Reviewer’s report

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

**Version:** 1  **Date:** 26 Jul 2017

**Reviewer:** Timothy Wood

**Reviewer's report:**

I have a better understanding of the design now and thank the authors of clarifying that. I still have some concerns but primarily around clarifying language.

I need some clarification around year. Under measures, third line it says that interview scores were converted to z-scores based on school and year. Does that mean that for school T11/2011 only those specific interview scores were converted to a z-score? And for T1/2012 only those scores were converted to a z-score etc. If that is the case then every year by test would have a mean of 0 and standard deviation of 1 so there would be no year differences. If my understanding is not the case and only school was converted to a z-score (ignoring year) then my concern about year still stands. If this is the case then I would ask that some kind of comparison be done to show that for example 2011 scores at T1 are no different than 2012 scores at T1 etc. I realize that the adjusted ICC shows no influence of year but that analysis is reported after all the correlations so comes too late.

Under analysis, the use of inter-interview and inter-rater terms is a bit misleading. If I look at T1, there is a faculty person doing one interview and a student doing a completely different interview. Because there is only one interviewer in each interview, the design means that interviewer and interview are confound and a pure interviewer reliability cannot be determined. Similarly for the MMI, there is one rater per station so rater and station are confounded and a pure inter-rater reliability cannot be determined. This confounding between rater and station is very common and does not reflect a weakness in the design. I am only suggesting some care in how the terms are used.

In the results section the authors convert TI correlations to a cronbach's alpha to facilitate comparisons to the MMI. They find that the alpha's are smaller for the TI than the MMI. That would be expected because for the TI's the alpha is based on two interviews but for the MMI, it is based on either 7 or 10 stations. To make that comparison, I think the alpha should be reported on the same number of stations/interview, either by reporting it as a reliability on a single station or applying the spearman brown formula to equate the number of stations. Similarly the reliability of the MMI with 7 stations is compared to the reliability of the MMI with 10 stations. These should equated for number of stations and then compared.
Page 13 third line says that the "...observed lower internal consistency..." Because no official test one done this is a difficult claim. Either add qualitatively as the authors have done elsewhere or equate the reliability for the number of stations and then report the comparison.

I note that one once the reliability have been equated for the number of stations/interviews, some of the differences mentioned in the discussion may no longer be there so it might change the discussion.

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

Yes

**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.

Yes

**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