



**HAL**  
open science

## Effects of linguistic variations of word problems on the achievement in high stakes tests

Sabine Schlager, Jana Kaulvers, Andreas Büchter

► **To cite this version:**

Sabine Schlager, Jana Kaulvers, Andreas Büchter. Effects of linguistic variations of word problems on the achievement in high stakes tests. CERME 10, Feb 2017, Dublin, Ireland. hal-01937143

**HAL Id: hal-01937143**

**<https://hal.science/hal-01937143>**

Submitted on 27 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Effects of linguistic variations of word problems on the achievement in high stakes tests

Sabine Schlager, Jana Kaulvers and Andreas Büchter

University of Duisburg-Essen, Germany

[sabine.schlager@uni-due.de](mailto:sabine.schlager@uni-due.de); [jana.kaulvers@uni-due.de](mailto:jana.kaulvers@uni-due.de); [andreas.buechter@uni-due.de](mailto:andreas.buechter@uni-due.de)

*In Germany, as in other countries, language proficiency impacts the achievement in mathematics. Linguistic features may constitute possible obstacles for students solving word problems. This study explores the interplay of language proficiency and achievement in mathematics with tasks in which linguistic characteristics were varied experimentally. 640 10<sup>th</sup> grade students solved the tasks. An analysis of the shift in difficulty of linguistically varied tasks indicates an irregular scheme: Only tasks with extreme variation of linguistic aspects, for example many nominalizations, show a significant shift. Data collected from student interviews using the think-aloud method show that linguistic aspects can be obstacles and therefore have an impact on the solution process, even if these linguistic aspects may not have an impact on the results in the quantitative part of the study.*

*Keywords: Language proficiency, difficulty level, item analysis, language obstacles, mathematics achievement.*

## Language and mathematics

### Impact of language proficiency on mathematics achievement

Research on the relationship between language and mathematics has a long tradition in Germany and internationally (Barwell et al., 2016; Maier & Schweiger, 1999). In Germany, research on differences in achievement in mathematics due to background factors is relatively new when compared to research in Anglophone countries. Internationally, the strong influence of language proficiency on mathematics achievement, especially for word-problems and tests that follow the literacy approach, has been an issue for decades (Abedi, 2006). International surveys, such as PISA, show that there is a strong connection between mathematics achievement and family background in Germany, especially regarding the socio-economic status and the first language (Gebhardt, Rauch, Mang, Sälzer, & Stanat, 2013). In a recent German study, Prediger and colleagues have identified that “language proficiency is the background factor with the strongest connection to mathematics achievement, among all social and linguistic background factors” (Prediger, Wilhelm, Büchter, Gürsoy, & Benholz, 2015, p. 77; see also Prediger, Wilhelm, Büchter, Gürsoy, & Benholz, 2013).

### Linguistic aspects causing difficulty in mathematics

For English, Abedi and Lord (2001) showed that word problems with low linguistic complexity can reduce the achievement gap evoked by differences in students’ language proficiency. However, Martiniello (2008) reports that research about effects of linguistic features is inconsistent. For example, items with many polysemous words, pronouns or prepositions are difficult only in some grades and the number of subordinate clauses and passive-voice sentences had no significant effects. In her own study, she highlights complex sentence structures with embedded adverbial and relative clauses, long noun phrases and limited syntactic transparency on the syntactic level and

unknown or polysemous words on a vocabulary level as linguistic features creating difficulties (Martiniello, 2008).

For the German language context, there are only a few studies about linguistic obstacles. For example, noun phrases and the number of academic language words seem to be difficult. The identified obstacles of word problems in the study by Prediger et al. (2015) included prepositions, complex syntax and nominalizations. The obstacles did not only occur in the context of reading, but more often they were related to conceptual understanding. In this study the specific linguistic features which made the text complicated could not be isolated as they interacted with each other as well as with the mathematical content. With the experimental approach in our study, we strived to address this issue. For the specific context of the German language, we wanted to specify to what extent linguistic aspects could explain the achievement gap between students of higher and lower language proficiency by investigating the questions:

Q1. What impact do different linguistic aspects have on the achievement in literacy-based tests?

