Reviewer's report

Title: On line clinical reasoning assessment with the script concordance test: results of a french pilot study

Version: 1 Date: 27 June 2006

Reviewer: Clarence Kreiter

Reviewer's report:

General

-----------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Review for Sibert et al. article.

This study examines SCT data collected from a Web-based SC exam administered to both an expert group (reference panel) and to examinees with varying levels of experience. Significant differences between scores were found only between the students (the lowest group) and two top qualified subgroups.

The study does not address new research questions.

On page 17 they say the “test can be used in situations where there is no consensus among experts, it the literature or in daily practice.”

I think it would be helpful for the authors to describe in what context a test composed of such items might be useful. They imply high-stakes exams on page 10. Do they suggest then using it in licensure or competency testing? It would be useful to discuss how these items might be used.

Page 5 “The authors state “most “(PBAs)“are only a measure of behaviour.” Do they implying that measuring behavior is less useful than measuring a precursor to behavior? Perhaps they meant to imply that current PBAs are measuring trivial behavior. If so, they might be right, but they need to elaborate.

Page 7 “I am unsure why the diffusion of the SCT on a large scale will permit the researchers to "confirm the utility as a strategy for investigating the process of decision making." However, it is reassuring that here they seem to imply that the SCT is a tool for researching or investigating decision making rather than educational assessment.

Page 10 “Why do the authors seek a “norm referenced database for the reference panel to derive a scoring key? For example seeking a broad range of training levels makes little sense. I am not sure but if I understand what they seek to do with an SCT aggregate scoring key, they should rather characterize a group of experts with consistently high levels of experience and training.

The new territory described in this research report relates to the new SCT web-based testing software. However this article and research design does not focus on evaluating or describing the impact of the new delivery mode.

Other concerns relate to the fact that 45% did not complete the exam. How were missing values scored? Was the percentage constant across training levels? If not, could the differences in scores by level be explained by level of volunteer examinee commitment to performing well on the exam? This is an especially important question since it was only students who scored significantly different and other groups.
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:
I declare that I have no competing interests