Reviewer's report

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

**Version:** 3  **Date:** 3 June 2013

**Reviewer:** Bernard Charlin

**Reviewer's report:**

Report

The title describes well the content of the paper.

This paper is well written and it addresses an important theme, the use of tools to assess the quality of clinical reasoning processes (and not clinical reasoning outcomes). That kind of tool has the potential to help teaching clinical reasoning and also diagnosing of students / physicians in difficulty of reasoning. The research theme is a welcome preoccupation given the current state of literature and ideas about clinical reasoning.

The abstract,
- the tools are first named CRP and SCT, then authors speak of SCT1 -2 or CRP 1 or 2, while these distinctions are not yet explained. This is confusing.
- It would be interesting to precise within the abstract the number of participants in each group.
- Method section: tools were administered on line. This should be specified
- The method to reach pass-fail decisions are not mentioned in the abstract, while this is an important part of the analyses
- Cronbach alpha: I suspect that testing time was very different for the two tools. Specifying it will be very important to interpret the values of reliability coefficients and the difficulty to detect differences between groups.
- I am not sure that the last sentence of the result section should be kept in the abstract as I suspect that the difference is an artefact related to the difference in testing time

The introduction

It presents well the two tools and sets well the research aim, but the research questions and the way the research aim will be addressed should be specified.

Methods:
- Subjects do not represent cohorts but rather groups.
- I don’t understand what is the CRP1/SCT2 stream and the SCT1/ CRP2 stream”. Explanations come later in the text. This should be explained sooner.
- Mean time needed by participants to complete both tests are needed, as reliability coefficients (and other conclusions of the paper need to be interpreted in light of testing time)
- Methods used for establishing the passing scores need to be better defined.
- What is meant by a “set” is unclear at the beginning. A set means for CRPs 10 scenarios. The authors use test material described in another study.
- The method of creation of SCTs in not specified and how they were split in two sets is not explained. This is needed.
- The way the authors will decide on the respective value of both tests should be better specified.
- My understanding is that the authors used material developed for a previously published study. They did split this material in two sets of 10 (CRP1 and CRP2). They then constructed SCTs from the CRPs scenarios. They then obtained a SCT 1 derived from CRP1 and a SCT2 derived from CRP2. They then delivered to participants CRP1/SCT2 or CRP2/SCT1 to avoid having participants answering the same cases in two different formats. If this is the case this should be better explained for readers.
- By way of test construction taking the SCT format for each case is much shorter for SCTs.
- Answer keys: 21 Australian GPs were asked to take the test. Their answers were used to construct the answer keys. The method used to get that is not explained.

Results:
Subjects: 17 GPs, 202 students within two Australian universities agreed to participate. 12 GPS and 119 students did participate. CRP1/SCT2 stream (8GPs, 20Y4 and 44 Y3); CRP2/SCT1 stream (4GPs, 22 Y4, 33Y3). In complete responses -# analyses on 131 Sets of CRPs and 74 sets of SCTs. My understanding is that the 131 sets of CRPS are made of the 12 GPs and the 119 students” It is unclear why there is only 74 sets of SCTs.
Data are not given for the exact testing time but an estimation is given: 90 minutes for CRPs and 30 minutes for SCTs. This is a huge difference and results should be interpreted in light of these differences (Cronbach alpha and capacity to detect differences in groups of experience means). This should be clearly expressed in the discussion (and even may be in the abstract).

Data analysis:
A statistician review might be needed as I wonder if correlation analyses should be adapted, given the value of reliability coefficients.

I don’t understand the procedure that was used to determine passing scores for both tests. This needs to be better explained.

In the data analysis I wonder why data were not regrouped to make the computations (CRPs1 + CRP2; SCT 1+ SCT2) as the main goal is to compare the tools. This would make the table of data analyses (including the ANOVA analyses) easier to understand for readers.

For instance when all cases were regrouped, Cronbach alpha reached 0.93 for the CRPs and 0.63 for the SCTs.

The whole discussion on the capacity of discrimination for CRP not present for SCTs should take in account that one test tooks 90' and the other 30'. Most publications about SCTs speak of tests lasting about 60'.

Discussion: I didn’t comment on the discussion as I think the huge difference in testing time has a huge impact of interpretations of data.

As a whole a very useful paper that published once important modification have been realized.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I declare I have no competing interests