Author's response to reviews

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

Authors:

    Anzalee Anzalee Khan (akhan@nki.rfmh.org)
    Christian Yavorsky (wcy@cronosccs.com)
    Stacy Liechti (stacy.liechti@prophase.com)
    Mark Opler (mark.opler@prophase.com)
    Guillermo DeClemente (gdc@cronosccs.com)
    Brian Rothman (brothman@prophase.com)
    Sofija Jovic (sjovic@prophase.com)
    toshiya Inada (toshiya.inada@gmail.com)
    Lawrence Yang (lawrenceyang@gmail.com)

Version: 4 Date: 18 January 2013

Author's response to reviews: see over
Editorial Board,
BMC Psychology

January 18th, 2013

Dear Editorial Review Board,

Thank you for your thoughtful review of the original manuscript submitted to BMC Psychology, “A Rasch model to test the cross-cultural validity in the Positive and Negative Syndrome Scale (PANSS) across six geo-cultural groups.” We appreciate the insightful comments of the three reviewers and have responded to each query and/or request for amendments to the manuscript. We appreciate your consideration of our manuscript to be part of your journal. The authors feel the feedback from the three reviewers and the Editorial Board of BMC Psychology has substantially strengthened our manuscript.

The subject of this study is important to the field of psychiatry, as being able to identify common problems that arise when estimating measurement error and suggest methods to overcome these problems, is important for future clinical trials. Using the technique utilized in this manuscript to identify the differences in item functioning across geo-cultural groups can help guide training and data monitoring of outcomes of clinical trials and prevent future failed studies.

The authors believe this manuscript is a good fit for BMC Psychology.

If you require additional information, please do not hesitate to contact me via email at akhan@nki.rfmh.org.

Sincerely,

Anzalee Khan, PhD
William Christian Yavorsky, PhD
Stacy Liechti, PhD
Mark Opler, PhD, MPH
Guillermo DiClemente, PhD
Luka Lucic, PhD
Brian Rothman, PhD
Sofija Jovic, PhD
Toshiya Inada, MD, PhD
Larry Yang, PhD

The authors have addressed all responses to reviewers and have categorized all responses into the following three categories:

1. Study specific and methodological issues
2. Revisions and Clarifications to text
3. Geo-Cultural Groupings and the relationship of Language

The table presents each Reviewer’s comments and a detailed description of how these comments were addressed by the authors.
<table>
<thead>
<tr>
<th>Reviewers comments</th>
<th>Authors responses to reviewers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study Specific and Methodology Issue</strong></td>
<td></td>
</tr>
<tr>
<td>Reviewer #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.</td>
<td>In response to the comments of Reviewer #1, the authors have included additional clarification on unidimensionality and the General Psychopathology subscale as follows: The authors wish to describe that although unidimensionality was not obtained with the General Psychopathology Scale, the items being examined are conventionally assigned to this subscale. Additionally, although unidimensionality was not obtained, internal consistency and correlation coefficients were adequate to validate items in this subscale. As such, the Results section was revised as follows: ‘Although the General Psychopathology subscale was not unidimensional, basic steps for validating items were met, i.e., intraclass correlations were all ≥ 0.90, and the items of the General Psychopathology subscale was evenly distributed and linear.’ Additional reference to the issue of unidimensionality and the General Psychopathology subscale was addressed in the Discussion section, Paragraph 3.</td>
</tr>
<tr>
<td>Reviewer #2: 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.</td>
<td>The authors concur with the comments of Reviewer #2 and have completed analysis to reduce possible Type I Error as a result of multiple testing. The authors have added analyses for multiple testing and revised the Statistical Methods section, Paragraph 7. In a typical differential item functioning (DIF) analysis, a significance test is conducted for each item. As the scale consists of multiple items, such multiple testing may increase the possibility of making a Type I error at least once. Current DIF studies regarding a Type I error rate have found that the latter rate can be affected by several factors, such as test length, sample size, test group size, group mean difference, group standard deviation difference, and an underlying model. Bonferroni corrections for DIF were further clarified in the Statistical Methods subsection and subsequent calculation of p values was performed. Tables 4 and 5 were also revised to reflect Bonferroni corrections for multiple testing, and the Results section was amended to reflect significant findings based on a Bonferroni corrected p &lt; .0017.</td>
</tr>
</tbody>
</table>
Reviewer #2: In table 3 several WMS are greater than 2, nevertheless the authors claim for all subscales that “Overall, the goodness-of-fit of the […] items was satisfactory across all geo-cultural groups.” This should be better explained and probably underlaid with literature. In response to the comments of Reviewer #2, the authors agree that a description of WMS in DIF analysis is warranted and the steps undertaken to obtain appropriate goodness-of-fit values should be further expounded upon.

