Author's response to reviews

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

Authors:

Yvonne Finn (yvonne.finn@nuigalway.ie)
Peter Cantillon (peter.cantillon@nuigalway.ie)
Gerard Flaherty (gerflah@gmail.com)

Version: 5 Date: 1 October 2014

Author's response to reviews: see over
Dear Editorial team BMC Medical Education,

Many thanks for the valuable feedback from both the editorial team and the 2 reviewers on my manuscript. I appreciate the time and attention paid to my work and the detailed communication of the review. I have spent some time addressing the points made and made appropriate changes to the resubmission. The remaining section describes the points made that I have addressed. Reviewer 1 made over 40 points and I have addressed the compulsory points and some of the discretionary points made.

Should you have any queries about my resubmission please do not hesitate to get in touch.

Yours sincerely,

Dr Yvonne Finn
Lecturer in Clinical Methods and Clinical Practice, School of Medicine, National University of Ireland, Galway, Ireland
Section 1: Editor’s comments

1. Research question stated clearly in last paragraph of introduction line 25/26 page 4 ✓

2. Methods fully described – completed under methods ✓

3. Full Ethics ref supplied ✓

4. Title used stringency and results reported using examiner Stringency (updated) ✓

5. Supplementary file contains
   (i) Examiner mean total scores and standard deviations (SD)
   (ii) Scatter plots of examiner Personality Domains against Examiner Mean Scores

6. Line numbering ✓

Reviewer’s Report (reviewer 2) Tim Heywood

1. There is inconsistency in the use of the terms “variance” and “variability” ..... 
   
   Response: I accept this completely and have adopted Tim Heywood’s suggestion to 
   only use “variability”

2. “Some examiners consistently mark low compared to other examiners...” reviewer’s 
   suggestion to change to “Some examiners consistently give lower scores to some 
   candidates compared to other examiners...” has been adapted.

3. Methods section doesn’t provide enough detail to allow other researchers to check 
   results or repeat the study.
   
   We have accepted and have clearly explained in detail the methods section.

4. 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 or station-by- 
   station..analysis of
5. 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.

The results present the mean scores, SD and range at each of the 5 stations (table 1). This clearly demonstrates the presence of examiner variability. Maybe just report the mean scores at each station. Variability at each station.....Focus on reporting the results section reports on the analysis of variance at station level and the contribution of examiner variability to this variance. Which correlation stats have been used and justification for the choice? Methods section has been rewritten to give a full description of methods used and statistical analysis adopted with justification.

Make sure that the “Conclusions” section does acknowledge the uncertainty of the outlier and limits ability to draw conclusions accepted

Reviewer’s Report (reviewer 1) (Peter Yeates)

1. Refer to evolving literature on assessor variations/assessor cognition. This is now included in the introduction section.
2. Expand methods section as recommended by reviewer 2 above. YES – see methods
3. Agree to present study as a pilot study – this we think is the case already, as stated in the title and in the discussion and limitations. Also mentioned twice in conclusion of abstract.
4. Is the question posed by the author well defined?
   We have made the RQ clear in the last sentence of abstract background and in last sentence of Introduction.
5. Expand on the sources which contribute to examiner variability i.e. random influences and fixed tendencies to be stringent/lenient. The authors argue that an examiner is thought to be lenient or stringent based on a comparison of his/her judgment of candidates performances to other examiners’ judgement of candidaes’ performances. A person’s personality traits are
likewise described in terms of having a degree of each trait compared with others people amended in introduction – see point 1 above
6. Note on fixed differences compared with idiosyncratic or random differences – amended introduction to include same
7. Referring to mini-CEX state that examiner variability has been reported in WPBA and this study is of interest to those using OSCEs and workplace based assessments such as the Mini-CEX in assessments – amended - this is stated when discussing mini-CEX

Are the methods appropriate and well-described?
1. Methods – amended to give a complete description of the methods employed.

8. coding – examiner packs coded in advance and stated in amended methods
13. consent – has been addressed in amended methods
14. removed number in methods section as suggested
16. results should start with summary data – this is now included in amended results & discussion section
17. Descriptive data to determine that data is appropriate to parametric analyses - included
18. Mean total scores – has been amended and results are presented as examiner mean scores per station
21. Means with SD – amended to present means AND SDs
22. point to compare domains - Personality factors, as in the big 5 as described by Costa and McCrae, are not expected to be at the same percentiles in any one person. For example, a person may be low on extraversion and high on conscientiousness or visa versa. It would not be helpful to do an analysis differences between domains/factors

23. Correlation coefficients labelled as Pearson’s correlation coefficients
24. data presented in line with recommendation from reviewer – see results
26 summary of results has been amended to clearly state the a priori analysis was non-significant but post hoc showed 3 significant relationships (as advised)

29. data on examiner ethnicity and seniority is captured and reported at start of results & discussion section

31. revision point to remove claim on effect of examiner training has been removed

33. accepted and amended

34 accepted and amended