Nina Eliasson is a Ph.D. student in Science Education at the Department of Science Education and Mathematics (DMA) at Mid Sweden University, Sweden. Her main study interests include gender differences in science education. Member of the science expert team, PISA National Center, Sweden.

Karl Göran Karlsson is Professor in Physics Education at the Department of Science Education and Mathematics (DMA) at Mid Sweden University, Sweden. His research is mainly based on large scale international studies with special focus on equity. Member of the Science Expert Team, Senior advisor in science, PISA National Center, Sweden.

Lena Lenner, Faculty Administrative Officer at Faculty of Science, Technology and Media (NMT). Former employee at the Department of Science Education and Mathematics (DMA), Mid Sweden University. Member of the Swedish PISA, National Center, Project Ass.

Maria Lundgren, Investigator at the Department of Management Accounting and Control, The National Agency for Special Needs Education and Schools. Former employee at the Department of Science Education and Mathematics (DMA), Mid Sweden University, Member of the Swedish PISA National Center, Data Manager and Ass. National Project Manager.

NINA ELIASSON
Mid Sweden University, Sweden, nina.eliasson@miun.se

KARL GÖRAN KARLSSON
Mid Sweden University, Sweden, kg.karlsson@miun.se

LENA LENNER
Mid Sweden University, Sweden, lena.lenner@miun.se

MARIA LUNDGREN
The National Agency for Special Needs Education and Schools, Sweden, maria.lundgren@spsm.se

Boys' and girls' written responses to PISA science questions

Abstract
For the first time student responses to science questions from the Swedish PISA 2006 Main Study and the PISA 2015 Field Trial have been used in order to investigate differences in boys' and girls' written responses. Students' correct and incorrect answers to the science questions are studied with respect to response length, the number of everyday words used, and the inclusion of nouns and long words in the responses. The results reveal that girls give longer and denser correct responses to most of the questions, compared to boys. The difference in response length cannot be explained by girls' excessive use of the most common Swedish words, since boys and girls use the same proportion of these words. For incorrect answers the only difference between boys and girls is in the response length, since girls give longer answers than boys.

Introduction

Boys participate in the communication that occurs in the science classroom to a greater extent than girls (Eliasson, Sørensen & Karlsson, 2016; Kahle & Meece, 1994; Kelly, 1988; Wernersson, 2006). This finding has inspired us to further investigate written science communication. Therefore this study explores how boys and girls, at the end of the final year of the Swedish lower secondary school, respond to written science questions. To find out if there are any gendered differences in response patterns, questions from the Programme for International Student Assessment (PISA) have been used. Results from 1350 students in an unpublished pilot study on a publically released PISA science item from 2006, revealed that on average boys used fewer words than girls in order to correctly answer the two questions included. No difference was found between boys and girls with respect to the proportion of everyday words used. Here these results are presented together with new results from 27 open response science questions, included in the PISA 2015 Field Trial. Together the written responses to these questions will broaden our knowledge about any potential differences in boys' and girls' responses to science questions in large scale assessments, and possibly to science questions in general.

The empirical material in this study originates from the PISA 2006 Main Study and from the PISA 2015 Field Trial. Henceforth these studies will be referred to as PISA 2006 and PISA 2015 FT.

Learning science as a social activity

Mortimer and Scott (2003) argue that learning science involves being introduced to concepts, conventions, laws, theories, principles and ways of performing scientific work. They state that learning science can be considered to be a question of learning the social language of science, or at least learning a specific form of this social language. According to Vygotsky's (1978) socio-cultural view of learning, language development and learning both take place during interaction with others. Similarly, Lemke (1990) believes that the use of language should be considered both as a social, and as a cultural process. He states that the way we understand our world is built upon jointly shared meaning making activities, which are mediated within and through language. Halliday expresses similar thoughts when he says that "... language is the essential condition of knowing, the process by which experience becomes knowledge" (Halliday, 1993, p. 94). These ideas about the role of language in science learning are all in line with the socio-cultural viewpoint which sees meaning making as a dialogic process, itself an idea that draws on a Bakhtinian perspective on the dialogic nature of understanding (Mortimer & Scott, 2003).

Language activities are hardly possible without communication, something that is more than an oral activity that takes place between two or more interlocutors. Linell (2009) states that communication always occurs in any form of interaction. This interaction might be in the form of a direct dialogue between two or more individuals, but may also take place indirectly through different texts. Science learning can be described as a joint process involving all participants, where the focus is on learning how to master this specific scientific language by understanding it, speaking it and by writing it (Lemke, 1990; Schleppegrell, 2004; Wellington & Osborne, 2001).

The communication addressed in the present study is indirect. It is assumed that students interact in dialogue by reading the different texts that constitute the science questions, and by producing written responses to these questions.

Students and "School Languaging"

Some general linguistic differences in essays written by Swedish high school boys and girls have previously been described by Hultman (1990). He has shown that when they produce written texts, girls tend to use everyday language more often than boys, with a large proportion of short and commonly

used words. Girls' essays are often longer than boys', they tend to be more naïve but also to have relatively few errors. Boys' language in written essays is closer to what Hultman describes as an official language, their essays include more frequent spelling errors, and often have a more condensed syntax compared to girls' written texts.

Frändberg (2012) has shown that underachieving students in science tend to use less science-specific language. Different abilities to master a specific scientific language, or different parts of it, may imply different abilities to learn science.

Characteristic Features of Scientific Language

Features that are especially pronounced in written science texts have evolved over time. There are examples of grammatical features which are considered to make scientific texts difficult to comprehend. A high level of abstraction, and dense texts with a high proportion of technical terms are for example considered to be such features (Edling, 2006; Nygård Larsson, 2011). One direct consequence of the use of technical terms is that they condense information (Edling, 2006). Johansson Kokkinakis and Frändberg (2013) have defined technical terms as being subject-specific words which are used with a specific meaning in one subject area, and a different meaning in another subject area.

Viewing scientific language as being a technical language can be explained by a high inclusion of uncommonly used words according to af Geijerstam (2010). The number of uncommonly used words is particularly high in scientific language, compared to other subject languages (Eriksson, 2015).

Nation and Chung (2009) have shown that the 2000 most commonly used words in English, cover between 80 to 90 percent of the words used in a text, depending on the type of text. In newspaper texts approximately 85 percent of the words used come from the list of the 2000 most frequently used words (ibid.). The amount of technical vocabulary in specialized texts is higher and varies due to the subject (Nation & Chung, 2009). The amount of non-everyday words in the language of natural science is particularly high compared to other subject specific languages (Edling, 2006).

The use of the passive voice is another common trait which is considered to create more demanding texts for the reader, due to the suppression of agency (Persson, 2016). A further feature of scientific language is the use of grammatical metaphors, a concept which is used in Systemic Functional Linguistics (SFL). A grammatical metaphor is a generic name for non-typical ways of expressing meaning, e.g. nominalisations (Nygård Larsson, 2011). Halliday (1998) writes about nominalisation which in this case is the shift from one grammatical category to another, usually verbs or adjectives being made into nouns. This practice gives a high proportion of nouns in science texts (Persson, 2016) and is widely recognized as an important and common feature of scientific writing (Banks, 1999; Banks, 2005; Halliday, 1998). Persson (2016) also states that this nominalisation process often creates long words, which is a reason why the proportion of long words in a scientific text can be used to measure its complexity. The frequent use of nominalisations in scientific language makes it possible to compress information, and density in information being most pronounced in written science texts (af Geijerstam, 2010). *Packing,* as described by Persson (2016), refers to the proportion of nouns and long words in a text, and is one of three meta-functions developed within the framework of SFL.

Swedish Research on Science Questions in Large Scale Studies

Some Swedish research has been reported on scientific language in large scale studies such as Trends in International Mathematics and Science Study (TIMSS) and PISA.

Persson (2016) has examined scientific language in the different science subjects by analyzing all grade 8 science items in TIMSS 2011. He shows that different readability measures are correlated with test performance of different student groups.

Frändberg (2012) has described the degree to which students use linguistic features typical of scientific language in their written explanations to questions included in TIMSS 2007. She found that students with high total test-scores use significantly more technical terms than the students with low scores.

Serder (2015) has shown that the language used in PISA questions influences students' chances for meaning making. In group discussions on science questions in PISA, it has also been shown that these can be interpreted in more than one way (Serder & Jacobsson, 2015).

Regarding questions in chemistry from TIMSS 2007, Johansson Kokkinakis and Frändberg (2013) found that high achieving students used a higher rate of technical terms, wrote longer responses and used more words with a context-specific meaning compared to low achieving students. One of their conclusions was that there is an obvious positive relationship between successful students and their ability to express themselves scientifically.

