Author's response to reviews

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.

Authors:

Anna C Phillips (anna.phillips@mymail.unisa.edu.au)
Lucy K Lewis (lucy.lewis@unisa.edu.au)
Maureen P McEvoy (maureen.mcevoy@unisa.edu.au)
James Galipeau (jgalipeau@ohri.ca)
Paul Glasziou (Paul_Glasziou@bond.edu.au)
Marilyn Hammick (m.hammick@gmail.com)
David Moher (dmoher@ohri.ca)
Julie K Tilson (tilson@usc.edu)
Marie T Williams (marie.williams@unisa.edu.au)

Version: 3
Date: 16 May 2014

Author's response to reviews: see over
16 May 2014

Dear Lauren Maggio,

Thank you for your response and feedback regarding our recently submitted manuscript. The reviewers raise some very good points which we believe we have addressed and have added value and clarity to our manuscript.

Please find the detailed responses to each of the comments provided by the reviewers below.

I look forward to hearing from you.

Kind regards,

Anna

---

Editor Lauren Maggio

**Comment 1:** Please review your references to ensure that there are no typos and that the numbers correspond with the correct in-line citation.  
**Response:** We have reviewed the reference list and the citations in the manuscript to ensure that all typos have been corrected and the numbers correspond to the correct in-line citations.  
**Action taken:** Referencing software was used for the updated manuscript to ensure the correct in-line numbering for citations and to eliminate the typos from the reference list. All previous typos for the journal names have been corrected.

**Comment 2:** For example, does the below sentence on page 5 correspond with reference #7? “The first stage in this development process required a systematic review of the literature concerning EBP educational interventions for health professionals [19,20].”  
**Response:** The purpose of the sentence, ‘The first stage in this development process required a systematic review of the literature concerning EBP educational interventions for health professionals..’ was to introduce the concept that the Delphi survey is part of a three stage process planned for the development of GREET (as outlined in the published study protocol) (Reference 19). Reference 19 is the study protocol for GREET which describes the intended development process in detail, and Reference 20 is the guidance statement for developers of reporting guidelines in health research, upon which the study protocol was founded (Reference 20). Our systematic review (currently under review with BMC Medical Education) (reference 7) completed prior to undertaking the Delphi survey was not included in this sentence as the intent was to flag the planned development steps for GREET rather
than the completion of our systematic review. We recognise that this dual referencing may not be necessary and this sentence has been updated to include only the reference to the study protocol for GREET.

**Action taken:** Previously read: The first stage in this development process required a systematic review of the literature concerning EBP educational interventions for health professionals [19,20].

Now reads: The first stage in this development process required a systematic review of the literature concerning EBP educational interventions for health professionals [19].

**Comment 3:** The very brief mention of the systematic review at this point feels distracting. Perhaps it would flow better if introduced in relation to the data extraction template [page 8]?

**Response:** The flow of the manuscript has been improved in line with the comments provided by yourself and the reviewers. The intent for this sentence was to introduce the concept of the three stage development process planned for the GREET and to illustrate that the Delphi survey forms part of the second stage in this development process.

**Action taken:** Two paragraphs in this introductory section have been re-worded to assist in clarifying the intended development process for GREET and where the systematic review and the Delphi survey fit in this development process.

Previously read: To enable the standardised and transparent reporting of educational interventions in EBP, we have proposed the development of a reporting guideline [19]. Rather than replicate reporting guidelines for study designs, the intent of this reporting guideline is that it should be used in conjunction with existing reporting guidelines for study design, and provide detailed information for describing the intervention only. The first stage in this development process required a systematic review of the literature concerning EBP educational interventions for health professionals [19,20]. Reviewing randomised and non-randomised studies that investigated an educational strategy for facilitating knowledge and skills of EBP resulted in a list of items which have been commonly reported when describing educational interventions.

The next step in the development process for the reporting guideline was to undertake a Delphi survey to elicit a prospective expert consensus opinion regarding the information items that should be included in the reporting guideline [20].

