Reviewer’s report

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

Version: 2 Date: 3 August 2011

Reviewer: Niels Smits

Reviewer’s report:

The authors describe the evaluation of an adaptive version of an existing test (originally designed by some of them). I agree that for any test comprising a lot of items, a CAT version is appropriate, especially in a population such as veterans.

I have several problems with this paper, however.

Major Compulsory Revisions

The title suggests a psychometric evaluation of the CRIS-CAT. Because it is stated that "In previous work we developed the CRIS CAT...", and since there is no reference to this I would expect that this paper was in fact THE psychometric evaluation. But it is not. It is a study of how a CAT, which has not yet been validated model-wise, functions in practice: it studies the utility of the CAT scores with reference to external criteria, and how stable CAT scores are from one instance to another. A real psychometric (fitting psychometric models!) evaluation of each of the scales is needed.

I think that a thorough study of the psychometric evaluation of the CRIS-CAT, such as described by Reeve et al. (2007) is needed before a study of the external validity is conducted. For each sub-scale it needs to be shown that it is unidimensional, and meets other requirements for IRT. The three scales of CRIS-CAT, of which no examples were provided, seem to consist of items which, in my opinion, are causal indicators instead of effect indicators (Fayers and Hand, 1997). So maybe CAT is not appropriate at all. The authors may be much more interested in short assessment not
assuming latent constructs such as, for example, Finkelman et al. (2011).

So what in my opinion is needed, is (i) an exposition of the appropriateness of the CAT machinery for the items of CRIS (are these effect indicators?), (ii) if appropriate, an evaluation of the scales like in Reeve et al. Note that for a thorough psychometric evaluation of an item pool, and the estimation of items parameters of an IRT model a large sample (N>1000) is needed; the larger the item pool, the higher N needs to be.

The external validity of the CRIS-CAT can only be evaluated after it has been shown that the meaning of its scales are unambiguous, i.e. are unidimensional. (What would it mean if we would compare two veterans who had CAT scores, obtained using different sets of items from a single multidimensional item pool?)


P. M. Fayers and D. J. Hand (1997). Quality of Life Research Volume 6, Number 2, DOI: 10.1023/A:1026490117121 Factor analysis, causal indicators and quality of life.


In addition, I think the paper has structure which makes it very difficult to read. Often it is not explained why samples, external measures etc are used. For example, it is not really clear why the CRIS-CAT needs to be able to differentiate between the three groups. In addition, it is not made clear why a study of the minimum detectable change is necessary; I can make up these things
myself after thinking for a while, but it is the work of the authors to do that.

Some statistical/measurement concepts are described improperly. On page 13, the authors talk about true variance; variance of what?, a true score, or a latent trait? They go on to introduce an equation which has the word "standard error" in it. A standard error of what parameter, using which model? Same holds for equations on pages 14-15.

Likewise, on pages 6-7 (and figures 1-2), reliability appears in a CAT discussion. What type of reliability do you mean, marginal reliability, or internal consistency reliability such as Cronbach's? IRT uses information, which is a conditional estimate of measurement precision.

Minor Essential Revisions

References [37], on page 13, and [30] on page 19 are not in the reference list.

There is a 30% dropout rate for the longitudinal study. Please state why missings need not be dealt with. Alternatively, use ML estimation or an alternative to deal with this.

**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.

**Declaration of competing interests:**

I declare that I have no competing interests