Author's response to reviews

Title: Reliability, validity and administrative burden of the community reintegration of injured service members computer adaptive test (CRIS-CAT)

Authors:

Linda Resnik (Linda_Resnik@Brown.edu)
Matthew Borgia (Matthew.Borgia@va.gov)
Pengsheng Ni (psni@bu.edu)
Paul Pirraglia (Paul.Pirralgia@va.gov)
Alan Jette (Ajette@bu.edu)

Version: 6 Date: 6 August 2012

Author's response to reviews: see over
Dear Editor,

Please find a point by point response to Reviewer 2’s comments on the second revision of this manuscript. We believe that we have addressed all issues satisfactorily and very much hope that this review can be done in-house. Thank you for your consideration of this revision.

1.a. Formula for conditional reliability.
The formula provided in our manuscript is correct. A better reference than the one previously used is:


In essence, the formula provided by the reviewer and the one we used in the manuscript are the same. The reliability could be defined as: (True variance)/(observed variance). In the formula referenced by the reviewer, the true variance is defined as ((variance of the observed person scores)-SE^2). The ‘hat’ above the theta in equation (15) means the variance is estimated from the person scores. In our paper and the Mâsse (2006) paper, the authors assumed the true variance is 1, so the observed variance is defined as 1+SE^2.

In our case, the conditional reliability difference between two methods are small, average difference is (0.013~0.02), standard deviation of the difference is (0.013~0.015). (3 scales)

1.b. Change the word estimators to estimates.
We have made this change on page 14.

1.c. Remove the formulas for SEM and MDC and use references only. Because these two terms do not stem from an IRT model, explicitly state where these come from.
We have edited the text on p. 16 and included new references.

In Wyrwich et al. (1999), the authors state: “The SEM is the standard error in an observed score related to measuring with a particular test that obscures the true score. It is estimated by the standard deviation of the instrument multiplied by the square root of one minus its reliability coefficient [10]. Furthermore, according to classical test theory, the SEM possesses the unique attribute of being sample-independent.”

In De Vet HC etc. (2006) paper, the authors state: “In statistical terms, the minimally detectable change (MDC), also called smallest detectable change or smallest real change (Beckerman (2001)) shows which changes fall outside the measurement error of the health status measurement (either based on internal or test-retest reliability in stable persons). It is represented by the following formula: MDC = 1.96 * √2 * SEM, where the 1.96 derives from the 95% confidence interval of no change, and √2 is included because two measurements are involved in measuring change”
From the definition of SEM, we know that the SEM assumes the measurement error is constant across the range of the possible score, which is the assumption in CTT but is different from IRT, and MDC is calculated from SEM. So both indices were developed under CTT.

Here are the references that we have added:


2. Handling missing values
Our independent variables were collected by interview at time 1. Therefore, we had no missing data for any independent variables. The multiple imputation (MI) we performed was on the outcome (dependent) variables (PCS and MCS at visit 2). Our main argument was that we were “uncomfortable imputing primary dependent variables, particularly categorical variables where we have little variance”. The reviewer’s argument is we should impute all the variables (independent/dependent variables). In the manuscript we describe that we have imputed missing values for the continuous dependent variables (PCS and MCS) and the results revealed that the difference between MI or not was small. The remaining question is whether we should do multiple imputations for those categorical dependent variables. Groenwold et al (2012) state that the difference between the MI and completed case analysis may not make much difference. The authors state, “Complete case analysis with covariate adjustment and multiple imputation yield similar estimates in the event of missing outcome data, as long as the same predictors of missingness are included. Hence, complete case analysis with covariate adjustment can and should be used as the analysis of choice more often”. So based on this argument we did not use multiple imputations with the categorical dependent variables.


3. Structure and text
a) In the abstract, what does `reliability of simulated CAT scores mean?`  This refers to conditional reliability. The text was edited to clarify.

b) Move details on the CRIS to methods section.
We have made this change, leaving a brief overview in the introduction. This change was not marked in bold font because of the large amount of text.
c) Intro, CAT is described without any reference.  

d) Edit introduction to make testing goals more explicit.  
Thank-you for this suggestion. We have made this change.

e) 'Longitudinal cohort study' vs. Cohort study'  
The word “longitudinal” was deleted throughout to make this consistent.

f) Edit the introduction “latent traits”.  
We have edited p. 6 per your suggestion.

4. Reliability plots  
We realized that the reliability plots in the manuscript were completed on the full item bank, prior to the CAT simulations before the CAT simulations were conducted. We have corrected this error and replaced the plots with the correct ones. The results look the same as earlier version and no changes were necessary in the text.