Generally speaking there are several studies of oral communication but few of students' scientific writing and even fewer with a gender perspective. Moreover, to our knowledge there is no previous research on students' written responses to science items in PISA, a research gap this study aims to fill. One advantage is the size of the material used in this study which includes a large number of student responses to questions in PISA 2006 and student responses to a broad spectrum of science questions in PISA 2015 FT. The variety of questions makes it possible to study if differences between boys and girls are different for different types of questions.

Aim of Study

In this study we focus on students' written responses to science questions in PISA. The aim is to find out if there are any differences in boys' and girls' written responses. In order to describe any possible differences in language use we have for PISA 2006 investigated the *Mean Response Length* and the use of the *Most Common Words* in the Swedish language. For PISA 2015 FT we have in addition also used the number of nouns and long words, referred to here as *Packing*. Differences are studied both for correct and incorrect responses, if the latter have been deemed by expert coders as serious attempts to answer the question.

Science Questions in PISA

The questions used in PISA 2015 FT are not released to the public since they will be used in forthcoming PISA studies. The science item *Acid Rain* (see figure 1 below) was released to the public after inclusion in PISA 2006. Therefore the two questions from Acid Rain will serve as illustrations of paper based PISA science questions, at the same time as the written responses to these two questions are part of the empirical material.
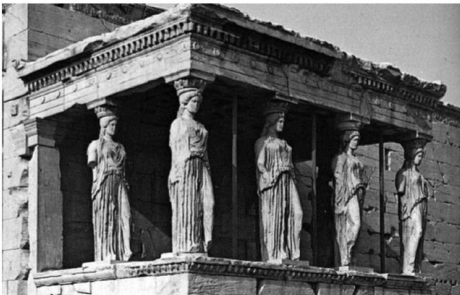
In PISA an item comprises of several questions on the same theme. The use of an introducing stem is more or less typical for all items used in PISA and Acid Rain is no exception. This stem consists of a short text about how acid rain eats away the marble statues of the Acropolis, and a picture of the Athenian statues which follows the text. The first of two open questions in this item asks students about the origin of Sulphur Oxide and Nitrogen Oxide. Two dotted lines are available for students' written responses. The general instructions to the coders of student responses in PISA are that spelling and grammar mistakes should be ignored, unless they seriously obscure the meaning and understanding of a student response. For more information about the coding procedures in PISA, please see the Technical Report (OECD, 2009).

To earn partial credit for this item, a code 1, the student needs to refer to some kind of pollution but there is no need to mention its source. A one word response such as "*Pollution*" is sufficient to earn partial credit, and to mention *the environment* in general with responses like "*Emissions to the atmosphere*" is deemed sufficient to earn partial credit. To earn full credit, a code 2, the student needs to mention any one of the following; *car exhausts*, *factory emissions*, *the burning of fossils such as oil* or *coal*, *gases from volcanoes* or other typical sources. Examples of full credit responses are "*Burning coal and gas*", or "*Oxides in the air come from pollution from factories and industries*".

Question two from this item is preceded by a text describing how the effect of acid rain on marble can be modelled in an experiment with water and vinegar. The question asks why students who performed this experiment also included a step where marble chips were placed in distilled water. A correct response can be rewarded with partial credit, a code 1, or full credit, a code 2. Code 1 is given for a response where the student states that "*This step is included to compare with the test of acid rain and marble*" but without any further explanation being given. A code 2 is given if the student responds that this step is included "*In order to compare with the test of acid and marble and to show that the acid (vinegar) is necessary for the reaction*". In this study partially correct responses and fully correct responses have both been treated as correct answers.

Figure 1 *Two questions in the PISA Science Item, Acid Rain*

Other publically released PISA science items in English and Swedish respectively can be downloaded from the OECD's website and from the website for the Swedish National Agency for Education.

Method

Participants and Materials

The domains tested in PISA are science, mathematics and reading. All the cognitive test material used in PISA 2006 was in the form of paper booklets, and was distributed as a paper and pencil test to each student. In Sweden a total number of 1362 students (644 girls and 718 boys) received the Acid Rain item. The students were sampled according to the PISA standard, in order to fully represent the total Swedish population of all 15- year old students. For a more detailed description of the sampling design in PISA we refer to the Technical Report (OECD, 2009).

