Author's response to reviews

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

Authors:

Geoff Wong (g.wong@pcps.ucl.ac.uk)
Trisha Greenhalgh (p.greenhalgh@pcps.ucl.ac.uk)
Ray Pawson (r.d.pawson@leeds.ac.uk)

Version: 3 Date: 16 September 2009

Author's response to reviews: see over
Dear Dr Norton

Re: Manuscript number: 2453225492755020
Internet-based medical education: A realist review of what works, for whom and in what circumstance

Many thanks for sending us the two sets of peer review comments. We found these most constructive and helpful and wish to thank them for the time and effort they have made to provide these. We have read the peer reviewers’ comments and addressed these in detail in our revised manuscript. Our detailed response to their comments follows this cover letter.

I hope that you will now find our manuscript suitable for publication and want to thank you for giving us the opportunity to revise and submit our manuscript.

Yours Sincerely

Dr Geoff Wong
Professor Trisha Greenhalgh
Professor Ray Pawson
Detailed response to reviewers
Our responses are in italics

Reviewer: Jarmila Potomkova

Reviewer's report:

- Title & structured abstract
well define article type („a realist“ review), target group (doctors, medical students), purpose of the review (help online course developers what offer online and how; end-users how assess e-learning supports; methods are appropriately explained; results show two types of findings (quantitative and qualitative); conclusions emphasize the main message of the review (technology cannot govern the educational process, but rather serve it).

Changes have only been made in order to address the comments by Dr Cook (second reviewer).

- Methods
A systematic „realist“ review seems to be a suitable and all-encompassing tool to answer the posed question, based on the authors´ past experience and other published literature. The methodological issues can be found in sections Background and Methods. Maybe, it would be better for readers to have substantial and detailed information about „realist“ review in Methods, not in Background. For the authors it would only mean moving some paragraphs to a new position.

We have moved the methodological sections from the Background to the Methods section.

- Results
The process of selecting relevant articles is demonstrated in great details, supported by a flowchart. Inclusion/exclusion criteria are evident, data management, analysis and synthesis are in concert with the goals of the review.
A particular value is the usage of „candidate theories“ and seeking for new ones, if the existing are not suitable. The authors clearly specify the number of included papers as well as total amount of participants and study designs. Surprisingly, 84% of studies assessed learners´ satisfaction. Of particular importance is identification of „perceived usefulness/relative advantage of online teaching & learning“. This point of view is referred to many times within the manuscript. I would recommend to include another recent reference to
emphasize the significance of this issue (eg. Ledden et al., J Business Res 2007; 60:965-974. I am attaching full text of the article).

We take on board this comment on the need to emphasise the significance of this issue and have amended our revised manuscript accordingly.

- Discussion, Conclusions
In my opinion, this is a most practical and attractive section to read. It is summing up the 2 key findings about internet-based courses, emphasizing the interactivity.
It also shows strengths and weaknesses of the review. One of the main achievements seems to be the first attempt to use „realist“ systematic review for medical education research.
The challenge for internet course developers is extremely thought-provoking: to develop internet courses as part of the curriculum, not its substitute.
Limitation of the review are clearly stated: as one may expect in the context of systematic reviews – it is always heterogeneity of primary studies included in the review.

Note: the text would be easier to read with fewer phrases between parentheses.

We have rewritten sections of the manuscript which contain ‘too many’ phrases between parentheses.
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?).

Additional detail is given in the revised manuscript. We respectfully suggest that this reviewer may be applying inappropriate criteria for rigour, based on positivistic assumptions. In the revised manuscript we have explained that in realist review (an interpretivist method), the criteria for rigour are different. Editors please note that realist review is a very new science and in our experience reviewers who are unused to this approach tend to view it as “unrigorous”. There is actually a very strong tradition of rigour within realist and other qualitative systematic review methods, and we ask that editors seek input from experts in this tradition if they remain in doubt about the rigour of the methodology. As Prof Cook correctly observes, Ray Pawson is the international expert in this method and we are confident that the approach used was rigorous.

The following comment from Professor Maggie McClure illustrates the counter-argument to the suggestion that exploratory, interpretive methods lack rigour (MacLure 2005 - ). Whilst we do not condone the confrontational tone of this quote, we hope the editors can see that conventional systematic review methods, based on a series of formal technical checks, might also be seen as 'unrigorous':

“[conventional systematic review] assumes that evidence can be extracted intact from the texts in which it is embedded, and ‘synthesised’ in a form that is impervious to ambiguities of context, readers’ interpretations of writers’ arguments (i.e. bias). Most significantly of all, [conventional] systematic review systematically degrades the central acts of reviewing: namely, reading and writing, and the unreliable intellectual acts that these support, such as interpretation, argument, and analysis. By replacing reading and writing with an alternate lexicon of scanning, screening, mapping, data extraction, and synthesis, [conventional] systematic review tries to transform reading and writing into accountable acts. It tries to force their clandestine operations – the bits that happen inside people’s heads – or in the incorporeal gaps between decoding and comprehension, thought
and expression – up into plain view, where they can be observed, quality-controlled and stripped of interpretation or rhetoric.”


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.

We accept that in the original version of the manuscript we gave the impression that the two theories offered provided a comprehensive explanation of mechanism, and we failed to engage with a number of other theories that are referred to in the literature. In the revised version of the manuscript, we have made the following points:

(1) Any systematic review is only as good as the primary data on which it is based. You wouldn’t criticize a Cochrane systematic review for not coming up with particular findings (you’d ask whether the primary studies were properly identified and analysed, and if the findings are negative or ambiguous, that would be the state of knowledge in that field). The ‘findings’ in realist review are the explanatory middle-range theories, and what this study has found is that not many theories have been systematically tested in sufficient detail to allow firm conclusions.

