What is so special about mathematical texts? Analyses of common claims in research literature and of properties of textbooks

Magnus Österholm[1,2] & Ewa Bergqvist[1]

[1]Umeå University, Sweden
[2]Monash University, Australia

Dept. of Science and Mathematics Education

Umeå University

SE – 901 87 Umeå

Sweden

+46 (0)90 786 71 31

magnus.osterholm@umu.se

Keywords: communication, history, language, linguistics, reading

This study surveys claims in research articles regarding linguistic properties of mathematical texts, focusing on claims supported by empirical or logical arguments. It also performs a linguistic analysis to determine whether some of these claims are valid for school textbooks in mathematics and history. The result of the survey shows many and varying claims that mainly describe mathematical texts as highly compact, precise, complex, and containing technical vocabulary. However, very few studies present empirical support for their claims, and the few empirical studies that do exist contradict the most common, and unsupported, claims, since no empirical study has shown mathematical texts to be more complex than texts from other subjects, and any significant differences rather indicate the opposite. The linguistic analysis in this study is in line with previous empirical studies and stands in contrast to the more common opinion in the unsupported claims. For example, the mathematics textbooks have significantly shorter sentences than the history textbooks.

1 Introduction

Reading mathematical texts in textbooks is an important part of many students' learning and doing mathematics. For example, in most countries participating in TIMSS 2007, "the textbook remains the primary basis of mathematics instruction at both the fourth and eighth grades" (Mullis, Martin, & Foy, 2008, p. 288). The reliance on textbooks is particularly strong in Sweden, which is the context of the empirical part of the present study, since 93–95% of students in grades four and eight are taught by teachers that use textbooks as the primary basis for lessons (Mullis et al., 2008). However, we have found few empirical studies that in a direct manner examine to what extent *students* use textbooks. We know that Swedish students work with written mathematical tasks during 59% of the mathematics lessons and during more than half of this time they work with their own textbooks (Bergqvist et al., 2009). Similar results exist for other countries, for example China where "for students, textbooks were their main learning resource for both in-class exercise and homework" (Fan, Chen, Qiu, & Hu, 2004, p. 186).

Many research articles discuss traits of, and difficulties with reading, mathematical texts. Some claim that mathematical texts are especially difficult and differ from texts in other subjects to such an extent that students need to develop a particular kind of reading ability for them. However, through the years, several researchers have pointed out that empirical research on mathematical texts is scarce (e.g., Burton & Morgan, 2000; Love & Pimm, 1996) and a recent literature review shows that research on mathematical discourse seldom focuses on textbooks (Ryve, 2011). More generally, another recent literature review shows that there is a lack of empirical research focusing on issues of "academic language," which includes all subject areas and all modalities, in particular both written and oral language (Anstrom et al., 2010). Therefore, this study surveys claims regarding mathematical text properties, especially claims in articles presenting empirical studies on the matter. Furthermore, we examine whether some of these claims are valid for school textbooks in mathematics and compare with history textbooks. The overall purpose of this study is to critically analyze and problematize the discourse regarding properties of mathematical language, in particular mathematical texts in comparison with other texts. We start with some research background before we present our literature survey and textbook analyses.

2 Background

2.1 Reading mathematical texts

The notion of content literacy highlights the potential need for different types of reading abilities in different content domains. McKenna and Robinson (1990) present a framework where they define *content literacy* as consisting of three components: general literacy skills, content-specific literacy skills, and prior knowledge of content. Similar divisions are made by, for example, Behrman and Street (2005). In many research publications it is stated that reading mathematics demands a specific type of reading ability (e.g., Burton & Morgan, 2000; Fuentes, 1998; Shanahan & Shanahan, 2008), that is, that students need to develop content-specific literacy skills in mathematics. Existing empirical studies of students' reading comprehension mainly produce evidence of general literacy skills, in particular through strong or moderate correlations between different tests of reading comprehension: between social studies and general reading comprehension (r=0.79) (Artley, 1943); between reading comprehension in an anatomy course and general reading ability (r=0.72) (Behrman & Street, 2005); and between reading comprehension for a mathematics text and a history text (r=0.47) (Österholm, 2006).

Instead of focusing only on the reading activity when examining content literacy, it is possible to examine texts from different domains, since different text properties can create a need for different literacy skills. To be familiar with a certain genre or linguistic register could be seen as a part of content-specific literacy skills. However, it is difficult to find a common description of all kinds of mathematical texts, since, as in any text genre, there are different sub-genres. For example, Richards (1991) describes four such sub-genres: research math, inquiry math, journal math, and school math. He argues that these are different linguistic domains with different cultures regarding "assumptions, goals, and underlying methodologies" (p. 16). Even within sub-genres there are large variations: Burton and Morgan (2000) find a large variety of writing styles in mathematics research articles. Therefore, it seems difficult to make claims that are valid for all mathematical texts or for mathematical language in general, and it is important to critically examine such claims.

The use of symbols in mathematical texts can be seen as the subject's most distinctive feature (Pimm, 1989). It has been highlighted as the most important potential cause for a need of content-specific literacy skills, through analyses of students' reading comprehension of different types of mathematical texts compared with a history text (Österholm, 2006). Even if a text does not have to use symbols to be mathematical, there is sometimes an "identification of mathematics with its symbol system" (Morgan, 1998, p. 13). We argue that such a view mistakenly reduces the relevance and importance of natural language in mathematics, for example the use of English or Swedish, and we agree with Burton and Morgan (2000, p. 430): "Although sometimes seen to be peripheral to the main mathematical content, natural language serves in the construction of the identities of the author and reader and of the epistemological and ontological assumptions underlying the writing." Besides these direct connections between properties of natural language and aspects of content, it is also relevant to focus on aspects of complexity of the natural language used in mathematics, for example by studying issues of readability for mathematical texts.

In summary, evidence exists that different forms of representations can create a potential need for content-specific literacy skills in mathematics, but we have not found clear evidence that the same need exists for mathematical communication using natural language. Therefore, it is relevant to focus on properties of the natural language in mathematical texts, especially to examine if and how mathematical texts put different demands on readers than texts in other subjects.

