Reviewer's report

Title: Psychometric evaluation of the community reintegration of injured service members computer adaptive test (CRIS-CAT)

Version: 4 Date: 17 January 2012

Reviewer: Niels Smits

Reviewer's report:

Although most of my previous points have been addressed I still think the paper has many flaws. I have many points, but unfortunately limited time to address them all.

Major Compulsory Revisions

*The title must be changed. Many clinicians use the term 'Psychometric' incorrectly. Psychometrics is about measurement (see http://en.wikipedia.org/wiki/Psychometrics), and you are obviously interested in prediction (outcomes, change, etc.). Please change 'psychometric evaluation' into something like 'external validation' or 'predictive utility'. Maybe look at McDonald's 1999 book (Test Theory) for a good term.

*The article does not read easily. Please provide more structure to it. Explain why you do things; explain the results to the audience instead of solely referring to the table. In your discussion, please discuss the outcomes in the light of your research goals. What can we learn from your study? Try to arrange the sections of your discussion according to content.

*The authors did not deal with the missing values point I made. They are 'unaware of methods for dealing with missing values in small longitudinal studies'. Instead they used complete cases. That this is very problematic is acknowledged in the discussion, and very obvious from Table 2. I would suggest that they multiply impute their data, and use relevant covariates (such as those in the table) in the imputation model to deal with this self selection. Under MAR the assumptions concerning the missingness are less strict then under MCAR (which is done here).

*I don't think the authors dealt sufficiently with my points concerning statistical concepts. The equations are a mixture of symbols and text, which makes it very messy (are they at all necessary to be presented?). Please provide a better and more extensive introduction of all statistical and psychometric terms. Under what model are things obtained? Standard errors of what parameter?!

*Likewise, please say something about latent variables earlier on in the paper. The first reference to them is on page 14, where is just suddenly appears.

*The end of the introduction does not all justify this study! The main point should
be that now that you have a psychometrically sound CRIS-CAT (which has a small respondent burden), is it useful in practice (or as useful as the full CRIS)? To explore this, you study if you can predict all kinds of (external) measures which are relevant for the clinical field (change, outcomes, concurrent measures etc.). If you can, the CRIS-CAT can be used in the field to select subjects who are at a high risk of showing or developing pathology! This should be your main selling point. Please extend your intro.

*The structure of the paper is very awkward (methods per sub-study, results per sub-study). Better describe all three sub-studies separately (Study 1: methods, results, small conclusion; Study 2, methods, etc.)*

*I truly don't understand what you did in the 'likelihood of change in marital status analysis'! As I understand it, having two times three values, would give nine possible combinations. Why would this give three levels of change? How would a multinomial regression model incorporate this change?*

*Your prediction of change in marital status cannot be assessed if 121 of 135 do not change. Your variance is almost zero; therefore a trivial outcome!*

Minor Essential Revisions

*There are still many typos (to assess to assess?), and inconsistencies [(within brackets Perceived mn 56.6 ±11.5) and outside brackets Satisfaction (mn 54.9 ±11.0), what is mn?].*

*If all p-values in a table are equal, say so in a note an delete column!*

*What is ability (p.6) in the context of community registration? It is based on self-report, and therefore incomparable with proficiency testing interpretations!*

*Page 7, 'reliability equal to 0.9. IRT has no single reliability outcomes. It has conditional reliability (SE of latent trait, and related Item Info) which is sometimes transformed to a more intuitive 0-1 scale (marginal reliability?). Please mention this!*

*Page 7, How can simulations tell something unidimensionality, and model fit? An IRT model is usually fit before CAT construction! Please change in a proper fashion!*

*Page 8, please change term 'psychometric characteristics' into something appropriate (see above).*

*Page 9, acronyms of questionnaires, please give full name at first reference for all scales.*

*Page 14, true variance should be true score variance; what is THE latent variable. The latent trait under what IRT model (the one you used, so GRM of PCM or other?).*

*Same page, person standard error. I guess this should be SE of the persons estimated latent trait? As above, reliability as you use it is something that is transformed from the real thing (SE?) to provide non-technical users with. Explain this!*
*For all kinds of technical measures you refer to papers of applied researchers using them (e.g., references 25, 26, 21, 27, 28). Instead, please refer to the original methodological/statistical sources/references where these measures were developed. Because it is uncertain if statistics were assessed properly in applied papers (and we therefore do not know if they were used properly), it is a good habit to refer to the original sources.

*Page 19 and Table 6. What do the MDC values tell us? Please provide explanation.

*Please, for each table, state for which study (1, 2 or 3) it is relevant.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.