Identifying Translation-Related Sources of Differential Item Functioning in International Reading Literacy Assessments

Inga Arffman

Background

International reading literacy assessments pose an extra challenge to validity: The different-language versions of the test have to be equivalent to each other; they must not unfairly favour some language groups over some others.

To ensure equivalence, several quality control procedures are implemented. Significant among these are statistical differential item functioning (DIF) analyses, conducted to detect the items that function differentially across countries. These are followed by judgemental item analyses, where the flagged items are reviewed for bias and those interpreted as biased are removed from the final analyses. In international assessments an important potential source of bias are translation-related differences, or translation errors and translation problems caused by differences between languages.

However, research suggests that making judgemental item analyses and identifying sources of DIF is extremely challenging and that the analyses have often not been able to identify sources of DIF consistently and accurately.

Purpose

The study explored whether it would be possible to improve the ability of judgemental item analyses conducted in international reading literacy assessments to identify translation-related sources of DIF.

Research design

Data: Three items which in the PISA 2000 reading test were identified as DIF (harder or easier than expected) in Finland but which were not interpreted as being biased because of translation-related differences and which were not deleted from the final analyses.

Data analysis: A fresh judgemental item analysis was conducted of the three items. The items were compared linguistically to the corresponding items in the English and French source versions, taking into account the cognitive processes required to answer them.
Results

Several translation-related differences were found which may have affected students’ performance on the items (Table 1).

Table 1. Summary of the translation-related differences found in the three items

<table>
<thead>
<tr>
<th>Item</th>
<th>Task</th>
<th>Harder/Easier in Finland?</th>
<th>Source of DIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>Inferring a problem-solution relationship</td>
<td>Harder</td>
<td>Language-specific differences in grammar</td>
</tr>
<tr>
<td>Item 2</td>
<td>Interpreting nuances: pity, compassion</td>
<td>Harder</td>
<td>Language-specific differences in meaning and connotations</td>
</tr>
<tr>
<td>Item 3</td>
<td>Retrieving conditional information</td>
<td>Easier</td>
<td>Ignorance of translation instructions and unduly free translation Differences in grammar and meaning; differences between the English and French source versions</td>
</tr>
</tbody>
</table>

The results suggest that it is possible to make judgemental item analyses better able to identify translation-related sources of DIF, especially those deriving from differences between languages, but also those resulting from differences between the source versions and ignorance of the translation instructions.

Conclusion

A more rigorous and standardised judgemental item analysis procedure should be developed in international reading literacy assessments which should be followed consistently in all participating countries. Attention should be paid to e.g. the following:

- The Project Centre should develop clear criteria as to what is meant by translation-related differences: not only translation errors but also differences between languages.
  The Project Centre should provide judgemental analysts with instructions and training on the most typical translation-related sources of DIF.
- The Project Centre should provide instructions and training on how to conduct the analysis. The analysis should involve not only a careful examination of the target versions in relation to the reading strategies assessed but also a close inspection of student responses and thorough linguistic comparisons between the target and source versions. In addition, not only the DIF items but also the stimulus texts with which they occur should be compared.
- The Project Centre should develop a rating scale and criteria as to how to estimate the significance of translation-related differences for performance.
- The Project Centre should develop a special sheet by means of which national teams could report their judgement results – the ratings and the justifications – in a clear and unequivocal manner to the Project Centre.
- The Project Centre, when making final decisions on the items, should communicate online with the national teams and native speakers of the languages.
Judgemental analysts should have not only knowledge of reading strategies, response strategies and factors affecting item difficulty, but also linguistic and translation expertise. Two analysts may be needed.

Measures to develop other quality control procedures:

- Item generation: Collecting data on the typical translation-related sources of DIF and writing them into item generation guidelines; involving translators in item generation; translating the items into other languages.
- A more extensive use of translation notes to alert translators to translation problems.
- Ensuring the equivalence of the two different-language source versions.

Contact information:
Inga Arffman
Finnish Institute for Educational Research
P.O. Box 35
FI–40014 University of Jyväskylä
Finland
E-mail: inga.arffman@jyu.fi

The study has been supported by the Academy of Finland (Grant No. 126855).