Reviewer’s report

Title: A randomised trial of the influence of racial stereotype bias on examiners' scores, feedback, and recollections in undergraduate clinical exams

Version: 1 Date: 08 Aug 2017

Reviewer: Karen Stegers-Jager

Reviewer’s report:

In my opinion, the authors have done a great job in responding to my own and the other reviewer's comments. I have a few comments:

- With respect to my comment on contextual differences: overall I am satisfied with the explanation added by the authors (lines 5-12). However, research has shown - at least in the job application context - that more experienced raters tend to rely more on intuitive judgments and hence might be more at risk for bias rather than less compared to less experienced raters. Therefore, I'm not sure whether more experience per se might lead to less stereotyping (although a tendency to individuate might do so).

- With respect to my comment on the reliability of the mark sheet: I was referring to inter-rater reliability - the degree to which different judges or raters agree in their assessment decisions. My comment was invoked by the data presented in the table at the bottom of page 24, which suggest that the panel of 8 reviewers rated the videos as less similar with respect to communication than with respect to knowledge. This finding would be in line with other studies that have suggested that the inter-rater reliability for communication scores is (often) lower than for knowledge scores. So, I was wondering whether the fact that raters tend to disagree more on scores for communication could somehow have influenced the findings of the current study.

- With respect to my comment on the test of stereotype activation: I am very happy with the addition of this issue to the limitation section. I also agree with the authors that this does probably not explain the lack of an effect of stereotype activation on scores or feedback. However, I still think that it might explain the lack of differences between the two groups. The authors hypothesized that Group A would have faster response times to stereotypical words than Group B, but this hypothesis was not confirmed. If the stereotypes are already active before completing the task, it might be that the exposure to three student performances is not enough to make a difference between the two groups.

- With respect to my comment on the description of the results on the influence on examiners' memories: to me this section is clear now.

- With respect to my remark regarding BAPIO vs RCGP and GMC: although the sentence is much clearer now, I would still suggest the authors to explain the acronyms for non-UK readers.
As an additional question: what happened with the data of the participants who guessed the study's true purpose? It appears as if no action was undertaken. Would it not have been better to exclude their data?

As a smaller remark: the authors may want to take another look at lines 31-34 at page 4, since both sentences start with "as a result".

**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.