Reviewer's report

Title: Mokken Scale Analysis: A Practical Guide to the application and interpretation of a Non-Parametric IRT method in Empirical Research with health and well-being questionnaires

Version: 1 Date: 19 October 2011

Reviewer: Per Bech

Reviewer's report:

Stochl J et al: Mokken Scale Analysis: A Practical Guide to the application and interpretation of a Non-Parametric IRT method in Empirical Research with health and well-being questionnaires

The authors have found the Mokken Analysis to be a less well-known item response theory model than the Rasch model and have therefore found it important via their manuscript to provide the readers of this journal with a practical guide on how to use the Mokken Analysis.

As their first example the authors have illustrated the Mokken Analysis by General Health Questionnaire (GHQ-12 data from a Scottish Health Education Population Survey (SHEPS) including 355 males and 418 females from 16 to 74 years of age.

This example is, however, not appropriate for either Mokken Analysis or Rasch Analysis. The authors consider the GHQ-12 to measure distress or anxiety/depression (page 10 at the top). On page 12, line 13 from the top, the authors seem to be aware of the fact that 6 items in the GHQ-12 with a negative phrasing cover symptoms of anxiety and depression (items 2, 5, 6, 9, 10, 11) whereas the other 6 items in the GHQ-12 with a positive phrasing cover positive, psychological well-being (items 1, 3, 4, 7, 8, 12).

Before making a Mokken analysis we need an evaluation of the clinical validity of such a scale. Hence, the clinical validity of the GHQ-12 is that the scale consists of 2 subscales: an anxiety/depression or distress scale (6 items) and a well-being scale (6 items).

The invariant item ordering of the scale (the inborn principle of the Guttman scales on which the Mokken Analysis is based) is only meaningful when these two GHQ-12 subscales are analysed separately. In the Mokken Analysis we do not as in the Rasch analysis have the ordering of the items by their locations on the latent dimension of either distress or well-being. The discrimination between the items in the Mokken Analysis is the raw scores (means). In this situation the original Likert scoring of GHQ-12 from 1 to 4 is much more informative than the re-coding from 0 to 1 used by the authors. An advantage of the Mokken Analysis is its application for multi-category Likert scales.

The authors need to provide a much more illustrative item ordering of the two
GHQ-12 subscales.

The authors have, unfortunately, used the Mokken Analysis in a way similar to factor analysis, i.e. without a clear hypothesis. The advantage of the Mokken Analysis over factor analysis is that it can test the hypothesis of dimensionality. The advantage of the Rasch Analysis over the Mokken Analysis is that here we can evaluate to what extent the invariant item ordering is found across gender and age.