Reviewer's report

Title: A Delphi survey to determine how educational interventions for evidence-based practice should be reported. Stage 2 of the development of a reporting guideline.

Version: 2 Date: 15 April 2014

Reviewer: Daniel Banks

Reviewer's report:

I have reviewed the manuscript entitled A Delphi survey to determine how educational interventions for evidence-based practice should be reported. Stage 2 of the development of a reporting guideline. By Anna C. Phillips, et al. Briefly, this manuscript reports the results of a Delphi study soliciting opinions from authors of a systematic review and relevant journal editors regarding the items that should be considered in a guideline for educational interventions.

Reviewer's report

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

Major Compulsory Revisions. I ask that the authors respond to these comments.

I offer the following comments:

I have reviewed items 1 – 9 as suggested in the invitation to review. For these questions, I have no specific comments on these individual items other than what is noted below.

1. Is the question posed by the authors well defined?
2. Are the methods appropriate and well described?
3. Are the data sound?
4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
5. Are the discussion and conclusions well balanced and adequately supported by the data?
6. Are limitations of the work clearly stated?

I should note that no limitations of the work are discussed.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
8. Do the title and abstract accurately convey what has been found?
9. Is the writing acceptable?
I have cited specific considerations that I ask that the authors address below.

10. On Table 4, the sum of comments for “Face to face contact with learners” should be 7 (not 8). On page 11, 27 of 36 responders yields 75%, not 79%. In addition, there is no “asterisk” or “cross” for “What method used to decide content” in Table 3 despite this achieving content. There are typographical errors in the Reference section.

11. I note that there two information items in Table 3 (“Timing of intervention” and “Extent of peer interaction”) which do not achieve consensus but have comparable mean, median and MAD-M scores to items which have achieved consensus. This leads one to wonder whether the empiric RAND approach to consensus is “the most reasonable.”

To compare to clinical medicine, until relatively recently, if a commonly used lung function test was greater than 80 % of predicted, then this was considered “normal.” Yet, it was recognized that “normal” could be much better expressed as a S.D., and the definition of “normal” changed from what was really an arbitrary cut-off to a statistically justifiable number.

I participated in a Delphi project where we attempted to use a “scientific method” to address agreement. We recognized that many of the methods are not statistically rigorous. We developed a statistically based definition of consensus and degree of consensus, based on probabilities (p value) and interquartile range (IQR). The Wilcoxon signed rank test was used to test the null hypothesis that there was an even distribution of scores throughout the range. Four degrees of consensus were defined: very good, p < 0.05 and IQR < 2; good, p < 0.05 and IQR < 3; some, p < 0.05 and IQR < 4; and none, p > 0.05 or IQR > 4.

I ask that the authors address these two outliers and the validity of their “empiric” categories.

12. In the ABSTRACT, consensus in the Methods Section is reported as at least 80% participant agreement. Two items reached consensus. In the Results section, 18 items reached consensus agreement when the four categories were merged into two categories (<7 and > 7). I ask that the ABSTRACT be modified in the Methods section to reflect the two ways consensus was able to be reached.

13. On page 10, the authors write that “items with the greatest participant agreement in the very high importance category (Likert scores >8) should be included in reporting.” First, this is not clearly a sentence. Next, it is not clear what the cut-off should be for inclusion as greatest participant agreement. Is there a cut-off? For example, in the information item “Supporting structures in organization to maintain behaviours targeted by the organization,” mean score is 7.9 and approximately 1/3 are in the very high importance category, yet this is included. Please comment.
Level of interest: An article of outstanding merit and interest 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'