Reviewer's report

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

Version: 5  Date: 15 October 2014

Reviewer: Peter Yeates

Reviewer's report:

Major compulsory revisions:
None

Minor Essential revisions:
Line 10: should the author's name be “Peter” rather than “Peer”

Page 7, lines 1-2: I am not sure it is appropriate to use the phrase "scatter plots demonstrated a positive pattern". Graphical data is illustrative; relationships are either statistically significant or they are not. I would suggest something like "Figures x to y show Percentile personality traits plotted against examiners' mean scores for (and then list relevant traits). I would then move on to line 9 and start describing what you actually found. This will help the flow of the results section. You could say something like visual inspection of these data suggested relationships may exist, but I'm not sure that really adds much to just saying “here are the graphs, and then here are the stats”.

Page 7, line 12: spelling error “xtringency”

Page 7, lines 22 - 28: The word "trend" is generally used when a relationship is approaching statistical significance - for example when the p value is between 0.1 and 0.05. I don't think it is reasonable to say that there is a trend in the overall data; I think they only thing you can say about your overall data is that there was no significant relationship. I would recommend doing that briefly, and then moving on to the data with the excluded outlier.

Page 8, line 23-26: I think this statement should be qualified, in case a reader reads your conclusions without reading the rest of the paper. This could be done by extending the first sentence of the paragraph to say "Ours was a small study ... that overall showed no significant relationships" or alter the 2nd sentence to say something like I "subset analysis of our data suggests that a relationship exists between ... " or “following exclusion of an outlying case our data suggested ...”

Page 8, lines 27-29: This conclusion seems to go a bit far. At best you could use personality testing to predict who might be a hawk or a dove. With an r of 0.76, about 57% of score variations (i.e. the r2) can be explained by differences in this trait, whereas the remaining 43% of variability remains unexplained. As such it is
a reasonable predictor, but it is not the whole story. A much more robust way of determining hawks and doves is to look at old exam data! So, this would really only work for new examiners, and then incompletely. I would recommend concluding that it if confirmed these findings may offer some explanation (remember you can’t attribute causality based on correlations) of hawks and doves, and could help to predict scoring behaviours in new examiners.

Tables and figures:
3 tables are presented. These are all formatted differently. It would help the appearance of the paper if they were formatted consistently.

Supplementary materials:
The table titled “Examiner mean total scores and standard deviations (SD)” adds little further to the data in table 1. I would suggest removing it, or including it as a final column in table 1.

The labelling of the x axes of scatter plots could be clearer. Agreeableness should say something like “Percentile agreeableness score”.

I’m not sure how many tables and figures the journal allow in the main manuscript, but if there are sufficient remaining, I would suggest including 1 or two of the scatter plots in the main paper. The illustrate the data nicely, and the outlier looks very clearly like an outlier; the scatter plots help to visually justify the exclusion of the outlier, so the more readers that see the scatterplots the better.

Discretionary revisions:
Line 1: in the title rather than saying “possible association” I would recommend using the term "relationship"

Page 6, lines 28-30: This sentence is unnecessary; you've already said in the methods that you performed correlations. The results will read more clearly if you simply state your results.

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

Quality of written English: Acceptable

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

Declaration of competing interests:
I declare that I have no competing interests