The participating schools in PISA 2015 FT were sampled from two different Swedish regions selected for convenience. Within these regions, schools were sampled to reflect the demographical structure of the country. Students from 39 schools, 15 from each school, were sampled for the study. Different science questions were included in two thirds of the booklets. (One third of the booklets contained only questions about maths and reading). 29 science questions of the open question format were allocated to six science clusters, and each student in this study received two science clusters. Two of the open questions should be solved by choosing the right combination of different options. The students were not required to write an answer, so these two questions have been removed from the dataset. With this design each question in the study has been allocated to between 112 and 121 of the students sampled. A total number of 347, 178 girls and 169 boys, completed the science test in the paper-based assessment. Nearly 1280 correct responses from PISA 2015 FT are included in this study, as well as about 1000 incorrect responses. The proportion of questions left without responses in the material is equal for boys and girls ($\approx$23%) and so is the proportion of responses that were not deemed to be serious attempts to answer the question ($\approx$3%).

Results from the cognitive test in the PISA Field Trials are never reported to the general public.

Statistical analysis

Student responses to PISA 2006 questions and to PISA 2015 FT questions have been transcribed into an excel spreadsheet as far as possible in the same way that the written responses appear in the booklets. For example this means that students' spelling mistakes, the writing of split compounds of words or an inconsistent use of lower or upper case letters have been transferred into the spreadsheet.

In order to quantify the number of words used an algorithm has been used for each student response. The algorithm singles out all letter and character combinations that are preceded and followed by a space, as one word. This does not necessarily mean that all sequences of characters enclosed in this way should be considered as real words that can be found in a dictionary. Examples of exceptions are numerical expressions, abbreviations and proper names (Järborg, 2007). If one or more of these sorts of exceptions have been used in the response, they have been counted as words according to the algorithm used.

Deciding what should be considered as a technical term or a scientific term is somewhat subjective, no unequivocal definition exists (Johansson Kokkinakis & Frändberg, 2013). In this article, to avoid subjectivity we measure the proportion of the *Most Common Words*, or everyday words, in the responses instead. Our assumption is that a low degree of everyday words means a high degree of technical and scientific words in the texts, which makes the language more scientific. To measure the use of the *Most Common Words*, text files with the answers for each science question in PISA have been created and used in *AntwordProfiler*. In these text files single numbers have been rewritten with letters since

single numbers are not counted as words by this software. *AntWordProfiler* is a freeware and a multi-platform tool for carrying out corpus linguistic research on vocabulary profiling (Anthony, 2013). This program is used to generate vocabulary statistics and frequency information about a corpus of texts. The target texts have been checked against a word list. The word list consists of the 2000 most common Swedish words. This list is used in order to find out to which extent boys and girls use the *Most Common Words* when they respond to the questions. In an email to author (N. E) Johansson Kokkinakis gave details about the origin of the 2000-word list. This list was created in 2012 and is based on 200 million words from a blog corpus (Johansson Kokkinakis, personal communication, December 10, 2015). If a word is used by a student it is identified only if it conforms to the words found in the word list used in the analysis. A word that is misspelt by the student or has been incorrectly transcribed to the data file will not be identified as one of the most common words.

The text files for the PISA 2015 FT material have also been used to find the number of nouns used and the number of long words used (words with six or more letters) in the answers. Nouns are found by inspection of the responses and long words are identified by an algorithm. In order to add nouns and long words together these values first need to be normalised by having its standard score, (z-score) calculated. Normalisation is done in order to avoid a dis-proportionate impact by the number of nouns, since nouns exceed the number of long words in the material (Persson, 2016, p. 59). The normalised values are then added together into the dimension *Packing* and are thereafter divided by two as described by Persson, (2016, p. 65). The packing variable used thus consists of the proportion of nouns and long words in relation to the total number of words used in each answer, in accordance with Persson's definition of the variable (2016, p. 59).

Separate methods of measuring the strength of group differences have been applied to PISA 2006 and PISA 2015 FT due to different sampling methods. The Swedish result for PISA 2006 could be generalized to the Swedish population and therefore Chi-2 tests and *t*-tests can be used to estimate the strength of group differences. According to Neill (2008) significance tests should be used in an attempt to generalize the sample's result to a population, but if the results are not to be generalized to a population, the strength of group differences can be examined by the use of effect sizes. Therefore, effect size measures for the two independent groups are used for PISA 2015 FT with respect to response length and packing. The results in PISA 2015 FT cannot be fully generalized to the Swedish population.

For the Effect Sizes, values of Cohen's *d* are used as the descriptive measure of group differences to questions in PISA 2015 FT. According to Cohen (1988), the standard deviation of either group could be used when the variances of the two groups are homogeneous. Since there are some differences in the standard deviation between the two groups, the pooled standard deviation has been calculated, a commonly used practice according to Rosnow and Rosenthal (1996). This practice is also recommended by the PISA Technical Report (OECD, 2009).

