Reviewer’s report

Title: Reliability of Multiple Mini-Interviews and traditional interviews within and between institutions: a study of five California medical schools

Version: 0 Date: 27 Mar 2017

Reviewer: Barbara Griffin

Reviewer's report:

This study presents an impressive dataset of interview scores from multiple institutions. It is a much needed study. That said, I found some aspects of the introduction and results difficult to follow and think readers will need more information. I trust my comments are helpful in developing your paper.

1. Definitions. Given the focus is on reliability, the authors need to define the various aspects and terms used (e.g. internal consistency, interrater reliability, test-retest reliability) and give at least a short explanation as to how alpha differs from G coefficient and correlations coefficient (especially as you compare MMIs and TIs using two different statistics).

2. There were a number of terms and statement that were unclear or even incorrect:
   i) low inter-interview reliability (i.e., consistency) - do you mean intra-interview?
   ii) "the between-school reliabilities of MMIs or Tis" - do you mean correlations?
   iii) "Raters at both MMI schools were blinded to participant applications" - this was not clear until later. You could just say, 'raters were not given any details about applicants prior to interviewing'

3. "This is an important research gap since at any given school, most applicants are not interviewed" - the logic of this argument is not clear. The fact that most applicants are not interviewed seems to have more to do with the tests that are used for short-listing interviewees.

4. The authors recommend in the Discussion that, "the reliability of long-form one-on-one interviews can be improved through relatively minor enhancements to interviewer training and structural aspects" It is not obvious what structural aspects these might be or how this follows from your analyses.

5. There was insufficient information about the interview ratings:
i) TIs. How many dimensions/criteria were assessed in each or did the rater only give one overall rating? It sounds like the TI consisted of two interviews, each with one interviewer (generally "traditional" interviews use a panel of at least two interviewers), but did the two interviews assess the same criteria?

ii) MMIs. Its important to know if each station was designed to assess one dimension and whether that dimension (e.g. teamwork) was the only thing assessed at the station or whether raters gave scores for multiple dimensions/criteria at each station.

6. What was the overlap in the dimensions being assessed by the MMIs? Even though "individual station content differed between schools" is sounds like the one company developed both interviews and that there may have been considerable overlap between rating methods, criteria assessed etc. How does this increase associations between the MMIs and did the three TIs have similar overlap?

7. I was a little confused by the sample numbers - how many were interviewed at all 5 universities? Why would you include those who only attended one interview when you argued in the Introduction "Importantly, no studies have directly tested whether MMIs are more reliable than TIs by examining a common pool of applicants completing both interview types". Then in the Discussion, the statement is made that "the current study was the first to examine the reliabilities of unstructured TIs and of MMIs in a common pool of applicants". However, it was unclear how many were in the "common pool". If there was not a group of applicants who were interviewed at all five schools, then we need to know the n for each correlation provided.

8. I am not sure that you can compare the demographics of those interviewed by MMI vs TI because, presumably there are individuals common to both types

9. Table 3. I think the different years needs to be separated - you likely have repeat candidates from one year to the next.

10. You compare correlation values (or intrarrer reliabilities - kappa?) for the traditional interviews - that is the relationship between two ratings, with MMI alpha reliabilities. Alphas increase with every score that contributes to the alpha and there were 10 (or 7) station scores. Correlations and alphas are two different statistics so please comment on why this is a legitimate comparison.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No
Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests
I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal.