and images that have no words, only instances that can be labelled with words representing objects are included. For example the word 'circle' and an image representing a circle would be included in the analysis.

Essential within Hasan's (1989) framework of cohesion are the notions cohesive tie and cohesive chain. A cohesive tie is a semantic relation between two spatially separated instances in a text. For example, in the phrase "the square is biggest since it…" there is a tie, a semantic relation, between *square* and *it,* since *it* refers to *square*. When more than two instances in the text are related with ties, that is by Hasan (1989) referred to as a cohesive chain. For example, in the sentence "a square is a rhombus since it…", the words *square*, *rhombus*, and *it* are tied together in the same cohesive chain.

In the analysis of multisemiotic cohesion, the notion cohesive network is used instead of cohesive chain. The word *chain* gives associations of linear reading and since linear reading cannot be assumed in multisemiotic texts, *network* is a more suitable word. Whenever at least two entities in the task text are related through a cohesive tie, those related instances are analysed as one cohesive network. In the analysis of cohesion, the categorization of the text in different semiotic resources (see 3.2.2) is used to distinguish when a cohesive tie is intersemiotic, that is, when there is a cohesive tie between two *different* semiotic resources. Cohesive ties within only natural language are intrasemiotic. This focus on cohesive ties, cohesive networks and the distinction between intersemiotic and intrasemiotic ties enables an analysis of exactly what the analysis is supposed to capture: meaning relations of importance in relation to the reading and solving of mathematics tasks. Those meaning relations are important since they make the text coherent (see also 2.2.4).

An analysis of the task in Figure 3 is presented in Table 2 and Table 3 to illustrate the analysis of cohesion in task text. There are four different cohesive networks in the task, and each network is given one colour in Figure 3. The lines that connect instances in the task text represent cohesive ties. Every word underlined with the same colour should also be attached to each other, but not all of those cohesive ties are drawn in Figure 3 because the illustration would become too messy.
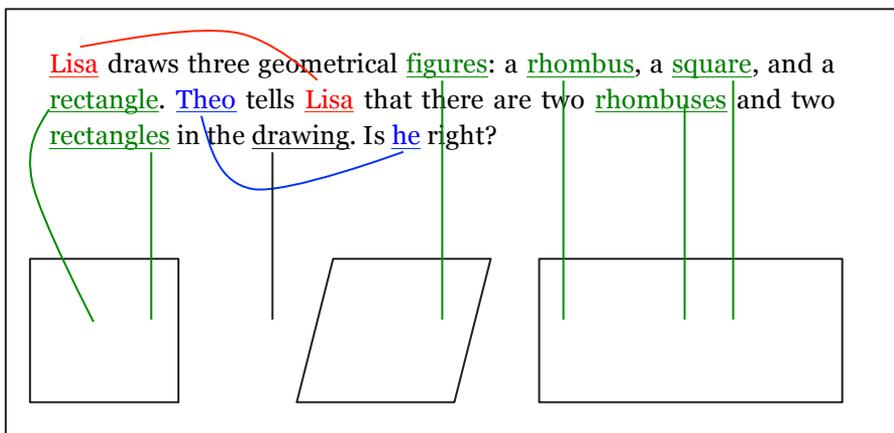

Figure 3: *Analysis of cohesion in the mathematics task "Figures".*

It is apparent from the example that the cohesive ties are different types. For example, the two instances of 'Lisa' are tied cohesively through repetition, whereas the tie between rhombus and square are a hyponymy. The different cohesive ties are explained in Article 4.

According to the analysis of the example there are four different cohesive networks in the task shown in Figure 3. When a word is tied to a word in a network, that new word belongs to the whole network. For example, in network 2, any word cohesively tied with even one word in network 2 becomes part of network 2; this inclusion in the network does not require the word to be cohesively tied to each individual word within network 2.

Table 2: *Cohesive networks in the task "Figures".*

Network	Related instances in natural language	If related to the schematic image[a]
1	Lisa, Lisa	no
2	Figures, rhombus, square, rectangle, rhombuses, rectangles	yes
3	Theo, he	no
4	Drawing	yes

[a]In the analysis of the tasks it is also coded whether the networks are related to pictorial images or mathematical notation.

Five different variables are analysed, an explanation of those variables are presented, followed by the result of the analysis of the task "Figures", presented in Table 3. All three research questions posed in this thesis concern potential difficulties related to textual features in task text. In this particular analysis, intrasemiotic and intersemiotic meaning relations in task text are considered and variables are chosen to represent different aspects of those relations. The analysis starts in natural language, since natural language has a prominent role in the reading of the tasks. Therefore, every network has at least one instance in natural language. Images can also be more open for different interpretations than words. The analysis is therefore likely to be more reliably conducted when starting in natural language and looking for instances in other semiotic resources cohesively tied to a particular word than if the analysis is done the other way around. For example, an image depicting a circle can represent a ball, a ring, a sphere or the equation $x^2+y^2=r^2$, whereas the word 'ring', for example, cannot represent a ball.

A cohesive network is intersemiotic when there are at least two different semiotic resources in the network. When all instances in a network are natural language, the network is intrasemiotic. For the task in Figure 3 the information coded for the five variables are presented in Table 3. Those five variables facilitate an analysis of the textual features of interest.

Variables 1-3 concern textual features that might demand the reader to connect different semiotic resources in the text and variables 4-5 concern relations between instances in the text without separation between intersemiotic or intrasemiotic relations. The separation between ties and networks enables an analysis of whether the number of meaning relations (cohesive

ties) or different types of objects (measured as cohesive networks), presented in the task text are demanding.

Table 3: *Coding on each variable for the task "Figures".*

Variable	Presence or number of
1. If any intersemiotic tie is present	yes
2. Number of words with an intersemiotic cohesive tie	6
3. Number of intersemiotic cohesive networks	2
4. Number of words with a cohesive tie	10
5. Number of cohesive networks	4

Essential for the analysis is the identification of concrete and abstract objects and the distinction between different semiotic resources. Objects are identified as nouns, pronouns or–in other semiotic resources–instances that can be labelled with a noun. However, the distinction between image and natural language is not always straightforward. For example, a table can consist only of lines and words; words are natural language, but at the same time, words in a table are something very different to read compared to words in sentences. If words in such a table are analysed as natural language, a task text consisting of a table and a few sentences would be analysed as if it consisted only of natural language. To avoid such a misleading analysis, words in tables and diagrams are analysed as part of the images. For example, to avoid a misleading analysis, the table heading "Length" is analysed as length represented through the semiotic resource schematic image.

3.3 Statistical analyses
Three research questions are posed in this thesis. All questions address relations between textual features and how difficult the tasks are to read or solve. A few statistical methods are used to answer the research questions. Principal component analysis (PCA) is used to obtain a value for how difficult tasks are to read, which is described in 3.3.1. In addition to reading demand (DRA), task difficulty is analysed; the proportion of correct answers on a task is used as a measure of how difficult the task is to solve. Those two measures, the tasks' demand on reading ability (DRA) and task difficulty are analysed in relation to different textual features. The relations between textual features and tasks' both DRA and task difficulty, are analysed using different types of correlation analyses and t-tests. T-tests and correlation analyses are commonly used methods to statistically analyse relations, and are therefore not described in more detail here.

Obtaining a measure for demand on reading ability (DRA)

In this thesis the existence of a particular type of reading ability, apt for mathematics (section 2.1.2) is assumed. Such an assumption implies that there is also a type of reading ability that is *not* relevant from a mathematics perspective. When such a reading ability is needed in the reading and solving of a task, the task has a demand on a non-mathematics specific reading ability. From the perspective of validity of assessments conducted using mathematics tasks, it is this type of demand on reading ability that should be avoided in mathematics tasks (see also section 3.6 regarding validity in relation to assessments). In the current research, a measure named *demand on reading ability* (DRA) is used to analyse the demand on a non-mathematics specific reading ability, a measure originally introduced by Österholm and Bergqvist (2012).

An evaluation of four statistical methods (two variants where correlations were used, a regression analysis, and a PCA) used to characterize mathematics tasks' demand on reading ability found that using a principal component analysis (PCA) was the best method when both reliability and validity was taken into account (Österholm & Bergqvist, 2012). Based on the results of Österholm and Bergqvist's study, in this thesis PCA was used as the statistical method to obtain a relevant measure for demand on non-mathematics specific reading ability, from here on shortly labelled demand on reading ability (DRA).

The data analysed with PCA is students' results on PISA mathematics tasks and PISA reading tasks from year 2003 and 2012. A simplified version of the data for only 3+3 tasks is presented in Table 4. The table illustrates one important characteristic of the data, namely that a student's results on both the reading and the mathematics tasks are connected to that individual.

Table 4: *A simplified example of the data analysed with PCA.*

	Scores on mathematics tasks			Scores on reading tasks		
	Math 1	Math 2	Math 3	Read 1	Read 2	Read 3
Student 1	2	1	1	1	.	2
Student 2	2	1	0	1	.	0
Student 3	2	1	.	1	2	0
Student 4	0	0	.	1	1	0

An "." in the table represents missing data. In the original data file retrieved from OECD (http://www.oecd.org/pisa/) there was a specific coding for tasks students had not tried to solve; this coding refers to the tasks after the last task a student had solved in the test. Such cases were re-coded to missing. This change in the coding is done since only results from tasks a student has tried to solve are reasonable to use in an analysis of aspects of difficulty.

The purpose of the PCA on this particular data is to investigate the correlation between the students' scores on both the mathematics and the reading tasks. The PCA results in a few principal components that represent something more general than the original variables. For example, if a detailed written visualisation were needed for full score on many of the tasks, a principal component representing a kind of drawing skill would be expected. Since the principal components are based on correlations between the scores on all tasks, the PCA also gives loading values for each task on the extracted principal components. A loading value describes the strength of the relationship between the original variables and the principal components. The first two principal components extracted in the PCA conducted on the PISA data are interpreted as a reading ability component and a mathematical ability component. This interpretation is based partly on the data (tasks that assess reading and mathematics) and partly on the pattern for the loading values. Examples of loading values for only six tasks are presented to illustrate the basis for why one of the principal components is interpreted as explaining reading ability in Table 5. The loading values are values from the real data, but in order to illustrate the pattern for the whole sample with only a few values the tasks for which the loading values clearly follows the overall pattern are specifically chosen. In the whole sample there are also examples of tasks for which the pattern is more ambiguous than presented here, but the overall pattern is shown in the following table.

Table 5: *Loading values on the first two principal components for three mathematics tasks and three reading tasks.*

Task	Principal component 1: mathematical ability	Principal component 2: reading ability
Math 1	**0.35**	0.26
Math 2	**0.41**	0.27
Math 3	**0.66**	-0.08
Read 1	0.08	**0.48**
Read 2	0.07	**0.34**
Read 3	-0.23	**0.70**

The first two principal components the PCA results in are interpreted as components representing mathematical ability and reading ability. The pattern for the loading values on the first two principal components can be seen in Table 5. The first principal component has high loading values on most of the mathematics tasks but low loading values on the reading tasks, whereas the second principal component has high values on the reading tasks (emphasized in boldface) with a few exceptions. Based on these loading values and an expectation of the first two components to represent mathematical ability and a reading ability, the first component is interpreted as represent-

ing mathematical ability and the second a reading ability. The interpretation of the components was originally done by Österholm and Bergqvist (2012), based on PCA on another sample of tasks. The method was adopted in this thesis and used on another sample with the same interpretation.

