Author's response to reviews

Title: A Randomized Trial Comparing Digital and Live Lecture Formats

Authors:

David J Solomon PhD (dsolomon@msu.edu)
Gary S Ferenchick MD (Gary.Ferenchick@ht.msu.edu)
Heather S Laird-Fick MD (Heather.LairdFick@ht.msu.edu)
Kevin Kavanaugh MD (kavanaugh@kcms.msu.edu)

Version: 2 Date: 10 September 2004

Author's response to reviews: see over
Reviewer 1

Major Compulsory Revisions

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)?

Response - The reviewer is correct that the two groups could differ due to chance on any number of characteristics however the fact that subjects were randomized to either the treatment or control group controls for potential biases or confounding. Campbell and Stanley address this issue very clearly.

“For psychological reasons it is difficult to give up 'knowing for sure' that the experimental and control groups were 'equal' before the differential experimental treatment. Nonetheless, the most adequate all-purpose assurance of lack of initial biases between groups is randomization. “


The reviewer’s concern is legitimate however it is an issue of sampling error not bias or confounding. The power analysis suggests the study is sensitive enough to likely detect real population differences between the performances of students viewing digital versus live lectures of approximately one standard deviation. Unfortunately that is a large difference and the lack of power certainly limits the value of this study. This is a chronic problem both in educational and clinical research. The best way to address this problem is by aggregating the data from small scale studies through techniques such as meta analysis. This is why we feel this study should be published even though the conclusions that can be drawn from it alone are quite limited.

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.

Response - Scores are likely low due to the nature of the examination, e.g., these were just difficult items. It would be nice to be able to confirm this as suggested by the reviewer; however, given that the students who did not participate in the study specifically chose not to volunteer to participate in the study, including their responses as part of this research report would be both unethical and illegal.
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)?

Response - A one-tailed test was chosen because the goal of the study was to see if students who viewed the presentations on CD-ROMs could perform as well on the examination as students who viewed the live lectures. We were not specifically interested in if the students viewing the CD-ROMs would perform better than the students who viewed the live lectures and particularly given the small sample size preferred to narrow the research question and increase the power.

The listing of two-tailed p values in Table 1 was an oversight and has been corrected.

--------------------------------------------------------------------------------------------------
Minor Essential Revisions

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).

Response - The discussion of the differences was shortened to just state there were no statistically significant differences.

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?

Response - Labels were added for each category.
Reviewer 2

Major Compulsory Revisions

Assuming that the examination used to assess learning outcomes has been used with a much larger sample of learners, the authors should be able to do an item analysis and say something about the performance characteristics of the examination. I believe they should report this information in a revision of their paper. If the examination is sound, my recommendation would be to publish this paper, despite its limitations.

Response - The reliability of the whole examination as well as the individual subscales is reported in the manuscript. The last academic year, when this study was conducted, was the first time this examination was used. The only other data that would be available to assess the psychometric characteristics of the examination is from students who chose not to participate in the study. As noted above, using their data for research purposes would be both unethical and illegal.

Traditional item analysis e.g. assessing item to-total test correlations and/or assessing how high/low score groups responded to the items would be useful in revising the examination but I don’t believe would provide significantly more insight into the quality of the examination beyond the coefficient alpha reliability that was reported.

Students took the examination anonymously and were only asked to indicate whether or not they chose to be included in the study. This precluded correlating their performance on the examination with, for example, their performance on the NBME shelf examination as a measure of concurrent validity. The items however were written by the presenters to specifically target the major objectives of each lecture. This would tend to support the content validity of the examination.

As noted in the original article (this section has been dropped per request of reviewer 3) we seriously considered using scores from the NBME subject examination as the outcome measure for the study. Our concern was that since the subject examination covers a much broader range of material then was covered in the lecture series, it would not be a very sensitive measure of the knowledge gained through viewing or attending the lecture series. Additionally, since the subject examination is a major portion of the clerkship grade and the lecture series is given about two weeks prior to taking the subject examination, students likely studied intensively for the subject examination further obscuring any impact of the lecture series.

I am not sure what else can be done to assess the quality of the examination but would be open to suggestions from the reviewers.
Reviewer 3

Major Compulsory Revisions

ABSTRACT/INTRODUCTION

In the background section of the abstract, the stated goal of the study is too vague. It should be reworded with more specific outcomes related to knowledge and attitudes.

