

# Sources of inequivalence in translated mathematics tasks identified with students' reflections

**Frithjof Theens**

Department of Education, Umeå University

*In multilanguage assessments, the validity of the results is threatened if the different language versions are not equivalent. In this study, task-based interviews with German and Swedish students were analyzed to identify possible sources of inequivalence between the language versions of mathematics PISA tasks.*

## **Introduction and background**

Multilanguage mathematics assessments are used in different environments for different purposes, for example, in international student assessments like PISA or TIMSS. Lack of equivalence between the different language versions of the tasks threatens the validity of the results of multilanguage assessments. For example, linguistic, curricular, and cultural differences have been identified as possible sources of inequivalence in earlier research.

To find sources of inequivalence in translated tasks, firstly, it has to be determined whether the tasks' language versions are equivalent or not. A common way to identify tasks that are not equivalent is to investigate whether a task displays Differential Item Functioning (DIF). A task displays DIF if participants with the same level of underlying ability that is supposed to be assessed, but belonging to different subgroups of participants, have different probability to answer the task correctly (Dorans & Holland, 1992). That is, in the case of multilanguage mathematics assessments, a task displays DIF if students with the same level of mathematical ability but belonging to different language groups have different probability to solve the task.

The purpose of this study is to deepen the understanding about issues that may be sources of inequivalence in mathematics tasks in multilanguage assessments. I intend to identify possible sources of DIF in different language versions of tasks. The research question in this study is: What possible sources of DIF between German and Swedish versions of PISA mathematics tasks can be found using students' reflections?

## **Method**

To answer the research question, students from Germany and Sweden were asked to read and think aloud while working with some mathematics tasks of the PISA 2012 assessment in German and Swedish, respectively. In these tasks, DIF

between the language versions was found in a previous study (Theens, Bergqvist, & Österholm, 2019). Afterwards, the students were interviewed about their work with the tasks and the difficulties they encountered. Transcriptions of the interviews were analyzed to find similarities and differences between the two language groups that may relate to inequivalence between the language versions of the tasks.

## Results and discussion

The analysis of the interviews revealed several possible *linguistic* sources of inequivalence between the language versions of the tasks, such as differences in the use of uncommon words and complex sentences. A possible non-linguistic source of inequivalence that came to light is the difference in how German and Swedish students dealt with information in the task text, not necessary to solve the task. The German students mentioned this information more often as disturbing or time consuming. This difference may occur due to *curricular* reasons, such as the type of tasks the students are used to work with. However, the analysis of the interviews showed that *cultural* differences between the German and Swedish students do not seem to cause inequivalence in these tasks.

Another interesting result is, that differences in the use of active and passive voice between the language versions do *not* seem to be a source of inequivalence for the tasks investigated in this study. The use of passive voice is often seen as a linguistic issue making texts more difficult to read and understand. For example, Abedi, Lord, and Plummer (1997) showed that eliminating passive voice from mathematics tasks resulted in higher scores for students. Also, in the PISA translation guide, it is recommended not to change voice when translating the PISA tasks (OECD, 2010), but in some tasks the translators did make changes anyway. This recommendation is challenged by the result of this study.

The results of this study can help to enhance equivalence between different language versions of mathematics tasks. When translating tasks, focus can be put on issues that were identified as possible sources of DIF instead of on features that do not seem to have impact on equivalence, such as changes in voice.

## 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.
- OECD. (2010). Translation and Adaption Guidelines for PISA 2012. In. Budapest, Hungary: OECD.
- Theens, F., Bergqvist, E., & Österholm, M. (2019). *The relation between linguistic features and DIF in multilanguage mathematics assessments*. Umeå.