Reviewer's report

Title: Internet-based medical education: A realist review of what works, for whom and in what circumstances

Version: 2 Date: 12 August 2009

Reviewer: David A Cook

Reviewer's report:

The authors have embarked on an ambitious and laudable task to qualitatively analyze the literature on Internet-based medical education, seeking answers to "what works, for whom, and in what circumstances." To accomplish this goal they have employed the methods of "realist review." While I admire this objective, and for the most part agree with their conclusions, I have some concerns with what came between (i.e., how they got from objective to conclusion). Since the third author is an expert in realist review, my hope is that my concerns are largely a result of (correctable) reporting issues rather than methodological flaws.

My first main concern regards the identification and selection of the "candidate theories." This process is described in very vague terms ("Browsing and snowballing" does not seem very rigorous. What sources were browsed? Were any experts contacted?). More importantly, it appears that several important theories were never considered – a major omission. The failure to consider important theories such as Mayer's Theory of Multimedia Learning, Sweller's work on cognitive load, Spiro's Cognitive Flexibility Theory, and Jonassen's work on problem solving (to name but a few) limits the confidence I have in their conclusion that Laurillard and Rogers are the only theories with explanatory relevance. Since the explanatory theories form the cornerstone of a realist review, the likelihood that these authors have inadvertently overlooked one or more key theories calls into question most of what follows. They authors must describe in much greater detail how they identified the theories they list. Also, I would strongly encourage them to include the theories and frameworks I have listed above (and others I mentioned when I reviewed this manuscript for BMJ), and to expand their search for additional candidate theories. I would not be surprised if they identify one or more additional theories with explanatory power; if so, they would need to repeat some elements of the qualitative analysis to incorporate this/these theory[ies]. However, this would greatly strengthen the rigor of the manuscript and the credibility of the conclusions. New theories might also expand the scope of the conclusions to provide greater guidance to developers.

Furthermore, as a minor point, it is confusing to separate out the initial 4 candidate theories and then state immediately thereafter that these theories were insufficient and that additional theories were required. Since per methods these theories were identified using "snowballing" it would seem appropriate to simply state that "In an iterative process of exploration, comparison with the data, and
re-exploration, we identified a large number of candidate theories including x, y, z" and include Rogers (and any new theories) in this list.

My second main concern regards the presentation of the results of the qualitative analysis: I simply want more detail. As it stands, the authors make large intellectual leaps. These leaps may be justified, but more data should be presented to help the reader understand this. First, more detail is needed to explain how the list of candidate theories was narrowed to two. Second, the section "Course-context interaction" is nice – gives several contrasting examples that show how the data support the model. The other two major sections (Technology Acceptance, and Interaction) do not do this nearly as well. These sections could be expanded without an overall increase in manuscript length by shortening the Discussion (which currently is much longer than the Results). It would also be helpful to explicitly develop the questions in Box 1 in the Results.

My third concern is the repeated claim that there are no laws of nature governing the learning process. This claim itself is a proposition that cannot be proven, and is fundamentally untenable. I certainly agree that learning, as with most other human interactions, is exceedingly complex. But this does not mean that such interactions are not governed by laws of nature. On the contrary, the laws of nature are inescapable. The limiting factors in understanding complex interactions are the man-made theories (not laws of nature) that explain these interactions, and the nature of empiric evidence that supports these theories. A more robust theory will be able to explain or predict the results of a greater number of interactions (but always imperfectly). (Einstein's theory explains more than Newton's; and while "truth" is relative [subjective] in Einstein's universe there are still underlying laws.) Thus, I wholeheartedly agree with the authors that no single theory is able to predict results across all contexts and learners, that "demi-regularities can only be explained by middle-range theory", and that "factors … cannot be built into courses independently of a consideration of learners' needs and priorities or … the course's context." I suggest the authors emphasize these points, and avoid their claim that fundamental principles do not exist (they exist, even if humans do not completely understand them).

I have a number of other, more minor concerns and suggestions. I group these as requested by the journal.

