Reviewer's report

Title: Analysing clinical reasoning characteristics using a combined methods approach.

Version: 3 Date: 20 May 2013

Reviewer: Mike Tweed

Reviewer's report:

The major revisions required are:
1. an explanation of how the questions were presented and responses analysed in light of this.
2. an conclusion related to the use and development of a 2 step procedure

1. Is the question posed by the authors well defined?
The authors use the introduction to discuss clinical reasoning, its assessment and the use of SCT and CRP and asking whether a two stage approach would prove "valid, reliable and targeted".
In the intro para 3 the statement that "both methods have demonstrated reliability and validity" needs reference and explanation. As validation is the summary of evidence supporting the use of results for a purpose, a test can produce results with validation evidence for one purpose and not another.
In the intro para 5 sentence 2, what does “them” refer to? If the questions- the question do not identify the weaknesses, but interpretation of results.
It appears that the 2 stage process is using one test as a screen and another test and confirmation.
In intro para 8, the last statement which is vital to the research needs justification.
In intro, last para- does this valid refer to validation as a tool to guide learning or as a tool to make progress decisions?

A comparison of responses to questions of different format with evidence to support the use in the assessment of clinical reasoning is a topic worthy of research. Also the use of a 2 step assessment process with a mass screen and a more accurate assessment is worthwhile.

2. Are the methods appropriate and well described?
I was not fully able to follow the method and hence the results.
I believe that for 20 topics a SCT and an equivalent CRP were developed. Candidates at different levels of training than sat exams made up of these questions. I thought that there were 2 sets of questions, one were the odd ones were SCT and the even CRP and the other visa versa. However on reading the results I cannot see how this was the case.
There are different scoring systems for SCT and this will influence internal consistency and correlations. Could this have affected the results?

Construct validity was evidence by correlation between total scores and subscale scores. It would have been useful to analyse the d-mark to demonstrate that the correlation with the f score was greater than that for the d score. How much of variance in the total score is due to variance in d and f score? These are unlikely to be independent.

The use of the analysis to demonstrate use in terms of a 2 step assessment (the purpose the research) is based on the application of a notional pass mark. If this mark is to be used it must be justified. Different cut scores will produce different results in terms of utility of the 2 step process.

3. Are the data sound?

The presentation of the results is where my interpretation of the method is unclear.

I thought that all candidates did CRP "and" SCT not "or".

How many incomplete responses were there?

Was the data normally distributed within group or across all groups combined?

One might have expected a bimodal- students and practitioners distribution.

The results for total scores and sub score correlations demonstration as noted already would be helpful.

Given the aim to develop a 2 step test as an aim, presentation of results in view of a 2 step test would be helpful.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?

I would need to understand how the questions were presented and therefore how the correlations and analysis could be interpreted.

5. Are the discussion and conclusions well balanced and adequately supported by the data?

Discussion of content specificity- weren't the questions the same? Or does this mean extrapolation beyond the 20 topics.

Although the results correlate given the notional nature of the pass mark means that pass fail decisions might not be meaningful.

The authors conclude that SCT is a useful screening test and CRP would make for a useful comprehensive test, but I cannot see how this can be concluded from this data.

6. Are limitations of the work clearly stated?

Limited numbers and the lack of use of a measure of clinical reasoning are mentioned. Limitations in terms of the notional passmark, and investigating the 2 step proposal is not mentioned.
7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?

Research into clinical reasoning assessment. The use of SCT and CRP especially with regard to pass-fail assessments rather than purely formative assessments might be added. As the aim was related to development of a 2 step process, this might be expanded further. Hence the use of a screen tests and confirmatory test would be required.

8. Do the title and abstract accurately convey what has been found?
Yes, once the issue of how the methods and results link is explained.

9. Is the writing acceptable?
There are some minor typographical errors that will no doubt come to light in editorial review. Eg missing bracket, missing “s”.

**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