Reviewer's report

Title: Modification of an OSCE format to enhance patient continuity in a high-stakes assessment of clinical performance

Version: 1 Date: 25 January 2011

Reviewer: Matt Homer

Reviewer's report:

This is an interesting, well-written, concise and informative study. There are one or two major issues which I think can be overcome, but otherwise I think it is a useful contribution to the research base.

Major Compulsory Revisions

1. Although the psychometric analysis as presented appears to add to knowledge, the method of ‘linked’ stations is certainly not new. In my own medical school, particularly in final year OSCEs, linked stations have been used for a number of years for the reasons described in the article. The authors should make explicit exactly what this study adds to the literature (i.e. not the linked stations per se, but the analysis that they have done of the outcomes).

2. The paper says nothing about the representativeness of the recruited sample. Do they, for example, tend to be the best students? In other words, is there a large risk of bias in the data? At the very least, some commentary on this issue would be useful.

Minor Essential Revisions

3. It might help the reader if the language used to describe the set-up consistently – does ‘scenario’=’station’ throughout? If so, then why not use one term or the other consistently. Compare Table 1, Table 2 and Figure 1 in this respect.

4. The use of language could, at times, be more precise. For example, in the abstract methods section (and later) what is the precise meaning of ‘independent’? I assume this means ‘independently marked’ or ‘separate’. Perhaps it is the statistician in me being a little pedantic but I think this is a key philosophical issue with the research and should be clarified.

5. It would be helpful to be more explicit about the inter-rater reliability calculations, perhaps with a reference. Was this Cohen’s Kappa or something else, and was SPSS used to calculate it?

6. In the results section (Table 2), were the estimated variance components for ‘station’ and ‘section’ not explicitly included in the model (i.e. set to zero in advance, presumably due to the earlier main effects ANOVA results)? Or did they calculate to exactly zero? Or were they not significantly different from zero?
This clarification might help others trying to replicate such an analysis.

7. In the discussion, wide confidence intervals for the variance components are alluded to – do the authors have any results for these?

8. I could not see the relevance of reference 8 to the rating scale description.

Discretionary Revisions

9. Does the phrase ‘patient continuity’ have a shared meaning amongst potential readers? If not, then the title of the paper might be a little misleading or confusing. How about ‘case continuity’?

10. In the Background section (final paragraph) there is reference to a ‘steering’ effect. I assume this is ‘steering’ the participants, but this could be clarified.

11. I assume that the satisfaction survey was exactly the same for both sets of respondents? This could be made explicit.

12. Are the correlations found statistically different from zero? Some p-values might help, or some discussion of why they were not included.

13. Why not spell out in full the inter-rate reliability results, perhaps in a small table?

**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.