Reviewer’s report

Title: Predictive validity of the UK Clinical Aptitude Test in the final years of medical school: a prospective cohort study

Version: 2 Date: 31 January 2014

Reviewer: Chris Dewberry

Reviewer’s report:

This paper examines some interesting new data on the criterion-related validity of some methods, and a psychometric measures used for the selection of UK medical students. This data is from two Scottish medical schools. The paper has a number of strengths: it is generally well-written, the data presented appear sound, and the methods adopted are for the most part appropriate. However, in my view the paper would benefit considerably from a clarification of the research question (or research questions) being addressed here, and almost certainly to some revisions of the paper in the light of this.

I have also drawn attention to some other issues which, if addressed, would in my opinion significantly improve this article.

Major Compulsory Revisions

1. I think the authors need to clarify the question they are addressing in this paper. Is the paper about whether or not UKCAT can be recommended for medical selection in the UK? Is it about the efficacy of some (but not all – note the absence of A Levels and Scottish Highers) selection methods which are widely used in medical schools in the UK? Is it about what predicts medical school performance in the later years of training? Is it about predicting a performance criterion (i.e. the OSCE) which has been neglected in previous studies of the criterion-related validity of medical school selection procedures? Is it about the extent to which UKCAT predicts performance better than UCAS forms and interviews do, and if so is this about the particular ways that interviews are carried out, and UCAS forms evaluated at Dundee and Aberdeen medical schools – or more generally? At the moment the paper, at different points, seems to be concerned with any and all of these issues. I am concerned that in attempting to address so many issues, many of them very broad, with what is a small-scale study with modest power, this paper ends up addressing none of them properly.

This issue also means that there is a disconnect between certain critical aspects of the paper. For example, the primary justification for the study, discussed in the Introduction, is that it uses a performance criterion (the OSCE) on which there is little previous validity research. But the paper then goes on to include written examinations as a criterion – a criterion on which there is a great deal of existing validity research.
2. In addition, and this is related to the first point raised in this section, I believe the authors should show some caution in generalizing their results, and that they should be careful not to overstate the implications of their findings. For example, a small-scale study such as this cannot, in itself, support the use of UKCAT – not least because the use of this selection method depends not only on its criterion-related validity but also on a host of other issues including cost, fairness, perceived acceptability, the availability of alternative selection methods and so on. A full evaluation of these issues is beyond the scope of this article. Much safer ground is to suggest something like this - that although the study is small in scale, and focuses on just two UK medical schools, it provides some evidence that UKCAT may offer better prediction at later stages of medical school training than it does at earlier stages (or, depending, on the research question, perhaps that it predicts OSCEs better than written exams), and that larger scale research would be helpful to examine this issue further.

Minor Essential Revisions

Page 2, para 3,

3. Replace “UCAS form and interviews scores did not predict examination performance” with “Neither UCAS form scores nor interview scores were statistically significant predictors of examination performance” (or something similar). Note: UCAS and interview scores did predict examination results in the sample data.

Page 3, para 1

4. Change “confirmed that the UKCAT was the most consistent predictor of assessment performance” to “showed that UKCAT made a statistically significant unique contribution to variance in examination performance, whereas neither UCAS form scores nor interview scores did so” (or something similar).

Page 3, para 2

5. The first sentence under Conclusions appears to imply that previous studies of the extent to which UKCAT predicts performance in later years of medical school showed that it did not do so, whereas this study does not. I don’t think that this is what the authors mean, and I would suggest that this sentence is modified.

Page 3, para 2

6. The claim that “this justifies UKCAT’s role in student selection” is, it seem to me, inappropriate for several reasons. To give thee examples of why (1) this is a very bold claim to make on the basis of just two small-scale studies in just two medical schools; (2) unless we no more about the alternatives which might be used (better-designed interviews for example?) the claim is unjustified (3) the extent to which the authors are generalizing their statement too is unclear. Are they saying that this study shows that UKCAT’s role in selection is justified just at Dundee and Aberdeen? At all UK medical schools? At all UK medical schools
irrespective of whatever other selection methods they are currently using? Throughout the world? I think this claim should be withdrawn or considerably modified.

Page 4, para 1

7. Replaced “skewed by” by “associated with”. Cite one or more references supporting this.

8. The sentence beginning “Other traditional methods…” implies that ALL alternatives to A levels in medical school selection have been found to be associated with socio-economic class and have no predictive power. Is this literally true? If so, I think it needs to be carefully referenced.

9. The sentence beginning “Moreover, A-levels” is questionable as it implies that the purpose of A-Levels is to discriminate between the performance of medical students rather than to discriminate between people who do, or might, apply to medical school. I think this should be revised.

Page 5 Para 3

10. The large-scale UKCAT-12 study (McManus, Dewberry, Nicolson, and Dowell, 2013, BMC Medicine) should be added – it has much greater power than the relatively small-scale studies cited here.

Page 6 Para 1

11. Please provide more information about the Yates and James study. For example, where was it carried out, and what year were the medical students in, how many students were included in the study etc.?

Page 7 Para 3

12. I am not convinced of the reason for excluding the portfolio assessment as a criterion in the case of Dundee. In what ways is this portfolio assessment more “inherently different” from written exams and OSCEs than written exams are from each other? And even if it is inherently different (whatever that means) if it is nevertheless contributing to the assessment of the students, surely it should be included here as it provides further evidence of the criterion-related validity of UKCAT.

