Author's response to reviews

Title: Script Concordance Tests: Guidelines for Construction

Authors:

Jean Paul Fournier (Fournier.jp@chu-nice.fr)
Anne Demeester (Anne.Demeester@mail.ap-hm.fr)
Bernard Charlin (bernard.charlin@umontreal.ca)

Version: 3 Date: 17 March 2008

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

You will find in attached file a new submission to the paper "Script Concordance Tests: Guidelines for Construction".

The paper has been revised. The present letter addresses the comments in a revised manuscript and gives a point-by-point response to the concerns made by the reviewers.

Reviewer 1
Background
Sentence 1: substitute "uncertain" for "ill-defined"
Done
Sentence 2: suggest "It makes possible inclusion of ..."
Done
Sentence 3: add "experts" after "reference"
Done
Purpose
Sentence 1: suggest replace "ground" with "acceptance"
Done
Discretionary Revision
Background
3rd paragraph, 3rd sentence replace "though scarcely" with "nevertheless poorly"
Done
Minor essential revision
Figure 1: typo? "n" in Case description doesn't appear
This is an error related to the printing by Adobe,
The paper needs some language correction before publication.
The paper has been send to an English editor for English polishing

Reviewer 2

One of the weakest aspects of the paper relates to the advice regarding score scales and scoring of the SCT. The recommended scoring, using the aggregate method, has many implications for implementing a classical theory test model for instance. The role of uncertainty and the response scale have not been well researched and the guidelines should indicate this. Also, the current status of the research regarding panel size, level of consensus, score scale, and the relationship with a discrimination index does not provide a solid evidence basis for making test construction decisions.

It is important to acknowledge this so that those developing new SCT-types of tests understand that experimenting with variations on the SCT are justified, especially since current 'best' methods have not been proven. I think the article does a good job of detailing how the SCT can play an important role in medical education testing, but I would encourage the authors not to present the guidelines as the product of a strong research tradition. It seems more accurate to state that these are the current practices,

However much psychometric research remains to be done. Also, they should state that validity research is required to understand the SCT relationship with more traditional knowledge item formats, performance assessments, and clinical reasoning.

The following paragraph has been added at the end of the scoring section

The aggregate scoring method described above is the most commonly used method. It is though important to acknowledge that what constitutes the SCT scoring optimal method is still debated. The aggregate method has many implications for implementing a classical theory test model and much psychometric research remains to be done on level of consensus, score scale, and the relationship with a discrimination index for instance. Also, validity research is required to understand the SCT relationship with more traditional knowledge item formats, performance assessments, and clinical reasoning.

Reviewer 3

One question:

Given that the test guidelines call for having a 15-member panel. I think there are situations where this number may not be met. Are there some minimum standards when following the optimal strategy may not work?

The following sentence has been added at the end of the first paragraph of the panel size section

For lower stake examinations, formative assessment within a clinical rotation for instance, smaller panels can be used. Though, panels with less than 10 persons are associated with more reliability estimates error.
We hope these corrections and improvements address reviewers' concerns.

With kind regards

Bernard Charlin
Research Director
CPASS