

**A47**

Kaftandjieva Felianka

**Flaw & Order: World 1 and World 2 of Item Difficulty – Synergies and Tensions**

The research findings in the last 50 years reveal that the panelists' perception (World 2) of item difficulty differs from the empirical data (World 1). At the same time, most of the existing test-centered standard setting methods require, in one way or another, judges to make estimations of item difficulty. The degree to which their estimations correspond to empirical results is one of the basic evaluation criteria for the quality of the standard setting.

Even though the importance of intra-judge consistency is well recognized, there are only a few methods for its analysis which are applicable for a limited number of test-centered standard setting methods and usually require IRT approach.

To overcome the deficiency of appropriate methods a new method for the analysis and quantification of intra-judge consistency was developed. The method does not require IRT approach and is applicable to standard settings, where the ratings of items are on the ordinal scale of measurement (like the ratings obtained in the Basket procedure – the prevalent standard setting method for linking language examinations to Council of Europe Framework of Reference).

This paper aims to:

- introduce the new method,
- describe the main features of the Misplacement index (MPI),
- summarize and discuss the results of the analysis of intra-judge consistency for eight different standard settings in reading and listening comprehension in English, German and Swedish language tests, involving more than 500 items and over 100 judges.