

Linguistic properties of PISA mathematics tasks in different languages

Ewa Bergqvist, Frithjof Theens, and Magnus Österholm

Umeå Mathematics Education Research Centre, Umeå University

Introduction and background

The results of international comparative studies are discussed broadly in media and are often used to influence the public opinion. To solve PISA tasks, students have to read and understand the task text. The mathematical tasks are primarily supposed to measure mathematical ability and not reading ability, so it is important to avoid unnecessary demands of reading ability. Since different natural languages have different inherent properties, the “same” task in different languages might vary in difficulty. Many commonly used readability formulas are using both *word length* and *sentence length* as indicators of a text’s difficulty (Lenzner, 2014). Earlier studies have shown that word length can be connected to difficulties when reading and solving mathematical tasks (e.g., Österholm & Bergqvist, 2012). Word length can be measured in different ways, usually through the number of syllables, which addresses the decoding of sounds, or the number of letters, which addresses the decoding of the visual representation. Sentence length is in most studies measured by words per sentence (e.g., Lenzner, 2014). Both word length and sentence length often differ between different natural languages. For example, three rather short words in one language (e.g., “the bus station”) can in another language be written as one long word (Swe. “busstationen”). This type of long compounds is common in both Swedish and German. In addition, sentence length gets directly influenced by the different ways of building compound words. The research question that this study will aim to answer is: What are the differences and similarities between English, German, and Swedish mathematics tasks in PISA regarding word length and sentence length?

Method

In this study we analyze word length and sentence length in all mathematics PISA tasks from 2000-2013, in total 146 tasks, in English (the USA version), German, and Swedish. Word length is measured in two ways, as number of letters per word (physical length) and as number of syllables per word (length of pronunciation). We only include such words that are encoding sounds (i.e., built up by phonemes), and exclude numerals or other types of symbols. Sentence length is measured as number of words per sentence. We only include complete sentences and the same type of words as when measuring word length. Our unit of

analysis is the task, since a focus of PISA is to create comparable tasks when translating them and since we plan for further analyses where we want to relate variation in word length and sentence length to aspects of students' reading and solving each task. Several tasks often share a common leading text, that is, a common background. For each task we include this leading text as part of the task text, since the information in the leading text can be needed when solving each task. For each task we then create three measures: two of *mean* word length (letters and syllables), and one of *mean* sentence length. In order to analyze if there are any differences between the three language versions of the tasks, we use a total of nine t-tests when we for each of the three measures compare the three languages pairwise. We use the Holm-Bonferroni method in order to counteract the problem of doing multiple statistical significance analyses. For differences that are statistically significant we calculate the effect size using Cohen's d.

Results and discussion

All nine t-tests show statistically significant differences between the languages, at significance level $p=0.05$. For eight of the differences, all except the difference in mean sentence length between German and Swedish, the t-tests show statistically significant differences also at level $p=0.01$. The results are:

- Word length (both measures): German > Swedish > English. These differences correspond to very large effects ($d=1.08-2.30$), except between German and Swedish for syllables where the effect is small/medium ($d=0.37$).
- Sentence length: English > German > Swedish. These differences correspond to small/medium effects ($d=0.31-0.42$), except between German and Swedish where the effect is very small ($d=0.14$).

We conclude that there are clear differences between all three languages regarding word length and sentence length, but with largest differences concerning word length. The main difference is between English and the other two languages. This is reasonable, since German and Swedish tend to have compounds, which makes the words longer and the sentences shorter. Does this mean that the reading of PISA tasks is more difficult for students in different countries or could the students be used to these linguistic properties so that there is no effect on the reading difficulty? We will continue our analyses in order to examine if and how these differences are related to the reading and solving of the tasks.

References

- Lenzner, T. (2014). Are readability formulas valid tools for assessing survey question difficulty? *Sociological Methods & Research*, 43(4), 677-698.
- Österholm, M., & Bergqvist, E. (2012). What mathematical task properties can cause an unnecessary demand of reading ability? In G. H. Gunnarsdóttir, et al. (Eds.), *Proceedings of Norma 11, The Sixth Nordic Conference on Mathematics Education in Reykjavík, May 11-14, 2011* (pp. 661-670). Reykjavík: University of Iceland Press.