Author’s response to reviews

Title: Predictive validity of a new integrated selection process for medical school admission

Authors:

Paul L Simpson (psimpson@kirby.unsw.edu.au)
Helen A Scicluna (helens@unsw.edu.au)
Philip D Jones (philip.jones@unsw.edu.au)
Andrew MD Cole (a.cole@unsw.edu.au)
Anthony J O’Sullivan (a.osullivan@unsw.edu.au)
Peter G Harris (p.harris@unsw.edu.au)
Gary Velan (g.velan@unsw.edu.au)
H Patrick McNeil (p.mcneil@unsw.edu.au)

Version: 2
Date: 26 March 2014

Author’s response to reviews: see over
26 March 2014

Dear Editor,

We thank the reviewers for reviewing the manuscript ‘Predictive validity of a new integrated selection process for medical school admission’ by Simpson et al for publication in BMC Medical Education. Our responses to the reviewer’s reports are attached.

Yours sincerely,

Dr Paul Simpson
Reviewer: Barbara Griffin

Major Revisions

1. Because this is framed at testing the predictive validity of a new system, then that system needs to be explained in more detail than is currently provided. How is shortlisting for interview made? How is the interview conducted? On a number of occasions the authors mention that tests were used as "filters" - this needs more detail

The authors thank the reviewer for bringing this to our attention. Text has been added (pages 5 and 6) to explain in more detail the new selection system and how the UAI, UMAT and interview contribute to final selection score. The term ‘filtering’ has been replaced with the idea of ‘differentiation’ in order articulate how the selection tools serve to distinguish between applicants who meet the criteria of a high UAI.

At UNSW, an integrated multi-dimensional selection system has been adopted in which performance of applicants in each of the three components of UAI, UMAT and interview, are equally weighted in the calculation of a final ranking (see Table 1). During the year prior to entry, applicants submit an application form and structured curriculum vitae, which asks applicants to respond preliminarily to material that would be explored in greater detail at interview, and which requires the applicant’s school to provide a predicted UAI. Interviews are offered based on the applicants’ known UMAT and predicted UAI outcomes. A second round of interviews is held for applicants whose actual UAI outcomes were much better than predicted by their school. Once all interviews are completed, scores for all three variables (actual UAI, UMAT and interview) are standardised and combined to determine a ranked order of offer for places in the UNSW program.

The interview instrument developed at UNSW is designed around a biographical assessment of the applicant’s life context and experiences which had led to a desire to pursue a medical career, and the motivation for this. Interviewers score an applicant in a range of dimensions, including for example empathy towards others,
communication skills, and coping with uncertainty. An identical process is utilised to select students who meet criteria for entry as ‘rural entry’, although these applicants compete for dedicated places reserved for applicants with a rural origin.

2. The claim the student demographics "should be considered" when interpreting predictive validity needs more explanation. Why would this be so? Rather than just controlling for these factors, perhaps they actually moderate the validity of the three selection tests?

The authors agree that the section on why demographics should be considered when interpreting predictive validity should have been explained better. The following text has been added (page 5):

   Student demographics should also be considered to determine their role in predicting selection tool performance and Medical Program outcomes.

3. The participants section was a little unclear - I was unsure why the 2003 cohort was compared to the 2002 cohort but then the 2004/5 cohorts used for the remainder of the testing. Why not combine the 2003 cohort with the other two where the new selection process was used?

The authors have provided additional text has been added on page 7 to clarify why the 2002 cohort was compared to the 2003 cohort:

   Demographic, UAI, and educational outcome data were also available from 304 students who studied the previous discipline-based medicine program in 2002 and 2003. The 2002 cohort (145 students) were selected through a ranking based on UAI scores alone, whereas the 2003 cohort (159 students) were selected using the new selection process. Analyses of these two cohorts provide some insight into effects of different selection processes for students undertaking the same discipline-based program.
4. Related to 1) above, the issue of direct and indirect range restriction needs greater attention. Most importantly, if there were "filters" used in the new process then arguably there was less range restriction for the TER than there was in the 2002 cohort, which would entirely explain the stronger correlations seen in the 2003 data. A table with means and SDs for all tests for all cohorts is essential for readers to assess range issues. Another possibility for assessing the utility of the new method would be to compare drop out rates from the two cohorts.