Q2. Which linguistic aspects are difficult for (lower language proficient) students?

## **Research design and methods for the mixed-methods study**

### **Context of the study**

The mathematical tasks of the study were oriented towards the literacy-based high stakes exam “Zentrale Prüfungen am Ende der Klasse 10 (ZP10)” [central examinations at the end of 10<sup>th</sup> grade] of North-Rhine-Westphalia (NRW), the most populous federal state of Germany. The ZP10-exam is designed by QUA-LiS NRW, the “Qualitäts- und UnterstützungsAgentur – Landesinstitut für Schule NRW“ [Quality and support agency – Institute for school NRW] that supports the Ministry of Education and lifelong learning in NRW. The exam does not only assess mathematical competencies acquired in 10<sup>th</sup> grade, but also the competencies acquired during grade 5 to 10. For that reason, our tasks are required to illustrate different mathematical levels.

### **Design for the quantitative part of the study**

The sample consists of 640 students from ten different comprehensive schools in the metropolitan area of Rhine-Ruhr. As independent variables, we collected information about students’ language proficiency, cognitive capabilities and family background (cf. Table 1). Another important variable was the linguistic variation in the tasks. The dependent variable was mathematics achievement.

Language proficiency was measured using a standardized C-Test. Cognitive capabilities were measured by the extended standardized German adaption CFT-20R of culture fair intelligence tests. Information about family background (socio-economic status, immigrant status, languages spoken at home) was collected by questionnaire.

The mathematics test was designed referring to the ZP10-tests. Members of QUA-LiS NRW assured that our test would be acceptable for the high-stakes exam. The 90 minutes long test contained the typical topics of ZP10 – functions, descriptive statistics and percentage calculation – which appear in the exam every year. These tasks were also chosen since the study of Prediger et al. (2015) revealed many language obstacles in tasks about these topics. These observations are supported by results of Martiniello (2008) who identified data analysis, statistics and probability as topics that were difficult for English learners with Spanish as native language in the US.

The test was presented in three versions (A, B and C) with six identical anchor tasks (21 items) and six tasks with linguistic variation (13 items). The tasks with variation always existed in three different versions; the context and mathematical content remained unchanged at all times. In the initial version, we avoided difficult linguistic structures in the tasks, especially those structures that were varied in the other versions. Besides this initial version, the tasks were varied using two of the four linguistic features context vocabulary, verbs followed by prepositions or separable verbs, nominalization and density of the text, and reference structure, which were chosen out of possible linguistic obstacles in word problems reported in Prediger et al.'s study (2015).

On the word level, we chose the criterion “context vocabulary”, which often causes problems but – in contrast to technical terms – can be changed without having influence on the mathematical content. Concerning the syntactical level, interviews in the study of Prediger et al. (2015) confirmed the linguistic supposition that it is difficult for students to connect information given in sentences with separated separable verbs or between a verb and its preposition. Since Uessler, Runge and Redder (2013) also state that separable verbs influence the understanding negatively, we selected these two features as a variation-category. As mathematical texts are often very dense and their length and density play an important role in student's understanding, the feature “density of the text” was chosen. This goes hand in hand with the feature “nominalization”, because a dense text implies (in German) the use of many nominalizations which have a negative effect on the understanding (Uessler et al., 2013). Therefore, these two features form one variation-category on the text-level. As mathematical texts try to avoid repetitions, the reference structure is often unclear. This led to consideration of “reference structure” as another category on the text-level. An example of a varied task will be presented later.

The groups of students sitting for the tests A, B or C were systematically formed based on the results from the C-Test and the CFT-20R. The variations were dispersed equally to the three versions of the test. There were no statistically significant differences between the groups concerning language proficiency, cognitive capabilities and mathematics achievement.

For data analysis, different statistical analysis procedures were applied. We split the group in half depending on students' results on the C-Test into groups of students with lower (C0) and higher (C1) language proficiency. The mathematical items were scaled using a Rasch model. An analysis of variance (univariate ANOVA) and a regression analysis were used to identify the background factors with the highest impact on mathematics achievement and to determine the explained variance. The shift in difficulty from the initial version of a word problem to its linguistic variation was determined by an analysis of the change of the WLE for these items on the Rasch scale.

