Reviewer’s report

**Title:** A rasch model to test the cross-cultural validity in the positive and negative syndrome scale (PANSS) across six geo-cultural groups

**Version:** 3  **Date:** 22 November 2012

**Reviewer:** Michael Obermeier

**Reviewer’s report:**

The authors report on the results of a Rasch-analysis using PANSS-data of several actors being rated by raters with different cultural backgrounds. The paper is nicely written and deals with an important topic in schizophrenia research: worldwide Schizophrenia research should be comparable and therefore its most popular rating instrument PANSS must be comparably rated over cultural boundaries! Therefore an analysis dealing with cross-cultural rating-differences seems to be necessary.

In the current analysis a Rasch rating scale model has been applied using cultural grouping as independent variable and presenting average item calibrations as results. Accordingly the groups have been compared with respect to differential item functioning. The methods in general are well chosen from a statistical point of view and the conclusions emphasizing the necessity of attending potential cultural bias in training curricula are reasonable.

Only some minor points should be addressed.

1) The article in general is very long and probably to detailed in some points. Especially the discussion is very long and in the results-part more links to the tables (and less text) and there the highlighting of important results might contribute to a better clearness.

2) Using the criterion that the ratio of the first two eigenvalues in a PCA should at least be three to assume unidimensionality the General Subscale is not unidimensional and therefore DIF about it seems to be questionable. This should at least be discussed.

3) The authors present a number of “significant” differences according DIF in tables 4 and 5: using the term “significant” and presenting p-values requires an additional remark about the problem of multiple testing. Authors should either control for this problem adequately or explain why they do not or probably even remove p-values and base their interpretations on effect sizes.

4) At the beginning of the discussion the authors describe their sample as “representative”, while in the limitations it is described as “not representative”. Please explain.

5) In table 3 several WMS are greater than 2, nevertheless the authors claim for all sub-scales that “Overall, the goodness-of-fit of the […] items was satisfactory across all geo-cultural groups.” This should be better explained and probably
underlayed with literature.

6) The terms “difficulty” or “functions better” are probably too abstract for people not familiar with Rasch-modeling. Therefore they should be better explained in very few and simple words.

**Level of interest:** An article of outstanding merit and interest in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests