Reviewer's report

Title: A Randomized Trial Comparing Digital and Live Lecture Formats

Version: Date: 11 August 2004

Reviewer: Peter W Callas

Reviewer's report:

General

This paper is very clearly written and well organized. The topic is important since attending lectures via modes other than in person is only useful if there is not a degradation in learning.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. My main question is the comparability of the 12 students who attended live lectures and 17 students who attended digital lectures. Although the students were randomized, with such a small sample size the two groups could easily differ on potentially confounding variables simply due to chance. If, for example, the better students happened to end up in the digital group, the lack of difference in scores could be because attending digitally brought down their scores. Although I realize that the exam forms were anonymous, is there any information available on the characteristics of the students in the two groups (e.g., NBME scores)? Also, were the scores on these exams similar for students who were not study participants? The scores presented in Table 1 seem quite low, with means around 50%. This again leads me to wonder about the characteristics of the students in each study group.

2. Results, paragraph 2: Why is the power analysis based on a one-sided type I error rate when the analyses for the study use two-sided hypotheses tests (Table 1)?

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

3. Results, paragraph 3: Re: “… the students who viewed the digital lectures performed…slightly worse for the items for the other four lectures." This is true based on the means but not on the ranks (and presumably the medians).

4. Table 2: For the final item, response categories are labeled only as “Agree” and “Disagree” but there are four responses. Should these be Strongly Agree / Agree / Disagree / Strongly Disagree?

Discretionary Revisions (which the author can choose to ignore)

5. Abstract, Results paragraph, second sentence: This sentence implies the variances differed for all of the lectures, which was not the case. As noted below, the lack of difference for four of the six lectures indicates that the difference for two of the lectures may just be due to chance, so perhaps this sentence should be deleted from the abstract.
6. Methods, final paragraph: Why was coefficient alpha computed? This statistic is used to measure the internal consistency of items. Why would it be expected (or matter) that questions on exams that cover different topics are internally consistent?

7. Also Methods, final paragraph: It is unusual to present mean ranks rather than medians as a nonparametric measure of central tendency.

8. Results, paragraph 5: The interpretation and usefulness of the coefficient alpha findings are unclear. It is not surprising that coefficient alpha for the combined is higher than for the 20-item exam, which is higher than for the 9-item exam, since this statistic increases as the number of items increases. Also, the use of the word “scales” seems odd – the exam questions are not intended to be scale components, are they?

9. Discussion, paragraph 3: The difference in variances for the CAD/renal failure lectures is perhaps overemphasized. As the authors point out, the dispersion in scores for the digital group is not what would be expected if it was due to technical difficulties. And, one might expect a larger difference in variances for the other four lectures, since these lectures were not the same for the two groups, yet these variances did not differ at all. The lack of consistency for the finding of greater variance suggests that this finding may be due to something other than lecture format, such as chance or different characteristics of the students in the two groups. Thus, perhaps less weight should be given to this finding.

10. Table 1: Why are mean raw scores presented rather than mean percentage correct, which would put the CAD/renal lecture results on the same scale as the other four?

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

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

**Quality of written English:** Acceptable

**Statistical review:** No

**Declaration of competing interests:**

None