### **Design for the qualitative part of the study**

The purpose of this part of the study was to gain a deeper understanding of the quantitative results through a qualitative approach. For this reason, four tasks were presented to N=32 students with different levels of language proficiency from different comprehensive schools. The students were required to solve the designed tasks independently using the think-aloud method. Once the students solved the task, there was a discussion about the task between interviewer and student. All processes and discussions were videotaped, transcribed, and analyzed interpretatively with respect to whether the linguistic variations created difficulties for the students.

## Examples of linguistic variation of tasks

An example of the linguistic variation of a particular task is represented in the translated initial version (1) and its variation “nominalization/dense structure” (2) of the task “Bathtub”.

(1) A bathtub has one cold water tap and one hot water tap. The bathtub can be filled with 135 liters of water. If both water taps are open, it takes 9 minutes until the bathtub is filled completely. If only the cold water tap is open, it takes 7.5 minutes more than with both water taps open. How much water runs out of the opened cold water tap per minute? Note your calculations.

(2) A bathtub **with one cold and one hot** water tap can be filled with 135 liters of water. **Opening** both water taps, the **filling** of the bathtub takes 9 minutes; **exclusively opening the** cold water tap, it takes 7.5 minutes more than **by opening** both water taps. Which **amount of water** runs out of the opened cold water tap per minute? Note your calculations.

The bolded words or phrases show the variations in the task for this article. The translation of the tasks can only give an idea of the linguistic aspects that have been changed during variation, because most characteristics are inherent to the German language and sentence structure. For example, if you try to nominalize “the cold water tap is open [der Kaltwasserhahn ist geöffnet]” into “**opening the** cold water tap [Öffnung des Kaltwasserhahns]”, in German, it implies the genitive case of “the cold water tap” tagged by an additive “s” at the end of the word [“Kaltwasserhahns”].

The variation of the items was not possible in every case, because of the structure of the German language. Sometimes it was impossible to use alternative formulations using the defined categories and employing them at positions in the text that are significant for the mathematical solution. The variations were limited due to the common use of language, the context and the linguistic realization of mathematical concepts. In addition, we had to accept the fact that variation of isolated linguistic aspects is almost impossible. In the German version of “If only the cold water tap is open [Wenn nur der Kaltwasserhahn geöffnet ist]”, we find the participle II of the verb “open [geöffnet]” (in the English sentence it has the function as an adjective) because of the use of passive voice “is open [ist geöffnet]”. This has to be nominalized. In the nominalized version, it is linguistically not possible to continue to use the familiar adverb “only [nur]” but necessary to use the less frequently used adjective “exclusively [ausschließlich]”, which imposes another possible lexical obstacle.

## Selected results

### Impact of background factors

To check if linguistic features can explain the achievement gap due to different levels of language proficiency, we first analyzed the impact of background factors in our study. The results provide additional support for evidence of the impact of language proficiency on mathematics achievement. Students with higher intelligence or higher language proficiency, without immigrant status, and students who only speak German at home had statistically significant better results in the mathematics test (see Table 1, the number of the test persons varies because not all students gave the questioned information).

Regression analysis shows that language proficiency and cognitive capabilities as isolated background factors have the highest impact on mathematics achievement. They both explain about

15 % of the variance. The other background factors explain much less: Univariate variance analysis showed only low explanation potential of immigrant status (12 %), languages spoken at home (8 %) and socio-economic status (1 %). As language proficiency has such a high impact in our study, we investigated if it is possible to explain the achievement gap by the introduced linguistic obstacles.