2.2 Textbooks in mathematics education

Textbooks play a central part in mathematics education since teachers in many countries, specifically in Sweden, tend to base their teaching primarily on a textbook (Mullis et al., 2008). There is a lot of textbook research, and it often contains different types of content analyses focusing the mathematical content, for example how it is presented (e.g., see Haggarty & Pepin, 2002) or what skills the tasks demand of the students (e.g., see Lithner, 2004). There are also many studies on socio-political aspects of textbooks, such as studies of how mathematics textbooks represent different political agendas (e.g., see McBride, 1994) or sociological studies (e.g., see Dowling, 1996). Fewer studies exist that perform some type of linguistic analysis of mathematics textbooks, which is evident from the literature survey conducted in the present paper (see Sect. 4) and also from a recent review of discourse research in mathematics education (Ryve, 2011). The present study focuses on linguistic properties of mathematics textbooks.

One type of linguistic analysis of textbooks uses readability formulas to determine whether textbooks are suitable for the intended grade level. Several studies have shown that the readability level is too high in mathematics textbooks, but also that great variations exist within single textbooks (e.g., Smith, 1969; Wiegand, 1967). A limitation of these types of studies is that they do not focus on characterizing textbooks concerning certain specific properties, but use formulas that are linear combinations of different properties. Analyses of that kind are not of interest for our present purpose; rather, we are interested in studies that focus on specific linguistic properties of mathematical texts.

3 Purpose

The overall purpose of this study is to critically analyze and problematize the discourse regarding properties of mathematical language, in particular mathematical texts in comparison with other texts. We aim to answer two research questions:

1. What are common claims in research literature about linguistic traits of mathematical texts and which of these claims are supported by empirical or logical arguments?
2. Are common claims regarding linguistic traits of mathematical texts (from question 1) valid for Swedish school textbooks in mathematics?

In connection to both questions, we include the comparing of mathematical texts with other types of texts. For the first question, claims that compare mathematical texts with other types of texts are examined in more detail. For the second question, linguistic traits of mathematics textbooks are compared with linguistic traits of textbooks in another subject.

In both questions we focus on linguistic properties of the natural language, such as syntactic or grammatical properties, and exclude the special properties of the symbolic language. In Sect. 2.1 we argue for the relevance of focusing on natural language when analyzing properties of mathematical texts. Furthermore, since symbolic languages, unlike natural languages, consist of ideograms and also have their own syntax and grammar, analyses of properties of symbolic languages would demand different methods, as also becomes evident in Part 2 of this study.

The research questions in our study are discussed separately in the next two sections; in Part 1 we present a literature survey used to answer Research question 1 and in Part 2 we present an empirical study of textbooks used to answer Research question 2.

4 Part 1: A literature survey

This first part aims to answer the research question: What are common claims in research literature about linguistic traits of mathematical texts and which of these claims are supported by empirical or logical arguments?

From here on, we use the word *claim* to refer to those claims found in the survey presented in this section, not to any claims we make ourselves.

4.1 Method for Part 1

The method used in this literature survey can be divided into three steps: (1) Search for and selection of literature to include in the analysis; (2) Analysis of each individual reference to find relevant claims about mathematical texts; and (3) Analysis of the claims found in the literature.

4.1.1 Selection of literature

In our survey we include literature that *focuses on* properties of mathematical texts, since it certainly should include relevant claims for our analyses. Primarily we look at literature that reports on empirical studies or logical analyses of properties of mathematical texts. Secondarily we include other types of literature in our survey, such that most likely includes claims about properties of mathematical texts. In addition to this basic description of relevant literature to include, we also specify the overarching focus of our survey, that is, research literature where we can find claims about properties of mathematical texts, as follows:

- *Research literature*: We use a broad meaning of research literature, where we not only include peer-reviewed, high-ranked scientific journals, but also include any reference that can be found in the well-known databases MathEduc and Education Resources Information Center (ERIC). This broader inclusion is used since we knew a priori that few empirical research studies exist and since we then get a broader characterization of existing perspectives on properties of mathematical texts.
- *Texts*: We focus on written natural language, and thus exclude literature that focuses on oral discourse or on the symbolic language, or on other forms of representation. Besides including literature that focuses specifically on texts, we also include literature that in some way focuses on properties of "the mathematical language," since we see this notion as a broader category that can include texts.
- *Properties*: We focus on linguistic properties of mathematical texts, and thus exclude literature that focuses on other types of properties, in particular textbook research where any type of content analysis is performed.

Our search for literature consists of two parts: a structured search using specific search words in databases, to fulfill our primary aim of finding almost all research literature that focuses on linguistic properties of mathematical text, and a more unstructured search for literature that focuses on aspects of reading mathematical text.

In the structured search for literature we search for specific words in the title of references, using the databases MathEduc and ERIC. Three different truncated search words are used in three separate searches in each of the two databases: "text*," "lang*," and "ling*." For ERIC, in each of the three searches, "math*" is added as a second search word to find references about mathematics. All searches are limited to references written in English. We use these types of general search words to find as much relevant literature as possible: the search resulted in a total of 1511 hits in MathEduc and 776 hits in ERIC. Due to the many hits in the searches, the potential relevance of a reference is primarily judged based only on the title, and when in doubt, the abstract is also used.

The aim of the more unstructured search for literature is not to find all potentially relevant literature, but to add a complementing type of literature compared with literature found in the structured search. This search was performed over several years, based on the authors' personal research interests about reading and texts in mathematics education. In particular, we add literature to the survey that includes a treatment of how or why the reading of mathematical texts can be seen as a special type of activity, potentially based on certain properties of mathematical texts.

4.1.2 Analysis of literature

The basic method of analysis of literature consists of reading the selected references and highlighting claims about any type of linguistic property of mathematical texts. When deciding whether to highlight a certain claim or not, we rely on the same specifications for inclusion and exclusion as when choosing what literature to include. That is, we include only statements about a linguistic property of written natural language, primarily concerning some aspect of syntax or grammar.

There are of course situations when it is unclear if a claim should be included or not. In particular, it is sometimes unclear if statements actually refer to aspects of the symbolic language or to oral discourse. We include a claim if it is not clear from the claim itself or the context that it is specifically about either the symbolic language or oral discourse. That is, we interpret such unspecified statements as concerning the mathematical language more generally. Similarly, we include rather than exclude also when there is uncertainty whether a claim can be interpreted as saying something about a linguistic property or, for example, about the content.

4.1.3 Analysis of claims

Due to the nature of found claims, there are limitations to what type of analysis can be made using our data of claims about mathematical texts. For many claims, no clarification is given about the content of the claim, primarily with two problematic issues. First, it is often not specified what type of text a claim is about. In particular, it is sometimes difficult to determine if a claim regarding properties of "mathematical language" is meant as a statement about mathematical language in general, the natural language used in mathematics, or the symbolic language of mathematics. Second, descriptive words are often used without explaining their exact meaning, for example when characterizing texts as "artificial," "sophisticated," or "complex."

The method of analysis of the highlighted claims in the literature consists of two parts. First, we create a structured description of what all claims are about, that is, focusing on the essence of the statements, through a small set of broad categories that describe different types of linguistic properties. The categories are created based on patterns found among the specific claims analyzed in this study. The main purpose with these categories is to include types of claims that are most common, rather than including all analyzed claims.

Second, all claims are analyzed concerning two aspects:

- Whether and how the claim includes any *comparison* between mathematical text and some other type of text. We place each claim in one of the following categories:

- *Inside school*: Claims that compare with other types of school text or academic text, either through another specific subject or more generally about “other” or “different” subjects or academic texts.
- *Outside school*: Claims that compare with texts outside the school or academia, using labels such as everyday, general, common, daily, natural, ordinary, or normal.
- *Mathematics*: Claims that compare different mathematical texts, for example by comparing between applied and pure mathematics, between students and teachers, or between geometry and algebra.
- *Unclear*: Claims that do not explicitly mention other types of text, but mathematical texts are either described as having more or less of some property, or are compared only with “other” texts.
- *Other*: Claims with comparisons that do not fit the above categories. In our data there are comparisons focusing on different types of texts not based on the context where the texts are used, for example inside or outside school and different school subjects, but based on structural or stylistic properties, such as “narrative,” “expository,” “parable,” or “prose.”
- *None*: Claims that make no comparisons, but only describe mathematical texts.

- Whether and how the claim contains any type of *argument* for the validity of the claim. We place each claim in one or more of the following categories:
  - *Empirical*: Claims for which a specified empirical material has been analyzed and the claim comes as a result of this analysis.
  - *Logical*: Claims for which an argument is presented focusing on the necessity for a certain property of mathematical texts, in particular when a (more general) property of mathematics as a subject is used as argument.
  - *Example*: Claims for which a specific example is presented in an ad hoc manner and not based on the analysis of some specified empirical material.
  - *Reference*: Claims for which a reference to another publication is given. We also note whether the reference is described as presenting empirical evidence or not.
  - *None*: Claims for which no direct argument is presented.

An example of claims and how they are categorized can be found in Table 1. We focus our analysis on those claims that (a) have empirical or logical arguments, which are seen as good types of arguments for potential generalizations about mathematical texts, and (b) include some type of comparison between mathematical texts and other types of texts, in order to focus on types of properties that can potentially create a need for content-specific literacy skills. These types of claims are compared with the common types of claims found in the first part of the analysis, to examine if there are common conceptions about mathematical texts that might be invalid.

Table 1 Examples of claims from Brunner (1976) and how they are categorized

Quote	Type of claim[a]	Comparison	Argument
“In mathematical language, quantification plays a	Complexity	Unclear	None

more sophisticated and complex role." (p. 208)			
"Borrowed natural language words in the object language are used in a special sense. In geometry, 'point' and 'line' do not carry all the connotations of these same words in English." (p. 209)	Vocabulary	Outside school	Example
"Because of the compactness of mathematical language owing to its hierarchical development of levels of definitions and to its use of symbolism, mathematical exposition communicates a great deal in a short statement." (p. 212)	Compactness and Structure	None	None

[a] See Sect. 4.2.1 for description of categories

4.2 Results from Part 1

We found in total 58 references to include in the search for relevant claims about mathematical texts, and 50 of these references include relevant claims, with a total of 311 claims. We treat these claims in the 50 references as the data for analysis. One of the 50 references was published in 1928, while the rest were published between 1968 and 2012, with a median publishing year of 1996 and mean of 1992. Based on the Norwegian system for the ranking of scientific publications,[1] 10 of the 50 references are at the highest level and 23 at the ordinary level, while there are two other types of references and 15 references are not included in the Norwegian list.

The number of claims reported should not be seen as exact numbers, in particular since it can be decided in different ways what to count as *one* claim. However, our focus is not on the exact number of claims, but we use these numbers as a way to show roughly how many times something is mentioned.

4.2.1 Categories of types of claims

Table 2 describes the categories of different types of properties included in the claims. As mentioned above, we often need to interpret a claim based only on a single word or short phrase that is used in a claim to characterize mathematical texts. These types of words are listed in Table 2, labeled as key words.

As seen in Table 2, based on the different categories, claims focusing on aspects of vocabulary are most common, in particular by addressing issues of technical vocabulary. The issue of complexity is also often included in claims, primarily described through its own category, focusing on the use of certain key words, but other categories can also be seen as highlighting a specific aspect of complexity, in particular Compactness. Furthermore, claims included in another category than Complexity sometimes also include issues of complexity, as is evident in Table 2 from the summary of categories Compactness and Relations. Therefore, as seen from the summaries of categories, claims about the high level of complexity of mathematical texts are common.

