Reviewer’s report

Title: Making Medical Student Course Evaluations Meaningful: Implementation of an Intensive Course Review Protocol

Version: 2 Date: 16 October 2014

Reviewer: Mike Tweed

Reviewer’s report:

Making Medical Student Course Evaluations Meaningful: Implementation of an Intensive Course Review Protocol

1. Is the question posed by the authors well defined?
   No. The objective was to describe. However “can intensive course review, in part flagged by student feedback, improve student course feedback?” might be a question that is asked.

2. Are the methods appropriate and well described?
   Yes although there could be some improvements.

3. Are the data sound?
   Some improvements or justifications needed.

4. Do the figures appear to be genuine, i.e. without evidence of manipulation?
   Yes.

5. Does the manuscript adhere to the relevant standards for reporting and data deposition?
   Yes, with some modifications as above.

6. Are the discussion and conclusions well balanced and adequately supported by the data?
   Partially. Some developments would help.

7. Are limitations of the work clearly stated?
   Yes, but not allowed for in the conclusion.

8. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
   Yes.

9. Do the title and abstract accurately convey what has been found?
   Mostly.

10. Is the writing acceptable?
    Yes.
This is a description of using student feedback forms as one of 2 sources of information for implementing a process of course review. The authors present a change in student feedback scale results and number of courses below a benchmark as the outcome. As mean scale result increased and number below a certain scale result dropped this was presented as a positive result, with some mention of the limitation due to response rate.

Course evaluation is to be encouraged and literature on this subject equally to be encouraged. Hence this study could be looked at as a means to investigate the process of course evaluation and development and also the actual results generated.

However, there are some areas that could be developed within the study and the manuscript.

Introduction

A clearer research question would help. Although this is described as a descriptive article, before and after figures are presented.

A change in feedback on the scales can be due to: a change in the scale or use of the scale; a change in those completing it, either intake or response rate; a change in the course due to review or another reason such as staff change; or chance. This does need to be considered especially given the change in response rate. A description of how the admission criteria and learning and preparation prior to this section of the course had remained constant, perhaps with a brief description would add to the context.

Page 4, para 2

The rating of <3.5 is justified by observation of negative comments. Why was a change of 0.5 chosen?

Given there are 3 criteria for being recognised as needing review, there are 7 combinations of criteria that could lead to the need for review. I could postulate that low results, falling results and critical incidents might have very different causes and potentially outcomes, so knowing how many were identified by each would be helpful. Although small numbers may preclude analysis by identification criteria this may be helpful, especially if those non-responders are from one type. This could add to the results and discussion.

Methods

Page 4, para 1 of methods

How many were excluded because of course restructuring or incomplete data?

This needs to go in results.

Analysis

A 5 point likert scale is used and parametric analysis is undertaken. There is some controversy over the uses of parametric rather than non-parametric analysis for likert scale and some justification is needed.

A copy of the questionnaire as an appendix is required.
Results
As noted above- analysis by identification criteria might add to this.
As noted above- numbers excluded.
Did different course get identified through the process? Did all those that were OK at baseline stay OK and did the improved stay improved.
Were the 3 that remained identified for the last few years the same 3 each year?
Discussion
If the results are re-analysed, with some of the suggestions above, this may add to the discussion.
There is a risk of regression to a mean effect. If we only intervene on those who score low then the improvement in their result might be explained by this. Hence a better description of the data for all the courses would help control for this.
Although the limitation of the falling response rate is identified, the conclusion and title raise that meaningful change has occurred. This is difficult to substantiate given the falling response rate.
What about other measures of QA for course delivery other than student satisfaction and staff identified critical incidents?
One thing that has interested me in analysis of student satisfaction is, was some of the improvement due to better explanation of the course rather than significant changes to the course? Ie the staff did not need to make significant changes to the course, just explain things better. Some improvements can be achieved by doing the same course, explained better, rather than wholesale significant changes to the course and all that entails.

References
Formatting needs checking eg 11.

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:
I declare that I have no competing interests