In PISA 2015 FT all longer responses, that is outliers with a length that exceeds the mean by two standard deviations, have been removed from the material. This is done in order to prevent these long answers affecting the mean in a dis-proportionate way (Djurfeldt, Larsson & Stjärnhagen, 2009). Outliers due to responses that are too short have been left in the material since it is often possible to gain a score with only one word or a few words in response to a question.

Chi-2 tests, *t*-tests and Effect Size tests are used to estimate the strength of the differences between the two groups studied. The selected level of significance for differences at group level in the material is at the 95 percent level ($p < .05$). The Effect Size differences between group means of 0.2 of a standard deviation should be deemed as small, 0.5 as medium and 0.8 as a large effect according to Pedhauzur and Pedhauzur Schmelkins' recommendations (1991).

Results

The results for PISA 2006 and PISA 2015 FT are presented separately because of the difference in methods used.

PISA 2006

A total of 753 correct responses to Acid Rain 1 and 849 responses to Acid Rain 2 are presented in Table 1 together with the *Mean Response Length* in the number of words used to answer these questions.

Table 1. Mean Response Length, correct responses, Acid Rain. (PISA 2006)

Question	Correct Response (N)	Mean Response Length	Correct Response (N)	Mean Response Length	Difference, Mean Response Length*	p-value
	Girl		Boy			
Acid Rain 1	358	7.6	395	6.4	**1.2**	<0.001
Acid Rain 2	421	12.2	428	10.9	**1.3**	<0.001

*Bold figures indicate significant differences in *t*-tests

The table shows that on average girls use 7.6 words in their correct responses to the first question while boys use 6.4 words on average. The responses to the second question are longer and on average girls use 12.2 words in their responses while boys use 10.9 words on average. On average, the difference in correct *Mean Response Length* between boys and girls is 1.2 words to the first question about Acid Rain, and the difference is 1.3 words to question two. These differences are both significant. Table 2 presents the number of the *Most Common Words* used for correct answers to the both questions about acid rain.

Table 2. Boys' and girls' use of the Most Common Words, correct responses, Acid Rain. (PISA 2006)

Question	Most Common Words (N)	Other Words (N)	Most Common Words (N)	Other Words (N)	Chi-square
	Girl		Boy		*p*-value
Acid Rain 1	1594	1127	1441	1087	0.25
Acid Rain 2	4121	1015	3768	897	0.51

The Chi-2 test of group differences in number of *Most Common Words* used show that the difference between boys and girls is not found to be statistically significant (*p* < .05) for neither Acid Rain 1 nor Acid Rain 2. Boys and girls use a smaller proportion of the *Most Common Words* to answer Acid Rain 1 (57% and 59%) compared to Acid Rain 2 (81% and 80%).

PISA 2015 FT

There are fewer responses to each question in PISA 2015 FT. As mentioned earlier, between 112 and 121 students have responded to each question in PISA 2015 FT. The number of correct responses for boys and girls in this material is about the same, 646 for boys and 632 for girls. For PISA 2015 FT we have summarized the results for all questions, and also calculated the mean for this overall result, as well as calculating the result for each single question.

The *Mean Response Length* for all the correct responses to the 27 questions is 16.1 words for girls and 12.6 words for boys. For each of these 27 questions the difference in *Mean Response Length* between boys and girls has also been calculated. The difference in boys' and girls' response length to each question varies, and Effect Size is used here to represent the strength in this difference.

Table 3 presents the results for questions which have received correct responses, and the number of questions with a difference in *Mean Response Length* for boys or for girls. The number of questions with a different magnitude in strength of difference between boys and girls with respect to response length are presented in Table 3.

Table 3. The number of questions where boys or girls have longer Mean Response Length. Correct responses (PISA 2015, FT)

Number of questions at different Effect Size levels, Mean Response Length, Correct responses			
	Effect Size ≤ 0.2	0.2 < Effect Size < 0.8	Effect Size ≥ 0.8
Boys have longer responses	1	0	0
Girls have longer responses	1	17	8

Table 3 shows that boys do not use more words to answer any of the questions correctly in PISA 2015 FT compared to girls. On average girls use more words to answer a question correctly in 25 questions. Thus, girls' longer responses are not limited to specific types of questions but is a general pattern.

The number of incorrect responses in the material are for boys 484 and for girls 521. One of the questions is not included in the analysis of incorrect responses since the number of incorrect responses to this item was too few. Therefore only 26 questions with incorrect responses are included. For questions where the students give incorrect responses, girls use 13.1 words on average and boys use 11.2 words.