The authors agree with the reviewer’s comment regarding the issue of range restriction differences between 2002 and 2003 cohorts. Consequently, Table 4 has been removed and UAI ranges for 2002 and 2003 have been reported on page 13. Discontinuation (drop-out) rates for the two cohorts have also been included (page 12).

5. I do not agree with the author's ranking of schools where selective schools are placed as "moderate" educational advantage. They may have lower SES than private schools, but given that students from selective schools have overwhelmed the top ranking students in NSW for many years, and are even in the majority of selected students in this study, suggests that they are actually the most educationally advantaged.

The authors have reconsidered the categorisation of schools and consequently, have re-categorised schools in light of the reviewer comments (see table 2). The data analysis was revised using SPSS (version 22).

6. Not sure I understand why coefficient alphas are used to examine divergent validity - this is not common practice. A factor analysis of all components would be a better option or just observation of the correlations between the tests.

The authors agree and have replaced the term coefficient alphas with Pearson correlation coefficients (page 12).

7. It does not appear that the authors have corrected for unreliability of the criterion (despite referring to the need for it in the text). Given that knowledge tests are likely to have much
greater reliability than assessments of clinical skills, correction would give a better comparison. Furthermore, in light of the likely significant range restriction of all predictor variables, the size of the correlations is actually quite good and if interpreted as per other selection research (see authors such as Hunter & Schmidt, Lievens, Ones), probably have more utility than the authors indicate.

We agree with the comment that given the complexity of assessing clinical skills (discussed on page 16), the correlations we have reported are quite meaningful.

8. The Discussion would benefit from reference to the need for construct matching when undertaking selection research. In other words, there is no conceptual or theoretical reason for an interview that assesses the constructs this one does to have any relationship to knowledge tests or WAM in phase 1 for instance. It might even be argued that there is less reason for cognitive ability/TER to relate to performance of clinical skills.

The authors thank the reviewer for highlighting this issue. Reference to construct (matching) validity has been added to the discussion (page 18).

Minor Revisions:

1. P.5 first para, "the UMAT ...was designed at UNSW" - might be misleading for some readers less familiar with the context.

Revision has been made to address any potential misinterpretation here.

2. I think TER was no longer used after 2000, when it was replaced by the UAI.

TER has been replaced with UAI.
Reviewer: Annette Mercer

Reviewer's report:

1. I am unsure of any value added to the study by the inclusion of results from the 2002 and 2003 cohorts. This part could be excluded and the paper would still make its main points. The lack of correlation with TER for the 2002 cohort is to be expected due to a severe TER range restriction in this cohort.

The authors thank the reviewer for bringing this to our attention. Additional text has been added on page 7 to explain why the 2002 cohort was compared to the 2003 cohort:

Demographic, UAI, and educational outcome data were also available from 304 students who studied the previous discipline-based medicine program in 2002 and 2003. The 2002 cohort (145 students) were selected through a ranking based on UAI scores alone, whereas the 2003 cohort (159 students) were selected using the new selection process. Analyses of these two cohorts provide some insight into effects of different selection processes for students undertaking the same discipline-based program.

The authors agree regarding the issue of range restriction differences between 2002 and 2003 cohorts. Consequently, Table 4 has been removed and UAI ranges for 2002 and 2003 have been reported on page 13.

2. I consider the small but significant increments in variance from the communication score to be important. This has been seen in at least one other study. The authors have played down this contribution; however as a proponent of the interview I consider this increment to be valuable. I agree with their final sentence.

No revisions undertaken.
3. Interview 6 - cope with uncertainty makes a significant negative contribution to WAM phase 1 and 3 and Knowledge phase 3. This is not commented upon, however it does contribute to the variance.

