Reviewer's report

Title: Simulated Patients' ratings of empathy in medical student final year clinical examination. Are they useful?

Version: 2 Date: 6 May 2014

Reviewer: Matt Homer

Reviewer's report:

Major Compulsory Revisions
Could there be some additional discussion of how these empathy scores should be used – as feedback to students, as part of summative judgments or what?
Wherever summed scores are used it would help to have a value of Cronbach’s alpha – to help judge the extent to which the items are internally consistent (although I suspect some of the values might be low due to the relatively short length of some of these scales).
Page 5, line 204 – can we have more details of how the cut-off is made between those having to return and those not?
Page 6, line 258 – with regard to the intra class correlation coefficient – can we have more details of this? There are many types of ICC and it would help the reader to know which was used and how it was calculated.

Minor Essential Revisions
The five stations were an OSLER I believe. Can this term be used throughout – e.g. in abstract under methods – this will help the reader understand the design?
Page 5, line 188 – why no empathy ratings for the OSLER?
Page 5, line 219 – it would be more consistent (and better practice) to give the n and the p-value for all correlations.
Page 6, line 252 - Whilst the correlations are in themselves effect sizes, I would welcome the addition of effect size measures for the t-test (e.g. a point-biserial correlation or Cohen’s d).
Page 7, line 291 – ‘some indication of construct validity’ - it is not clear to me (but I could guess) which construct is being validated (OSCE, OSLER, empathy?). Could this be clarified?
Page 7, lines 293-308 – it would be nice to have some indication as to why a standardised measure was not used alongside the empathy items employed in this study. Then you could have been really sure that you were measuring empathy, or at least, the same thing that others had measured and labelled as such. Whilst we are on design issues, why did you not use the same assessors for the OSLER and the OSCE – this would have, to an extent, controlled for some additional sources of error in the study?
Page 4, line 159 – ‘fo’ should be ‘of’. 
Figure 4 should be Figure 3 throughout.

Discretionary Revisions

Nothing else to add.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests