Reviewer’s report

Title: The Modified Essay Question: does it still have an effective role in assessment? Research Paper

Version: 1 Date: 5 June 2007

Reviewer: Martin Fischer

Reviewer’s report:

1. Is the question posed by the authors new and well defined?

Irwin and Bamber state in 1982 “The study shows however the potential of the MEQ to measure the entire span of Bloom and Buckwalter's levels.” Med Educ. Nov;16(6):326-31.

I believe that there is enough evidence that both MEQ’s and MCQ’s have the potential to cover higher levels of cognitive taxonomies depending on how well they are constructed. Therefore, I am not sure whether this study adds new relevant information.

2. Are the methods appropriate and well described, and are sufficient details provided to replicate the work?

I believe that the main weaknesses of the study are sampling and item quality: The comparison between MEQ and MCQ is based on the assessment at one university. It remains unclear why the nine-week surgical attachment was chosen (plus final MB BS-exam for boosting the number of MEQ items). It remains unclear to me, how adequate quality control for both MCQ and MEQ items was provided: Only one person did the final review of test materials and NO instruction on how to construct an MEQ was provided for item authors. I think this is a major confounding factor concerning the conclusions of the paper. It is furthermore unclear, how the judges for the classification of test items to the respective cognitive skill level were instructed. How was the item classification exactly done? What was the background of the judges? Did they discuss their results?


3. Are the data sound and well controlled?

I have problems with figure 1: There are 18 MEQ’s with a total of 145 stages - one should expect interdependencies with such a high number of stages per MEQ. The three MEQ’s from surgery only had 2.7 stages per MEQ compared to 9.2 for the
MEQ’s from the MB BS exam. Can you really pool those two MEQ sources? What kind of MCQ’s were analysed? Best answer from one-out-of-five options? Other types? I think more information is needed on the items and the comparison of such heterogeneous items is problematic as described in figure 1.

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

See 3.

5. Are the discussion and conclusions well balanced and adequately supported by the data?

I think that the discussion is a strong part of the paper. However, I believe that due to the methodological criticism, it is not sufficiently supported by the data.

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

I think the title should encompass the comparison with the MCQ-format and include the context (4th year surgery attachment).

7. Is the writing acceptable?

Yes.

In summary, I would not recommend the paper for publication due to the methodological shortcomings and its questionable novelty.