Reviewer's report

Title: Evaluating professionalism in medical undergraduates using selected response questions: findings from an item response modelling study

Version: 1 Date: 18 April 2011

Reviewer: Sandra Gibson

Reviewer's report:

Review of manuscript
Authors: Paul A Tiffin, Gabrielle M Finn and John C McLachlan
Title: Evaluating professionalism in medical undergraduates using selected response questions: findings from an item response modelling study

General Comments
I agree with the authors methodological approach and think that the reporting of the use of IRT in medical knowledge assessments is an area that warrants publication and further study. With the upcoming importance and use of SJTs for undergraduates it is timely that this paper has been written to question the validity of assessing professionalism using current item types. Although this paper does not directly evaluate the use of SJTs it does raise fundamental questions relating to the application new item types in high stakes assessments and without fully exploring their validity in new populations

I have outlined a number of questions and comments below for the authors.

• Major Compulsory Revisions
  1. Background; Although the authors have discussed observation of professionalism, I was unclear of context. Is this workplace based observation or/and in OSCE? It would be helpful if this could be expanded as the relative reliability and validity will vary between the assessment modalities.
  2. Data analysis; please can the authors comment on the format of the EMQ and whether the stems are locally independent and how this may impact on the underlying assumptions of the Rasch model.
  3. Discussion; para 1. I agree with the finding that SRQs were not in this case appropriate measures of measuring (or predicting performance) but feel that the argument to support this would be strengthened by comparison to an additional item type, assessing integration of knowledge as well as “pure” knowledge. Although this may be because of my unfamiliarity with the anatomy questions.
  4. The conclusion in the abstract and the main paper place different weight to the findings as to whether SRQs ‘have’ or are ‘likely to have' poor psychometric properties. I would suggest as there are a number of limitations in the study and that the issue of sample size, although well justified, remains contentious that the later is more appropriate.
• Minor Essential Revisions
1. Reference 5, Yates and James requires the year to be added 2010.
2. Background para1 line10 accurate to accuracy
3. Data collection para 1 line 6. Is the “classical MCQ” a best of five single best answer MCQ or a variation on the true/false MCQ?
4. Data para 3, line 13, word group missing after tutor?
5. Discussion para 4 line 4, do you mean EMQ or EMI?
6. There are a number of inconsistencies in the formatting of the reference list that needs addressing.

• Discretionary Revisions
1. Background; para 1 line 11 do the authors wish to give an example of the on the implications the think are most likely to arise as a direct result of accuracy of the evaluations?
2. Discussion para 2; Could it be that the lack of clinical exposure in the first 2 years may be influencing the students responses to the professionalism questions, in that they have not had the opportunity to observe professional behaviour in the workplace and so are uncertain of the “right” response in unfamiliar scenarios? This could explain the larger variation in responses given by the students.
3. I suggest it would be useful to repeat the study with and include at least one other medical school that has early clinical exposure to explore the issue of role modelling and situational learning in learning and assessing professionalism.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

'I declare that I have no competing interests'