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

Authors:

Paul Orrock (paul.orrock@scu.edu.au)
Sandra Grace (sandra.grace@scu.edu.au)
Brett Vaughan (brett.vaughan@vu.edu.au)
Rosanne Coutts (rosanne.coutts@scu.edu.au)

Version: 3  Date: 20 August 2014

Author's response to reviews: see over
Developing a viva exam to assess clinical reasoning in pre registration osteopathy students

Response to reviewers

Reviewer's report 1
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.

Author response: The oral exam developed is similar to the script concordance test in that it is case scenario based, cases were developed from authentic clinical experience, has uncertainty embedded, focuses on data interpretation and follows the basic graduated levels of a case with examiner prompts. Because it is not using the full structure of the SCT as the reviewer mentioned, the statement has been removed from the manuscript.

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?

Author response: The study did have the limitation of two cases, but the cases chosen have more than one “content area” – there are cues in each case leading to divergent issues (e.g. psychosocial aspects). This required students to be able to reason in complexity. We agree that more cases would improve breadth of content. This is planned for the next stage and will be noted in manuscript.

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?
Author response: A bank of cases had been developed by SCU and VU. At SCU the content validity of cases developed by osteopathic staff was confirmed in a series of focus groups of experienced clinicians; at VU via a reference group from multiple clinics and universities. The cases used in this study were chosen from the bank of cases based on their complexity and presence of cues to other systems and possibilities. Simple acute or uncomplicated red flag cases were avoided (this will be added to manuscript).

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 (eg. 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?

Author response: The partial eta squared for the between year levels result has now been reported and interpreted. The year level accounts for approximately 11% of the variance in the total score. Further work is required to establish whether this result is generalisable. The ability of the examination to discriminate between competence levels will be the subject of future research.

Minor essential revisions
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?

Author response: There is now a report on these observations in the manuscript.

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?

Author response: Yes, the plan is to use this as one of the summative clinical exams – following further development in validity testing. The rubric performance attributes were based on the established learning objectives of 4th and 5th year clinical students in a previous benchmarking exercise and this has now been added to the manuscript. The marking standards for each attribute were established during the
case development. Content validity of the cases was established through the development phase of each scenario (reviewed by experienced osteopaths and educators) as well as through further refinement after each case had been used in medium or high-stakes clinical competency assessments.

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?

Author response: The design of the study was such that it does not allow for the calculation of inter- or intra-examiner reliability. This aspect of the exam’s development will be addressed in subsequent research. The comparison is based on the aggregate scores.
Reviewer’s report 2
Minor Essential Revisions
In the abstract, the background and methods sections are too long and not entirely well focused. They would benefit from being written in a more concise way; this would enable readers to have a clearer idea of the background, gap in the literature and core utilised methods and methodologies.

*Author response: Abstract has been rewritten to be more concise.*

In the introduction, there are a number of points which benefit from editing; appropriate terminology should also be consistently used. For example,

in line 3, page 5 it would be better to refer to clinical reasoning as core capability.

*Author response: done*

In line 6, it is important to consider that clinical reasoning is not a cognitive process but involves a series of cognitive processes such as reasoning, problem solving and decision making.

*Author response: this line has been altered*

In line 9, the authors refer to metacognition as reflective thinking; this is not entirely correct and it is suggested that the authors consider the work of e.g., Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American Psychologist, 34(10), 906-911 - on metacognition.

*Author response: Thank you - we have now read Flavell’s work, and the implication that metacognition equates to reflective thinking in line 9 has been removed.*

In lines 19-22, the concepts of HD reasoning and pattern recognition are not appropriately introduced; it may be useful to link HD to analytical processing and pattern recognition to non-analytical reasoning i.e. system 2 and system 1 in the dual processing model.

*Author response: As suggested, lines 19 – 22 have been amended to clarify the links between Systems 1 and 2 in dual process theory and the two approaches to clinical reasoning.*

In lines 17-18, page 6, it reads as if Vygotsky advocated that. Although he clearly talked about scaffolding and zones of proximal development, it is important that the way in which this author is cited does not lead to a misattribution of this claim/argument to this author. In the general, the interdiction could be more concisely written to enhance clarity and readability.
Author response: We agree that the way in which the reference to Vygotsky was used was misleading and consequently been removed.

In the methods section, it is unclear whether (on page 12, lines 1-3) students may have seen these cases as examples of past papers. This needs to be clarified because it would negatively impact on the validity of this assessment approach.

Author response: The following has been included in the manuscript: The cases used in this study were new to the students who were examined. There were example cases given to students so they understood the examination process.

In the discussion section, page 20, lines 23-25, the authors argue that students at point of graduation are more likely to have developed metacognitive (typo there – author note - corrected) skills. This argument needed to be linked to relevant literature. For example, Spadaccini and Esteves (in press) in a small scale study, have recently found that although graduating osteopathy students displayed significantly more analytical decision-making than their novice peers; however, reflective thinking dispositions did not change with increased exposure to osteopathic education.

Author response: The Spadaccini and Esteves paper has been added. The following has also been added to manuscript: There is debate regarding whether metacognition is developed from general experience or from specific contexts of learning (Krätzig GP, Arbuthnott KD. (2009) Metacognitive learning: the effect of item-specific experience and age on metamemory calibration and planning. Metacognition and Learning 4(2);125-144.),