Reviewer's report

Title: Cumulative assessment: strategic choices to influence students' study effort

Version: 2 Date: 25 October 2013

Reviewer: Godfrey Pell

Reviewer's report:

Before I start my review, I have a long held belief that in many programmes students are over assessed, this view is evidence based, and I have published on the issue, albeit some years ago. So in essence I am saying that I generally disapprove of the policy of frequently repeated testing; this review should therefore be read bearing in mind my stated viewpoint.

W.r.t the suggestion that score is proportional to effort p4 our experience is that effort is only part of the issues, study strategy is an important part of the mix; we regard this as so important that this topic now forms a significant part of the early syllabus at our institution.

To state that as a result of frequent testing of the same topic mediocre students will improve their score is stating the obvious. In stating that this methodology benefits poorly performing students, the authors have not demonstrated that this improvement in score represents deep learning, and improved performance as a medical practitioner. Further more it is not possible to state that this method is not detrimental to the better performing students, there is an opportunity cost to all this testing, what would students do with their time if they were not being tested? What is the effect of repeated s.b.a. testing on learning strategies? None of these has been discussed, so in the absence of any sort of control the authors’ assertions cannot be justified.

I think that to be publishable the authors need a much wider discussion as outlined above.

Secondly it would be of interest if the authors separated out student performance of the different phases of the programme. For example how do both groups of students perform for the repeated topic questions vis a vis the new questions. In other words is the improvement of the scores of the lower quartile students purely in the repeated questions, or is there some improvement in the performance in the responses to questions on the new topic? Does the performance on the former questions improve and the latter decline relative to the top quartile?

The issue of question difficulty is vexed, I assume here that the authors are giving the facility index. The same question if given a position in a different test, say T1 with one cohort athen T2 with a different cohort would have different facility metrics. So the concept of difficulty expressed by the authors is relative, this needs more discussion
I have no problem with the stats analysis as far as it goes.

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

The paper is well written and the English and the development of the topic are fine.

I declare that I have no competing interests