Overall, the claims within a certain category are consistent with each other, with the category of Precision as an exception, since there are contradictory claims within this category: claims mostly highlight an

[1] Consists of three levels: highest international level (2), ordinary level (1), and others (-); see http://dbh.nsd.uib.no/kanaler

unambiguous property of mathematical texts but there are also several claims that highlight an ambiguous property.

Table 2 Most common types of claims about properties of mathematical texts

Category	**Refs (of 50)**	**Claims (of 311)**	**Types of claims**[a]	**Summary of most common claims about math texts**
Vocabulary	38	111	About vocabulary in general (types of vocabulary) or about specific word classes, words, or phrases.	Exist (many) technical/specialized words, i.e., words that exist only in mathematics or borrowed/familiar words that are used in a special sense or manner.
Compactness	24	60	Key words: compact, dense, non-redundant, less context, concise. About nominalizations or noun phrases.	High level of compactness, in particular density of many types of technical words, many nominalizations, many/long/complex noun phrases.
Precision	20	35	Key words: precise, exact, unambiguous, clarity.	Mostly that math is precise and unambiguous (16 refs) but sometimes that math is ambiguous (8 refs). Mostly about words.
Complexity	17	38	Key words: complex, complicated, unintuitive, sophisticated, difficult, high demand, challenging, confusing, perplexing, misplaced, illogical, haphazard, simple.	High complexity of words, of sentence constructions and of relationships between sentences/statements.
Structure	15	26	About organization of texts or ordering of things.	Existence of a specific or clear type of structure, in particular hierarchical, sequential, or logical. Certain ordering of specific type of content (e.g., definitions and illustrations).
Relations	10	19	About connections between statements, sentences or other parts of texts.	Many relationships, characterized as complex and/or logical. The importance of certain words or word classes, in particular prepositions and conjunctions.
Style	9	29	About a more overarching characterization of a text, such as prosaic or descriptive.	Little connected prose. Narratives are seldom/never used. Variety of styles/genres.
Persons	8	26	About whether or how persons are addressed, existing or depicted.	Persons are often non-existent, in particular first person (author), but more variation whether and how second person (reader) is included or addressed.

[a] Key words are types of words we interpret as characterizing a property within the given category, and are often used (more or less) in isolation in the analyzed literature, without any explanation

4.2.2 Comparisons and arguments

As seen in the last column of Table 3, regarding different types of comparisons, most claims are descriptive, without making any type of comparison between texts (category "None").

Regarding different types of arguments, relatively many claims (38% of all claims) are given without any argument (see the last row in Table 3). When limiting our focus to the 10 references from the highest scientific level (i.e., level 2 in the Norwegian list), the proportion of claims given without any argument is of similar magnitude (34%; 26 of 76 claims).

In almost all cases when references are given as arguments, they only refer to someone who has made the claim, and not to any other type of argument. Examples are common to use as arguments, which is relevant at least when arguing for existence, but as mentioned before, claims are often formulated as claims about mathematical texts *in general*, making many of the arguments using examples difficult to interpret.

Table 3 Number of references and claims that exist in combinations of categories of different types of comparisons and different types of arguments

	Argument					
Comparison	**Empirical**	**Logical**	**Example**	**Reference**	**None**	**TOTAL**
Inside school	2 refs 14 claims	0 refs 0 claims	1 refs 2 claims	4 refs 5 claims	7 refs 11 claims	**12 refs 31 claims**
Outside school	0 refs 0 claim	2 refs 3 claims	11 refs 17 claims	5 refs 5 claims	9 refs 19 claims	**21 refs 42 claims**
Mathematics	8 refs 17 claims	1 ref 1 claim	3 refs 6 claims	2 refs 2 claims	7 refs 11 claims	**18 refs 36 claims**
Unclear	0 refs 0 claims	0 refs 0 claims	4 refs 4 claims	2 refs 2 claims	5 refs 7 claims	**10 refs 16 claims**
Other	0 refs 0 claims	1 ref 2 claims	4 refs 7 claims	0 refs 0 claims	5 refs 6 claims	**9 refs 15 claims**
None	10 refs 31 claims	4 refs 4 claims	20 refs 50 claims	18 refs 29 claims	25 refs 63 claims	**43 refs 173 claims**
TOTAL	**13 refs 63 claims**	**8 refs 10 claims**	**28 refs 85 claims**	**21 refs 43 claims**	**30 refs 117 claims**	**50 refs 311 claims**

4.2.3 Comparative claims based on empirical or logical arguments

As described in the method section, focus in this study is on empirical or logical arguments and direct comparisons between mathematical texts and some other type of text; that is, focus is on the shaded cells in Table 3. The five references included in the shaded cells are described in more detail below, where claims from each reference are compared with the common types of claims described in Sect. 4.2.1.

Remmers and Grant (1928) present an empirical study where the vocabulary of 12 secondary mathematics textbooks is analyzed and compared with other types of textbooks from the same publisher. Claims included in our analysis are two comparisons. First, the vocabulary range (i.e., the number of different types of words) is smaller in all examined mathematics textbooks compared with other textbooks, for example from history and language studies. Second, the proportion of technical words (which in the study is defined as words that are not included in a given list of generally common words) for a history textbook is of equal magnitude to the mathematics textbooks, while a chemistry textbook has a higher proportion of technical words compared with all mathematics textbooks. These claims relate to categories Vocabulary and Compactness. In general, claims in the category Compactness focus on a high density of technical vocabulary (see Table 2), while results from Remmers and Grant show a relatively lower density in mathematical texts.

Butler et al. (2004) present an empirical study where the data consists of word problems from three mathematics textbooks and text segments from three science textbooks and three social studies textbooks, all textbooks from grade five. They perform different types of linguistic analyses and produce claims about similarities and differences. All texts show a low level of complexity due to simple sentence constructions and low vocabulary range. The mathematics textbooks have generally lower (sometimes equal, but never higher) complexity than the other textbooks, due to fewer complex sentences and words, lower density, fewer nominalizations, fewer passive constructions (almost non-existent in the mathematics textbooks), fewer participial modifiers, fewer academic words, and fewer different types of clause connectors. These claims relate to categories Vocabulary, Compactness, and Complexity. In general, claims in these categories focus on a high level of complexity regarding words and sentences, while results from Butler et al. in general show a relatively lower level of complexity.

