Reviewer's report

Title: Joining the dots: programmatic assessment enhances recognition of factors hampering student progress

Version: 1 Date: 12 April 2011

Reviewer: Adrian Hastings

Reviewer's report:

The paper is a useful contribution to the scholarship around student assessment as it describes implementation of a move towards programmatic assessment and the benefit that it contributes. The summary of the required features of programmatic assessment that they describe in the Background section is very relevant and worthy of wide dissemination.

My interpretation of their paper is that the major intervention has been replacement of a borderline pass grade with conditional pass. The data they present suggest that over time there has been a change in the willingness of their assessors to use this grade. I was unclear whether the development of standards based assessment preceded the introduction of conditional pass (and was therefore in use in the old system) or whether this happened in a parallel process and therefore faculty were adapting to a change in assessment method as well as grade description.

Major Compulsory Revisions

1. Please explain your reasons for choosing one campus to do the before and after analysis, rather than for example a random sample of one third of students from each campus.

2. The authors do not explain what processes they used to inform and/or train faculty to use the new conditional grade. It is necessary to know this as it is possible that greater use of the conditional grade was produced by the training process rather than by the name change.

3. Page 15: First paragraph. Please confirm that the fail rate in year 5 high stakes exam of 7/128 was under the old system and 4/156 under the new one. Also please state percentages as well as numbers; I calculate these to be 5.5 and 2.6%. Also is this comparison valid? If the new system is extracting more weak students prior to their peers sitting the exam then one would expect there to be a difference in the fail rate between successive cohorts.

4. It would be helpful if the authors could be more consistent in the use of the terms professionalism and professional attitudes. At times it appeared that professionalism was about attendance and meeting deadlines but also there is a reference to 'Other professional attitudes'. It would be helpful to know what these are. In Wilkinson's blueprint paper the authors describe a framework for assessment of professionalism (Professionalism can be assessed using a combination of observed clinical encounters, multisource feedback, patients'
opinions, paper-based tests or simulations, measures of research and/or teaching activities, and scrutiny of self-assessments compared with assessments by others.) that does not include what I would prefer to label 'diligence' (attendance and meeting deadlines). Also, the professionalism framework includes a key item that in this paper has been separated away from professionalism (patient interactions).

Minor Essential Revisions

1. Page 7. Final paragraph – briefly outline which recommendations the new system does not act on and why.

2. There are a small number of typos scattered through the paper which need careful proof reading to remove.

Discretionary Revisions

1. Please provide in an appendix a brief description of how you previously aggregated results from assessment of very different natures into a single grade/percentage and how you set a pass/fail threshold for each student’s performance in the yearly in-course assessment.

2. My reason for this request is that one way of addressing the ‘failing to fail’ problem would simply be to raise the threshold. It would also be of interest for the authors in the discussion to explain whether they believe that raising the threshold under the old system would have identified the same, similar or a different group of students who had failed the year.

3. In the background they refer to ‘feed forward’ controversy and allude to it in the discussion. What has been their experience of informing faculty in advance about problems with students and has it allowed them to develop an opinion in favour or against ‘feed forward’?

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests