Author's response to reviews

Title: Mokken Scale Analysis of Health and Well-being Questionnaire item responses: a Non-Parametric IRT Method in Empirical Research for Applied Health Researchers

Authors:

Jan Stochl (js883@cam.ac.uk)
Peter B Jones (pbi21@cam.ac.uk)
Tim J Croudace (tjc39@cam.ac.uk)

Version: 2 Date: 18 November 2011

Author's response to reviews: see over
Dear Editorial office,

Thank you for sending us the referees’ comments on our manuscript. They have allowed us to amend and improve the manuscript that we hope is now acceptable for publication. Our responses are detailed, below:

Referee 1:

1. Rephrase technical terms in the abstract e.g., monotonicity for a non-technical reader.
   We rephrased few technical words in abstract. However, technical terms which are part of the name of the model (such as “double monotonicity model”) are kept in the abstract.

2. The following sentence is noted without first explaining to the reader that the Guttman model is deterministic, what that means and why it’s an unrealistic model- “Mokken models belong to the class of statistical models called non-parametric item response theory (NIRT) and extend the simple Guttman scaling model.”
   Brief explanation of Guttman model is now provided in the introduction.

3. Reference the more established IRT methods- “performance of more well established parametric IRT methods”.
   The sentence now reads “performance of more well established parametric IRT methods such as Rasch models”.

4. Typo - some sense „person free“
   The typo has been corrected.

5. Supply references to paragraphs in the introduction 1-3 that are not referenced.
   Paragraphs 1-3 are now supplied with references.

6. Please briefly detail key differences between the Mokken and Rasch models.
   We believe we have described key differences already in the manuscript, see first page in introduction. We also refer to paper discussing the differences at the end of the same page.

7. To increase readability in the introduction and make the manuscript accessible to a general audience include a summary table in the introduction. Consider including- key aspects of what will be analyzed, terminology, symbols that are relevant and then analyzed.
   We added such summary table in introduction (Table 1).

To this end, see and cite- Levine SZ, Rabinowitz J, Rizopoulos D (2011): Recommendations to improve the Positive and Negative Syndrome Scale (PANSS) based on item response theory. Psychiatry Research 188:446-452.
We have consulted and read this interesting paper and are grateful to the referee for bringing it to our attention. It gave us inspiration regarding how to set out some of our tables from a stylistic and content point of view. However, we do not think it adds to our argument where our analyses concern two completely different scales. Therefore, we have not actually cited this – we have amended Table 1, as above.

Also it may be informative to generate curves to illustrate sigmoid shapes and so on for clarity.

Examples of sigmoid shaped curves are now provided in Figure 1.

8. The hazards of dichotomization aside, does dichotomization of responses change the picture of future development of this measure and perhaps a different (polychotomous) model would have been altogether more adequate? I am not saying the approach adopted is wrong (given the skew) merely pointing out choice of statistic shouldn’t drive the analysis, rather the data being polychotomous responses may have performed differently given a different technique.

We have chosen the binary coding of data for two reasons. First, this coding is traditional (although has been criticized [1, 2]). The other reason was that we wanted to show Mokken analysis for binary scored items. In general, dichotomization of polytomously scored items is usually considered as negative (reducing the variability and therefore information that the items carry). On the other hand it may be necessary if the distribution of item responses is highly skewed (as referee correctly pointed out). We therefore also performed Mokken analysis on original scored data and the results were identical to those obtained for dichotomized data.

9. Appendices please provide meaningful variable labels

Short meaningful labels are now provided in Appendices and also in all tables.

10. Please ensure the Tables are self-standing e.g., Insert the meaning of Pmatrix in Table 4.

We have considerably simplified the majority of the tables and now also provide meaningful labels for better clarity.

11. Why Backward stepwise removal of items?

Backward stepwise removal of items is meaningful, because we check for invariant item ordering initially for the complete set of items. When violations are identified then the corresponding item is removed and the remaining items are checked again. Also, it is the only method currently available in R software package “mokken”.

12. Where is the Discussion header, a paragraph or so on limitations, and implications?

Given the “educational” nature of the manuscript we do not provide standalone discussion but we rather discuss interpretation and related limitations throughout the manuscript. There is also brief section entitled “limitations and practical considerations” where we discuss general limitations of Mokken models.
13. correct citation of R

The citation has been corrected. Please note that we had to take into account the required citation format of the journal.

Referee 2:

The main (and only) concern of this referee is with example 1, where we illustrated Mokken scale analysis for binary scored GHQ-12 data. The referee comments Mokken scaling for this scale as inappropriate because: a) this scale measure 2 distinct dimensions, namely distress and well-being, and b) binary coding is much less informative than original polytomous coding. In essence, we disagree with the referee, otherwise we would not have written the paper in the first place. Our arguments are as follows and are, we believe, clear in the paper and backed up with relevant literature.