Response - The discussion of the goal of the study has been expanded and clarified.

Focusing on a difference in variances is not appropriate for an abstract and distracting to the central purpose of the study.

Response - I disagree with the reviewer on this issue. When an outcome measure is more dispersed in the treatment group than in the control group it suggests the subjects respond differently to the treatment. This is an important finding that I believe should be noted in the abstract. I will defer to the judgment of the editor but if I was searching the literature in distance learning, I would definitely want a finding such as this noted in an abstract so it wouldn’t be missed.

Background - the last sentence of the first paragraph of the background section “although didactic…” does not make sense.

Response - It’s been dropped from the manuscript

You thought the value of video was important enough to comment on in the introduction, but you did not ask the students questions about it (too late now), nor do you discuss it in the discussion section (not too late). Please address this.

Response - We did not address this issue on the feedback form to keep the instrument as short as possible and focused on the major issue of the study. We are very interested in value of including video of the presenter in these digital presentations and are in the process of conducting another study comparing including/not including video of the presenter in the digital presentations. This is now mentioned in the discussion section.

METHODS

State in the methods section how many students are in the class and what proportion of the class actually were randomized? In fact, the first whole paragraph of the discussion should be in the methods section instead.

Response - The actual number of students who completed the clerkship as well as the number who volunteered and participated in the study is probably more appropriate in the results than the methods section. These values are data from the execution of the study rather than describing how the study was conducted. This
information along with what was the first paragraph of the discussion has been moved to the results section.

*Please state clearly in the methods section what the questions were to which they were seeking agreement on advantages and disadvantages of CD-ROM lecture from the students.*

*How did the students view the CD-ROMs? What kind of computers? Does the school issue standard laptops?*

**Response** - The questions have been added to the methods section. Each of our community campuses provides access to computers in some type of laboratory in the community campus office or one of the hospitals. The type of computers varies across the campuses. Students are not issued laptops at MSU. This has been clarified in the methods section.

**DISCUSSION**

*The comparison of the NBME versus the “examination” may confuse readers because some readers are not going to know what the NBME is or when it is administered in your clerkship timeline. The fact the students agreed that the live lectures could be replaced with the digital format could suggest that neither format had value for the student.*

**Response** - This section was dropped.

*Characterizing the students view as very positive is too strong, given the acknowledged problems with the study. It is valuable to have the Michigan State web site for others to view.*

**Response** - The wording has been changed.

Some discussion of the lag between video and audio is warranted.

**Response** - I not sure what you are referring to, please clarify.

*In my view, it is not proper to advertise proprietary software in a research article, nor is it necessary. In fact, the last paragraph of the discussion should be removed as it is a discussion of different data, and a different experience. The data from your study should stand on its own.*

**Response** - First of all, none of the authors have any relationship what so ever with Impatica. Name of the software and the URL of their Web site were included because we thought readers might be interested in knowing what software we are using. It’s been dropped from the manuscript.

*Again in the conclusion, you bring up a new subject altogether, that of web-based testing. The conclusion should focus on the data initially presented, not start new ideas.*
Response - This has been dropped.

-----------------------------------------------------------------------------------

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

RESULTS

There is much statistical jargon in this manuscript in the methods section and on Table 1 that is not reader friendly. Do you really need degrees of freedom on the main table? Have a clinician-educator edit this section to improved readability?

Response - The degrees of freedom have been dropped from Table 1. It is not clear what jargon you are referring to in the methods section. The only technical terms used are the three statistical tests, t-test, Mann-Whitney, Levene’s test. The readability of a research report has to be balanced with clearly describing what was done.

Table 2 - Just not like being in the room with the presented “ must be a typographic error that must be corrected. Very confusing to the reader. Need the headers (strongly agree, agree,... On the table 2

Response - The typographical error is in the quote above, not in the manuscript. What was stated in the manuscript and the original instrument is:

“Just not like being in the room with the presenter”

What we were trying to ascertain were the more intangible aspects or gestalt of viewing the presentation on a computer rather than in the room with the presenter. Whether it is a sports event or lecture, there is something different about viewing video as opposed to physically being at the event. We were trying to determine the extent the students felt this was a disadvantage of the viewing the lectures on a computer.