Reviewer’s report

Title: Assessment of Junior Doctor Performance: A validation study

Version: 1 Date: 2 June 2013

Reviewer: Nick Sevdalis

Reviewer’s report:

The paper reports a 2-year prospective psychometric evaluation of a tool (Junior Doctor Assessment Tool-JDAT) to assess junior residents in Australian hospitals – based on a UK assessment model. Reliability and validity analyses are reported.

The study is interesting and the paper reports an interesting dataset – the 2 year duration, reasonable sample and psychometric analyses used are all strengths of this study which should be of interest to the Journal audience. The manuscript, however, has significant weakness in the reporting in its present form, which I outline below – these are mostly methodological and technical, but also in the Discussion.

MAJOR COMPULSORY REVISIONS

Methods: the assessment framework is important for this type of study, yet very underdeveloped in the paper. The Australian junior doctor curriculum framework should be described in detail, it deserves at least a paragraph on its own – for the readers to understand what the JDAT is based on).

Scoring: there is some justification regarding why ‘not observed’ was scored as 1 - but it appears rather like a retrospective thought. Given the nature of the tool, I found it hard to see how this scoring would have been used in practice – for example, what instruction was provided to the assessors in order for them to use this anchor? This point requires further clarification in the description of the JDAT (p 6) and also in the authors’ response letter.

Data analysis: it is unclear to me why you computed an overall Cronbach since the tool is theoretically meant to assess 3 categories of competences. 3 alphas should have been reported in the first instance. Further within this section, the sentence “an item-total correlation...hence reliability” needs some further explanation for the journal readers.

Results: the descriptive data on page 9 should be reported within a separate table (currently within table 2) – and ideally an analysis (within subjects ANOVA or MANOVA) reported as well to show which elements were significantly better scored. Table 2 should reported the 2 factors identified in different columns, with the rotated factors loadings reported for each item – so readers can see which items load onto which factor. Further, the text should clarify that the results on this table and also on pp 10-11 refer to the varimax rotated loadings – at present
it is vague. The comments on page 11 regarding what was found compared to what was expected should be entirely removed – these are discussion points. I could not find table 3 attached, or in the supplementary materials.

Discussion: the sentence (p 13) “however this remains...reducing the 9 items to subscales” does not make sense – the aim of the PCA would not be to reduce the already small no of items to 2, but to see what the scoring actually means at a higher level once the scores have been aggregated. Limitations and also ‘where next’ now that this study has been completed need to be included – there are lacking.

Level of interest: An article of importance in its field

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