Reviewer's report

Title: Selecting Optimal Screening Items for Delirium: An Application of Item Response Theory

Version: 2 Date: 20 August 2012

Reviewer: Joel Coste

Reviewer's report:

The objective of the paper was to identify a short set of indicators to assess the four features of delirium in the Confusion Assessment Method diagnostic algorithm using modern measurement theory. The paper also aims illustrating modern measurement theory applied to clinical assessment.

A pool of 135 indicators from established cognitive tests and delirium interview tools administered to 4598 older adults assessed for delirium has been considered. Modern measurement methods have been used, together with classical test theory methods (EFA/CFA) to assess dimensionality assumptions and identify the best indicators for screening for each delirium feature in the perspective of developing a short assessment for detecting delirium.

The paper deals with the important topic of modern measurement input to the process of shortening composite measurement scales (CMS). It could turn out to be an important paper for those who embark on shortening such scales. However, there are some areas for improvement, detailed below.

Major Compulsory Revisions

1. The role of modern psychometrics and IRT in item reduction should be further investigated and discussed. Following the application of such methods, items can be selected or removed not only on the basis of their information content ("discriminatory power") but also on the basis of the level of fit to their respective dimension, and on the basis of closeness or overlap with other items of the dimension. IRT methods allow for a precise diagnosis of the functioning of every item or even every response category of every item. IRT provides a set of interpretive tools (information curves but also item characteristic curves, response probability curves, differential item functioning, etc.) that can be useful for scale refinement or shortening. These tools encourage an interpretation of the results, which is a key aspect for selecting the items to be retained. Furthermore, most of IRT models benefits of the property called specific objectivity (item and person invariance), which allows independent comparison of instruments across several samples and symmetrically independent comparison of individuals across several instruments. This property is particularly precious for CMS shortening, as two instruments (the original and the short-form CMS) have to be used in several samples (development and validation samples at least).

2. IRT methods and factorial analyses, especially confirmatory factorial analysis
(CFA), are usually considered complementary to check the effect of the removal of items on the dimensional structure of the CMS. CFA has an important advantage over exploratory factor analysis in that it allows testing whether the initial model remains intact when items are removed. It is important to note here that if the removal of an item alters the dimensionality of the scale, the item should be kept rather than the dimensions modified, in order to respect the underlying conceptual model, which should always be kept in mind during the shortening process. In practice, however, IRT methods and CFA can rarely be simultaneously applied since IRT methods are appropriate to dichotomous/ordinal items and CFA requires normal distribution of the scale items. If the authors wish to continue presenting both IRT and CFA results, they should clarify which technique they used for which scale and item (indicator) distribution of responses and detail how the hypotheses and assumptions of techniques were checked.

Minor Essential Revisions

1. Whereas pen and paper is still the dominant mode of administration of composite scales, the increased use of electronic administration using Computer Adaptive Testing (CAT) changes the role of IRT methods in the context of selecting items: the number of items that may contribute to a scale score might even be increased with CAT without any corresponding increase in respondent burden. This development should at least be discussed.

2. Check the references -- some seem incorrect eg ref 24 page 27; and spelling of names, eg Gibbon page 16

Suggested additional references


Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res. 2007;16:5-18.


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