Reviewer's report

Title: Effective In-Service Training Design and Delivery: Evidence from an Integrative Review of the Literature

Version: 1 Date: 11 September 2012

Reviewer: Christopher Carroll

Reviewer's report:

The topic is interesting. I appreciate that the approach taken was not systematic, that this is not a systematic review. However, I see no reason why systematic review methods should not have been adopted. Indeed, the review's findings - in terms of their internal and external validity - are weakened as a consequence of the absence of any transparency or reproducibility of many of the processes used. Hence, the authors' own Grade 1 evidence is systematic review, not integrative review. The failure to use systematic review methods adversely affects the validity of this review's findings. I see no reason why a systematic review was not performed instead; there are published methods for conducting systematic reviews of reviews:

Smith V et al. Methodology in conducting a systematic review of systematic reviews of healthcare interventions, BMC Medical Research Methodology 2011, 11:15

I found the amount of data in the tables overwhelming; it was unclear how far the findings in the text reflected the findings reported in the tables (the text and its "findings" appear to be highly selective); it was unclear what was meant by "outcomes" - this could range from "learner satisfaction", through "prescribing behavior" to a highly specific "post test score".

This issue is only finally addressed on p.17.

Lumping all such outcomes together as a measure of "effectiveness" is highly questionable. There is also no attempt to control for or acknowledge the effect of other factors on these outcomes (which might be different in each different RCT or review - just as a very obvious example, are findings for Kenya transferable to the United States and vice versa?).

The review's conclusions are therefore greatly open to debate, as the extrapolations that are taking place are not obviously supported by the evidence.

Indeed, no caution is expressed about the conclusions at all, which is very surprising, given the approaches used.

The Limitations section was brief - it could have been three times as long. I understand the choice not to perform a meta-analysis; but there is no justification for the failure to conduct a systematic review. The data of the included studies could have been synthesised using an appropriate technique. The claim is that
no quality assessment of included studies was performed, but there is frequent reference to the "quality of the evidence" or "studies" being limited or weak or "low quality", but it is not apparent how this was determined (e.g. pp.8, 11, 17, 19) - are some of these the authors' assessments? Some (eg. p.11) are clearly those of the included publications, but how valid are these assessments, how were they determined? Tools for the critical appraisal of systematic reviews are available (AMSTAR, PRISMA) - why were they not used to gain a greater sense of the strength of the review evidence?

All problems identified might have been addressed at least to some extent with the application of a transparent systematic review methods.

A major compulsory revision would be that the authors conduct a systematic review; I see little merit in the current approach. This is obviously a massive task and rewrite, hence my reluctant decision to reject.

Other minor/major revisions:

p.4 references to support the statements in the first paragraph would have been helpful.

b) "timing of instruction" - is this variable not "frequency" rather than "timing"? The term is used throughout and is arguably not representative of the data it purports to reflect.

p.5 - but what were the outcomes? Knowledge improvement? Knowledge retention? Experience of learning? All these? I think this needs to be made more explicit at this stage.

p.5 - " A research assistant searched the ... literature published in English between May and June 2011". This is surely not correct. The research assistant conducted the search between these dates, not the literature published between these dates.

p.5 the criteria for inclusion appear to have been developed in response to the citations retrieved, rather than a priori, which introduces bias into the selection process. Table 1 is also a standard hierarchy of evidence table; some reference to one of the many established such models is probably due here.

p.6 - What journals or sources were hand-searched?

p.6 - This appears to be a form of framework analysis or synthesis - perhaps relevant methods references should be cited here?

p.9 "You found improved performance" - highly colloquial

Fig.2. - the "second review" numbers are incorrect - should have excluded 102

Level of interest: An article of limited interest

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