

Linguistic Features as Possible Sources for Inequivalence of Mathematics PISA Tasks

Frithjof Theens, Ewa Bergqvist, and Magnus Österholm

Umeå Mathematics Education Research Centre, Umeå University

Introduction and background

In the last decades we have seen a rising of different international student assessments such as TIMSS, PIRLS and PISA. The results of these assessments are discussed in media and have impact on political decisions and educational systems. But as the tasks have to be translated to many different languages, there is a risk that the different language versions are not equivalent.

To be able to solve a written mathematics task, students have to read and understand the task text, that is, they need some reading ability. This demand of reading ability might differ between different language versions of a task and by that be a source of differential item functioning (DIF) of the task. DIF occurs when students with the same underlying mathematical ability but from different language groups have different probability to answer the task correctly. Earlier studies used statistical methods to detect DIF between different language versions of items and expert reviews to identify the sources of DIF. By this, amongst others, linguistic issues were identified as possible sources of DIF in multilanguage assessments (e.g., Roth, Oliveri, Sandilands, Lyons-Thomas, & Ercikan, 2013). To be able to avoid DIF caused by linguistic issues it is necessary to know more explicitly which linguistic features can be a source of DIF. In this study, we investigated whether differences in some linguistic features are related to DIF between the English, German, and Swedish version of mathematics tasks of the PISA 2012 assessment. The linguistic features chosen in this study have shown connections to the difficulty of solving mathematics tasks in earlier studies (e.g., Abedi, Lord, & Plummer, 1997) and are:

- Grammatical person, that is, if the text is written in first, second, or third person.
- Voice, that is, if active or passive voice is used in the text.
- Sentence structure, that is, how the sentences are built of main and subordinate clauses.

The research question in this study is: Are differences in these linguistic features between the different language versions of PISA tasks connected to DIF?

Method

We analyzed the English (USA), German, and Swedish language versions of 83 mathematics tasks of the PISA 2012 assessment in three steps. First, we searched for differences in the three linguistic features between the tasks and calculated the

amount of differences. Then, we used the PISA item results of about 5000 students in each language group to calculate DIF, using the Mantel-Haenszel procedure pair-wise for two language versions at a time. This is a common method to detect DIF and to determine its degree as negligible, intermediate, or large (Dorans & Holland, 1992). The last step was to search for correlations between the amount of linguistic differences between the language versions of the tasks and DIF between the versions.

Results and discussion

The first step of the analysis showed that differences in linguistic features occurred between the language versions. Most frequent were differences in the use of active and passive voice. The least differences were in grammatical person. The DIF analysis revealed several items with both intermediate and large level of DIF. The third step of the analysis then revealed no statistical significant correlations between the differences in linguistic features and DIF between the language versions.

Since the results show that DIF occurs but is not correlated to differences in the three linguistic features included in this study, there must be other sources for DIF. This might be other linguistic issues, cultural differences, or other reasons.

To further search for possible sources of the DIF revealed between different language versions of mathematics tasks of PISA 2012, we will do a study using think aloud protocols (TAP) of students working with these tasks. We plan to let a number of students try to solve the tasks that showed large DIF with disadvantage for the Swedish version compared with the English version (5 tasks) or the German version (1 task). By analyzing the TAPs we can either find possible linguistic sources of DIF or exclude linguistic issues for these tasks. Of course, it would be desirable to also let English and German speaking students work with these tasks to be able to compare their perceived difficulties with the ones perceived by the Swedish students. Due to limitations in this study, we plan to do this in a later study.

References

- Abedi, J., Lord, C., & Plummer, J. R. (1997). *Final report of language background as a variable in NAEP mathematics performance (CSE Report 429)*. Los Angeles, CA: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Dorans, N. J., & Holland, P. W. (1992). DIF Detection and Description: Mantel-Haenszel and Standardization. *ETS Research Report Series, 1992(1)*, i-40.
- Roth, W.-M., Oliveri, M. E., Sandilands, D. D., Lyons-Thomas, J., & Ercikan, K. (2013). Investigating Linguistic Sources of Differential Item Functioning Using Expert Think-Aloud Protocols in Science Achievement Tests. *International Journal of Science Education, 35(4)*, 546-576.