Reviewer’s report

Title: Developing a viva exam to assess clinical reasoning in pre-registration osteopathy students

Version: 2  Date: 17 June 2014

Reviewer: Matthew Sibbald

Reviewer’s report:

This is a well written paper describing a test of clinical reasoning for osteopaths. The authors test fourth and fifth year students, doing careful statistical analysis describing the reliability and correlation of testing content. It is a multicenter study and well grounded in the literature. While this certainly adds to the field, I highlighted a few areas which could be improved, described below.

Major compulsory revisions

1. The authors describe their objective as the creation of a test of clinical reasoning based on a script concordance test approach. As described by Charlin in the reference provided, the SCT is a written test framed by an ambiguous clinical stem, wherein candidates are asked to make microdecisions about small pieces of additional information in a hypothesis driven manner, which are compared to a panel of experts. Eg. If you were thinking of heart failure, then finding an elevated JVP would make this diagnosis much less likely / less likely / unchanged / more likely / much more likely. The authors described long and short cases which are assessed on a 7 attribute rubric of clinical reasoning, which seems quite different than a SCT. I would suggest either removing the description of the test as based on the SCT or making a clearer argument for why it resembles this form of testing.

2. The authors point out the lack of ‘generalizeable problem solving skills’, noting that clinical reasoning is content dependent. However, as far as I can tell, students were only tested on two different content areas. Can the authors comment on whether or not this is sufficient this in the limitations?

3. As this paper explores the validity of clinical reasoning testing based on cases, establishing the content validity of the case material would seem important. The authors refer to adaptation of pre-existing cases used at two Universities. How were these chosen? Were these cases commonly encountered in clinical practice? Would experienced clinicians be able to navigate them without difficulty? Could the authors describe the development of the content in more detail?

4. While the degree of consistency between scores on questions is reassuring, it does not speak to the ability of the assessment tool to discriminate clinical reasoning ability - separate those with good abilities from those with poor abilities. Could the authors estimate the magnitude of the difference in scores
between fourth and fifth year students (e.g., calculate an effect size from the F value) and comment on its clinical significance? Are there any other comparative standards which could be used to assess the test's ability to discriminate the failing student from the marginal student from the competent student?

Minor essential revisions

1. The authors mention that "a non examining member of the research team acted as an observer at each location and recorded comments reflecting the consistency of the process". Can they report on what comments were made, and how consistent the process was?

Discretionary revisions

1. The scoring rubric table mentions "failing" score for several questions. The idea of using this type of assessment for a "high stakes" judgment is also alluded to in the conclusion. Given that the validity of assessment tool is tied to its purpose, it would be important to identify in the manuscript the purpose of the assessment. Were the authors intending to use the assessment for summative decisions, and if so, how did they propose to set and defend a standard?

2. Determining reliability frequently involves calculating intra-rater and inter-rater reliability statistics. Can the authors comment on either of these attributes? Is the comparison between examiners in the results section a comparison of aggregate scores, or actual duplicate assessments of the same student and same case?

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

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I declare that I have no competing interests