Abstract

Given that a key function of tests is to serve as evaluation instruments and for decision making in the fields of psychology and education, the possibility that some of their items may show differential behaviour is a major concern for psychometricians. In recent decades, important progress has been made as regards the efficacy of techniques designed to detect this differential item functioning (DIF). However, the findings are scant when it comes to explaining its causes. The present study addresses this problem from the perspective of multilevel analysis. Starting from a case study in the area of transcultural comparisons, multilevel logistic regression is used: 1) to identify the item characteristics associated with the presence of DIF; 2) to estimate the proportion of variation in the DIF coefficients that is explained by these characteristics; and 3) to evaluate alternative explanations of the DIF by comparing the explanatory power or fit of different sequential models. The comparison of these models confirmed one of the two alternatives (familiarity with the stimulus) and rejected the other (the topic area) as being a cause of differential functioning with respect to the compared groups.