Brunner (1976) discusses issues of reading mathematical texts where two claims about properties of mathematical texts are included, comparing with natural language and presented together with logical arguments that relate these properties to more general properties of mathematics. A first claim is that mathematical language is relatively unambiguous which “is necessary because very often contexts in mathematics are short and in order to use ambiguous words meaningfully there must be a way to resolve ambiguity from context.”(p. 209) A second claim is about complexity; that “multiple quantification within a sentence is the usual situation in mathematics,” which is not described as necessary but is possible since “mathematics has, in principle, an infinite supply of pronouns; namely, variables, which are used to control quantification.”(p. 210) These claims relate to categories Precision and Complexity. In general, claims in these categories focus on a high level of complexity and precision, although several claims also highlight a lack of precision through ambiguities. Claims and arguments from Brunner are thus in line with common claims.

Nesher and Katriel (1986) focus on the linguistic function of numbers in mathematical language compared with natural language. Although the function of numbers is perhaps seen as primarily referring to aspects of mathematical symbols, it is clear throughout their study that their focus is not on a specific property of the symbolic language but more generally about the language of mathematics. Their claim is that numbers function as objects in mathematics while they function as adjectives in natural language systems. In essence, their entire article comprises the argument for this claim, in particular through logical arguments that utilize theories of learning and observations of students’ use of numbers. The claim relates to the category Vocabulary, in which claims often focus on the special use of ordinary words in mathematics. The claim from Nesher and Katriel is thus in line with common claims.

Solomon and O’Neill (1998) focus on the cohesion and style of mathematical texts, where they include two interrelated claims with logical arguments. A first claim is that mathematical texts have a logical and not temporal cohesion, since “mathematical argument is itself atemporal” and cannot be “reducible to a temporal sequence of events.”(pp. 216-217) A second claim is that mathematical discourse cannot be narrative, since “it is structured around logical and not temporal relations.”(p. 217) That is, their first claim is used as argument for their second claim. These claims relate to categories Relations and Style. Common claims in these categories include both claims given by Solomon and O’Neill.

In summary, regarding relations between claims based on empirical or logical arguments and common claims from the different categories of claims, there are some contradictions between claims based on empirical arguments and common claims, while claims from logical arguments tend to be in line with common claims. Contradictions between claims based on empirical arguments and common claims concern mainly levels of complexity, measured through properties of words and sentences.

At this point, since it is outside the scope of the present paper, no in-depth critical analysis of the presented arguments is performed, such as an analysis of the methodology in empirical studies or validity of supporting claims in logical arguments. In particular, for logical arguments, a supporting claim can in itself be interpreted as a claim about mathematical texts, for example the argument that contexts are short by Brunner (1976), or it can be based on a more fundamental conception of mathematics, for example seeing mathematics as either a human activity or an abstract entity, which could allow mathematical texts to be narrative or not, in relationship to arguments by Solomon and O'Neill (1998).

5 Part 2: Analysis of textbooks

This second part of the study aims to answer the research question: Are common claims regarding linguistic traits of mathematical texts (from Part 1) valid for Swedish school textbooks in mathematics?

5.1 Method for Part 2

In Part 2 we use Swedish textbooks from different school years and different subjects to examine the validity of a few common types of claims from Part 1. We compare textbooks in mathematics with textbooks in history, since history can be seen as a very different type of subject, as part of humanities and/or social sciences. This limited selection of textbooks, including the specific language that is analyzed (Swedish), cannot be used to fully validate any claims, since the claims often concern mathematical texts in general. It is however possible to examine whether the claims are valid for at least this particular selection; if they are not, the selection functions as a counter-example against the general types of claims about mathematical texts. At the same time, we have selected common Swedish textbooks in order not to focus on finding a, perhaps obscure, counter-example, but to have a reasonably representative sample of Swedish textbooks. In the following sections we present details of the selection of data, present and argue for the selection of claims, describe methods of analysis, and present the results of Part 2.

5.1.1 Selection of data

We analyze Swedish textbooks for school years 4 and 7, and the first course in upper secondary school for the natural science program. Two large Swedish publishers (*Natur & Kultur* and *Sanoma*) that produce textbooks in both subjects for all levels are selected. One book for each subject (2), each publisher (2), and each year (3) is chosen, in total 12 textbooks. To choose comparable textbooks, in the sense that they have been created for the same national curriculum, we use textbooks based on the previous Swedish curriculum documents, since full sets of new textbooks have not yet been produced for the 2011 curriculum documents.

From each book, four disjoint passages are randomly selected, containing at least 100 words each, in complete sentences. The suitability of this sample procedure for all statistical measures is addressed below. Mathematical symbols (e.g., +) or groups of symbols (e.g., $3x + 4$) are not included as words, except for single numbers (e.g., 419) and the % sign (which is handled as a unit just as, for example, km/h).

5.1.2 Selection of claims

Different principles are used to select claims found in Part 1 as the basis for analysis of the textbooks. In particular, we choose claims that are:

1. common and dealing with a general type of property, and we therefore focus on the different categories of types of claims and not singular specific claims; and
2. operationalizable and possible to examine within the scope of the present paper, and we therefore limit the number of different properties to examine and also focus on somewhat simpler types of properties. For example, we exclude the common types of claims about technical vocabulary, which is a type of property more complicated to operationalize (e.g., see different methods used by Butler et al., 2004; Remmers & Grant, 1928).