Now reads: To enable the standardised and transparent reporting of educational interventions in EBP, we have proposed the development of a reporting guideline, the guideline for reporting evidence-based educational interventions and teaching (GREET) [19]. Rather than replicate reporting guidelines for study designs, the intent of GREET is that it should provide detailed information for describing the intervention only and would be used in conjunction with existing reporting guidelines for study design. The development process for GREET included three stages, the first of which required a systematic review of
the literature concerning EBP educational interventions for health professionals in stage 1 [19]. Reviewing randomised and non-randomised studies that investigated an educational strategy for facilitating knowledge and skills of EBP resulted in a list of items which have been commonly reported when describing educational interventions.

The second stage in the development process for GREET was to undertake a Delphi survey to elicit a prospective expert consensus opinion regarding the information items that should be included in the reporting guideline [21].

Comment 4: To further support the argument on page 4 for the need for this study you might consider Illic's recent systematic review in Medical Education or Maggio's in Academic Medicine that both identified a need for increased detail in the description of EBM educational interventions.

Response: We are grateful for your input on this point as the systematic reviews by Maggio (2013) and Ilic & Maloney (2014) provide further support for the need for improvements in the consistency in reporting for EBP educational interventions.

Action taken: These references have been incorporated into the introduction and discussion to strengthen the justification for the need for the development of GREET.

Previously read: Despite this investment of time, effort and expertise from researchers and educators in EBP education, little is known about the most effective teaching “prescription” for educational interventions [8].

Further development of the evidence-base for education in EBP is needed, however educational interventions are complex, as are the systems in which they are conducted and these complexities pose significant challenges in the design, evaluation and reporting of educational interventions [8]. Olson and Bakken (2013) recently highlighted the need for standardised and transparent reporting in educational interventions.

Now reads: Despite this investment of time, effort and expertise from researchers and educators in EBP education, best practice for the teaching of EBP remains unknown [8].

Further development of the evidence-base for education in EBP is needed, however educational interventions are complex, as are the systems in which they are conducted and these complexities pose significant challenges in the design, evaluation and reporting of educational interventions [9]. In two recent systematic reviews of EBP educational interventions, Ilic & Maloney (2014) [8] and Maggio (2013) [10] highlighted the need for improvements in the level of detail provided in the reporting of the description of the intervention to enable conclusions to be drawn regarding the efficacy of EBP educational interventions.

Previously read: The standard of reporting for educational interventions in EBP remains inconsistent [6]. Olson and Bakken (2013) [8] list poorly described interventions as “a common complaint of investigators undertaking systematic reviews on the effectiveness of educational interventions”.

3
Now reads: The standard of reporting for educational interventions in EBP remains inconsistent [8-10]. This means the most effective intervention for increasing competency for EBP is not able to be determined despite the extensive investment of time and resources spent on educational interventions for EBP [8]. Olson and Bakken (2013) [9] list poorly described interventions as “a common complaint of investigators undertaking systematic reviews on the effectiveness of educational interventions”.

**Comment 5:** Please identify the limitations of this study.

**Response:** The discussion section for the manuscript has been updated to include the heading ‘limitations’ to enable the limitations for this study to be clearly acknowledged.

**Action taken:** Amended as requested see response to Comment 6.

**Comment 6:** For example, as pointed about by Dr. Mi what are the implications for the small number of physician participants?

**Response:** This Delphi manuscript was submitted as a companion paper with the systematic review manuscript. We recognise the potential difficulty for the reviewers reading and interpreting this Delphi manuscript without the context of the systematic review manuscript.

To provide context for the reviewers, the systematic review manuscript (currently under review with BMC Medical Education) has been provided.

The participants planned for inclusion in the Delphi survey were the corresponding authors of the studies included in our systematic review, and the journal editors from the journals in which these studies were published.

As the intent for this Delphi survey was to determine the key reporting requirements for describing the intervention undertaken in the learning of skills and knowledge of EBP, the stakeholders were considered to be the researchers and educators undertaking the EBP educational interventions and the journal editors and peer reviewers of these EBP educational interventions. Thus it seemed reasonable to include educators, clinicians, researchers and journal editors as the front-line for undertaking, reporting and reviewing EBP educational interventions.