One important property of the principal components is that they are uncorrelated; this is achieved by an oblique rotation of the fist solution of the PCA (see also Tabachnick & Fidell, 2007). This property is important since when uncorrelated, the loading values represent the unique contribution of each principal component to a variable (for example a mathematics task), but without inclusion of variance coming from the overlap between correlating principal components.

The values used in this thesis are positive *loading values* for the component interpreted as a "reading ability component" on the mathematics tasks. These positive loading values can be seen for the first two mathematics tasks in the example in Table 5. The loading value for the reading ability component on a mathematics task represents how much of the variance is *uniquely* explained by a particular principal component (the reading component) for that variable (in this case the mathematics task). Therefore each mathematics task's loading value on the reading component should be interpreted as a non-mathematics specific reading ability. This measure for reading demand was originally given the name demand *of* reading ability (Österholm & Bergqvist, 2012), but is here referred to as demand *on* reading ability (abbreviated DRA). Ideally mathematics tasks should test mathematical ability and nothing else; therefore the loading values representing DRA are evidence for the presence of some type of *unnecessary* demand on reading ability in mathematics tasks (see also section 2.1). However, the DRA is not considered unnecessary, per se. In reading tasks, which are supposed to assess reading ability, high loading values on DRA indicates a good construct validity.

4. Results and conclusions

The result section is structured in relation to the three research questions posed in this thesis. This structure has been chosen in order to facilitate a summary, or even better, a synthesis based on results from several studies. Section 4.1 concerns task difficulty in relation to textual features (RQ 1), section 4.2 focuses on the tasks demand on reading ability (DRA) in relation to textual features (RQ 2), and the last part, section 4.3, presents results and conclusions concerning whether textual features' potential difficulties are mathematics relevant difficulties or not (RQ 3). The conclusions presented in section 4.3 are based on an interpretation of the combined results regarding task DRA and task difficulty (the results presented in sections 4.1-4.2).

For all statistical analyses, the limit p=0.05 is used for significance. The statistics not presented here since they can be found in the articles. Therefore, in the tables where the results are presented, the results are only referred to as significant or not.

4.1 Textual features in relation to task difficulty

The results and conclusions presented in this section concern research question one: *Are there any particular textual features in mathematics tasks that are related to task difficulty, and if so, how?* This question is answered based both on the results of a literature review (Study 1) and on different types of analyses of textual features in task text. Two different types of textual features are analysed: vocabulary (Study 2) and multisemiotic features of the task text (Study 3-4).

Results
None of the textual features that have to do with vocabulary are related to difficulty. Some of the results regarding different word aspects are based on a sample of studies, since it is a review (Study 1), and the results on word commonness are based on correlations tested in one study (Study 2). Since Study 1 is a research review, separate studies reviewed at times show results that indicate that a particular textual feature is difficult. However, the results presented here regard the review as a whole. Therefore, only results about a particular word aspect where the results are consistent in at least 2/3 of the studies are reported. Based on the review, it is only possible to draw reliable conclusions regarding word commonness and word length, because of differences in methods among the studies for the other features. The result from the review then shows that these particular word aspects are not related to task difficulty.

The analysis of commonness of different categories of words (Study 2) reveals that no category of uncommon words is related to task difficulty.

The analyses regarding multisemiotic features (Study 3-4) reveal several textual features that are related to task difficulty. The analysis of the presence and co-occurrence of different semiotic resources (the tasks' semiotic characteristic) resulted in four different categories that differ significantly in task difficulty from the rest of the tasks. Common to these four categories (tasks with different semiotic characteristic) is that pictorial images are present in every group of tasks that is more difficult. Another important result is that the number of semiotic resources in a task is not related to difficulty. This lack of correlation between number of semiotic resources and difficulty suggests that it is actually the *type of* semiotic resources that are present in the task text that are connected to difficulty, not simply the presence of several semiotic resources in the task text. All significant results relating to the presence of semiotic resources (semiotic feature and number of semiotic resources) in task text stem from the analyses of the PISA sample. For the analysis of SweNT, no group of tasks of a particular semiotic type was significantly more difficult than the rest of the tasks.

The other analysis of multisemiotic features focuses on cohesion in the task text. Analyses are conducted on all instances of cohesive features in the task text, and also separately on the intersemiotic cohesion. The results on all instances of cohesive features are presented first, and then the results on intersemiotic cohesion. Two different measures for the extent to which there is cohesion in the text are analysed, namely *number of words with cohesive ties* and *number of cohesive networks*. Both variables turned out to be significantly correlated to difficulty for both test samples in the analyses where no distinction is made between intersemiotic or intrasemiotic cohesion. Recall that the cohesion is intersemiotic when more than one type of semiotic resource is tied together cohesively. These results are the same for both test samples.

Analyses are also conducted to test whether intersemiotic cohesion is in any way more or less difficult than intrasemiotic cohesion. There is only one significant result relating to intersemiosis: tasks for which there are at least one intersemiotic cohesive tie have a higher mean difficulty than the rest of the tasks. However, the difference is only significant in one sample (SweNT) and the effect size is small (Cohens' d= 0.257). The unique correlation (without intrasemiotic cohesion) between number of intersemiotic cohesive ties as well as number of intersemiotic cohesive networks and difficulty is also analysed, which is a partial correlation using the number of *intra*semiotic cohesive networks and words with *intra*semiotic cohesive ties as control variables. None of these features are correlated to difficulty in either of the test samples.

The results from all analyses of textual features' relations to task difficulty are presented in Table 6. For the analyses where two test samples were used (Study 3-4), it is noted which results are obtained only for one sample and

which results are the same for both test samples. In Study 2 only the PISA sample was analysed.

Table 6: *Results on the relation between textual features and task difficulty in the four different studies in this thesis.*

Study	Sample	Textual features related to difficulty	Textual features not related to difficulty
1	36 articles	none	•length (letters or syllables) •general commonness
2	PISA	none	•globally uncommon •uncommon only in math •uncommon only in everyday language
3[a]	PISA uniquely	four semiotic characteristics where pictorial images are present in the group of more difficult tasks	none
	SweNT uniquely	none	•three semiotic characteristics that includes pictorial images •one semiotic characteristics with all semiotic resources except pictorial images
	Both tests	none	•four semiotic characteristics that as a group do not say anything particular about pictorial images •number of semiotic resources
4	PISA uniquely	none	•intersemiotic cohesive tie present
	SweNT uniquely	•intersemiotic cohesive tie present	none
	Both tests	•number of words with cohesive ties •number of cohesive networks	•number of words with intersemiotic cohesive ties •number of intersemiotic cohesive networks

Note. A significant relation (correlations or significant mean difference) is positive when the presence of or more of a particular textual feature is correlated to more difficult tasks. All correlations are positive. For the differences in mean difficulty, the textual feature that are more difficult are presented.
[a]The semiotic characteristics that are significantly correlated to difficulty in the PISA sample are the same as those not significantly correlated to difficulty in SweNT. They are differently described since when there is no significant correlation it is not meaningful to say anything about the group of tasks that do not have the particular semiotic characteristics.

Common to the semiotic characteristics (Study 3) of the tasks that are more difficult in the PISA sample is that pictorial images are either present in the

more difficult tasks, or the group of tasks that are less difficult (negative relation) do not include pictorial images. Exactly which textual characteristics that are related to difficulty are explained in Article 3.

Conclusions

Based on the results of the separate analyses, some conclusions can be drawn regarding which textual characteristics are related to difficulty in any way.

None of the investigated word aspects (Study 1 & 2) are related to difficulty. The conclusion drawn from this result is that none of the investigated word aspects can reliably be said to be particularly difficult in mathematics tasks. This conclusion is drawn based both on the review of 36 studies of potentially difficult vocabulary and on the nonexistence of a relationship between any of the categories of uncommon words and task difficulty.

For textual features concerning semiotic resources other than natural language, conclusions can be drawn regarding difficulty in relation to both presence of semiotic resources and cohesive relations in the task text. One conclusion concerning the presence of semiotic resources is that particular semiotic characteristics of task seem to be difficult, specifically that combinations of semiotic resources where pictorial images are present are more difficult. Another conclusion is that the different semiotic resources seem to have different roles in the different tests; this could imply that the tasks are different from a multisemiotic perspective and that this difference is related to task difficulty. These conclusions are based on two results: first, the significant correlations in the PISA sample and, second, that the results are different between the two test samples. The common factor for the more difficult tasks, the presence of pictorial images, is hard to interpret and further analyses are needed to be able to draw any general conclusions regarding particular semiotic resources and combinations thereof (the tasks' semiotic characteristics).

Regarding cohesion, the conclusion is that the tasks' cohesion is important for how difficult the tasks are, since both number of words with cohesive ties and number of cohesive networks are variables positively correlated to difficulty. For intersemiotic cohesion, the conclusion is that the intersemiotic feature seems to be one part of the complexity, but only as one aspect among several. This conclusion is based on the weak relation between the presence of intersemiotic cohesive ties and difficulty, and on the significant correlations between the combination of intersemiotic and intrasemiotic features and difficulty.

In conclusion, the multisemiotic features investigated in this thesis seem to be more strongly related to task difficulty than the examined aspects of vocabulary.

4.2 Textual features in relation to task demand on reading ability

The results and conclusions presented in this section correspond to research question two: *Are there any particular textual features in mathematics tasks that are related to task demand on reading ability, and if so, how?* This question is answered based on different types of analyses of textual features in task text. Two different types of textual features are analysed: different categories of uncommon vocabulary (Study 2) and multisemiotic features of the task text (Study 3-4).

Results

Recall that the statistical variable demand on reading ability (DRA) measures a demand on a reading ability that is not part of a mathematics specific reading ability. DRA does therefore represent a reading demand that a mathematics task does not aim at assessing (see section 3.3). Study 1 does not focus on reading demand and therefore no results from that study are presented here.

Two variables are significantly correlated to DRA: the amount of globally uncommon words and the number of cohesive networks in task text. For the different categories of vocabulary included in the analyses, only globally uncommon words are significantly correlated to DRA. The globally uncommon words are words that are uncommon both in a mathematical context and in an everyday context.

The analyses of cohesion in task text reveal a relation between cohesive features of the text and DRA. The number of cohesive networks is negatively correlated to DRA. In addition, the number of words with cohesive ties is also correlated to DRA, but for cohesive ties the relation is only nearly significant ($p=0.081$). The analyses show no relations between intersemiotic cohesion and DRA. The mean task DRA is not higher for tasks with any intersemiotic cohesive tie. Also, neither the number of intersemiotic cohesive ties nor the number of intersemiotic cohesive networks are significantly correlated to DRA.

For the different semiotic characteristics of tasks, that is, which semiotic resources are present, only seven of the twelve different semiotic characteristic that were included in Study 3 were possible to test in relation to DRA, since for the other five semiotic characteristic there were too few tasks for the statistic analysis to be meaningful. For the tested features there was no group of tasks with a particular semiotic characteristic that has a significantly higher mean DRA than the other tasks. Table 7 present the results concerning the measure DRA.

Table 7: *Results on the relation between textual features and the tasks' DRA in three different analyses.*

Study	Features related to DRA	Features not related to DRA
2	globally uncommon words	•words uncommon only in math •words uncommon only in everyday language •words common in both math and everyday language
3	none	•all seven tested semiotic characteristics •number of semiotic resources
4	number of cohesive networks (neg)	•presence of intersemiotic cohesive ties •number of words with cohesive ties (neg)[a] •number of words with intersemiotic cohesive ties •number of intersemiotic cohesive networks

Note. A significant correlation is positive when the presence of or more of a particular textual feature is correlated to more difficult tasks. Negative relations are marked "(neg)".
[a]Number of words with cohesive ties is negatively correlated to DRA and the correlation is almost significant (p=0.081).

In Study 3 and 4 analyses were conducted both on the PISA sample and on the SweNT sample. All results were the same for both test samples in those analyses. The seven semiotic characteristics tested in Study 3 are explained in Article 3

Conclusions

Three main conclusions can be drawn about textual features that may play a role for the tasks' (non-mathematics specific) demand on reading ability (DRA).

First, words that are uncommon both in an everyday context and in a mathematical context can contribute to a task's DRA. For words that are uncommon only in one of those contexts, or common in both, there is no such relation to DRA.

Second, the presence of semiotic resources other than natural language in task text is not related to the extent to which a non-mathematics specific reading ability can be used when solving the tasks. This conclusion is drawn based on the fact that there are no significant correlations between the presence of different combinations of semiotic resources and task DRA. Natural language is present in every task except one and therefore these conclusions regard combinations of semiotic resources in addition to natural language.

Third, the presence of more cohesive features in a task text means that the task demands less of a non-mathematics specific reading ability. Moreover, it is cohesion generally that is related to DRA, not any particular intersemiotic feature of the cohesion. Intersemiotic cohesive ties play no essential role for

the task's demand on reading ability: the intersemiotic cohesive ties are important as ties among all ties but they do not stand out in the results. These conclusions are based on the significant correlations between the number of cohesive networks and DRA, and also on the nearly significant correlation between number of words with cohesive ties and DRA. The almost significant correlation is considered relevant as a basis for the conclusion since the correlation follows a clear pattern among all correlations (see Table 7).

4.3 Conclusions based on results regarding difficulty and DRA interpreted together

The conclusions presented in this section correspond to research question three: *Regarding textual features that in any way are related to how difficult the tasks are to read or solve—is the particular textual feature's difficulty a mathematics specific difficulty or not?* This research question is answered based on the results from the analyses on the PISA sample that are conducted in relation to research question 2 & 3. Only results from the PISA sample are relevant since results for demand on reading ability (DRA) can only be obtained for the PISA tasks.

The existence of a reading ability that is specific for reading mathematics text, and therefore can be seen as part of a mathematical competence, is relevant in relation to the third question posed in this thesis. This mathematical reading ability is represented by field 2 in the theoretical model presented in section 2.1.2 (Figure 1). To be more precise, the research question concerns the difference between a reading ability that is perceived as a mathematical reading ability, and reading ability that is reading ability generally but from which the mathematical reading ability has been excluded. This difference is also clearly visualized in the theoretical model, as the difference between field 2 and field 3.

By using the statistically computed measure DRA, it is possible to analyse textual features in relation to a reading ability that does not include mathematics-specific reading ability, that is, that are contained in field 3 of Figure 1. The reading ability that is specific for reading mathematics text is not analysed separately; only task difficulty (solution frequency) and task DRA are analysed. However, by interpreting the results regarding task difficulty and task DRA for the same textual features, it is possible to draw conclusions about whether the difficulty identified in relation to particular textual feature is mathematics specific.

In Study 2-4, various textual features are analysed in relation to both task difficulty (solution frequency) and task DRA. Three conclusions are drawn based on an interpretation of both those types of analyses. First, a reading ability that is not specific for mathematics can be used when solving mathematics tasks with globally uncommon words (words uncommon both in an

everyday and a mathematical context). This means that tasks with many globally uncommon words suffer the risk of low construct validity, since potential reading demands related to those words are construct irrelevant factors in a mathematics test. This conclusion is based on the significant correlation between globally uncommon words and DRA.

Second, the difficulty that is related to the presence of particular semiotic resources in task text (different semiotic characteristics) is not unwanted in the test, because the ability needed to deal with this particular difficulty is part of a mathematical ability. This conclusion is based on the pattern for the relations between the different semiotic characteristics and difficulty as well as DRA: every textual feature that is significantly correlated to difficulty is uncorrelated to DRA.

Third, the presence of more cohesive ties and networks in task text means that the task is likely to assess exactly the desired construct, namely mathematical ability. It is not important that there are many intersemiotic ties in the task; it is the total number of cohesive networks and number of words with cohesive ties that is related to the mathematical difficulty of tasks. This conclusion is based on the fact that the relations identified in the statistical analyses have operate in opposite directions.

For the number of cohesive networks in tasks, there is a significant positive correlation to difficulty and a significant negative correlation to DRA. The correlations are in the same direction for the number of words with cohesive ties as for cohesive networks (positive for difficulty and negative for DRA). For number of words with cohesive ties the correlation to DRA is only nearly significant (see Table 8). Interpreted together, these results indicate that tasks with more words with cohesive ties and more cohesive networks are more difficult to solve but the same tasks also demand less non-mathematics specific reading ability. For non-mathematics specific reading ability the relation is the opposite: tasks with few cohesive networks and cohesive ties demand more of the non-mathematics specific reading ability.

Table 8: *Results on significant, or nearly significant, correlation between textual features and both difficulty and DRA in the analyses of cohesive ties and networks.*

Independent variable	Data sample and dependent variable		
	SweNT difficulty	PISA difficulty	PISA DRA
number of words with cohesive ties	0.648 (**0.000**)	0.244 (**0.005**)	-0.171 (0.081)
number of cohesive networks	0.130 (**0.014**)	0.347 (**0.000**)	-0.242 (**0.013**)

From a validity perspective, the conclusion is therefore that tasks with more cohesive ties and cohesive networks are preferable, since for those tasks there is less risk of assessing a non-mathematics specific reading ability. A

simplified visualisation of the relation between the two measures DRA and difficulty and the number of cohesive networks and/or words with cohesive ties that are present in a task is presented in Figure 4. On the y-axis, a higher value represents tasks that are more difficult to read (DRA) and to solve (difficulty).

Value for the tasks'
difficulty and DRA

DRA   Difficulty

                                                              Number of
                                                              cohesive ties
                                                              and/or networks
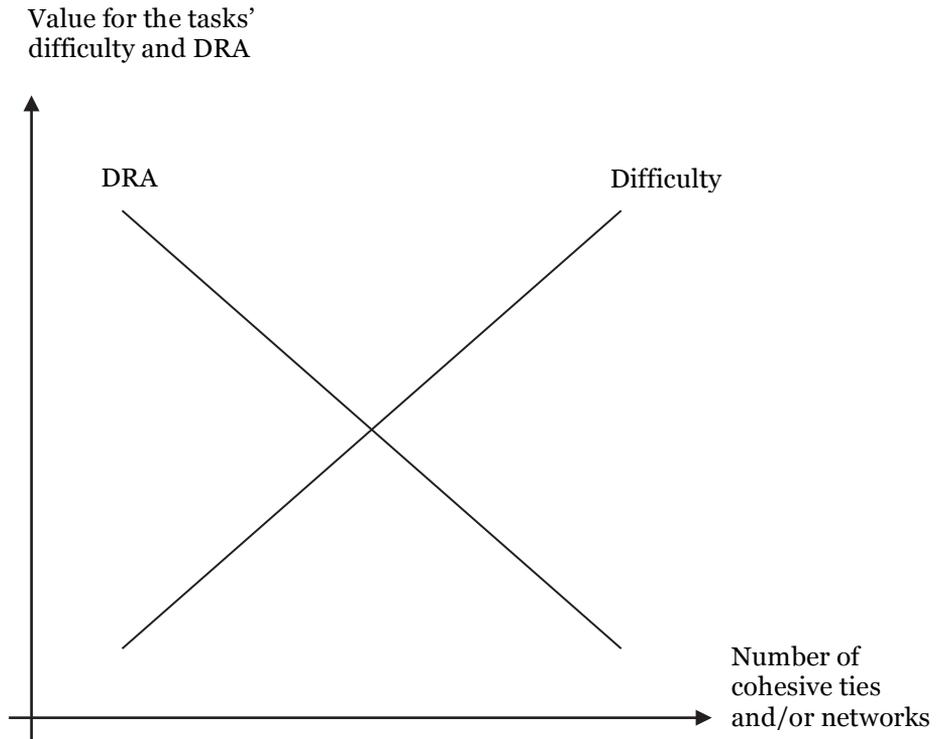
Figure 4: The relation between *different aspects of cohesion and both task demand of reading ability (DRA) and task difficulty.*

Figure 4 is, of course, just an illustration to visualize that the correlations to DRA and to task difficulty have opposite signs. It is important not to interpret the figure as an exact graph. However, the figure supports the interpretation of Table 8 since it makes the signs of the correlation coefficients easier to grasp.

5 Discussion

The discussion is structured based on a comprehensive view of the results presented in this thesis. The first two sections address *difficulty* from two different perspectives, difficulty that is unwanted and difficulty that is relevant in mathematics tasks. In essence, a key claim is that difficult textual features in a mathematics task are not problematic in themselves. The difference between a problematic difficulty and a reasonable difficulty is addressed. In section 5.1, unwanted difficulty is discussed and it is argued that it is only one of the textual features investigated in this thesis that enhances the risk of assessing something else than mathematical ability. In section 5.2, difficulty is addressed from another perspective. It is argued that there are particular textual features that are both difficult and reasonable to be part of what a mathematics task assesses. Both those textual features have to do with the multisemiotic task text. In section 5.3 a few issues related to research about textual features are discussed, and the necessity of diversity in the research field is argued for. The chapter ends with some implications from the research and some suggestions for future research.

5.1 Unwanted difficulties that can be attributed to textual features

In this section, textual features that are both difficult and unwanted in a mathematics task are discussed. It is argued that despite the results that indicate that several textual features are difficult, it is only *one* of those that are problematic from a validity perspective.

The literature review (Study 1) did not reveal any reliable results pointing to any word aspect as particularly difficult. It is worth noting that it was not possible to draw any conclusions about many of the word aspects included in the review because of the diversity of methods used in the reviewed studies. It was only possible to draw conclusions regarding word length and general commonness, which were that those categories of words are not difficult. For the textual features that involve multisemiotics, results in this thesis reveal that several features are difficult, but this difficulty is not considered problematic (see section 5.2). Therefore among the results there is only one textual feature that is both difficult and unwanted in mathematics tasks, namely globally uncommon words. Globally uncommon words are words that are uncommon both in an everyday context and in a mathematical context. Those words are considered to be potentially problematic based on the results regarding demand on reading ability (DRA) (see also section 3.2.1). This means that the presence of globally uncommon words in mathematics tasks should be avoided since tasks with such words enhance the risk of assessing a non-mathematics specific reading ability.

