Reviewer's report

Title: Exploration of possible association between Examiner stringency and Personality factors in clinical assessments: a pilot study

Version: 2 Date: 14 July 2014

Reviewer: Tim Heywood

Reviewer's report:

1. The paper is an interesting and useful pilot study into an area not well understood, and is welcomed.

2. MINOR ESSENTIAL REVISION: there is inconsistency in the use of the terms "variance" and "variability", (e.g. in the abstract "Background" section, the authors state, "Inter-examiner variance is a key source of this variance...", while in the "Results" section, they state, "...variability due to the examiners was greater than candidate (true) variability..."). The paper would read more clearly if one or the other were used throughout, and the suggestion would be to use "variability" as a default unless the specific statistical meaning of "variance" is required. While the inconsistency is probably not strictly wrong, it creates a sense of slight confusion that would be easy to correct.

3. DISCRETIONARY REVISION: abstract "Background" section states, "Some examiners consistently mark low compared to other examiners...". "Low" is an adjective and reads badly when used as an adverb. Replace with "Some examiners consistently give lower scores to some candidates compared to other examiners...".

4. MAJOR COMPULSORY REVISIONS: the methods section doesn't provide enough detail to allow other researchers to check results or repeat the study. Some questions left unanswered, and some assumptions have to be made:
   a) it's ambiguous whether the analysis was done purely on mean total score (as suggested by the opening line of the "results and discussion" section and the y-axis labeling of figure 1) or station-by-station (as suggested by the second sentence of paragraph 1 in the "Results and discussion" section). Given that the stations chosen represent (probably) several discrete constructs being measured, then it is possible that stringency is construct-dependent, and so mean score variability may not truly reflect the impact. On the other hand if the hypothesis is that stringency effects are fixed across all constructs, then the second sentence of paragraph 1 of the "results and discussion" section is redundant. Would be better to have consistency of level of reporting.
   b) which correlation stats have been used and justification for the choice?
   c) given this is a pilot study, the handling of the outlier has been managed well and the limitations on ability to draw conclusions is adequate, although perhaps slightly understated in the "Conclusions" section (as the suggestion of an
association is only made when the outlier is removed, and that may or may not be the right thing to have done). However, as the possibility that the outlier is not really an outlier is raised, it would be good to have comparison correlations with and without exclusion of the outlier, to allow the suggested further study to be planned more effectively.

Summary: it's a pilot study and presented as such. The only realistic conclusion is drawn, i.e. further study is needed. As a result, impact is limited, but there is definitely a need to understand this area better, and so there is no doubt that this would benefit from publication. However, more details are required about the specifics of the methodology. Additionally, while the authors appropriately accept the handling of the outlier is a potential weakness, the analysis would be much more impressive if equal weight was given to the results both with and without the outlier's data being included, direct comparisons drawn, and guidance therefore regarding direction and methodology of future study given. The paper would also be improved by greater consistency of terminology (esp with respect to "variance" vs "variability" and consistency of outcome measure.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare I have no competing interests.