Revisions were made to the Statistical Methods subsection to further describe WMS and to the Results section describing the procedures undertaken to assess model fit as follows: ‘Most items showed high mean squares (WMS > 2.0 or low discrimination). Poor fit does not mean that the Rasch measures (parameter estimates) aren’t additive (or appropriate). So a WMS > 2.0 suggests a deviation from unidimensionality in the data, not in the measures. Therefore, values greater than 2.0 (see Table 3) indicate unpredictability (unmodeled noise, model underfit). Items with high WMS were examined first (influenced by one or two responses or outliers), and temporarily removed from the analysis, before investigating the items with low’ until unidimensionality was achieved with WMS values closer to 1.0.

Revisions and Clarifications to text

**Reviewer #2:** 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.

The authors recognize the comment of Reviewer #2 of the discrepancy, and see this as an error in editing. The first sentence in the Discussion section was revised to read as follows: “This article is the first to publish a cross-cultural comparison of the psychometric performance, mean scale scores, and item and scale-summary for the PANSS using qualified raters who rated one of 13 standardized training videos of a patient in the United States.”

**Reviewer #2:** 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.

The authors agree with the comments of Reviewer #1 and have substantially edited and reduced the text in the background. For the Discussion section, Paragraphs 3, 6 and 8 were revised. Additionally, the paragraph on the contribution of the SCI-PANSS was removed as it was not deemed relevant to the discussion. Paragraph 9 was also revised to be more succinct and repetitive comments were removed.

The Limitations and Conclusion sections were also edited, leading to reduction in text which contributed to clarity of the important findings.

**Reviewer #2:** 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.

The authors agree with the comments of Reviewer #2 that the statistical terms can be better explained. As a result, ‘functions better’ was revised to ‘the item was easier to score for’ and ‘difficulty,’ where possible, was revised to ‘significant DIF.’
Reviewer #3: The authors write in the introduction: "Although, many assessments used in psychopathology of schizophrenia have examined construct, functional, translational and metric categories of rating scales, the significance of rater differences across cultures in schizophrenia rating scales has rarely been investigated." I would encourage the authors to situate their excellent study in the literature on cross-cultural rater differences with the PANSS. I have published two papers explicitly on this issue comparing Chinese and American raters with the PANSS which the authors can cite and from which they can also review the bibliography for relevant references.

The authors concur with Reviewer #2 that reference to recent literature would strengthen the manuscript. At the time of preparation of this manuscript, the available literature on one of the articles mentioned was not available. We have reviewed and referenced the study accordingly and believe these articles add substantially to the literature and to the Background section of our manuscript. The following were also added to the Reference section:


Reviewer #3: Please provide references for the methodology explained in these sentences:

"Translation standards followed internationally recognized guidelines with the objective to achieve semantic equivalence. Semantic equivalence is concerned with the transfer of meaning across language, and aims to achieve a similar response from individuals with schizophrenia in different languages."


This was added to the manuscript along with a link to the MHS translation policy on the site. Another form of publication and/or journal/book reference is not available for this policy; therefore a link to the Policy via the MHS website was inserted.

Reviewer #3: Please check grammar throughout the article. This sentence could be substantially clarified from the abstract: "Lower item values reflects items in which raters often showed less difficulty scoring; higher item values reflects items more difficult to score." Noun/verb agreement and clarity of which items would be helpful.

The sentence in the Abstract was revised and items were clarified in the ‘Results’ subsection of the Abstract.

Reviewer #3: Please clarify this sentence in the article: "The goal of this study was to examine the cross-cultural validity of the PANSS scale across six geo-cultural groups (East & Maritime Asia, India, Northern Europe, Russo Europe, Southern Europe, and the United States of America) using Rasch analysis of data United States training video (the video was translated and subtitled for other languages)."

This sentence was revised as follows: ‘The use of this scale in different countries could be influenced by the different perception of symptoms from different cultural backgrounds. The goal of this study was to examine the cross-cultural validity of the PANSS across six geo-cultural groups (Asia (excluding India), India, Northern Europe, Russia & Ukraine, Southern Europe, and the United States of America) for data obtained from United States training videos (translated and subtitled for other languages).’

### Geo-Cultural Groupings and the relationship of Language

Reviewer #3: The authors compellingly advocate for The authors agree with the Reviewer’s comments and have
a position that appreciates cross-cultural differences in the PANSS. However, the presentation of the data uses broad regional groups such as "East and Maritime Asia" and "Northern Europe," among other problematic labels. At the same time, India and the United States are distinguished as countries. I recommend that the authors list all countries individually to avoid stereotyping cultures within broad regions.

Moreover, the terms "Maritime Asia" and "Russo Europe" are unconventional without clearly specifying countries from the outset of the article with the abstract.

Further described the reasons for the categorization of each country. ‘East & Maritime Asia’ was revised to ‘Asia (excluding India)’ and ‘Russo Europe’ was revised to ‘Russia & Ukraine.’ Revisions to the labels listed above were made throughout the manuscript, including within Tables and in the Abstract.

Although it would be optimal to assess responses from each country separately, the authors realize a large amount of data would be required from each country and there are also many cultural differences within countries. The groupings of countries were based solely geographical and cultural similarities noted in these countries. The authors understand that arguments can be made for some grouping and have attempted to further describe the rationale for groupings and provide additional references.

Reviewer #3: The explanation of the authors in the methods section is currently insufficient: for example, the authors write: "Malaysia, Singapore, Taiwan, Japan and South Korea were grouped together. An argument could be made to have included Malaysia and Singapore with India instead. However, the large Chinese populations in these two countries as well as historical interactions and broad similarities in social interaction across all of these countries were deemed adequate to group living in these five countries raters together." This assumption prioritizes similarities over differences when opposite arguments could be made. It may be better to avoid this issue entirely by presenting data by country.

The authors agree with the reviewer that there are somewhat confounding cultural differences in these groups of raters. Additional published articles were reviewed and referenced in the text and in the Reference section of the manuscript.

Reviewer #3: Under the methods section, this sentence references culture: "An attempt was made to include raters who were likely to share more culturally within each group." The authors should explain this more clearly. How did they make a determination about which raters would be considered to share more culture? Since culture is a key variable of interest in the paper, it should be defined explicitly with references to the cultural mental health literature. The authors should also be explicit about how they decided to group raters together based on their definition of culture. Otherwise, the reader is left wondering whether the groupings were created after data analysis.

The Reviewer’s comments are similar to comments addressed in above relating to the selection of geo-cultural groups. - The subsection ‘Selection of Geo-cultural groups was revised to address these concerns.

Additionally, it should be noted that, in order to control for any investigator bias, ‘the resultant groups were defined prior to considering the amount of available data for each group.’

Reviewer #2: Under limitations, the authors should better incorporate the extensive literature on the relationship of language and affect. A substantial scholarship has emerged in cultural psychiatry and medical anthropology on this topic. Given the fact that the raters had variable English fluency, this is a serious limitation. Similarly, the authors could also discuss how the PANSS itself is a cultural creation.

The authors agree with the Review’s comments and have reviewed and discussed relevant literature on language and affect in cultures that may affect response patterns in rating scales. The Limitations section was briefly revised accordingly. The following articles were referenced:

Guarnaccia PJ, Angel R, Worobey JL. The factor
which may or may not match the realities of patients with schizophrenia outside of the United States. For example, an alternate reading of why international PANSS raters scored items differently may be due to differences in clinical experience: they may be trained with the PANSS as an American cultural artifact, but their actual experiences with schizophrenia in the local population may be very different. It would be interesting for the authors to discuss how cross-cultural training could consist of presenting raters with local language, culture, and affect taken into account through patients from dominant cultural groups in their societies.


The authors have also made additional edits not recommended by the reviewers. The authors have added Dr. Luka Lucic (affiliation Pratt Institute) as a co-author for his significant help in editing the revisions based on his expertise in Socio-Cultural and Language Development. The authors also edited all Tables to reflect additional requested analysis or revision of labels to the selected geo-cultural groups. All data and preparation of tables were double entered for accuracy.

We thank the reviewers for their introspective and considerate review and comments of our manuscript and feel the revised and condensed manuscript is significantly improved as a result of the recommendations of the reviewers and the BMC Psychology editorial board.