Reviewer's report

Title: Integrated online formative assessments in the biomedical sciences for medical students: effects on learning

Version: 1 Date: 19 August 2008

Reviewer: Edward Palmer

Reviewer's report:

The paper is well written and addresses important and topical issues in assessment. Arguments are made convincingly and are well supported by literature and the evidence from the study.

Tables would typically have a summary of the statistical tests used to analyse the data as part of the legend.

Comments on the paper are below.

p4 Web-based formative assessments also support equity and inclusiveness by allowing students to attempt each assessment anonymously on multiple occasions, without fear of adverse consequences on account of any misconceptions that might have been revealed.

Minor Essential Revision
The authors could elaborate on the stated benefits for students above the fear of being found wrong. Equity and inclusiveness go beyond this issue.

P6. One of the authors (GMV) developed integrated formative assessments in the biomedical sciences with automated individualized feedback, which were embedded in each of the sequential 8-week courses in the Medicine program

Minor Essential Revisions
How many exercises for each week? This is an important point, as it illustrates the amount of effort students may feel the assessment requires. It also allows for comparison with other studies, where the number of formative assessments has been stated.

Additional information that would be helpful summarized here is the total number of students in each year and each course. Percentages of students who completed the formative assessments would then be useful additional data. The response rate to the surveys implemented is essential information in the legend for the figures.

P8. in Course 3 there was a trend towards a positive effect, p=0.077

Minor Essential Revision
This is a convention that is tempting to use and is often used to suggest that “it
nearly made it” I am not comfortable with this use of significance and would recommend the authors simply state that it is not significant.

Discretionary Revision
Table 1 displays all results in the negative. My personal preference, and one supported by the way the text is worded is to use positive numbers comparing use of formative assessment vs no use rather than vice versa.

Minor Essential Revision
The multiple attempt data is interesting and is unfortunately not elaborated upon. Perhaps the authors could devote a sentence or two suggesting why second attempts provided no additional benefit.

P9. For all courses, performance in formative assessments had highly significant correlations with EOC examination marks (p<0.001)

Minor Essential Revision
Significant but moderate correlations would be better phrasing.

P9. Improvements implemented as a consequence of student feedback obtained in 2005 and 2006 are likely to have led to significant improvements in the positive perceptions of formative assessments

Minor Essential Revision
perhaps the authors could summarise these improvements

p10. However, there was a tendency for students in the first year of their program to derive greater benefit from participating in formative assessments.

Minor Essential Revision
The authors support this by stating four examples from a potential 14 courses. If these are the only four instances of positive results, then the tendency is in the other direction. Was the formative material reflected to a greater extent in the EOC exam for these courses, more than in the others? Either way, a table summarizing these results would be helpful.

p10. Better performance in each of the formative assessments was also consistently associated with higher marks in the respective EOC examinations

Discretionary Revision
I believe this is an important statement as it provides evidence to support the assertion that the better students were doing the cases, especially as second attempts at cases had no significant effect. Reference could be made to it in the summation at the end of p12 and beginning of 13.

The authors’ difficulties in being able to create a control group are understandable and a hazard of educational research, which must take a back seat to student welfare.
The second study was a randomized control trial of online formative assessments for medical students in clinical clerkships, which found no positive effect on learning [23].

Discretionary Revision
Perhaps a statement highlighting the fact that this study failed to gain significant numbers of participating students (far less than in this study) would be helpful in defining the contrast

Is the question posed by the authors well defined?
Yes
Are the methods appropriate and well described?
Yes
Are the data sound?
Yes, additional requests for further data are in the report above
Does the manuscript adhere to the relevant standards for reporting and data deposition?
Yes
Are the discussion and conclusions well balanced and adequately supported by the data?
Yes
Are limitations of the work clearly stated?
Yes. Limitations are acknowledged and discussed. Additional recommendations are provided above.
Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
Yes
Do the title and abstract accurately convey what has been found?
The title could be more positive, in order to attract researchers attention, but the title and abstract are accurate reflections of the content
Is the writing acceptable?
Yes

**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