Reviewer's report

Title: The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: An analysis of MRCP(UK) written examinations, 2002-2008, and Specialty Certificate Examinations

Version: 1 Date: 13 February 2010

Reviewer: Tanya Beran

Reviewer's report:

This paper compares the 'reliability' of reliability to standard error of measurement. The conclusions in this paper have significant implications for high stakes testing in medicine, and, thus, are very appropriate for this journal. Some suggestions are made below to strengthen the quality of the paper.

The title can be shortened.

The third line on page 5: candidates should be candidate's

 Conjunctive adverbs need commas around them when in mid-sentence.

In appendix a the sentence after the 3rd formula should read: As a result, the equation can be simplified to provide an approximation to SEM, which, to avoid confusion, is called SEM*.

There should be a space before and after all = signs.

In the first formula in appendix b there is a period after 10/9

In the fifth last line of appendix b there needs to be a space between remains non-zero

In the first paragraph on p.6 the authors state that reliability is not the sine qua non of an assessment. Given that reliability and validity are often presented hand-in-hand when discussing psychometric properties of tests, it is important to briefly mention the importance of validity in determining the 'quality of the assessment', as the authors indicate.

On p.7 'very variable' should be reworded to 'highly variable'. There should be no punctuation in the header on p.7.

On p.7 the first line of the last paragraph: 'radically' needs to be re-worded.

On p.10 in the second line: 'in figure' can be deleted. In the fourth line there are too many periods. In this same line there should be no comma after seen.

On p.10 fourth last line of first paragraph: (929 is an error. These types of errors occur throughout the paper, requiring a careful proof reading.

The statement on p. 10 that 'something has to be wrong' is overstated and needs to be clarified.

The first paragraph of the discussion is a run-on sentence.
Given the recommendation of the use of SEM over reliability in the conclusion section, it may be worthwhile in the discussion section to compare the interpretation of reliability and SEM. That is, people easily understand the range of reliability coefficients (0 to 1.00 with higher reliability closer to 1.00), but how can people interpret a SEM of 3.06 for example? Some guidance to the reader here may encourage its use.

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

**Quality of written English:** Needs some language corrections before being published

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

**Declaration of competing interests:**

'I declare that I have no competing interests'