Page 7 Para 4

13. It would be helpful to have some information about what “EMQs” and “SBA” questions are. It would also be helpful to know how many written questions there were, how many OSCE assessments there were, and other relevant details about the OSCEs such as the number of assessors etc. This is particularly important in relation to the reliability statistics reported in Table 1 as Alphas increase with the number of items as well as with the strength of inter-item correlations.
14. I am not sure why stepwise regression was used here. It is a controversial procedure (see the Tabachnick and Fidell Using Multivariate Statistics textbook for a helpful discussion of this), and it would be safer to use the standard/entry method.

15. It would be very helpful to have a full correlation matrix in Table 3 so that the correlations between the various predictor variables (within each medical school) can be examined as well as the correlations with the criteria. This table should also include gender – partly because at present there is no information at all about the effect sizes of the associations between gender and the criteria. This would make it unnecessary to report the results of the two t-tests on gender.

16. Are the correlations between S1 Written and UCAS form negative for both medical schools – or are the minus signs typos?

17. I also think it would be helpful to include the confidence intervals of the correlations between the primary predictor variables (i.e. UCAS forms, interviews, and UKCAT scores) and each of the criterion variables at the two medical schools. This will give the reader a helpful indication for the range in which the population correlations (these are, after all, the correlations we are really interested in here) fall.

18. I don’t think it is appropriate to exclude a multiple regression analysis (Dundee OSCE 1) in Table 4 just because the zero-order correlation between UKCAT and the criterion was not statistically significant. The strength of this correlation might be due to another predictor variable acting as a suppressor. I would suggest adding this multiple regression to Table 4 – it will add very little to the overall length of the article.

19. Information about the number of males and females (see the next paragraph also) are better placed in the method section rather than here. It would also be helpful to know the means and standard deviations of the ages of each of the groups of students studied.

20. If this study is about the extent that UKCAT makes a worthwhile contribution to the selection of medical students, why have Scottish Higher/A level results (e.g. average of three best grades) not been included in the analyses reported here? After all, A levels and Scottish Highers are regularly used in the selection of UK medical students, and if UKCAT is to be considered worthwhile it must surely show some incremental validity over and above A levels. The only counter I can see to this would be that A levels should be discontinued due to associations with socio-economic status. But that is a complex issue which cannot be properly addressed within the confines of a short empirical paper such
as this.

21. Given that on page 6 the authors criticize Yates and James study for not providing reliability data about examinations, it seems a case of the pot calling the kettle black when they provide no reliability information about their own examinations data!

22. A table with information about the means and standard deviations of all the predictor and criterion variables in each school/year/examination should be added.

23. I think the last sentence on page 10 is confusing. It should either be dropped or re-worded.

Page 11 Para 2

24. The statement beginning “We identified that” needs to be expanded and fully justified, drawing upon the articles cited in the introduction,

Page 11 Para 3

25. I don't think the statement that UKCAT accounted for between 6 and 15 percent of the variance alone or in combination with gender is helpful – for two reasons. First it is factually incorrect (see Dundee OSCE results). Second what we are interested in here is surely the unique contribution of UKCAT – not how well it predicts when combined with gender. Or perhaps not – it depends of course on what the research questions are, and this returns to the first comments and suggestion made in this review about the need to clarify the research questions.

Page 12 Para 2

26. I believe that the authors should draw attention, as a limitation, to the small sample sizes in both the Aberdeen and Dundee studies. The fact that the Dundee study had a smaller sample size than the Aberdeen one here is given too much prominence in my view. The important point is that both sample sizes were small.

Page 12 Para 3

27. Without information about the extent to which UKCAT has incremental validity over A Levels/Scottish Highers, I would suggest that the claim that the study supports the use of UKCAT is too strong. If data on the A level/Scottish Higher performance of the students is available I would strongly encourage the authors to incorporate this in their analysis and to examine the incremental validity of UKCAT. If such data is not available I think the authors should tone down their claim here and draw attention to the need to examine the incremental validity of UKCAT over A levels/Scottish Highers in the prediction of fourth/fifth year medical students in the future.
28. Please provide a reference for the predictiveness of MMI’s.

Page 12

29. If the authors decide, in a revised version of this paper, to continue presenting data prediction of written assessments, I think it would be helpful for them to discuss why the effect sizes of the correlation between UKCAT scores and written examinations found here are considerably greater than in the highly powered analysis reported by McManus et al. (2013) in their UKCAT-12 article in BMC Medicine.

Table 3

30. What does “S1” refer to?

Table 4

31. It would be helpful to include adjusted R squares (these are outputted by SPSS).

Discretionary Revisions

Page 2, para 1,

32. Replace “traditional selection methods” with “traditional UK selection methods”?

Page 2, para 2,

33. Replace “test the relationships” with “examine the relationships”?

Page 5 Para 2

34. Suggest replacing “aptitude required to perform may be very different ” with “knowledge, skills, abilities and other characteristics may different to some extent”.

Page 8, Para 3

35. I found Table 2 a little confusing. I think it would be helpful to amend the table so that the figures relating to students included in the study, and those relating to students excluded, is made clearer.

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