Based on the principles above, we examine the following types of properties, which are located in four of the categories of different types of claims (the properties are operationalized in the next section):

- *Compactness*: Many claims describe mathematical texts as compact, for example through the use of nominals in different ways. We choose to examine *nominalizations* and the *noun–verb quotient*, which both characterize texts regarding the emphasis on objects (nouns) in relation to processes (verbs), making texts more dense (see Einarsson, 1978).
- *Complexity*: A common claim is that mathematical texts are complex, either through a general and unspecified statement or more specifically about sentences and words. We choose to examine *sentence length* and *word length*, which are two measures of complexity commonly used in studies of readability (e.g., see Oakland & Lane, 2004).
- *Relations*: Several claims specify logical relations as most common in mathematics. In particular, Solomon and O'Neill (1998) present logical arguments for the impossibility to include temporal relations in mathematical texts and that logical relations are an inherent property of mathematics. Therefore, we choose to examine *logical relations* and *temporal relations*.
- *Persons*: Several claims describe the absence of human beings in mathematical texts, in particular that the author is hidden and that the inclusion of the reader varies. Therefore, we choose to examine the use of *pronouns*. Although a text can include or address human beings in different ways, pronouns are one important way (Herbel-Eisenmann & Wagner, 2007). We also examine

the use of *passive voice* since it can be seen as another way to hide human agents in mathematical texts (Burton & Morgan, 2000).

Based on these four categories of types of claims, we are interested in testing the validity of four overarching claims:

1. Mathematical texts have much information in a small space, that is, they are compact.
2. Mathematical texts have complex sentences and words.
3. Mathematical texts focus on logical relations and not temporal relations.
4. Mathematical texts hide human agents.

The above claims are very broad and we do not test the validity of each "whole" statement but, rather, more specific properties that can be placed within each claim, as described above. In particular, we examine whether mathematics textbooks and history textbooks differ with regard to these more specific properties. That is, our aim is to examine whether some specific parts of the four claims are valid for mathematics texts when compared with texts from another subject, that is, history.

5.1.3 Method of analysis

Each of the four claims described above contains different specific text properties that are analyzed. In this section, we operationalize these properties through different quantitative measures and describe the statistical methods used to test the validity of the four claims.

Claim number one, about compactness, contains two measures:

- *Proportion of nominalizations* (per total number of words): Nominalizations are nouns that have been created from a transformation of a verb (e.g., "calculation" from "calculate").
- *Noun–verb quotient*: Proper names and units are counted as nouns, where different parts of composite units are counted as different nouns (e.g., "m/s" counts as two nouns).

Claim number two, about complexity, contains two measures:

- *Sentence length* (number of words): A sequence of symbols, such as an algebraic expression or an equation, is counted as *one* word. Although such sequences could be regarded as corresponding to several words, the simplest view of this type of sequence is to regard it as "one thing" and therefore we count it as one word in our analysis that focuses on aspects of the natural language. Abbreviations (including units) and numbers are counted according to how they are pronounced (e.g., "km/h" counts as three words).
- *Proportion of long words* (per total number of words): A long word is defined as a word with three or more syllables (Butler et al., 2004). As with sentence length, for abbreviations and symbols the number of syllables is counted based on the pronunciation.

Claim number three, about relations, contains two measures:

- *Proportion of logical relations* (per total number of relations, counting only logical and temporal relations): A logical relation is an occasion when words are used that signal a logical implication in any direction between statements, through words like "therefore," "if... then," or "because."

- *Proportion of temporal relations* (per total number of relations, counting only logical and temporal relations): A temporal relation is an occasion when words are used that signal a temporal connection between text segments describing different situations or events, through words like “thereafter,” “first... then,” or “before.”

Claim number four, about persons, contains two measures:

- *Proportion of pronouns* (per total number of words): The pronouns we count are personal or possessive pronouns, first or second person, singular or plural; for example, me, mine, we, our, you, and your.
- *Proportion of passive constructions* (per total number of words): A passive construction is an occasion where a verb is written in passive form, such as “this can *be calculated*,” in contrast to the active form used in “we can *calculate* this.”

The analysis procedure includes tests of reliability. First, both authors analyzed two (one from each subject) randomly selected text passages from among the total of 48 passages. In this first step, there was an agreement of 92% for all judgments made, with a spread from 77% to 100% for different types of judgments. *One judgment* corresponds to *one specific count for one sentence*, for example, to count the number of nouns in one sentence. Differences in judgments were discussed until an agreement was reached and the coding instructions were adjusted and specified. Second, another four passages, spread over different subjects and grade levels, were analyzed by both authors. In this second step, there was an agreement of 96% for all judgments made, with a spread from 86% to 100% for different types of judgments. Again, differences were discussed and coding instructions were adjusted. Finally, the remaining text passages were randomly divided between the two authors for analysis.

The main reason for the lowest percentages of agreement noted above is the inclusion of numbers and units when analyzing words. It is not always evident how to make judgments about symbols concerning properties that are originally meant for natural language, for example to count syllables for “1945.” We still choose to include these parts of the texts to perform more complete descriptions of properties of the texts, aiming to exclude only symbolic (mathematical) expressions and not individual symbols that are of common type assumed to appear in many different subjects, for example “%.” This choice is made in order to be able to make more valid comparisons between texts, but with the apparent risk of reducing the reliability.

When comparing the texts from the two subjects, mathematics and history, different statistical methods are used for different measures. For measures that describe a proportion of a certain property (see list above), Pearson Chi Square Tests are used, while medians are compared for other measures, using Mann-Whitney U-tests. In all statistical analyses, we use two-tailed tests with level of significance $p<0.05$.

Although the noun–verb quotient is directly comparable between texts, it does not directly lend itself to statistical analysis based on random selections of texts. Therefore, to be able to perform statistical analyses also on the noun–verb quotient, we compare medians when the quotient is calculated separately for each text passage. Therefore, the noun–verb quotient relies on a sample size of 24 text passages for each subject, which is a relatively small sample size. However, in similar analyses, using the same sample procedure, this sample size has been enough to produce reliable measures most of the time (see Fitzgerald, 1980). Measures concerning logical and temporal relations are proportions of the total number of relations in the texts and therefore the sample sizes for these measures are a priori unknown. Exact numbers of sample sizes for the other measures are not determined a priori, but they are presented in Table 4 below: more than 200 sentences for each subject for analyses of sentence length and more than 2500 words for each subject for analyses of all other measures. The statistical methods we use rely on independent random selections, but we use cluster sampling, that is, each word and sentence is not chosen independently from all other words and sentences. However, other empirical results, specifically for text analysis, show that cluster selection of a total of 500–3000 words is enough for a representative sample for many types of linguistic analyses, both about words and about sentences (Tesitelová, 1992). Therefore, we draw the conclusion that our sample sizes are suitable for analyses of properties of words and sentences.