Compulsory revisions:

- The abstract needs a lot of editing work. The content is good, but awkwardly presented; notably second sentence of Background (esp. phrase "and learners on how") and second sentence of Methods ("references of references" is confusing; do you mean references of included articles?), among others. I would also suggest including more detail on methods, and more specific results.
- In Discussion, second paragraph: I do not recall being provided evidence in the Results to support the proposition that the same course could be effective in one context and ineffective in another. You present data showing that similar delivery modalities (but entirely different courses, developed by different people) had different effects in different contexts (e.g. light microscopy). But if these were
different courses then perhaps it was the course design, not just the context, that led to different outcomes. It's not that I don't believe what you state in this paragraph, I'm just not sure the data support it. It seems this is a hypothesis meriting further testing, not an empirically-justified finding.

Minor essential revisions:

- I believe you overstate the limitations of meta-analysis (second paragraph Background, fourth paragraph of Discussion, etc). Techniques such as meta-regression and subgroup analysis can offer tremendous insight into what works, for whom, under what circumstances. Of course, qualitative literature reviews are also exceptionally useful (see Cook DA. Narrowing the focus and broadening horizons: complementary roles for nonsystematic and systematic reviews. Adv Health Sci Educ Theory Pract. 2008;13:391–395) and, as you note in the Discussion, complementary. I think you can highlight the need for a qualitative review without disparaging the usefulness of meta-analysis.

- How is interrater agreement calculated? Is this raw agreement? Wouldn't kappa be better?

- In presenting results, please give actual number (and if needed denominator) rather than just percentage.

- You claim that this is the first realist review in medical education – which may be true. However, I do not think that this necessarily reflects a "new tradition in systematic review." Systematic narrative reviews accomplish essentially the same purpose, and are numerous. The realist review simply provides a defined set of methods.

- I like the Discussion paragraph beginning "The pursuit of rigor in realist review ..." but to be honest I do not see these key features of rigor reflected in the present manuscript. If you did all these things, you need to make it more clear – especially as you present the Results.

- There are other limitations inherent to a qualitative systematic review that should probably be mentioned, even if they are unavoidable; I believe the authors are aware of these.

- Box 1 seems to appear out of nowhere. The questions here – including the subquestions (a, b, c, ...) – should emerge explicitly in the Results. The questions could then be summarized as a final paragraph in the Results, and Box 1 referenced there.

- Figure 1 has some formatting problems ("496 full text obtained and re-").

- The search strategy in the Appendix is helpful. Would be important also to have a list of the articles included in the review (the full citation as a minimum; selected details if possible).

Discretionary revisions:

- When discussing my 2008 meta-analysis in the Background, you might point out the substantial unexplained heterogeneity, which suggests that these analyses fail to account for the complexity of the interactions. This would nicely
set the stage for a qualitative analysis exploring this heterogeneity. (Explicitly noting the concept of heterogeneity in your presentation of ref 14 [where it is currently implied but not stated] would help people understand this as well.)

To specifically address the questions requested by the editors:
1. Is the question posed by the authors well defined?
   Yes, and it is an important question.

2. Are the methods appropriate and well described?
   Methods are most likely appropriate, and reasonably well described.

3. Are the data sound?
   As noted above, I am concerned about the explanatory theories identified. The authors need to present additional data to support how their conclusions derive from the original reports.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
   Some room for improvement as noted above, but nothing major.

5. Are the discussion and conclusions well balanced and adequately supported by the data?
   This needs a fair bit of work; probably do-able.

6. Are limitations of the work clearly stated?
   Most limitations are noted.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
   As far as I am aware, yes.

8. Do the title and abstract accurately convey what has been found?
   Title is fine. Abstract needs work.

9. Is the writing acceptable?
   I think a bit long-winded, especially in the Introduction and Discussion which could be trimmed about 30% without sacrificing content. Otherwise acceptable.

In closing, I wish to commend the authors for their efforts. I can appreciate what a monumental undertaking this must have been. It was a pleasure to review this manuscript. I hope you find these suggestions helpful.

David A. Cook, MD, MHPE
Mayo Clinic College of Medicine

**Level of interest:** An article of importance in its field
Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.