Table 4 presents the results for questions which have received incorrect responses and the number of questions with a difference in *Mean Response Length* for boys or for girls.

Table 4. The number of questions where boys or girls have longer Mean Response Length, incorrect responses (PISA 2015, FT)

Number of questions at different Effect Size levels, Mean Response Length, Incorrect responses			
(n)	Effect Size ≤ 0.2	0.2 < Effect Size < 0.8	Effect Size ≥ 0.8
Boys have longer responses	3	1	0
Girls have longer responses	10	9	3

The difference in the number of questions where boys or girls use more words in their incorrect answers, is not as evident as is the case for correct answers. Still, girls use more words in 12 questions and boys in one question, when medium and large Effect Size differences are considered to be sufficient. There is also a difference between boys and girls with respect to the length of their correct or incorrect responses. For girls the difference in response length with respect to correct and incorrect responses is 3.0 words (*ES* = 0.89) and for boys this difference is 1.4 words (*ES* = 0.48).

There are no differences in the use of the *Most Common Words* in the two questions used in PISA 2006. Similarly there are no differences in PISA 2015 FT between boys and girls with respect to the use of the *Most Common Words*. An average of all questions shows no significant difference in boys' or girls' use of the *Most Common Words* (*p-value* = 0.71). There is no significant difference in the use of the *Most Common Words* in any of the single questions included in this study. However, both boys and girls use less of the *Most Common Words* in their correct responses (67% and 68%) compared to their incorrect responses (75% and 73%).

The extent to which boys or girls have a higher degree of packing in their responses to the questions is presented in tables 5 and 6.

Table 5. The number of questions where boys or girls have a higher degree of packing, correct responses (PISA 2015, FT)

Number of questions at different effect size levels, Packing, Correct Responses			
(n)	Effect Size ≤ 0.2	0.2< Effect Size <0.8	Effect Size ≥ 0.8
Boys, higher degree of packing	6	3	3
Girls, higher degree of packing	2	8	5

Table 6. The number of questions where boys or girls have a higher degree of packing, incorrect responses (PISA 2015, FT)

Number of questions at different effect size levels, Packing, Incorrect Responses			
(n)	Effect Size ≤ 0.2	0.2< Effect Size <0.8	Effect Size ≥ 0.8
Boys, higher degree of packing	4	4	7
Girls, higher degree of packing	2	4	5

There is a small difference in favor of girls (13) with respect to the number of questions where girls have a higher degree of packing in their correct responses, compared to boys (6). Boys and girls appear to have almost the same degree of packing (11 and 9) in their incorrect responses.

Discussion and conclusion

This study has been conducted in order to examine if there are any differences in Swedish boys' and girls' responses to a set of different science questions in PISA. More specifically, in order to find differences in how boys and girls respond to them, *written* responses have been in focus.

The results show that on average girls use more words than boys to answer a science question in PISA correctly. This was the case for the two questions about acid rain from PISA 2006. Results from PISA 2015 FT show that these earlier results were not a coincidence. For 25 questions out of 27 in PISA 2015 FT girls on average used more words than boys to answer the questions correctly. (16.1 words for girls compared to 12.6 words for boys). This difference also applies to incorrect answers, although the difference is not as great as that for correct answers (13.1 words for girls compared to 11.2 for boys). Therefore it seems that girls write longer answers to almost all kinds of questions, so the pattern is not limited to certain kinds of tasks.

Hultman (1990) has shown that on average in the school subject Swedish girls write longer essays, more often use everyday language with a large proportion of short and commonly used words and that their texts tend to be more naïve compared to boys' texts. On the other hand, boys tend to use a language which is closer to what can be described as an official language with a more condensed syntax compared to girls' written texts. Dense (or condensed) texts are considered to be a feature of scientific writing (Edling, 2006; Nygård Larsson, 2011) and are a result of the use of many technical terms (Edling, 2006). Could it be that girls' excessive use of words can be explained by more frequent use of the *Most Common Words*, and that boys' shorter responses are denser since they use less of these everyday words? However, none of the results in this study support these speculations.

No difference is found between boys and girls with respect to their use of *the Most Common Words*, regardless of whether their responses are correct or incorrect. Boys and girls use the *Most Common Words* to the same extent. What is shown however, is that the entire student group uses more of the *Most Common Words* in their incorrect responses (74%) compared to in their correct responses (68%). The proportion of everyday words in the incorrect responses is still lower compared to the 85 percent that Nation and Chung (2009) have reported for a newspaper text. Since previous research has shown that the number of uncommonly used words is particularly high in scientific language compared to other subject languages (Eriksson, 2015) there is reason to believe that students that give incorrect responses are not as skilled in scientific writing as their peers. This suggestion is supported by Frändbergs (2012) finding that high achieving students use a higher rate of technical terms, write longer responses and use more words with a context-specific meaning compared to low achieving students.