| Background factor                               | Specification of groups  | Distribution of groups                              | Mean score (WLE), m(SD)                                      | Significant differences   |
|---|--|---|--|---|
| <b>Students</b>                                 | 10 schools   | n=640   | -1.37 (1,16)   | -   |
| <b>Version of mathematics test</b>              | version A<br>version B<br>version C  | 219 (34.2 %)<br>214 (33.4 %)<br>207 (32.3 %)        | -1.28 (1,16)<br>-1.49 (1,14)<br>-1.34 (1,16)                 | -   |
| <b>Socioeconomic status (SES)</b><br>(n=626)    | low SES<br>medium SES<br>high SES  | 187 (29.9%)<br>174 (27.8%)<br>265 (42.3%)           | -1.52 (1,12)<br>-1.34 (1,17)<br>-1.24 (1,16)                 | high&low: 0.046   |
| <b>Cognitive capabilities (CFT-20R)</b> (n=577) | lower scores<br>higher scores  | 289 (50.1%)<br>288 (49.9%)                          | -1.68 (1,13)<br>-1.00 (1,09)                                 | < 0.001   |
| <b>Immigrant status</b><br>(n=568)              | 1 <sup>st</sup> generation<br>2 <sup>nd</sup> generation<br>3 <sup>rd</sup> generation<br>no | 42 (7.4%)<br>287(50.5%)<br>56 (9.9%)<br>183 (32.2%) | -1.64 (1,12)<br>-1.72 (1,07)<br>-1.17 (1,09)<br>-0.83 (1,09) | 2 <sup>nd</sup> & 3 <sup>rd</sup> : 0.007<br>no & 1 <sup>st</sup> : <0.001<br>no & 2 <sup>nd</sup> : <0.001 |
| <b>Languages spoken at home</b> (n=616)         | 1: German + x<br>2: no German<br>3: only German  | 269 (43.7%)<br>106 (17.2%)<br>241 (39.1%)           | -1.58 (1,17)<br>-1.75 (0,95)<br>-0.94 (1,08)                 | 2 & 3: < 0.001<br>3 & 1: < 0.001  |
| <b>Language proficiency</b><br>(C-test, n=578)  | low proficient<br>high proficient<br><i>all C-tests</i>                                      | 289 (50.0 %)<br>289 (50.0 %)<br>578 (100 %)         | -1.76 (1,08)<br>-0.94 (1,11)<br>-1.35 (1,17)                 | < 0.001   |

**Table 1: Distribution and differences among groups**

### Shifts in difficulty due to linguistic variation

All results for the different variations have been scaled in a Rasch model. Shifts in difficulty due to the linguistic variation were identified especially in Item 6 “Bathtub”. In the initial version, 69 % of the C1-students and 48 % of the C0-students solved the item correctly. This shows that the initial version was easy for C1-students and common for C0-students. In the nominalized version with dense structure, half of the C1-students solved the item correctly compared to only 37 % of the C0-students. The varied version was common for C1-students and difficult for C0-students.

Figure 1 depicts the shift in difficulty for the variations in comparison to the initial version. In comparison to the initial version, the denser text containing nominalizations became 0.6 WLE more difficult on the Rasch scale, comparable to three out of 34 correctly solved items in our study. Other items of this variation had no significant shift in difficulty (see Figure 1, left side). Students who only did tasks formulated like Item 2b would solve one item less correctly (0.2 WLE = 1 item); students who only had tasks formulated like Item 2a or 8a would solve the test basically the same. One explanation for the increasing difficulty of “Bathtub” is that this item is linguistically more difficult than others with the same linguistic variation, as the Flesch-Reading-Ease index shows.

The variation “reference structure” equally evoked a shift of 0.6 WLE for Item 6 “Bathtub”. Other items with variation showed inconsistent shifts. Some items became more difficult, other items less

difficult. A possible explanation is the higher relevance of mathematical content in contrast to language in particular items. For example, language has less effect in Item 8a, “What is the median?”. In this case it is more important that the concept of median has been taught in mathematics classroom. Similar irregular shifts are visible for the other variations.