5.2 Results from Part 2

Relations are excluded from the statistical analysis since logical and temporal relations are too uncommon, with a total of only seven logical and eight temporal relations found in all mathematics texts (15 and 21 are corresponding numbers for all history texts). Our main results are comparisons between all mathematics textbooks on the one hand and all history textbooks on the other hand, since our focus is general differences or similarities between the subjects. However, we also report comparisons for each grade level separately.

The results in Table 4 can be summarized as follows. Comparisons regarding compactness show only one occasion, nominalizations at the upper secondary level, where there is a significant difference between mathematics and history textbooks. The mathematics textbooks at this level include fewer nominalizations than the history textbooks, but overall no clear differences exist concerning the property of compactness. When focusing on sentences, the mathematics textbooks are clearly less complex than the history textbooks, while no differences exist when focusing on the complexity of words. Even if not all differences are significant, there is a tendency for the mathematics textbooks to include more human agents in the texts. The significant differences show that mathematics textbooks include more personal pronouns, overall and for grade 7, and less passive voice, only for grade 7.

Table 4 Statistical comparisons of linguistic properties between mathematics and history textbooks. Statistical significances are marked accordingly: *p<0.05; **p<0.01; ***p<0.001

				Compactness		Complexity		Persons	
Texts	**Text passages**	**Sen-tences**	**Words**	**Nomina-lizations**[a]	**Noun–verb quotient**[b]	**Sentence length**[b]	**Long words**[a]	**Pronouns**[a]	**Passive voice**[a]
Math, all	24	319	2645	1.89%	1.56	7	20.9%	1.78%	1.17%
History, all	24	219	2563	2.50%	1.62	10	20.9%	0.66%	1.48%
Difference				*–0.61%*	*–0.06*	*–3****	*0.0%*	*1.12%****	*–0.31%*
Math, grade 4	8	121	827	1.69%	1.36	6	20.1%	2.66%	1.21%
History, grade 4	8	85	830	1.57%	1.52	9	16.9%	1.45%	0.84%
Difference				*0.12%*	*–0.16*	*–3****	*3.2%*	*1.21%*	*0.37%*
Math, grade 7	8	114	909	1.76%	1.48	7	20.5%	1.87%	0.66%
History, grade 7	8	80	865	1.85%	1.42	10	21.7%	0.35%	1.73%
Difference				*–0.09%*	*0.06*	*–3****	*–1.2%*	*1.52%***	*–1.07%**
Math, upper secondary	8	84	909	2.20%	1.94	10	22.2%	0.88%	1.65%
History, upper secondary	8	54	868	4.03%	1.74	15	24.0%	0.23%	1.84%
Difference				*–1.82%**	*0.20*	*–5****	*–1.8%*	*0.65%*	*–0.19%*

[a] Pearson Chi Square Test is used to test statistical significance of difference

[b] Mann-Whitney U-test is used to test statistical significance of difference

6 Conclusions and discussion

The literature survey in Part 1 shows that there are many and varying claims about linguistic properties of mathematical texts. Overall, claims tend to be consistent, in particular since they have a common focus on issues of vocabulary and that mathematical texts are characterized as complex. However, many claims are presented without any argument, equally frequent in top-ranked scientific publications as in all examined references. There are also few actual *studies*, that use empirical or logical arguments, on linguistic properties of mathematical texts, and very few studies compare mathematical texts with other types of texts. In addition, many claims are vague, such as claims that mathematical language is "unintuitive," "synthetic," or "complex." In many cases it is also unclear what exactly is meant by *mathematical language*: is it the symbolic language, the natural language, or both? Similarly, it is often not explicitly stated whether the claim concerns a particular natural language (e.g., English) or type of text, for example a particular "level" (e.g., primary school textbooks or research articles) or topic (e.g., algebra). We conclude that there is a weak foundation for drawing any general conclusions about what, if anything, is special about mathematical texts.

Furthermore, there are contradictions between (the few) existing empirical studies and common claims about linguistic properties of mathematical texts, since there is empirical evidence that mathematical texts tend to be less complex than texts from other subjects. However, as noted above, there are several limitations when analyzing common claims. In particular, many claims are unspecific and vague, and the apparent contradiction could therefore be created by the focus on different types of more specific properties. Also, claims seldom focus on comparing mathematical texts with some other type of text, while empirical studies examine relative complexity, that is, whether one text is more or less complex than another text. Thus, even if mathematical texts, when compared with texts from other subjects, tend to be less complex, their level of complexity might still be regarded as high, either on some absolute scale (e.g., as shown by readability formulas) or in relation to some other type of text, for example in relation to ordinary, everyday texts.

We see it as problematic that the discourse on traits of mathematical language on the surface seems to be consistent, when in fact: the claims are rather vague; it is often not clear what is actually discussed; and claims are seldom supported by empirical or logical arguments or even contradict empirical studies.

The second part of the present paper consists of a contribution to comparative empirical research about linguistic properties of mathematical texts. At a general level, our results are in line with previous empirical comparisons of texts from different subjects (see results from part 1, and in particular Butler et al., 2004): mathematical texts have so far never been shown to be more complex than texts from other subjects, and any significant differences indicate rather that mathematical texts tend to be less complex. The notion of complexity here includes such properties as sentence length; that mathematics textbooks tend to have shorter sentences, and proportions of nominalizations and passive voice; and that mathematics textbooks tend to have fewer of these types of constructions. However, the empirical basis for the above conclusions about linguistic properties of mathematical texts is very limited, essentially consisting of only three studies: two found in Part 1 together with the empirical study in Part 2.