There is also a difference in use of the *Most Common Words* for all students if correct responses to different questions are considered. In Acid Rain 1 students use about 58 percent of the *Most Common Words* and in Acid Rain 2 they use about 81 percent. The latter proportion is close to what Nation and Chung (2009) have shown to be normal for a newspaper text, so the responses to this question are not as scientific as the answers to the first question about acid rain. This suggests that the type of question is one important factor that influences the degree of scientific language that students use when they respond to a question. However this needs further investigation.

Still boys' responses are shorter than girls'. Could it be that boys are able to express themselves more efficiently and therefore give answers in accordance with Hultman's (1990) description? If so, one possible explanation to girls' excessive use of words could be attributed to a lesser tendency to use packing compared to boys. But the results provide no support for such an assumption either. In order to answer a question correctly, girls have a higher degree of packing than boys in 13 questions compared to 6 questions for boys. This shows that girls' more extended responses are somewhat more densely packed compared to the responses given by boys. A high degree of packing implies that the text includes a high proportion of nouns and long words (Persson, 2016). Moreover, with a high proportion of long words in a text it is the number of letters that will be high, while the number of words will be relatively low. The results presented show that girls, compared to boys, write longer and denser texts compared to boys when they respond to the scientific questions in PISA. Since dense texts are a feature of scientific writing (Edling, 2006; Nygård Larsson, 2011) one conclusion is that girls are

able to respond to scientific questions in a more scientific manner better than boys. Girls also write longer responses than needed in order to earn a credit code as compared to boys. Thus girls give more information than necessary, a behavior which is familiar to many teachers.

For incorrect responses to the science questions no differences are found between boys and girls in the degree of packing of the text. As mentioned in the background material, the proportion of unanswered questions is high at 23 percent, but is equal for boys and girls. The proportion of nonsense responses, 3 percent, is also equal for boys and girls. The high proportion of unanswered questions could be an effect of this test not being a high stake test for the students, since the results have no effect on their grades.

Results in PISA 2006 and PISA 2015 have shown that girls are more likely to have low science self-efficacy than boys (OECD, 2016; Skolverket, 2016). Therefore one hypothesis is that girls, despite good results in school science, are also affected by a lower self-esteem in test situations compared to boys. This may be a reason why girls to a higher degree give longer scientifically well formulated responses with an excess of information as compared to boys. The excess of information is given "just in case ...".

On the other hand the lengths of correct answers are significantly longer than incorrect answers. This indicates that longer answers usually give better results, which emphasizes the importance of practicing scientific writing. The smaller use of everyday words in correct answers compared to incorrect answers points in the same direction. A science education that enhances and encourages scientific writing would thus be one way to strengthen students' ability to express themselves scientifically in written texts. At the same time girls who already have the ability to express themselves would be strengthened in their self-esteem.

The main study in PISA 2015 was carried out on computers. To investigate whether boys' and girls' scientific answers will be different when the test is computer-based is one suggestion for further research. Another suggestion is to investigate potential differences in students' longer scientific texts such as lab-reports.

References

Anthony, L. (2013). *AntWordProfiler* (Version 1.4.0.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

Banks, D. (1999). Aspects of the development of grammatical metaphor in scientific writing. *Cahiers de l´APLIUT, 19*(1), 6–25.

Banks, D. (2005). On the historical origins of nominalized process in scientific text. *Elsevier, English for Specific Purposes, 24*(2005) 347-357.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Djurfeldt, G., Larsson, R., & Stjärnhagen, O. (2009). *Statistisk verktygslåda – samhällsvetenskaplig orsaksanalys med kvantitativa metoder*. Lund: Studentlitteratur AB.

Edling, A. (2006). *Abstraction and Authority in Textbooks: The Textual Paths Towards Specialized Language*. Uppsala: Acta Universitatis Upsaiensis.

Eliasson, N., Sørensen, H., & Karlsson, K. G., (2016), Teacher–student interaction in contemporary science classrooms: is participation still a question of gender?, *International Journal of Science Education, 38*(10), 1655-1672. DOI:10.1080/09500693.2016.1213457

Eriksson, M. (2015). *Att kommunicera naturvetenskap i nationella prov: En studie med andraspråksperspektiv*. Linköping: Linköpings universitet.