These results regarding textual features that are unwanted in mathematics tasks contribute to the developing research field about textual features in mathematics tasks, informed by earlier research.

First, based on earlier results, the age of the test takers is an important aspect in relation to textual difficulties. Earlier studies have shown a decrease in how strongly reading ability is related to mathematical ability between earlier and later grades (Chen, 2010; Hickendorff, 2013). These studies found that reading ability was more important for performance in mathematics in earlier grades. In this thesis, the age of the participants was not considered in the literature review (Study 1), and in the study about different categories of uncommon words (Study 2) the results are based on 15-year-old students. It is therefore important to be aware that textual features that are not difficult according to in the results in this thesis can potentially be difficult for younger students.

Second, based on the results about globally uncommon words, more nuanced research is required concerning word commonness. The results suggest that it would be useful to continue research about *different categories* of uncommon vocabulary. Some of the studies that are reviewed (Study 1) focus particularly on, for example, technical vocabulary, but much of the research focuses on the broad category of uncommon words (without separation between contexts).

The results about globally uncommon words are also important in relation to research about language accommodations of tests. As argued in section 2.3, it is neither very surprising nor very enlightening that when task texts are extensively rewritten to be easier to read, the tasks are correctly solved to a higher extent. However, a question that often lies behind studies where such accommodations are tested is very relevant, namely a question about construct validity. Based on the meta-study conducted by Kieffer and colleagues (2009), it is very reasonable to argue for the need for more useful results regarding how to accommodate tests for students with some kind of language difficulties. They found that among the seven methods used to accommodate the tests, only one slightly reduced the performance gap between native speakers and students with limited language proficiency, which was the provision of a dictionary. Kieffer et al. (2009) argue that accommodations for second language learners are largely ineffective and that one reason is that the students have deficiencies in their construct relevant English, that is, the mathematical language. In relation to those results, the results presented here about globally uncommon words are relevant since a particular category of problematic words is distinguished. To obtain a high construct validity of an assessment it is reasonable to avoid words that are not just uncommon, but uncommon both in an everyday context and in the mathematical context. However such words should be avoided for the whole student group, not just as an accommodation.

5.2 Textual features important in relation to a mathematical competence

In this section, textual features that are difficult but *not* problematic from the perspective of construct validity in mathematics assessment are discussed. It is argued that some particular textual features are difficult but at the same time relevant in mathematics tasks. This argument is based on the statistical results, as well as on an understanding of an aspect of mathematical competence that has to do with the reading and solving of mathematics tasks. This section is structured in relation to the two textual features that are relevant in this section, the semiotic characteristics of the task and meaning relations in the task.

The results reveal that tasks with these two multisemiotic features are difficult and at the same time do not have a higher demand on reading ability (DRA). That is, these tasks do not demand more of a non-mathematics specific reading ability. It is therefore concluded that the ability to read and solve tasks with those features is part of a mathematical competence. The logic of this conclusion can be illustrated in the theoretical model of abilities (Figure 1, section 2.1.2). Based on what DRA represents, it can be assumed that a mathematical reading ability (field 2 in the model) is utilized in the solving of tasks with those textual features since the textual features are positively related to high difficulty but not to high DRA.

The semiotic characteristic of the task refers to the combinations of semiotic resources that are present in the task text. The results show that the presence of pictorial images is common to the more difficult tasks. The relation between a difficulty that is relevant within mathematics and pictorial images is a bit unexpected since pictorial images are seldom seen as particularly mathematical.

There is also the possibility that a covariate is the source of the correlation between the particular semiotic resource and difficulty, a covariate that occurs together with the semiotic resource. One example of such a possible covariate could be a particular kind of demanding content in geometry that in all cases is assessed by a particular combination of semiotic resources. Another factor that could cause a relation between the presence of pictorial images and difficulty would be if pictorial images in the PISA tasks were being used to make demanding tasks more appealing. Such a relation could explain the correlation between difficulty and pictorial images, but no such use of pictorial images was easily identified.

There are also some features of the data that are relevant to the difficulty relating to pictorial images. Such features support the claim that it is likely that this difficulty involves the combination of semiotic resources, not just the presence of pictorial images. Investigating this claim was limited by which combinations of semiotic resources could be tested. There were too

few tasks with only natural language and pictorial images, and no other semiotic resources, to test that particular semiotic characteristic. That also means that for all other results where tasks with pictorial images are in the group of more difficult tasks, very few of those tasks had only pictorial images and natural language in the task text. In almost all of the tasks with a higher difficulty there are combinations of pictorial images and two other semiotic resources. Therefore it is likely that the difficulty has to do with the co-occurrence of the semiotic resources. Since the co-occurrence of particular semiotic recourses is related to difficulty, but not to the tasks' demand on reading ability, the semiotic characteristic of the task is likely to be important for the tasks' construct validity.

Earlier studies reveal difficulties related to both images and mathematical notation in task text (e.g., de Kirby & Saxe, 2014; Driver & Powell, 2015) and also that the problem solving process is affected by which diagrams are used (Elia & Philippou, 2004). Those results contribute to explaining the difficulty revealed in relation to the tasks semiotic characteristics, but it is very likely that the interaction between the different semiotic resources also is part of the explanation. Several researchers emphasise some types of meaning expansion that occur when several semiotic resources are used together (e.g., Liu & O'Halloran, 2009; Lemke, 2002; Unsworth & Cléirigh, 2009). Lemke (1998) explains that meanings created by one semiotic resource can alter the meaning revealed by another semiotic resource and therefore the set of possible meanings are multiplied when several semiotic resources are used together. Such a meaning expansion can also help explain the difficulty that has to do with the tasks' semiotic characteristics. There are also other results that in different ways reveal difficulties that students experience when working with different semiotic resources. For example, students experience difficulties related to translations between different semiotic resources (see e.g., Capraro & Joffrion, 2006; Janvier, 1987; Lech et al., 1987) and difficulties in the construction of a useful model based on the multisemiotic task text (Hegarty et al., 1995). However, the results (Study 4) regarding the difficulty aspect related to particular combinations of semiotic resources add to these previous results by distinguishing this task feature as a mathematics specific difficulty.

The other textual feature that is difficult and relevant within a mathematics task is cohesion, specifically intersemiotic and intrasemiotic cohesion between concrete or abstract objects in the task text. This conclusion is based on results showing that the tasks' demand of a non-mathematics specific reading ability decreases as the presence of those textual features increases, but the general task difficulty increases. The interpretation of the results is that the number of words with cohesive ties and the number of cohesive networks in the tasks are textual features that take some effort from the stu-

dents in the solving of the tasks, and that a mathematical ability is utilized when dealing with those features.

There are also earlier results that in different ways reveal that task difficulty is related to presence of meaning relations in the text or to the need to relate instances in mathematics tasks are difficult. For example, Turner and colleagues (Turner et al., 2009) found a relation between task difficulty and the extent to which the reader needs to move backwards and forwards between and within different semiotic resources in the text. Other studies also show that students experience difficulties when they have to connect different semiotic resources during problem solving (see e.g., Hegarty et al., 1995; Moon et al., 2013). It has also been shown that difference in types of cohesive relations, that is, how explicit the reference is between different semiotic resources, influences the reading. Acartürk et a. (2013) found differences both in eye movement parameters and in success on a post-test on the content of the text, depending on type of reference (how explicit the reference is) between natural language and images. The differences in eye movement parameters, depending on reference used in the text, are interpreted by Acartürk et al. (2013) as differences in how difficult it is for the readers to integrate the information contributed by the text. Therefore, an aspect of difficulty related to cohesion in the task text could be expected in Study 4. However, the results in Study 4 show that cohesion between concrete or abstract objects is one such textual aspect that is difficult. Acartürk and colleagues' results relate to the results in this thesis and may have implication for the analyses of cohesion in mathematics tasks (Study 4), since a task text with many words with cohesive ties also has many references that potentially can affect the reading in ways that are either beneficial in the solving or not. Thus, it may be that students have difficulties in identifying less explicit relations in the task text, relations that may be important to correctly understand the task.

The results also contribute to what is known about difficulties that have to do with, for example, relations in the task text, since *cohesive networks* are included in the analyses. The number of cohesive networks in the task text indicates something about how many different types of objects the task is about, and therefore how many 'types of objects' the reader has to handle. For example, representations of persons and representations of money in a task text will be part of different cohesive networks.

The results regarding multisemiotic textual features contribute to earlier results by adding the aspect of unnecessary reading demand. Recall that the results show that task text with more cohesive ties and networks are more difficult but demand less of a non-mathematics specific reading ability. Therefore, it is concluded that those textual features are part of what a test should assess in order for the assessments to have high construct validity.

The sample of tasks analysed, namely PISA tasks and SweNT tasks, is also an important aspect in relation to the results. For example, if extensive text irrelevant for the task is included in the task text, this would reasonably result in an increase in cohesive ties and cohesive networks. But if this addition of text makes the task more difficult, it is not a mathematics specific difficulty. Since the measure for DRA could only be obtained for the PISA tasks, the results regarding demand on reading ability stems from the PISA sample. For the PISA tasks, irrelevant texts are rare and the cohesive ties can therefore be assumed to concern content relevant for the interpretation of the task. More words with cohesive ties and more cohesive networks then reflect an increase of features central for the interpretation of the task. With this interpretation of the relation, the results in Study 4 that reveal that tasks with a low DRA have more of cohesive ties and networks are also reasonable. The ability to read and interpret the cohesive ties in the task text is therefore likely to be part of a mathematical reading ability, not a general reading ability, at least to some extent. However, the documented relation between mathematics ability and reading ability is very convincing (e.g., Edge & Friedberg, 1984; Hickendorff, 2013) and therefore this conclusion must be drawn with caution.

It is also important to take the result revealed by Teich and Fankhauser (2005) into consideration when the results about cohesion, revealed in this thesis, are interpreted. Teich and Fankhauser found that that subject specific words are overrepresented among words that are cohesively tied, compared to all words in a particular text. The results in this thesis reveal a correlation between words with cohesive ties and task difficulty and it is reasonable that those words are to a large extent subject specific. The relation found by Teich and Fankhauser can be part of the explanation to the lower demand on reading ability for those tasks, but the analysis in this thesis is multisemiotic. Therefore, these earlier results are relevant but only to parts of the results in this thesis.