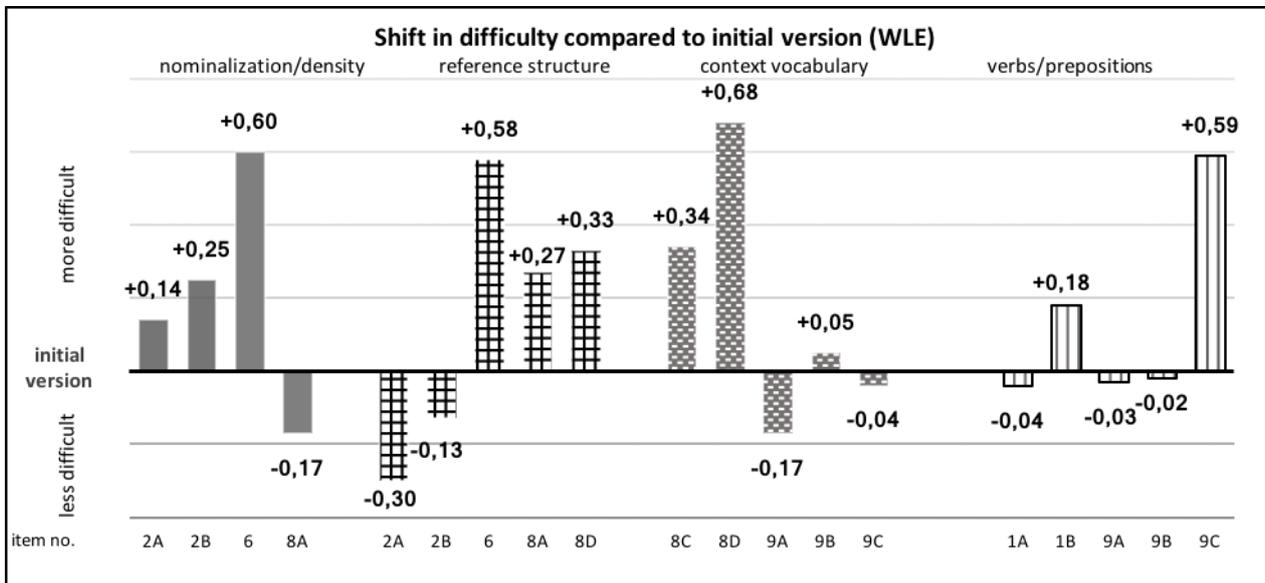


Figure 1: Shift in difficulty compared to initial version (in WLE)

In summary, linguistic variations make word problems more difficult, when applied often and when the general difficulty for understanding the word problem (cf. Flesch-index) increases greatly because of the application of these variations. In these cases, we talk about “extreme variations”. Slight variations with only few modifications, which do not increase the difficulty for understanding, do not have statistically significant shifts. As the effect fluctuates substantially between different items, we can presume that the topic of a task is very important.

### Linguistic variations as source for mathematical difficulty

The fact that the difficulty does not increase significantly for all items (with linguistic variation) does not imply that linguistic variation does not evoke obstacles for students, as our qualitative analysis highlights. Furthermore, we have observed that linguistic aspects do not create difficulty on their own, but in combination with other characteristics of the task as illustrated below.

In addition to the quantitative data, 16 tenth grade students solved the item “Bathtub” in the version “nominalization/dense structure” during the interviews. In particular, students with lower language proficiency had difficulties in understanding the word “exclusively [ausschließlich]”, which had to be used due to the nominalization. While reading the text aloud, several students were puzzled by this word and explained they would never use it. This problem has already caused wrong solutions in the tests. One student wrote: “I do not understand the question. Does ‘exclusively opening the cold water tap’ mean that only the cold water tap is used?” He knew the correct meaning of “exclusively” but he was not sure about it. Probably, this was the reason he did not solve this item. The nominalization “Opening [Öffnung]” of “to open [öffnen]” and the following genitive case of “the cold water tap [der Kaltwasserhahn]” also caused understanding problems, as the genitive case of “cold water tap [Kaltwasserhahn]” finishes with an ‘s’ “[Kaltwasserhahns]” which in German is sometimes also a cue for plural of a word. In this case, the plural would be [Kaltwasserhähne].

Student: I was confused by the word “Kaltwasserhahns”. Does it mean two water taps or only one?

Apart from the linguistic variation, we identified other difficulties in our interviews. In particular, students with lower language proficiency had problems in understanding the relation “more than” and made wrong calculations. One student thinking aloud said (after three minutes of task solving):

Student: Ah, now I made a wrong calculation [...] I thought that ... the filling takes 7.5 minutes, but it takes 7.5 minutes MORE. I have to do a new calculation.