Frändberg, B. (2012). *Elevers språk i naturvetenskapliga förklaringar. Analys av skriftliga svar från årskurs 8 i TIMSS 2007*. Göteborg: Chalmers University of Technology. http://www.gu.se/forskning/publikation/?publicationId=157683

af Geijerstam, Å. (2010). Om skrivande i naturorienterande ämnen: "Hon skrev upp vad vi skulle ha med på labbrapporten och sen så skrev vi det. Så det var inte så mycket mer än så.". In M. Olofsson (Ed.), *Symposium 2009: genrer och funktionellt språk i teori och praktik* (p. 176-188). Stockholm: Stockholms universitetsförlag.

Halliday, M. A. K. (1993). Towards a Language-Based Theory of Learning. *Linguistics and Education*, 5, 93-116.

Halliday, M. A. K. (1998). Things and relations: Regrammaticising experience as technical knowledge. In J. R. Martin & R. Veel (Eds.), *Reading science, critical and functional perspectives on discourses of science* (p. 185–235). London: Routledge.

Hultman, T. G. (1990). Språk och kön i skolan. *Tidskrift för genusvetenskap*. (1), p. 19-29.

Johansson Kokkinakis, S., & Frändberg, B. (2013). Högstadieelevers användning av naturvetenskapligt språkbruk i kemiämnet i TIMSS. *Utbildning & Demokrati, 22*(3), 53-68.

Järborg, J. (2007). Om ord och ordkunskap. In I. Lindberg & S. Johansson Kokkinakis (Eds.), *OrdiL – en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år*. Rapporter om svenska som andraspråk (ROSA). http://hdl.handle.net/2077/20503.

Kahle, J. B., & Meece, J. (1994). Research on gender issues in the classroom. In D. L. Gable (Ed.), *Handbook of research on science teaching and learning* (pp. 542–557). New York, NY: Macmillan.

Kelly, A. (1988). *Gender differences in teacher-pupil interactions: Meta-analytic review*. Manchester: University of Manchester, Department of Sociology.

Lemke, J. L. (1990). *Talking science: Language, learning and values*. Norwood, NJ: Ablex.

Linell, P. (2009). *Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making*. Charlotte, NC: IAP.

Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Great Britain: Ashford.

Nation, P. & Chung, T. (2009). Teaching and Testing Vocabulary. In H. L. Michael & C. J. Doughty (Eds.) *The handbook of Language Teaching*. Singapore: Blackwell Publishing.

Neill, J. (2008). *Why use Effect Sizes instead of Significance Testing in Program Evaluation?,* Retrieved 2016-05-03, from http://wilderdom.com/research/effectsizes.html

Nygård Larsson, P. (2011*). Biologiämnets texter: Text, språk och lärande i en språklig heterogen gymnasieklass*. Malmö: Lärarutbildningen, Malmö Högskola.

OECD (2016). *PISA 2015 Results. (Volume I): Excellence and Equity in Education*, PISA, Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264266490-en

OECD (2009). *PISA 2006 Technical Repor*t. Paris: OECD.

Pedhauzer, E. J., & Pedhazur Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach*. New York: Psychology Press.

Persson, T. (2016*). De naturvetenskapliga språken. De naturvetenskapliga uppgifterna i och elevers resultat från TIMSS 20011 år 8*. Uppsala: Acta Universitatis Upsaliensis.

Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods*, 1, 331-340.

Schleppegrell, M. J. (2004). *The language of schooling; a functional linguistics perspective*. London: Lawerence Erlbaum Associates.

Serder, M. (2015). *Möten med PISA: Kunskapsmätning som samspel mellan elever och provuppgifter i och om naturvetenskap*. Malmö: Studies in educational sciences. No 75.

Serder, M. & Jakobsson, A. (2015). "Why bother so incredibly much?": student perspectives on PISA science assignments. *Cultural Studies of Science Education, 10*(3), 833–853.

Skolverket (2016). *PISA 2015 – 15-åringars kunskaper i naturvetenskap, läsförståelse och matematik*. Stockholm: Skolverket.

Vygotskij, L. S. (1978). *Mind in Society. The Development of Higher Psychological Process*. Cambridge: Harvard University Press.

Wellington, J. & Osborne, J. (2001). *Language and literacy in science education*. Buckingham: Open University Press.

Wernersson, I. (2006). *Genusperspektiv på pedagogik,* Stockholm: Högskoleverket.