For natural language there are several studies that reveal that the ability to recognize or identify cohesive ties is positively correlated to reading comprehension (e.g., Bayraktar, 2011). The better the students are able to recognize the cohesive ties, the higher reading comprehension scores they have. Bayraktar's results may be interpreted as the opposite to the results in this thesis, where the presence of more cohesive relations in the mathematics task is related to a lower demand on a non-mathematics specific reading ability. However, the result from the Study 4 analyses makes sense if one takes into account what the cohesion in the mathematics tasks represents, namely meaning relations between objects important for the interpretation of the task text. Another difference is that Bayraktar's study focuses on general reading ability, whereas the present results considering DRA are only about part of the reading ability, the non-mathematics specific reading ability. The

theoretical model of abilities (Figure 1, section 2.1.2) is sufficient to illustrate the differences between what the results of Bayraktar's and Study 4 are about. While Bayraktar's results are about a general reading ability (field 2 & 3 in the model), the results from Study 4 about DRA are only about a non-mathematics specific reading ability (field 3 in the model). Aided by this model, a negative relation between a non-mathematics specific reading ability and a textual feature essential in the interpretation of a mathematics task makes sense.

Bayraktar's study is also rather specific in how the ability to recognize cohesive ties was measured. For example, students were explicitly asked to identify particular words, such as a synonym to a known word, in a text. It can be assumed that it is different to identify cohesive ties in the reading of a text and to answer questions about which words that are related by cohesive ties. Therefore, the textual feature analysed in this thesis and the cohesion analysed by Bayraktar are only to some extent similar, which can also explain why in this thesis the presence of more cohesive ties is related to a higher DRA.

The cohesive ties and networks consist of meaning relations between entities within natural language, and also between entities represented using different semiotic resources. The ability to correctly decode and interpret the text with those meaning relations is an essential part of the ability to solve the task. From that viewpoint it is reasonable to conclude that the presence of more of such cohesive ties is not related to a higher demand on reading ability that is unnecessary in the mathematical context.

In summary, the results show that the semiotic feature of a task and the amount of cohesive ties and networks in the task are features relevant in mathematics assessments, and that these results are reasonable from several perspectives. Both conclusions from earlier empirical results about semiotic resources in mathematics tasks and an evaluation of what the textual features represent are in line with the results.

5.3 Studying textual features in mathematics tasks

In this thesis, task text is studied and conclusions about potential difficulties are drawn based on solution frequencies. Studying task text instead of, for example, the actual solution process has its pros and cons, which is apparent in relation to limitations in scope and applicability of the results. However, the results in this thesis demonstrate a need of more studies on mathematics task text. Such research would ideally include both replication studies and innovative studies, for example, where methods are adopted from other research areas.

In the literature review (Study 1), in total 36 studies on potentially difficult vocabulary in mathematics task text were reviewed. In total, 16 different

word aspects were analysed in those studies. Despite this substantial sample of studies, conclusions about word difficulty could only be drawn regarding three of the word aspects. The main reason for this limited result has to do with method choices in the studies. The reviewed studies had a diversity of methods in combination with a frequent choice to not analyse individual word aspects separately. Hence there is a need for more replication studies, since such studies could substantially contribute to the possibility of making reliable conclusions based on many studies.

In the studies presented in this thesis, several different methods have been utilised, some which are somewhat explorative. Of course, the selection of possible methods is huge and there is no attempt to argue for some superiority of the methods used in this thesis. However, they are useful as examples in this argumentation for unconventional methods. Both the use of *two* corpora to distinguish particular categories of vocabulary and the use of the framework for cohesion to grasp issues about reading multisemiotic mathematics task text are somewhat new methods. None of these methods had been seen used previously in mathematics education research.

The use of two corpora to analyse task text enabled an analysis of different categories of uncommon vocabulary. The use of a framework for cohesion enabled an analysis of textual features important in the reading and solving of the tasks. The results from the studies contribute to what is known in the particular areas of interest and they provide a good starting point for future research in the same area. For example, the research suggests that it is relevant to take into account what vocabulary the students meet in their everyday school practice when judging and researching potentially difficult vocabulary in mathematics tasks. The results regarding meaning relations (cohesion) suggest that the meaning relations between objects in the task text are important to focus on in research about students' reading and solving of mathematics tasks. Therefore, the studies provide examples of the usefulness of new methods, particularly when new areas are researched.

In conclusion, studying mathematics task text in relation to aspects of difficulty is valuable in research about mathematics printed language. It is suggested that research within this area should continue, utilizing both exploratory methods and, in areas that are not quite new, conducting replication studies to strengthen the reliability of the emerging collection of results.

5.4 Implications for the research community and the school practice

The results of the research in this thesis have implications both for the research community and for school practice, in terms of both vocabulary and multisemiotic features of the tasks.

Regarding vocabulary in mathematics tasks, the results have implications for school practice, assessment practice, and the research community in different ways. In total, 16 different word aspects are analysed in the 36 reviewed studies. There are two reliable results from the literature review are relevant in relation to the school practice, namely that neither word length nor generally uncommon words in task texts seem to be related to task difficulty. This means that those word aspects are not particularly important to focus on in teaching about reading and interpretation of task text, at least for students after primary school. Also, in order to construct assessments that can be used to make valid assumptions about students' mathematics ability, it is not important to avoid those word aspects.

However, if words that are uncommon in both the mathematics context and an everyday context are present in the task text, that enhances the risk of assessing a general reading ability, not mathematics ability. It is worth noting regarding these words that *uncommon* does not mean that they must be *very* unusual words, since all analysed words are words that are present in the PISA tasks.

The results of the research review also have further implications regarding vocabulary. A rather common statement is that passive form and nominalisations (e.g., the use of "an addition" instead of "adding") are difficult and should be avoided in task text (e.g., Wolf et al., 2008). It was not possible to draw any conclusions regarding those word aspects in the review since the word aspects were not investigated separately in the reviewed research studies. Several studies reveal that tasks that are rewritten regarding, for example, *both* passive form and nominalisations (e.g., Lee & Randall, 2011) are solved to a higher rate than the original tasks. The method used in those studies is to rewrite the task text regarding several word aspects. Implications from the review are that i) there is a lack of knowledge regarding if/how those word aspects are related to difficulty, and that ii) it is of importance not to interpret studies where several word aspects are analysed together as results regarding the separate word aspects. Studies where task texts are extensively rewritten only give guidance about how a task text can be simplified regarding several word aspects, a simplification that is likely to result in a higher solution frequency. An essential question in relation to such simplifications is what part of the task text should be managed as part of a mathematical competence. If the task text is extensively simplified there is a risk of also simplifying textual features that are construct relevant.

Results that concern the multisemiotic feature of mathematics tasks also have implications for school practice and the research community. Results reveal that it is the presence of particular semiotic resources in task texts, not how many different semiotic resources that are present, that is related to task difficulty. These results imply that there are particular demands on the students that are associated with reading and solving mathematics tasks with

particular combinations of semiotic resources. It is therefore recommended to explicitly focus on multisemiotic features of the task text in teaching, especially since the difficulty aspect is considered to be a mathematics specific difficulty. Students may need guidance regarding which features are essential to focus on in the reading of task texts with several semiotic resources.

Another textual feature of importance for the reading and solving of mathematics tasks is the meaning relations (cohesive ties) in the text. Since the presence of more cohesive ties in task text is positively related to difficulty and negatively related to a demand on a non-mathematics specific reading ability, this feature seems to be mathematics specific. Identifying and correctly interpreting meaning relations in multisemiotic task texts is important for success in mathematics. Therefore, a focus on how to read task texts is recommended in teaching. For example, in developing a communicative competence or reasoning competence in mathematics, it can be valuable to focus particularly on meaning relations in task text.

Results regarding the multisemiotic task text also have implications for the research community. The results regarding difficulty related to the semiotic characteristics of the task text only occur for the PISA tasks, which is worth reflecting on. Pictorial images are present in every combination of semiotic resources in tasks that are more difficult to solve, but only for the PISA sample. These results have implications for research that either focuses particularly on the PISA test or just uses PISA tasks as data. This is true particularly since the SweNT sample used in the analyses was bigger than the PISA sample, which means that the difference in results should not be attributed to the size of the data. It is therefore reasonable to suspect that pictorial images play a slightly different role, or are used differently, in PISA than in for example the SweNT. This difference in the use of pictorial images is worth taking into consideration when conclusions are made based on research on PISA tasks. Schematic images can be difficult to interpret but the results imply that pictorial images should not be considered as trivial, rather, it is suggested that attention should be paid to the role of those images in the PISA tasks.

5.5 Further research

A single study contributes a little piece of information that adds to the developing knowledge within a particular research area. If, as for some of the research presented in this thesis, the particular research area is rather new, many different types of studies can contribute substantially to the existing research field. The implications for the research community suggested in section 5.4 are also implicit suggestions for future research. However, based on the results there are a few additional research areas that are identified as worth further investigation.

Regarding potentially difficult vocabulary, there is a lack of studies where word aspects are analysed separately, and therefore such studies are suggested. In Study 2, a rather small corpus representing the mathematics register is used, since that was the only corpus available. As new corpora are composed, it would be meaningful to continue the research about different categories of words in mathematics tasks, using corpora as reference. In particular, a larger mathematics corpus would be useful in analyses of technical vocabulary in mathematics tasks. The results presented that have to do with multisemiotics and cohesion contribute to the knowledge about particular textual features' role for task difficulty. However, much is still unknown and further analyses are needed to fully understand the features investigated. Both statistical analyses of task text and analyses of how students actually read the task text are important research within this area. Such analyses can contribute to an understanding of potentially difficulties regarding the text, difficulties that students experience when reading and solving mathematics task.

There is a lack of knowledge regarding multisemiotic task texts, in particular about how the semiotic resources interact in the presentation of the task. For example, within studies on only natural language, it is revealed that cohesive ties between paragraphs are more difficult than ties within the same paragraph (Bayraktar, 2011). It would be worth investigating whether this type of distance applies also to distance between different semiotic resources in mathematics tasks.

Sammanfattning på svenska

Bakgrund

Matematik examineras i stor utsträckning genom skriftliga test, något som innebär att elever för att lyckas i ämnet även måste läsa och tolka uppgifts-text. Matematiktext innehåller förutom naturligt språk även olika typer av bilder och det speciella symbolspråk som utvecklats inom matematiken. De särskilda skrivsätt som utvecklats för att kommunicera matematik (se t.ex. O'Halloran, 2005) är förstås en del i det som behöver läras som en del av matematikämnet. Således kan uppgiftstexten i test dels ses som en nödvän-dighet för att uppgiften ska kunna förmedlas skriftligt, dels som en del i det som faktiskt ska examineras, det vill säga att läsa och tolka uppgiftstexten. Det är därför viktigt att texten inte innehåller svårigheter utan direkt kopp-ling till matematikämnet.

Den forskning som presenteras i avhandlingen har genomförts i syfte att få kunskap om svårigheter elever har vid läsning och lösning av matematik-uppgifter. Mer specifikt fokuseras svårigheter relaterade till uppgiftstexten, något som är relevant på olika sätt. En aspekt handlar om att svårigheter relaterade till texten bör vara kopplade till matematikämnet, för de förmågor som testet examinerar. Om en uppgiftstext medför svårigheter som inte har med matematikämnet att göra riskerar detta att äventyra examinationens validitet, då något annat än matematisk förmåga testas. En annan aspekt handlar om att kunskap om svårigheter som har att göra med egenskaper i texten är av intresse från ett undervisningsperspektiv.