(2) Realist review entails coming up with candidate theories (i.e. ones that possibly explain the outcomes of interest) and then testing these against the empirical data in the primary studies. This is why we have listed the initial theories we were exploring and described how we rejected some of these through a process of progressive focusing.
The two theories presented as our ‘findings’ were the ones that emerged from our data. Whilst our study has demonstrated that these theories are important in explaining the fortunes of Internet-based courses, it has not demonstrated that other theories (as-yet untested in empirical studies – or at least, not written up in sufficient detail to allow testing via the realist method) are unimportant.

We hope that we have now made clear what we are claiming, and what we are not claiming, about the two theories presented.

Some further methodological detail: The theories are initially termed ‘candidate' because they may or may not be ones that eventually go on to best explain the patterns of outcomes observed in the included studies. It is common practice in realist reviews to initially start off with an ‘educated guess’ at what may be the middle-range theories that explain the data. This does not mean that we picked the theories out of a hat, but that through a process of discussion, brain-storming, browsing through library collections, discussions with fellow educators and pursuing references of references, theories were selected. Once a theory is selected, it is ‘tested’ to see if it can explain the findings in the included studies. This means that we would read full text articles in their entirety (often more than once) and then interpret if sections of texts within the included studies confirm, refute or refine the candidate theories. If a candidate theory offers no explanatory insights across more than one included study, then it is put aside and a new one sought. Realist reviews are rigorous because candidate theories are not taken for granted as being ‘operational’ in any family of similar interventions, but because the theories are deliberately tested to see if they can explain the empirical findings in the included studies.

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.

We have explained in the section above the methodological steps we took towards our analysis of the data. We have endeavored to provide the necessary explanations in our revised manuscript.
In response to “I simply want more detail”, we have provided further examples of verbatim texts that informed our interpretation in the form of an appendix – APPENDIX 2. We have also explained how the list of candidate theories was narrowed to two. We hope that our changes and additions will assist readers.

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.

The above points have been noted and we have

- further developed the Technological acceptance and Interaction sections
- shortened the Discussion
- explained the origins of the questions in Box 1 in the results section

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.

We appreciate Dr Cook’s comments and we believe there is much less difference between our positions than he implies. Social realism (the
Philosophical basis of realist review) strongly supports the notion of ‘laws’ (meaning patterns) in the social world, so we certainly do not hold that “there are no laws of nature governing the learning process”. We’ve gone through the manuscript and tried to remove any implication that we hold such a view!

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.

We have rewritten the abstract and hope that this reads more clearly.

• 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.

We have re-worded the relevant section to explicitly point out that both course design and context matter. Note, however, that it is accepted convention in qualitative research that a direct link between all findings presented and the underlying data from which those findings were derived is not always possible to demonstrate given the limitations of word count (since findings are often based on lengthy interpretive analyses). We ask editorial advice on whether you would like a detailed worked example for an appendix.

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.

We found this new reference most interesting and have now included it in the Discussion. We also agree that realist review (an essentially interpretivist approach) and newer, more sophisticated approaches to meta-analysis can both provide insights that add value to the old-style Cochrane review. We are not suggesting, however, that meta-regression should be abandoned in favour of realist review. Our claim is much humbler – that for some research questions, a realist approach is useful.

• How is interrater agreement calculated? Is this raw agreement? Wouldn’t kappa be better?

We calculated raw agreement. This has been stated in our revised manuscript. Given that a key dimension of rigour in realist review is discussion between reviewers and construction of interpretations, we do not believe that either raw agreement scores or kappa scores are intrinsically “better” – both are somewhat marginal to the central methodology, though useful tangential data when placed in the wider context of what was done.

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

This has been provided as requested.

• 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.

We do not agree that, “Systematic narrative reviews accomplish essentially the same purpose [as realist reviews]”. The main difference is the deliberate focus in realist review in using middle-range theory to make sense of the empirical observations. Furthermore, realist reviews are based on a specific set of ontological, epistemological and methodological assumptions, something which systematic narrative reviews may or may not be. See references 13 and 14 in our manuscript.
Incidentally, one could present a similar argument about meta-analysis (or meta-analysis using individual patient data, or any of the refinements in this method): it’s really just a subcategory of quantitative systematic review, so nothing new. In the end, such arguments boil down to whether you’re a ‘lumper’ or a ‘splitter’ and to the amount of ‘added rigour’ you attribute to the ‘new’ approach. Hence, lots to debate – but not a reason in and of itself to reject the paper. This is surely a topic for the correspondence columns?

• 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.

As explained above, and with the greatest of respect to this reviewer who has put a lot of time and effort into his review of this paper, realist reviews come from a different research tradition and philosophical paradigm from conventional Cochrane reviews. We have revised the manuscript to include more detail about the realist review process and exhort the editor to seek independent advice from an expert in realist review if s/he is in doubt about the level of rigour.

• 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.

We have added to this list and hope that our frankness is now sufficient.

• 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.

We have explained the origins of Box 1 In our Results section.

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

We have addressed this in our revised manuscript.
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).

An alphabetical list of all the full text articles we included has been uploaded as APPENDIX 3.

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.)

Thank you for this helpful point. It has been noted and incorporated into our revised manuscript.

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.

   We have noted this point and have shortened our Introduction and Discussion sections.

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.