In the final discussion, many students could identify the relevant information “more than” and “exclusively opening the cold water tap” on their own or after a question from the interviewer. This shows that, even if there are difficulties in understanding the text, most of the students could solve the task if they had enough time and could verbalize their thoughts aloud. In contrast, in the tests, students often stopped solving a task because of uncertainty concerning their understanding.

### **Conclusion and consequences**

Our outcome concerning Q1 is that linguistic variations had significant impact when the task varied in an extreme way. The fact that in German isolated variations have no overall significant impact evokes the hypothesis that they create difficulty in combination with each other or with other characteristics of the tasks. This hypothesis is supported by the qualitative results showing that not only the linguistic feature itself, such as nominalizations, but also the linguistic structures evoked by this feature, such as genitive cases, create difficulty. However, this has to be analyzed further.

In the context of Q2, in German, nominalization/dense structure, academic words such as “exclusively [ausschließlich]” or genitive cases are obstacles for several students, especially those of lower language proficiency. This was, for example, prominent in the results of lower language proficient students for the item “Bathtub”. Students struggled in understanding these structures, which sometimes lead to not solving the task. When students had a lot of time and could speak out their thoughts, they reflected on these difficulties.

Consequences of this study for test construction are that high stakes tests should try to avoid extremely difficult linguistic structures, e.g. due to the frequent use of nominalizations, to reduce the achievement gap. Our study also showed that linguistic difficulties cannot always be avoided because there are constraints in the language. A consequence for mathematics classrooms that has emerged from students’ ability to overcome linguistic difficulties by talking about them in our interviews, would be to explicitly address language issues and not to avoid linguistic difficulties.

The fact that our results pertaining to the effects of linguistic variations cannot explain the achievement gap, entails further and deeper research on the interplay between language proficiency and mathematics achievement in order to determine the nature of this correlation. A hypothesis that came up in our study and will be investigated is that the selection of strategies by students and their processes of solving tasks differ according to their language proficiency.

## Acknowledgment

We would like to thank Claudia Benholz, who was part of the project team as a linguistic expert, Susanne Prediger for the discussion and advice, and QUA-LiS NRW for funding this study.

## References

- Abedi, J. (2006). Language issues in item-development. In S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 377–398). Mahwah, NJ: Erlbaum.
- Barwell, R., Clarkson, P., Halai, A., Kazima, M., Moschkovich, J., Planas, N., ... Ubillús, M. V. (Eds.) (2016). *Mathematics education and language diversity: The 21st ICMI Study*. New York: Springer.
- Gebhardt, M., Rauch, D., Mang, J., Sälzer, C., Stanat, P. (2013). Mathematische Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund. In M. Prenzel et al. (Eds.), *PISA 2012: Fortschritte und Herausforderungen in Deutschland* (pp. 275–308). Muenster: Waxmann.
- Maier, H., & Schweiger, F. (1999). *Mathematik und Sprache: Zum Verstehen und Verwenden von Fachsprache im Mathematikunterricht*. Vienna: ÖBV & HPT.
- Prediger, S., Renk, N., Büchter, A., Gürsoy, E., & Benholz, C. (2013). Family background or language disadvantages? Factors for underachievement in high stakes tests. In A. M. Lindmeier, & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 49–56). Kiel, Germany: PME.
- Prediger, S., Wilhelm, N., Büchter, A., Gürsoy, E., & Benholz, C. (2015). Sprachkompetenz und Mathematikleistung–Empirische Untersuchung sprachlich bedingter Hürden in den Zentralen Prüfungen 10. *Journal für Mathematik-Didaktik*, 36(1), 77–104.
- Uessler, S., Runge, A., & Redder, A. (2013). “Bildungssprache” diagnostizieren. Entwicklung eines Instruments zur Erfassung von bildungssprachlichen Fähigkeiten bei Viert- und Fünftklässlern. In A. Redder, & S. Weinert (Eds.), *Sprachförderung und Sprachdiagnostik. Interdisziplinäre Perspektiven* (pp. 42–67). Muenster: Waxmann.