I avhandlingen antas läsförmåga kunna delas upp i, å ena sidan en *mate-matikspecifik* läsförmåga, å andra sidan en *icke matematikspecifik* läsför-måga. Den förra betecknar en läsförmåga som ingår i allt det som räknas som matematiska förmågor, medan den senare står för en läsförmåga som inte är specifikt knuten till matematikämnet. Den icke matematikspecifika läsförmågan är vidare att betrakta som irrelevant inom matematikkontexten men är nödvändig vid läsning av till exempel skönlitteratur.

Syfte och forskningsfrågor

Forskningens syfte är att bidra till kunskapen om vilka egenskaper i mate-matiktext som har betydelse för svårigheter som elever erfar vid läsning och lösning av matematikuppgifter. Genomgående i avhandlingen används ordet *text* i vidgad betydelse (se t.ex. Björkvall, 2010) som förutom naturligt språk (ord och bokstäver) även inbegriper matematikens symbolspråk (matema-tisk notation) och bilder.

De fyra studier som ingår i avhandlingen berör alla frågor om på vilket sätt uppgiftstext kan vara svår för eleverna, vilket medför att resultat från de enskilda studierna är relevanta att tolka tillsammans. För detta syfte har tre

övergripande forskningsfrågor formulerats. Var och en av dessa frågor besvaras av resultat från flera av de studier som ingår i avhandlingen.

1) Har några specifika textegenskaper betydelse för hur svåra matematikuppgifter är att lösa och i sådana fall, vilka?
2) Har några specifika textegenskaper betydelse för matematikuppgifters krav på en icke matematikspecifik läsförmåga och i sådana fall, vilka?
3) För textegenskaper som på något sätt har betydelse för svårigheter vid läsning och lösning av matematikuppgifter - är svårigheten matematikanknuten eller ej?

Metoder

Avhandlingens syfte är att bidra till kunskap om uppgiftstexters betydelse för svårigheter elever erfar vid läsning och lösning av uppgifterna. För att uppnå detta syfte används flera metoder. I tre av avhandlingens fyra studier (Studie 2-4) används olika typer av textanalyser och test av huruvida textegenskaper statistiskt kan påvisas ha betydelse för läsning och lösning av uppgifter. En studie (Studie 1) är en litteraturstudie där resultat från 36 tidigare studier sammanställs.

De data som används i uppgiftsanalyserna är 133 uppgifter från PISA matematik (Programme for International Student Assessment) från 2003 och 2012 samt 364 uppgifter från Skolverkets Ämnesprov i matematik för årskurs 9, åren 2004-2013. För PISA används elevsvar från 1500 elever per uppgift och för de svenska ämnesproven från 2000 elever per uppgift. Båda proven skrivs av elever som är 15-16 år gamla. Proven har delvis olika syften och kompletterar därför varandra. PISA syftar till att utvärdera trender över tid och ämnesproven i matematik syftar bland annat till att mäta elevers kompetenser såsom de är beskrivna i den svenska kursplanen i matematik. För PISA finns dessutom resultat för samma elev på både delprovet för matematik och för läsförståelse, något som behövs för att skapa ett mått på uppgifters krav på icke matematikspecifik läsförmåga (förklaras nedan i detta avsnitt).

Tre olika typer av textanalyser genomförs, en avseende ord och två avseende matematikens *multisemiotiska språk*: alltså, förutom naturligt språk även matematisk notation och bilder. När det gäller ord fokuseras tre olika kategorier som definieras av hur vanliga orden är i den vardagliga respektive den matematiska kontexten. De tre kategorierna av ord är *generellt ovanliga ord* (ovanliga i båda kontexterna), *tekniska ord* (vanliga i matematisk kontext men ovanliga i vardagsspråk), och *icke matematiktypiska* (vanliga i vardagsspråk men ovanliga i matematisk kontext).

De två textanalyser som fokuserar det multisemiotiska språket är olika så till vida att den ena enbart fokuserar vilka *semiotiska resurser* som finns i

uppgiftstexten medan den andra fokuserar betydelserelationer mellan textens alla delar. Fyra olika semiotiska resurser analyseras och alla analyserade uppgifter kategoriseras beroende på vilka, och hur många olika, semiotiska resurser som finns i texten. De fyra semiotiska resurserna är *naturligt språk*, *matematisk notation* (olika symboler) och två typer av bilder: *schematiska bilder* och *avbildningar*. *Schematiska bilder* är typiska i matematikuppgifter, t.ex. tabeller och diagram, men även andra bilder som är förenklade genom att bara illustrera de viktigaste delarna, är schematiska (t.ex. ritningar). *Avbildningar* är detaljrika bilder som i hög utsträckning liknar ett fotografi eller en detaljrik målning.

Textanalysen som fokuserar betydelserelationer (kohesiva länkar) i texten baseras på Hasan's (1989) ramverk för kohesion i naturligt språk men är utvidgad till att även omfatta bilder och matematisk notation. I analyserna görs åtskillnad mellan betydelserelationer som finns mellan olika semiotiska resurser (intersemiotiska) och de som finns inom samma semiotiska resurs (intrasemiotiska). Analysen är i naturligt språk begränsad till betydelserelationer mellan ord som representerar abstrakta eller konkreta objekt samt för bilder och matematisk notation sådant som kan namnges med dessa ord. Om en triangel omnämns i text både som *triangeln* och som *ABC* finns till exempel en betydelserelation mellan dessa båda ord. Innehåller uppgiftstexten även en bild av en triangel finns även en betydelserelation mellan bilden och orden *triangeln* och *ABC*. Flera instanser i texten som är sammanbundna med betydelserelationer analyseras som nätverk.

De olika textegenskaperna i uppgiftstexterna analyseras i förhållande till hur svåra uppgifterna är att läsa och lösa. Som ett mått på allmän svårighetsgrad används lösningsproportion: kvoten mellan elevernas totala antal poäng och det totala möjliga antalet poäng. Detta värde subtraheras från 1 så att ett högre värde representerar en svårare uppgift. Som ett mått på hur svår en uppgift är att läsa används ett värde som tagits fram med hjälp av en principalkomponentanalys. Värdet anger uppgiftens krav på en icke matematikspecifik läsförmåga. En viktig egenskap hos detta värde är att det enbart förklarar den läsförmåga som inte är del av en matematisk förmåga.

Två typer av statistiska analyser genomförs i syfte att besvara forskningsfrågorna. För textegenskaper som analyserats som närvarande eller icke närvarande i en uppgiftstext analyseras huruvida det är någon statistiskt säkerställd (signifikant) skillnad i medelvärde mellan uppgifter med och utan denna egenskap. Både uppgifters medelvärde för svårighet och för krav på icke matematikspecifik läsförmåga analyseras. För textegenskaper som analyseras utifrån i hur stor utsträckning egenskapen förekommer i uppgiftstext genomförs statistiska test av huruvida det är något signifikant samband mellan detta mått och uppgiftens svårighet. Både mått på hur svår uppgiften är att lösa och uppgiftens krav på icke matematikspecifik läsförmåga analyseras.

Resultat

Den första forskningsfrågan rör huruvida några specifika textegenskaper har betydelse för hur svåra matematikuppgifter är att lösa och i sådana fall, vilka dessa textegenskaper är. När det gäller ord analyseras ords vanlighet och i litteraturstudien även ordlängd. Resultaten visar att ingen av dessa egenskaper är relaterad till hur svåra uppgifter är att lösa. När det gäller vilka semiotiska resurser som finns i uppgiften finns däremot ett tydligt mönster för vilken typ av uppgift som är svårare. Uppgifter med vissa kombinationer av semiotiska resurser är signifikant svårare än de uppgifter som inte har denna specifika kombination av semiotiska resurser. Gemensamt för de uppgiftstyper som är svårare är att avbildningar finns i uppgiftstexten. Till exempel uppgifter med kombinationen: naturligt språk, matematisk notation och avbildningar, är svårare att lösa i jämförelse med uppgifter som inte har denna kombination av semiotiska resurser. Dessa resultat gäller dock enbart för PISA. För de svenska ämnesproven är det ingen skillnad i svårighet mellan uppgifter med vissa kombinationer av semiotiska resurser.

När det gäller betydelserelationer i uppgiftstexten visar sig både antal ord med betydelserelationer och antal nätverk där flera instanser i texten är sammanbundna med betydelserelationer vara relaterat till uppgifters svårighet. Fler betydelserelationer finns i svårare uppgifter. Inget övertygande samband finns dock mellan vare sig närvaro av eller antal betydelserelationer som är intersemiotiska och svårighet.

Den andra forskningsfrågan rör huruvida några specifika textegenskaper har betydelse för matematikuppgifters krav på en icke matematikspecifik läsförmåga och i sådana fall, vilka textegenskaper detta är. Resultaten visar att vilka semiotiska resurser som närvarar i texten inte har någon betydelse för uppgiftens krav på en icke matematikspecifik läsförmåga. Däremot har uppgifter med många globalt ovanliga ord högre krav på en icke matematikspecifik läsförmåga. För uppgifter med många betydelserelationer gäller att uppgifterna har ett lägre krav på en icke matematikspecifik läsförmåga. Sambandet är alltså omvänt i jämförelse med hur dessa egenskaper är relaterade till svårighet (lösningsproportion).

Den tredje forskningsfrågan rör huruvida de textegenskaper som på något sätt har betydelse för svårigheter vid läsning och lösning av matematikuppgifter går att karaktärisera som viktiga inom matematikämnet. Det vill säga om ett matematiktest bör testa dessa egenskaper och om det är egenskaper i texten som behöver beröras i undervisning. Slutsatser rörande denna fråga dras baserat på analyser av uppgifters svårighet och krav på en icke matematikspecifik läsförmåga tolkat tillsammans. De textegenskaper som är relaterade till uppgifternas svårighet och som förefaller vara matematikrelevanta är av två typer. För det första är den typ av svårighet som har att göra med närvaro av avbildningar i uppgiftstexten inte relaterad till en icke matema-

tikspecifik läsförmåga. För det andra är den svårighet som har att göra med betydelserelationer i texten positivt relaterad till svårighet och negativt relaterad till uppgiftens krav på en icke matematikspecifik läsförmåga. Således samvarierar dessa textegenskaper med hög svårighet utan att uppgiftens krav på en icke matematikspecifik läsförmåga ökar. I det senare fallet minskar till och med uppgiftens krav på en icke matematikspecifik läsförmåga när textegenskapen (många betydelserelationer) finns i hög utsträckning.

Slutsatser och diskussion

Baserat på avhandlingens resultat kan flera slutsatser dras rörande uppgiftstexters betydelse för svårigheter som elever har vid läsning och lösning av matematikuppgifter. De svårigheter som identifierats relaterat till textegenskaper är dock av två skilda typer: svårigheter som är matematikrelevanta och svårigheter som är oönskade i ett matematiktest. Denna skillnad är avgörande för resultatens implikationer och därför diskuteras de olika typerna av svårigheter separat. Avsnittet avslutas med några exempel på implikationer av avhandlingens resultat.