Based on the combined results from the two parts of this study, one conclusion is that there exist several types of claims regarding mathematical texts that do not seem to be valid for texts in mathematics textbooks. There are many possible reasons for this scenario, for example that textbook authors try to compensate a perceived, or actual, difficulty in the subject of mathematics by using simple natural language to present the subject matter. However, since very few empirical studies exist that have analyzed mathematical texts, it is of course possible that mathematical texts that have not been analyzed, such as higher level textbooks or texts on some particular topics, have properties that are more in line with common claims. We can only conclude that it is not clear how, or even if, mathematical texts in general can be described in common linguistic terms. Empirical research must take into consideration more types of linguistic properties of texts, as well as more natural languages, more topics, and more subjects.

7 References

Anstrom, K., DiCerbo, P., Butler, F., Katz, A., Millet, J., & Rivera, C. (2010). *A review of the literature on academic English: Implications for K-12 English language learners*. Arlington, VA, USA: The George Washington University Center for Equity and Excellence in Education. Retrieved August 21, 2012, from http://ceee.gwu.edu/sites/files/AcademicLitReview2011.pdf

Artley, A. S. (1943). A study of certain relationships existing between general reading comprehension and reading comprehension in a specific subject matter area. *Journal of Educational Research, 37*, 464-473.

Behrman, E. H., & Street, C. (2005). The validity of using a content-specific reading comprehension test for college placement. *Journal of College Reading and Learning, 35*(2), 5-21.

Bergqvist, E., Bergqvist, T., Boesen, J., Helenius, O., Lithner, J., Palm, T., et al. (2009). *Matematikutbildningens mål och undervisningens ändamålsenlighet. Grundskolan våren 2009*. Gothenburg, Sweden: National Center for Mathematics Education. Retrieved May 6, 2013, from http://ncm.gu.se/media/ncm/forskning/kunskapsoversikt_ncm_ufm_gr.pdf

Brunner, R. B. (1976). Reading mathematical exposition. *Educational Research, 18*, 208-213.

Burton, L., & Morgan, C. (2000). Mathematicians writing. *Journal for Research in Mathematics Education, 31*, 429-453.

Butler, F. A., Bailey, A. L., Stevens, R., Huang, B., & Lord, C. (2004). *Academic English in fifth-grade mathematics, science, and social studies textbooks* (CSE Report 642). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved August 21, 2012, from http://www.cse.ucla.edu/products/reports/r642.pdf

Dowling, P. (1996). A sociological analysis of school mathematics texts. *Educational Studies in Mathematics, 31*, 389-415.

Einarsson, J. (1978). *Talad och skriven svenska: Sociolingvistiska studier [Spoken and written Swedish: Sociolinguistic studies]*. Doctoral Dissertation. Lund, Sweden: Ekstrand.

Fan, L., Chen, J. A., Qiu, X. L., & Hu, J. Z. (2004). Textbook use within and beyond Chinese Mathematics Classrooms: A study of 12 secondary schools in Kunming and Fuzhou of China. In L. Fan, N. Y. Wong, J. Cai & S. Li (Eds.), *How Chinese learn mathematics: Perspectives from insiders* (pp. 186-212). Singapore: World Scientific Press.

Fitzgerald, G. G. (1980). Reliability of the Fry sampling procedure. *Reading Research Quarterly, 15*(4), 489-503.

Fuentes, P. (1998). Reading comprehension in mathematics. *Clearing House, 72*(2), 81-88.

Haggarty, L., & Pepin, B. (2002). An investigation of mathematics textbooks and their use in English, French and German classrooms: Who gets an opportunity to learn what? *British Educational Research Journal, 28*(4), 567-590.

Herbel-Eisenmann, B., & Wagner, D. (2007). A framework for uncovering the way a textbook may position the mathematics learner. *For the Learning of Mathematics, 27*(2), 8-14.

Lithner, J. (2004). Mathematical reasoning in calculus textbook exercises. *Journal of Mathematical Behavior, 23*, 405-427.

Love, E., & Pimm, D. (1996). 'This is so': A text on texts. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), *International handbook of mathematics education (Part 1)* (pp. 371-409). Dordrecht, The Netherlands: Kluwer.

McBride, M. (1994). The theme of individualism in mathematics education: An examination of mathematics textbooks. *For the Learning of Mathematics, 14*(3), 36-42.

McKenna, M. C., & Robinson, R. D. (1990). Content literacy: A definition and implications. *Journal of Reading, 34*, 184-186.

Morgan, C. (1998). *Writing mathematically: The discourse of investigation*. London: Falmer.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA, USA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Nesher, P., & Katriel, T. (1986). Learning numbers: A linguistic perspective. *Journal for Research in Mathematics Education, 17*, 100-111.

Oakland, T., & Lane, H. B. (2004). Language, reading, and readability formulas: Implications for developing and adapting tests. *International Journal of Testing, 4*(3), 239-252.

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

Pimm, D. (1989). *Speaking mathematically: Communication in mathematics classrooms (paperback edition)*. London: Routledge.

Remmers, H. H., & Grant, A. (1928). The vocabulary load of certain secondary school mathematics textbooks. *The Journal of Educational Research, 18*(3), 203-210.

Richards, J. (1991). Mathematical discussions. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 13-51). Dordrecht, The Netherlands: Kluwer.

Ryve, A. (2011). Discourse research in mathematics education: A critical evaluation of 108 journal articles. *Journal for Research in Mathematics Education, 42*, 167-198.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40-59.

Smith, F. (1969). The readability of junior high school mathematics textbooks. *The Mathematics Teacher, 62*(4), 289-291.

Solomon, Y., & O'Neill, J. (1998). Mathematics and narratives. *Language and Education, 12*(3), 210-221. Retrieved August 6, 2008, from http://www.channelviewpublications.net/le/012/0210/le0120210.pdf

Tesitelová, M. (1992). *Quantitative linguistics*. Philadelphia, USA: Benjamins.

Wiegand, R. B. (1967). Pittsburgh looks at the readability of mathematics textbooks. *Journal of Reading, 11*(3), 201-204.