Reviewer's report

Title: Investigating possible ethnicity and sex bias in clinical examiners: An analysis of data from the MRCP(UK) PACES and nPACES examinations

Version: 2 Date: 21 May 2013

Reviewer: I Bartman

Reviewer's report:

Review 'Investigating possible ethnicity and sex bias in clinical examiners: An analysis of data from the MRCP(UK) PACES and nPACES examinations '
I C McManus, Andrew T Elder and Jane Dacre

I enjoyed reading this article. I think that the study is very interesting. However, as mentioned by the authors, there are very few clinical skills examinations with two physician assessors. There are some oral examinations with more than one examiner but I am not aware of OSCE type of examination with two examiners present in the room. Since PACES use a mixture of surrogate patients and real-life patients it resembles more a traditional OSCE than an oral examination. However, this is one of my comments to the authors, the paper needs a better explanation of the PACES. PACES may be well known to readers in Europe but not outside of the continent. There were some gaps that need filling in.

The overall conclusion, i.e., “no found bias towards individuals with particular characteristics” is reassuring and important to all stake holders. Yet, the fact, that this study required two physician examiners and a specific program written in Matlab, means that the methodology described by the authors cannot be easily adopted by other testing institutions.

However, keeping in mind the limitations I believe that this is an article of outstanding merit and interest in its field.

Minor Revisions but needed:

Page 3: better definition of the level of knowledge tested at PACES. It says: “postgraduate clinical skills”, but what level: beginning of post grad training?, the end of it?

I had to research PACES to get a better understanding of it. It seems that this is an OSCE type of examination, i.e. an SP is present. A better description of the exam is needed.

Page 11: I find this statement too strong: “It is not at all clear how statistics solutions can be developed to the problem of identifying examiner biases with only a single examiner at each station.” The authors carried on two quick studies based on one examiner but did not consider longitudinal studies where examiners performance could be assessed through a number of years and grand
larger data sets.

- Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Page 8 second paragraph: the word “reached” is repeated.

Suggestion for the future.
The authors did not look at the ethnicity and gender of the patient. Perhaps characteristics of the patient play role in examiner bias?