Den slutsats som dras rörande svårighet som är matematikrelevant är att flera egenskaper som har att göra med den multisemiotiska uppgiftstexten är svåra och samtidigt relevanta i matematikuppgifter. En av dessa egenskaper är den specifika kombinationen av semiotiska resurser i uppgiftstexten. Tidigare studier har visat att de olika semiotiska resurserna kan vara svåra på olika sätt (t.ex. de Kirby & Saxe, 2014; Österholm, 2006), något som skulle kunna medföra att ökat antal olika semiotiska resurser i uppgiftstext leder till ökad svårighet. Mina resultat visar dock inget sådant samband utan resultaten visar att uppgifter med specifika kombinationer av semiotiska resurser är svåra. Det indikerar att svårigheten ligger i hur dessa semiotiska resurser i samverkan utgör texten. En annan matematikrelevant svårighet är de betydelserelationer som finns i texten. Svåra uppgifter har mer av betydelserelationer i texten samtidigt som en icke matematikrelevant läsförmåga är mindre användbar i lösning av dessa uppgifter. Därför förefaller mängden betydelserelationer (mellan konkreta eller abstrakta objekt) i texten vara en textegenskap som är matematikrelevant. Avseende intersemiotiska betydelserelationer i texten påvisas, något oväntat, inga reliabla samband till vare sig svårighet eller krav på icke matematikspecifik läsförmåga när dessa analyseras enskilt. Dock ingår intersemiotiska relationer (t.ex. ord-bild) tillsammans med intrasemiotiska relationer (ord-ord) i de analyser där samband påvisas till aspekter av svårighet. Slutsatsen blir därför att intersemiotiska betydelserelationer inte är krävande i sig men att de ingår som en del i den svårighet som har att göra med alla betydelserelationer i texten.

Den slutsats som dras rörande oönskad svårighet är att av de egenskaper som undersökts är det enbart globalt ovanliga ord som är både svåra och

oönskade i matematikuppgifter. Ord som är ovanliga både i en vardaglig kontext och i matematikkontext förefaller nämligen ha potential att påverka om uppgiften examinerar matematikförmåga eller läsförmåga. Ingen ökad svårighet kunde påvisas i förhållande till mängden tekniska ord eller mängden icke matematiktypiska ord. Således visar dessa resultat att när vokabulär i uppgiftstext beaktas i syfte att öka testets validitet är det av vikt att inte enbart fokusera på hur vanliga ord är generellt. Även ordens vanlighet i en matematikkontext bör tas med som en faktor.

De resultat och de slutsatser som presenteras i avhandlingen baseras på stora mängder data och är så till vida reliabla, men det finns flera andra faktorer som medför att resultaten och slutsatserna bör ses som rimliga indikationer och förslag på tolkningar snarare än fakta. De statistiska resultaten är förstås objektiva, men i tolkningen av dessa resultat är en medvetenhet om möjligheten att de signifikanta sambanden är ett uttryck för en tredje samvarierande faktor viktig. Till exempel är det rimligt att ställa sig kritisk till det identifierade sambandet mellan svårighet och avbildningar. Ytterligare analyser behövs för att bättre förstå vad detta samband är ett uttryck för. Resultaten från uppgiftsanalyserna gäller också en specifik åldersgrupp och textegenskaper som inte identifierats som svåra i avhandlingen kan så vara för yngre elever.

Forskningen som redovisas här har implikationer både för det matematikdidaktiska forskningsfältet och för skolpraktiken. Resultaten påvisar även nyttan av att inte enbart använda lösningsfrekvens för att analysera aspekter av svårighet. För skolpraktiken har resultaten betydelse ur flera perspektiv – dels som en fingervisning om vilka egenskaper i uppgiftstext som är av vikt att fokusera i undervisningen, dels i förhållande till examinationer. Resultaten ger kunskap om vilka egenskaper i uppgiftstext som rimligen testar matematisk förmåga och vilka egenskaper som bör undvikas i prov då de riskerar att bidra till att något annat än matematisk förmåga testas. Denna avhandling bidrar till kunskapen om uppgiftstextens betydelse för hur svåra matematikuppgifter är att läsa och lösa, men ytterligare studier inom området behövs, särskilt eftersom de samband som påvisats i avhandlingen inte är uttryck för kausalitet utan bara för samvariation.

References

Abedi, J. (2000). *Confounding of Students' Performance and Their Language Background Variables*. Research Report 143, UD 033 934.

Acartürk, C., Taboada, M., & Habel, C. (2013). Cohesion in multimodal documents: Effects of cross-referencing. *Information Design Journal (IDJ), 20*(2), 98-110. doi:10.1075/idj.20.2.02aca

Adu-Gyamfi, K., Bossé, M. J., & Faulconer, J. (2010). Assessing understanding through reading and writing in mathematics. *International Journal For Mathematics Teaching And Learning, 11*(5), 1-22.

Ainsworth, S., Bibby, P., & Wood, D. (1997). *Evaluating principles for multi-representational learning environments*. Paper presented at the th European conference for Research on Learning and Instruction, Athens.

Alshwaikh, J. (2011). *Geometrical diagrams as representation and communication: A functional analytic framework*. Institute of Education, University of London. London, UK.

Bagni, G. (2006). Cognitive difficulties related to the representations of two major concepts of set theory. *Educational Studies in Mathematics, 62*(3), 259-280. doi:10.1007/s10649-006-8545-3

Baiduri, B. (2015). Mathematics education students' understanding of equal sign and equivalent equation. *Asian Social Science, 11*(25).

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: the journey from features to meaning (a travel update). In M. J. Traxler & A. Morton (Eds.), *Handbook of Psycholinguistics* (2nd ed.). Amsterdam, Boston: Elsevier.

Bayraktar, H. (2011). *The Role of Lexical Cohesion in L2 Reading*. Germany: Verlag Dr. Müller.

Beitlich, J. T., Lehner, M. C., Strohmaier, A. R., & Reiss, K. M. (2016). *The relation of eye movements on mathematical task and task difficulty*. Paper presented at the 13th International Congress on Mathematical Education, Hamburg, 24-31 July 2016.

Björkvall, A. (2010). *Den visuella texten: multimodal analys i praktiken.* Uppsala: Hallgren & Fallgren.

Blatto-Vallee, G., Gaustad, M. G., Porter, J., & Fonzi, J. (2007). Visual–Spatial Representation in Mathematical Problem Solving by Deaf and Hearing Students. *Journal of Deaf Studies and Deaf Education*, 432-449.

Bolt, S. E., & Thurlow, M. L. (2007). Item-level effects of the read-aloud accommodation for students with reading disabilities. *Assessment for Effective Intervention, 33*(1), 15-28.

Bolt, S. E., & Ysseldyke, J. E. (2006). Comparing DIF across math and reading/language arts tests for students receiving a read-aloud accommodation. *Applied Measurement in Education, 19*(4), 329-355.

Bossé, M. J., Adu-Gyamfi, K., & Chandler, K. (2014). Students' differentiated translation processes. *International Journal For Mathematics Teaching And Learning, Web.*

Breland, H. M. (1996). Word frequency and word difficulty: a comparison of counts in four corpora. *Psychological Science, 7*(2), 96-99.

Brown, J. D. (1996). *Testing in language programs.* Upper Saddle River, NJ: Prentice Hall Regents.

Caponera, E., Sestito, P., & Russo, P. M. (2016). The influence of reading literacy on mathemtics and science achievement. *The Journal of Educational Research, 109*(2), 197-204. doi:10.1080/00220671.2014.936998

Capraro, M., & Joffrion, H. (2006). Algebraic equations: Can middle-school students meaningfully translate from words to mathematical symbols? *Reading Psychology, 27*(2-3), 147-164.

Chahine, I. (2011). The role of translations between and within representations on the conceptual understanding of fraction knowledge: A trans-cultural study. *Journal of Mathematics Education, 4*(1), 47-59.

Chen, C.-L., & Herbst, P. (2013). The interplay among gestures, discourse, and diagrams in students' geometrical reasoning. *Educational Studies in Mathematics, 83*, 285-307. doi:10.1007/s10649-012-9454-2

Chen, F. (2010). *Differential language influence on math achievement.* Unpublished dissertation. University of north Carolina.

Choi, J., Milburn, R., Reynolds, B., Marcoccia, P., Silva, P. J., & Panang, S. (2013). The intersection of mathematics and language in the post-secondary environment: Implications for English language learners. *Collected Essays On Learning And Teaching, 6*, 671-676

Cobb, P. (2004). Mathematics, literacies, and identity. *Reading Research Quarterly, 39*(3), 333-337.

cohesion, n. (n.d.). OED Online. Oxford University Press. Retrieved 1 july 2016
http://www.oed.com/view/Entry/35943?redirectedFrom=cohesion

de Kirby, K., & Saxe, G. B. (2014). Using geometrical representations as cognitive technologies. *Journal of Cognition and Culture, 14*, 401-414. doi:10.1163/15685373-12342134

de Lange, J. (2003). Mathematics for literacy. In B. L. Madison & L. A. Steen (Eds.), *Quantitative literacy. Why numeracy matters for schools and colleges* (pp. 75-89). Princeton, NJ: The National Council on Education and the Diciplines.

Delice, A., & Sevimli, E. (2010). An investigation of the pre-services teachers' ability of using multiple representations in problem-solving success: The case of definite integral. *Educational Sciences: Theory and Practice, 10*(1), 137-149.

Dimmel, J. K., & Herbst, P. G. (2015). The semiotic structure of geometry diagrams: How textbook diagrams convey meaning. *Journal for Research in Mathematics Education, 46*(2), 147-195.

Driver, M. K., & Powell, S. R. (2015). Symbolic and nonsymbolic equivalence tasks: The influence of symbols on students with matheamatics difficulty. *Learning Disabilities Research & Practice, 30*(3), 127-134.

Drouhard, J.-P., & Teppo, A. R. (2004). Symbols and language. In K. Stacey, H. Chick, & M. Kendal (Eds.), *The Future of the Teaching and Learning of Algebra. The 12th ICMI Study* (pp. 227-264). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Duval, R. (2006). A cognitive analysis of problems of comprehension i a learning of mathematics. *Educational Studies in Mathematics, 61*(1-2), 103-131.

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).

Edge, O. P., & Friedberg, S. H. (1984). Affecting achievement in the first course in calculus. *The Journal of Experimental Education, 52*(3), 136-140.

Elia, I., & Philippou, G. (2004). *The functions of pictures in problem solving.* Paper presented at the The 28th Conference of the International Group for the Psychology of Mathematics Education, Bergen, Norway.

Grimm, K. J. (2008). Longitudinal associations between reading and mathematics achievement. *Developmental Neuropsychology, 33*(3), 410-426. doi:10.1080/87565640801982486

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English.* London: Longman.

Halliday, M. A. K., & Matthiessen, C. M. (2004). *An introduction to functional grammar* (3 ed.). London: Arnold.

Hasan, R. (1989). The texture of a text. In M. A. K. Halliday & R. Hasan (Eds.), *Language, context, and text: aspects of language in a social-semiotic perspective.* Oxford: Oxford University Press.

Haworth, C. M. A., Kovas, Y., Hrlaar, N., Hayiou-Thomas, M. E., Petrill, S. A., Dale, P. S., & Plomin, R. (2009). Generalist genes and learning disabilities: a multivariate genetic analysis of low performance in reading, mathematics, language and general cognitive ability in a sample of 8000 12-year-old twins. *Journal of Child Psychology and Psychiatry, 50*(10), 1318-1325. doi:10.1111/j.1469-7610.2009.02114.x

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*(4), 684–689.

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.

Hickendorff, M. (2013). The language factor in elementary mathematics assessments: Computaional skills and applied problem solving in a multidimensional IRT framework. *Applied Measurement in Education, 26*(4), 253-278. doi:10.1080/08957347.2013.824451

Irwin, J. W. (1986). Cohesion and comprehension: A research review. In J. W. Irwin (Ed.), *Understanding and teaching cohesion comprehension* (pp. 31-34). Newark: Delaware: International.

Janvier, C. (1987). *Problems of Representation in the Teaching and Learning of Mathematics*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson, E., & Monroe, B. (2004). Simplified language as an accommodation on math tests. *Assessment for Effective Intervention, 29*(3), 35-45.

Jones, J. (2007). *Multiliteracies for academic purposes: A metafunctional exploration of intersemiosis och multimodalitey in univeristy textbook and ocmputer-based learning resources in science.* Unpublished EdD thesis. University of Sydney.

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis of effectiveness and validity. *Review of Educational Research, 79*(3), 1168-1201.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics, 12*, 317-326.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Kirshner, D. (1989). The visual syntax of algebra. *Journal for Research in Mathematics Education, 20*(3), 274-287.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal Of The Learning Sciences, 13*(2), 129-164.

Kovas, Y., Haworth, C. M. A., Harlaar, N., Petrill, S. A., Dale, P. S., & Plomin, R. (2007). Overlap and specificity of genetic and environmental influences on mathematics and reading disability in 10-year-old twins. *Journal of Child Psychology & Psychiatry, 48*, 914-922.

Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer/verbalizer dimension: Evidence for two types of visualizers. *Cognition & Instruction, 20*, 47-77.

Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition, 33*(4), 710-726.

Kress, G. (2007). Meaning, learning and representation in a social semiotic approach to multimodal communication. In A. McCabe, M. O'Donnell, & R. Whittaker (Eds.), *Advances in Language and Education* (pp. 15-39). London: Continuum.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Milton Park, Abingdon, Oxon: Routledge.

Kress, G., & van Leeuwen, T. (2006). *Reading images*. London: Routledge Ltd.

language, n. (n.d.). OED Online. Oxford University Press. Retrieved 19 May 2016 from http://www.oed.com/view/Entry/105582?rskey=yjxM4E&result=1 &isAdvanced=false

Lech, R., Post, T., & Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In C. Janvier (Ed.), *Problems of Representations in the Teaching and Learning of Mathematics* (pp. 33-40). Hillsdale, NJ: Lawrence Erlbaum.

Lee, M. K., & Randall, J. (2011). *Exploring language as a source of DIF in a math test for English language learners*. Paper presented at the In NERA Conference Proceedings 2011.

Lemke, J. L. (1998). Multiplying Meaning: Visual and verbal semiotics in scientific text In J. R. Martin & R. Veel (Eds.), *Reading Science* (pp. 87-113). London: Routledge.

Lemke, J. L. (2002). Travels in hypermodality. *Visual Communication, 1*(3), 299-325.

Li, X., Ding, M., Capraro, M. M., & Capraro, R. M. (2008). Sources of differences in children's understandings of mathematical equality: comparative analysis of teacher guides and student texts in China and in the United States. *Cognition and Instruction, 26*, 195-217.

Lin, Y., Wilson, M., & Cheng, C. (2013). An investigation of the nature of the influences of item stem and option representation on student responses to a mathematics test. *European Journal Of Psychology Of Education, 28*(4), 1141-1161. doi:10.1007/s10212-012-0159-9

Linn, R. L. (2014). Validation of the uses and interpretations of results of state assessment and accountability systems. In G. Tindal & T. M. Haladyna (Eds.), *Large-Scale Assessment Programs For All Students. Validity, Technological Adequacy, and Implementation* (pp. 27-48). Mahwah, New Jersey: Routledge.

Liu, Y., & O'Halloran, K. (2009). Intersemiotic texture: Analyzing cohesive devices between language and images. *Social Semiotics, 19*(4), 367-388.

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for english language learners in math tests. *Educational Assessment, 14*(3-4), 160-179.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

Messick, S. (1995). Validity of psychological assessment validation of Inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Moon, K., Brenner, M. E., Jacob, B., & Okamoto, Y. (2013). Prospective secondary mathematics teachers' understanding and cognitive difficulties in making connections among representations. *Mathematical Thinking and Learning, 15*(3), 201-227. doi:10.1080/10986065.2013.794322

Moreno-Armella, L., & Sriraman, B. (2010). Symbols and Mediation in Mathematics Education. In B. Sriraman & L. English (Eds.), *Theories of Mathematics Education: Seeking New Frontiers*. Berlin, Heidelberg: Springer.

Moschkovich, J., Schoenfeld, A. H., & Arcavi, A. (1993). Aspects of understanding: On multiple perspectives and representations of linear relations, and connections among them. In T. Romberg, E. Fennema, & T. Carpenter (Eds.), *Intergrating research on the graphical representation of function* (pp. 69-100). Hillsdale, NJ: Erlbaum.

NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA, USA: National Council of Teachers of Mathematics.

Niss, M., & Højgaard, T. (2011) Competencies and Mathematical Learning: Ideas and inspiration for the development of mathematics teaching and learning in Denmark. *Vol. 485*. Roskilde: Roskilde Universitet.

O'Halloran, K. (2005). *Mathematical Discource: Language, symbolism and visual images*. London: Continuum.

O'Halloran, K. (2008). Inter-semiotic expansion of experiental meaning: Hierarchical scales and metaphor in mathematics discourse. In E. Ventola & C. Jones (Eds.), *From Language to Multimodality. New developments in the study of Ideational Meaning* (pp. 231-254). London: Equinox Publishing Ltd.

OECD. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Scince, Problem Solving and Financial Literacy*: OECD Publishing.

Pimm, D. (1987). *Speaking Mathematically: Communication in mathematics classrooms*. London: Routledge Kegan & Paul.

Påsse, T. (2001). *An empirical model of glacio-isostatic movements and shore-level displacement in Fennoscandia*. Stockholm: SKB.

Qian, D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21*(1), 28-52.

Radford, L., Edwards, L., & Arzarello, F. (2009). Introduction: beyond words. *Educational Studies in Mathematics, 70*(2), 91-95. doi:10.1007/s10649-008-9172-y

Ribeck, J. (2015). *Steg för steg. Naturvetenskapligt ämnesspråk som räknas*. (Doctoral thesis, University of Gothenburg, Göteborg). Retrieved from https://gupea.ub.gu.se/handle/2077/40506

Rousselle, L., & Noël, M. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition, 102*(3), 361-395

Royce, T. (2007). Intersemiotic Complementarity: A Framework for Multimodal Discource Analysis. In T. D. Royce & W. L. Boucher (Eds.), *New Directions in the Analysis of Multimodal Discourse* (pp. 63-109). New York: Routledge.

Sato, E., Rabinowitz, S., Gallagher, C., & Huang, C. W. (2010). *Accommodations for English Language Learner Students: The Effect of Linguistic Modification of Math Test Item Sets*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schleppegrell, M. J. (2004). *The language of schooling: a functional linguistic perspective*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly, 23*(2), 139-159.

Schweiger, F. (1992). *Mathematics is a language*. Paper presented at the The 7th International Congress on Mathematical Education, Quebec, 17-23 August.

Sfard, A. (2008). *Thinking as Communicating: Human Development, the Growth of Discources, and Mathematizing*. NY: Cambrige Univeristy Press.

Shorrocks-Taylor, D., & Hargreaves, M. (1999). Making it clear: a review of language issues in testing with special reference to the National Curriculum mathematics tests at key stage 2. *Educational Research, 41*(2), 123-136.

Skolverket. (2013). *Ämnesproven 2012 i grundskolans årskurs 9 och special skolans årskurs 10*. Retrieved from http://www.skolverket.se/publikationer?id=2985

Skolverket. (2015). *Med fokus på matematik. Analys av samstämmighet mellan svenska styrdokument och den internationella studien PISA*. Stockholm: Skolverket Retrieved from skolverket.se/publikationer.

Susac, A., Bubic, A., Vrbanc, A., & Planinic, M. (2014). Development of abstract mathematical reasoning: the case of algebra. *Frontiers in Human Neuroscience, 8*, 1-9. doi:10.3389/fnhum.2014.00679

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (Vol. 5 rev. ed.). Boston, MA: Allyn and Bacon.

Teich, E., & Fankhauser, P. (2005). Exploring lexical patterns in text: lexical cohesion analysis with WordNet. In S. Dipper, M. Götze, & M. Stede (Eds.), *Heterogeneity in Focus: Creating and Using Linguistic Databases. Interdisciplinary Studies on Information Structure* (Vol. 2, pp. 129-145). Potsdam: Universität Potsdsam.

Tindal, G. (2014). Large-scale assessments for all students: Issues and options. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students : validity, technical adequacy, and implementation* (pp. 1-24). Mawah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Turner, R., Dossey, J., Blum, W., & Niss, M. (2009). Using mathematical competencies to predict item difficulty in PISA: A MEG study. In P. Manfred, K. Mareike, S. Katrin, & R. Silke (Eds.), *Research on PISA. Research Outcomes of the PISA Research Conference 2009* (pp. 23-37). London: Springer.

Unsworth, L., & Chan, E. (2008). Assessing integrative reading of images and text in group reading comprehension tests. *Curriculum Perspectives, 28*(3).

Unsworth, L., & Cléirigh, C. (2009). Multimodality and reading. The construction of meaning through image-text interaction. In C. Jewitt (Ed.), *The Routledge Handbook of Multimodal Analysis* (pp. 151-169). Abingdon, Oxon: Routledge.

Usiskin, Z. (1996). Mathematics as a language. In P. C. Elliott & M. J. Kenney (Eds.), *Communication in Mathematics, K-12 and Beyond. 1996 Yearbook* (pp. 231-243). 1906 Association Drive, Reston, VA 22091-1593: National Council of Teachers of Mathematics.

Wakefield, D. V. (2000). Math as a second language. *Educational Forum 64, no. 3*, 272-279.

Wolf, M. K., Herman, J. L., Kim, J., Abedi, J., Leon, S., Griffin, N., . . . Shin, H. W. (2008). *Providing validity evidence to improve the assessment of english language learners*. Retrieved from University of California, Los Angeles:

Yang, D., & Huang, F. (2004). Relationships among computational performance, pictorial representation, symbolic representation and number sense of sixth-grade students in Taiwan. *Eucucational Studies, 30*(4), 373-389. doi:10.1080/0305569042000310318

Österholm, M. (2006). Characterizing reading comprehension of mathematical texts. *Educational Studies in Mathematics, 63*(3), 325-346.

Österholm, M., & Bergqvist, E. (2012). 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. (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.