The authors thank the reviewer for pointing this out and have added comment regarding this significant negative contribution on pages 13 and 15.

4. The authors comment that their classification of school types into educational levels may not have been optimum and I agree with them. This is a weakness which may have had other outcomes, had the categorisation been more mainstream. As it stands it could not be replicated in other states of Australia. Most other states have only one or two selective schools, so NSW is unique in this categorisation.

The authors have reconsidered the categorisation of schools and re-categorised schools in light of the reviewer comments (see table 2). Analyses were conducted again in light of new school categories.
Reviewer: Ian Puddey

Reviewer's report:

1. Is the question posed by the authors well defined? This paper evaluates the predictive validity of an integrated selection process for medical students utilised at the University of NSW. The process comprises a cognitive skills test – the UMAT, a structured interview and prior academic achievement and replaced previous selection via prior academic achievement alone. A series of studies of varying quality and size have similarly evaluated the addition of UMAT and/or interview at other Australian and New Zealand universities but have not previously considered the potential confounding influence of socio-demographic factors. This is a unique contribution of this manuscript. Potential socio-demographic confounders clearly need consideration given the conclusions of 2 recently published articles in BMC Medical Education (1-2).

   The authors thank the reviewer for highlighting these studies. The references have been incorporated into the revised manuscript.

2. Are the methods appropriate and well described? Overall the methods are appropriate and well described. The authors indicate that the broadened selection process was introduced to filter out students with identical prior academic achievement but do not indicate whether threshold levels for entry were applied for each of the selection criteria. Were the approximate 25% of students admitted through the rural entry scheme selected utilising identical criteria to the rest of the cohort or with lower thresholds and lower overall scores? Some clarification on this will reassure the reader that this has not confounded the reported outcomes.

   The authors agree and have added text to clarify the criteria of rural student entry (page 5-6).

   The interview instrument developed at UNSW is designed around a biographical assessment of the applicant’s life context and experiences which had led to a desire to pursue a medical career, and the motivation for this. Interviewers score an
applicant in a range of dimensions, including for example empathy towards others, communication skills, and coping with uncertainty. An identical process is utilised to select students who meet criteria for entry as ‘rural entry’, although these applicants compete for dedicated places reserved for applicants with a rural origin.

3. Are the data sound? The data are sound and have been carefully analysed. However, I am troubled by the uncritical use of data from the 2002 and 2003 cohorts to imply that the integrated selection process introduced from 2003 is a better predictor of academic outcomes than the TER alone. No socio-demographic analysis comparing the 2002 and 2003 cohorts has been done and it is very likely that there will be substantial differences as the authors themselves note on page 4 (ref 18). Such potential differences (eg more students from an East Asian background or more males or fewer students of rural origin in the 2002 cohort) may have dictated the change in profile of TER as a predictor and a comparable regression analysis with correction for these potential confounders to that undertaken for the 2003 and 2004 cohorts is indicated.

The authors have considered the reviewer comments regarding the use of 2002 and 2003 cohort data in the following ways:

- Student demographic data for 2002 and 2003 cohorts has now been included (page 11).
- Table 4 has been removed and UAI ranges for 2002 and 2003 have been reported in the text (page 13).
- As an additional way to assess the utility of the new selection methods the authors have included a comparison of discontinuation (drop-out) rates for the two cohorts (page 11-12).

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
   Yes

   No response required.
5. Are the discussion and conclusions well balanced and adequately supported by the data? The authors reasonably conclude, based on their data, that the assessment of communications skills by interview together with prior academic achievement, are valuable components of an integrated selection process. However, the predictive validity of the UMAT for undergraduate medical school performance has been consistently disappointing in a number of studies in Australia and New Zealand. The authors acknowledge this and have replicated this finding in the current study. The conclusions in the abstract remain ambiguous in this regard, however, and could be read as continuing to support its inclusion in their integrated selection process.