Regarding dimensionality of GHQ-12: Many studies tried to reveal dimensional structure of this scale. Although there is evidence supporting referee’s view on 2-dimensional structure [3-5], there is also evidence showing unidimensionality of this scale and therefore supporting our approach. Also there are studies supporting even higher number of dimensions, typically 3 [e.g. 3, 6, 7-10]. In addition there are studies which (despite finding 3 factors underlying GHQ-12) doubt practical usefulness of consideration of this scale as multidimensional [11]. In general, factor structure of GHQ-12 is highly variable from sample to sample and not clearly established [12]. Therefore the sound way how to deal with dimensionality is to reveal the structure from the data. As we mention in the paper, Mokken scaling can be used in both “exploratory” and “confirmatory” manner (and the parallel with factor analysis the referee used is correct). But unlike in factor analysis, we do not agree that in this case that exploratory Mokken (or as referee says “Mokken without clear hypothesis”) is worse than confirmatory. In fact, this exploratory approach is much a core feature of Mokken scale analysis and is traditionally used for building unidimensional scales [13].

Our additional concern with this comment of the reviewer is the substantial interpretation of the dimensions (subscales) of GHQ-12. According to the referee they are anxiety/depression (or distress) and well-being. Although anxiety and depression items are usually identified in factor analytic studies, we are not aware of a single study identifying well-being as subscale of GHQ-12. The reported factors considered are quite consistently entitled as ‘social functioning’ and, in the three factor models, ‘loss of confidence’ (see [3] for an overview).

Beside the evidence for unidimensionality provided in our manuscript, we also crossvalidated dimensionality using exploratory factor analysis and confirmatory factor analysis (or 2-parameter logistic model as they are equivalent for binary data). Although majority of FA studies ignored the categorical character of the data and treat them as continuous, we took ordinal nature of data into account. All analyses supported unidimensionality of GHQ-12 (outputs are attached for the convenience of reviewers). Altogether with other papers supporting unidimensionality of GHQ-12 we believe our analysis is appropriate.
The second concern of referee is with binary scoring of the GHQ-12. We agree with reviewer that Likert scoring (1-2-3-4) is more informative than binary scoring used in our analysis (0-0-1-1). We have acknowledged that in the paper. Two contra-arguments can be used: First, binary scoring is traditional for this scale and it is still widely accepted when GHQ-12 is used as screening instrument. Second, the aim of our manuscript is to provide illustrative example of mokken scaling and its interpretation for binary coded scale responses, the results themselves are, to some extent, of secondary interest.

Referee 3:

1. The title of the manuscript should be changed. It said that it was a guideline. But the manuscript focuses on how to conduct non-parametric IRT based on Mokken method. I do not see any guidelines regarding any particular issues, such as a cut-off point, etc. since we know them already.

We thank the referee for this suggestion and have amended the title to better reflect the contents. This is now: Mokken Scale Analysis of Health and Well-being Questionnaire item responses: a Non-Parametric IRT Method in Empirical Research for Applied Health Researchers

2. The manuscript is too long, and it can be shortened quite a bit. This brings the angle issue of the manuscript. As it stands now, it discusses or describes Mokken model, comparison with parametric IRT models, and assumptions of Mokken model in greater details. It reads more like a textbook rather than an article.

We agree this paper is written as an “educational” piece and, as such, hope it will be popular and well cited. This approach was deliberate – we wanted to attract readers’ attention to the use of nonparametric item response models, explain advantages and disadvantages of this methodology and provide easy to understand tool to perform Mokken analysis with readers own data. We provide a general and brief explanation of the model but discuss also its limitation. We tried to be comprehensive, concise and not too much simplistic. This approach resulted in the manuscript which a bit longer than usually accepted criteria. We believe that the length is well balanced by comprehensiveness. We also believe that given the online format of the journal the length is less important issue.

But more importantly, I do not have a clear understanding of what issues or aspects of psychometric issues are discussed after reading the manuscript:
- Are we talking about assumptions of unidimensionality, monotonicity, etc. or are these aspects of the psychometric analysis we need to address?

Both are correct. These are assumptions (for psychometrical model) that need to be empirically checked. Mokken analysis is indeed all about empirical checking of these assumptions. When they are met, the scale has properties of this model described in introduction (such as ordering of person according to their sumscores or invariant item ordering).

- Is unidimensionality related to factor structure?

Yes, unidimensionality is related to common factor solution in factor analysis. But no factors are present in Mokken scaling per se.
- If we have an instrument with more than one factor, can we still use this method?

Yes and we discuss that in the paper. If more than one dimension is observed then Mokken scaling can be applied to each dimension separately.

4. It would be good if the manuscript could be presented in the following fashion. Of course, it is just a suggestion.
   a. Discuss psychometric issues related to instrument development or evaluation.
   b. Discuss methods are available for such investigation
   c. Focus on Mokken models and why such an approach is advocated.
   d. The aspect of psychometric issues can be investigated using Mokken model. If an instrument has more than one factor, what is the role of Mokken model?
   e. Then use examples to illustrate.

Thank you to the reviewer for these excellent suggestions. However, discussing all these aspects would elongate the paper still further (and length was one of the concerns of this reviewer)

5. There are some points needing clarification.
   - MHM and DMM needs be better explained for their role in psychometric study.

The whole paper is dedicated to explanation of MHM and DMM and their role in psychometrics.

   - Relation to reliability needs to be better explained.

The reliability within the context of Mokken scaling is now described in greater detail. We also provide references to other literature devoted to this issue.

References


