

IDENTIFYING CHARACTERISTIC TYPES OF DIF ITEMS IN THE NATIONAL EDUCATIONAL TESTING OF MATHEMATICS IN THE SLOVAK REPUBLIC

Michal Hajduk *, Martina Pigová **

** Department of Psychology, Comenius University*

*** National Institute for Certified Educational Measurements, Bratislava*

Keywords: national educational testing; item response theory; differential item functioning

The primary aim of the study was not only to identify particular items which seem to be differentially functioning (depending on the gender of the respondent), but also to specify the typical representatives of DIF items in educational tests.

For the purpose of the study we used data collected from two types of national measurements in the Slovak Republic: Matura (school leaving examination) in Mathematics in 2011 and 2012; and Testing 9 (external testing of primary schools ninth-graders) in Mathematics in 2011 and 2012. To provide simplified and more organized data, not all of the data in the datasets were used, only a random sample of 4,000 respondents was selected for each of the four tests.

Both the classical test theory (CTT) and the item response theory (IRT) were employed during analysis. CTT suggested only slightly significant differences in performance between girls and boys in some of the items. The results were similar when using IRT methods (parameter invariance; Mantel-Haenszel statistics), however, in each of the above mentioned tests we identified a larger number of items which appear to be measuring differently for boys or girls (differences were observed even in items which did not show significant differences in CTT).

Taking a closer look at particular differentially functioning items we have noticed several similarities among them. Results suggest that girls outperform boys in the equation- or mathematical-expressions- type of items. On the other hand, boys achieved moderately better results than girls in items which are contextually relevant for them (e.g. sport, computer or cars related topics).

Though a test might generally appear to be equally difficult for girls and boys, the detailed analysis revealed a number of differentially functioning items. Whereas CTT methods show only total differences in performance, employing IRT methods furthermore enable to observe the disparities in performance at different ability levels.

Implications for test designers are being discussed further.