The authors agree with the reviewer’s comments and have included the following conclusion regarding the UMAT in the abstract:

*Results indicate promising validity for an integrated selection process introduced for the Medicine Program at UNSW, with UAI and interview predictive of learning outcomes. Although not predictive, UMAT may have other useful roles in an integrated selection process. Further longitudinal research is proposed to monitor and improve the validity of the integrated student selection process.*

6. Furthermore, without the limited analysis of the 2002 and 2003 cohort that is presented in Table 4, it would be very difficult to support the conclusion that the integrated selection process is superior to the TER alone because the correlation of the final program WAM with the TER \( r=0.448 \) is considerably stronger than that with the integrated admission score \( r=0.151 \) (Table 3).

The authors have introduced a more critical use, and additional data, regarding the 2002 and 2003 data (see above response for reviewer comment 3).

7. Are limitations of the work clearly stated? A paragraph on the limitations of the study (those listed above as well as range restriction, unaccounted confounders, relatively small numbers etc) in the discussion should be considered.
The authors agree with the reviewer’s comment and have included the following text (page 18):

*Findings comparing the 2002 and 2003 cohort data provide tentative support for the use of the new selection methods. Correlation coefficients between UAI and performance scores among the 2003 cohort who were enrolled in the old Medicine Program via the new selection methods were statistically significant for all outcome measures, whereas the very narrow range of UAI in the 2002 cohort selected via UAI scores alone precluded examination of UAI with outcomes. This comparison between students enrolled in the same program yet through different selection methods, combined with lower year 1 discontinuation rates in students enrolled in the new Medicine Program compared to the previous Program, on face value suggests the differentiating function of UMAT and the interview is effective by improving access to a larger range of academically high performing applicants. However, due to the similar conceptual elements present within both predictive validity and correlation analysis, caution is also warranted with this interpretation. Further research would be valuable in exploring in depth the relationship between the original use of selection tools and prediction. Future research would also benefit from undertaking construct matching to assess the construct validity of selection tools [14]. Limitations of the study including, unaccounted confounders, range restrictions in the 2002 cohort, and the relatively small sample sizes within each cohort, should also be considered when interpreting predictive validity findings [21, 37].*

8. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished? Yes. Other recently published work on the predictive validity of the UMAT might also be considered for inclusion and discussion (5-6).

The authors thank the reviewer’s suggested published works and have referenced these publications in the manuscript.
9. Do the title and abstract accurately convey what has been found? Yes.

   No response required.

10. Is the writing acceptable? Yes.

   No response required.

Minor Essential Revisions:

1. On page 3 it is stated “the ability of prior academic achievement to predict performance in the later years of medicine drops to insignificant levels”. This is not supported by our own work (ref 5 in the manuscript)) or that quoted by the authors (ref 10). Most studies find that prior academic achievement remains the best predictor of subsequent academic achievement in medical school although its relative predictive strength may diminish as the course progresses.

   A revision has been undertaken.

2. On page 15 it is stated “One interpretation of this finding is that the university environment provides students with uniform learning opportunities that rescind any educational advantage that may come from secondary school attended. However, there is no evidence that the authors are aware of to support this interpretation.” In fact the phenomenon of a student’s performance being enhanced by immersion in an independent school but with subsequent under-performance relative to students from government schools on studying at university is well described in Australian settings (3-4).

   The authors thank the reviewer for highlighting this literature. We have included reference to these studies on page 17.

3. On page 15 it is stated “students who were born in European/European-derived countries outperformed students born in East Asian and other countries in overall
program and clinical skills outcomes.” For students of East Asian origin this is supported by the data presented in Table 5, but not for those from Other countries.

This has been corrected.
4. P values designated as 0.000 in Tables 3 and 5 need to either be defined or re-written as <0.001.

Revision has been undertaken.

Discretionary Revisions
1. The authors refer to unreferenced anecdotal evidence in relation to justification of the use of the interview in medical student selection, both in the background (P4) and the final paragraph of the discussion (P16). I think this weakens rather than strengthens the arguments put forward and would suggest their exclusion.

The paragraphs referred to have been deleted.