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: 22 February 2010

Reviewer: Kent Hecker

Reviewer's report:

This manuscript outlines an often misunderstood concept; the relationship between SEM, SD and reliability (in this case either coefficient alpha or a correlation coefficient).

Minor Essential Revisions

For completeness of the article I would strongly encourage the authors to:

1. Provide further information regarding the range of restriction problem, a restriction of range algorithm as well as a worked example showing the change in the reliability coefficient for the values from the Monte Carlo experiment. Given the values provided, I calculated an adjusted correlation coefficient as being 0.87. This should be included on page 10 and 11. It would nicely reinforce your arguments regarding the relationship between SD and reliability coefficients.

2. Provide a description of how to interpret the SEM. While most people understand that reliability is reported from 0-1, with a value closer to 1 as being more reliable, what do the percentages mean that are calculated from SEM?

3. Carefully proof read and edit the document, for instance:
   a. Being a non European I do not understand the term “diet” first found on page 7. I think that this paper will have a wide readership throughout the world and therefore would like to see this term explained.
   b. SCE’s? What are they? Specialty certificate examinations? Pg.8
   c. There are two page 2’s at the end of the manuscript. One before the reference list and one being the 2nd page of the appendix.

4. Finally, and this is a subjective comment, I would recommend that the point can be made for calculating and reporting SEM using the PMETB document as an example, not as a criticism of the PMETB document. Through the discussion, I found that I was distracted by the comments regarding the PMETB documents, which became discouraging. I would encourage the authors to stay true to the message of the use and rationalization of SEM and remove/rephrase paragraphs that are distracting, such as the paragraph at the bottom of page 15 and top of page 16 as well as the comments made in the final paragraph of Appendix 2.
Level of interest: An article of importance in its field

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

Statistical review: No, the manuscript does not need to be seen by a statistician.