Detecting Sources of DIF in Translated Verbal Items

Avi Allalouf

National Institute for Testing and Evaluation, Israel

Stephen G. Sireci

University of Massachusetts, Amherst

Abstract

Translated tests are being used increasingly in educational testing for assessing the knowledge and skills of individuals who speak different languages. A difficult and under-researched problem related to test translation is the identification of translated items that do not function equivalently across languages. Furthermore, there is little research exploring why translated items sometimes function differently across languages. If the sources of differential item functioning (DIF) across languages could be predicted, these sources could be eliminated at an early stage in the test development process, and this could have important implications for decisions regarding test development process, scoring and equating.

This study focuses on two questions: Is DIF related to item type? What are the sources of DIF? The data that were analyzed consisted of three forms of the Israeli Psychometric Entrance Test (PET) in Hebrew (source) and Russian (translated). The sources of DIF were assessed by analyzing the DIF direction with the aid of translators who judged the items. The results indicated that 42 out of 125 items (34%) functioned differentially across languages. The conclusions are presented by item types: Analogies - have the highest DIF rate (65%), with Russian-speaking examinees performing better on 82% of the analogy items with DIF; Sentence Completions - have relatively high DIF (45%), with no group performing better than the other; Reading Comprehension - has moderate DIF (23%), probably related to the content of the specific passage; Logic - almost no DIF was found (8%). The main sources of DIF were changes in word difficulty, changes in item format, translation errors and differences in cultural relevance.