

COMPARISON OF VARIOUS METHODS FOR DETECTING ITEM BIAS

By: Badrun Kartowagiran (2005)

ABSTRACT

The objective of this research study is to conceive: (1) Items of Junior High School mathematics national examination which statistically significantly contain bias (Differential Item Functioning/DIF) if detected by the Item Characteristic Curva (ICC) method; (2) Items of Junior High School mathematics national examination which statistically significantly contain DIF if detected by Raju's Area Measure; (3) Items of Junior High School mathematics national examination which statistically significantly contain DIF if detected by Lord's Chi-square; (4) Items of Junior High School mathematics national examination which statistically significantly contain bias DIF if detected by the likelihood ratio test; and (5) the most sensitive DIF detection method for the Junior High School mathematics national examination.

The data for the study consist of State Junior High School students' responses to the mathematics national examination of the 2003 academic year. Before subjected to the DIF analyses, test items were sorted according to the classical theories using the ITEMAN program package and three-parametre item response model using BILOG program. Good test items were then subjected to the DIF analyses using the characteristic curve method using BILOG, Raju's area measure and Lord's Chi-square using IRT-DIF program, and the likelihood ratio test using the MULTILOG program. The bias items were then plotted for the probability of the female or male group to answer correctly using the Maple program to determine whether the bias in the items was uniform or non-uniform. Two methods to find out the most sensitive DIF model, that are: to count the number of items which contain DIF, and to check the validity and reliability of DIF detection measurement result with confirmatory factor analysis by using the LISREL program package.

Results of the study show: (1) eight test items were detected by the ICC method as statistically significantly contain DIF, namely items 1, 3, 5, 9, 14, 17, 19, and 21; (2) four items were detected as containing DIF by the Raju extended model, namely items 3, 5, 9, and 12; (3) Lord's Chi-square

detected five items with DIF namely items 3, 5, 9, 17, and 20; (4) The likelihood ratio test detected nine items with DIF namely items 1, 3, 5, 9, 14, 17, 19, 20, and 21, (5) of the ten items with DIF, nine items favour male students and one item favour female students; and (6) the most sensitive model for the detection of DIF is the likelihood ratio test, both in terms of the number of items with DIF detected and the results of the confirmatory factor analysis. Accordingly, it is suggested that: (1) readers or interested researchers develop similar studies using different DIF methods or different bias sources; (2) the provincial or regional education offices make use of the results of this study by conducting training in of test item construction and analysis in order to produce test items which do not contain DIF; (3) the centre for educational of the national education department use results of this study as an input in item banking so that no bias items are filed; and (4) national examination bodies use results of this study as an input in determining the model for detecting item biases in order not to be faulty in the choice of the model.