The demographic data for these Delphi included the participants’ gender, professional role, professional discipline, highest qualifications, experience in their current role and the country of their work. The point made by Dr Mi that there were only three medical doctors or four clinicians who participated in the Delphi survey is based on a mixture of the demographic data for the participants’ highest qualification and their professional role. The highest qualification for medical doctors with a PhD was listed only as PhD which may have resulted in some confusion. However, in table 4, row four, under professional discipline there are 14 (39%) medical doctors who participated in the Delphi survey. We have amended the discussion section to reflect the proportion of medical practitioners in the Delphi survey.
Action taken: The discussion has been amended to read:
There are several potential limitations identified for this study. Firstly, despite the intent to invite a Delphi panel that was representative of authors who had completed an educational intervention study for foundation knowledge and skills in EBP, and journal editors from the journals in which these studies were published, the final Delphi panel was comprised of a predominance of authors (n=28, 78%) who were medical professionals (n=14, 39%), nurses and librarians (n=5, 17%). Most participants were North American (n=22, 61%), and there were no Delphi participants from developing countries. It should be noted that studies from developing countries were under represented in the systematic review undertaken in stage 1 of the development for GREET, with only one study (2%) from developing countries (Mexico – Sanchez-Mendiola 2004 [1]. The corresponding author of this study was invited to participate in the Delphi survey, but did not accept our invitation. It is unclear how input from authors and journal editors from the developing world may have impacted on the results of this Delphi survey.

Comment 7: Additionally, the majority of participants are from North America and there is no representation from authors or editors from the developing world. How might this potentially impact your findings?
Response: The majority of Delphi participants were from North America and there were no participants from developing countries. However, this was not our intent for the Delphi survey. All of the corresponding authors from the 61 studies included in our systematic review and the journal editors from the journals in which these studies were published were invited to participate in the Delphi survey. Studies from developing countries were under represented in the systematic review, with only one studies (2%) from developing countries (Mexico – Sanchez-Mendiola 2004). The corresponding author of this study was invited to participate in the Delphi survey, but did not accept our invitation. The impact of the predominance of North American participants on the outcomes of the Delphi survey is not clear and has been included as a potential limitation for this study.
Action taken: Refer to amendment for Comment 6.

Comment 8: Lastly since this manuscript heavily cites and includes methodology that relies on a paper under review (the systematic review), if possible please provide with your revision a copy of the manuscript clearly marked that it is under review. This will be helpful in clarifying components of the manuscript.
Response: A copy of the systematic review paper marked as ‘currently under review’ has been included to assist in clarifying components of the current manuscript.

Comment 9: Authors’ contributions: In order to give appropriate credit to each author of a paper, the individual contributions of authors to the manuscript should be specified in this section.
Response: The authors’ contributions have been described in greater detail to enable the identification of the individual contributions made by each author.
Action taken: Previously read: AP, LKL, MPM, MTW, planned and undertook the Delphi survey, analysed the results and completed the first draft of the manuscript.
PG, DM, JG, JKT, MH, contributed to Delphi survey development, analysis of the results and drafting of the manuscript. All authors read and approved the final manuscript.

Now reads: AP planned and carried out the Delphi survey, completed the analyses and drafting of the manuscript.
LKL contributed to the planning stages for the Delphi survey, participated in the entire Delphi process including the analyses and writing of the manuscript.
MPM participated in the planning and development for the entire Delphi process, including reviewing of the results, assisting with the analyses and drafting of the manuscript.
JG and DM made substantial contributions to the analysis of the Delphi survey, particularly with respect to the determination of consensus.
PG provided considerable input into the analysis of the Delphi survey and the determination of the Delphi intervention items.
MH participated in the planning phase for the Delphi survey, assisted in reviewing the results from each round, the analyses and writing of the manuscript.
JKT contributed extensively to the drafting and critical revision of the manuscript.
MTW contributed the original concept for the Delphi process and contributed to the undertaking and analysis of the Delphi process and helped write the manuscript.
All authors read and approved the final manuscript.
Reviewer 2 Daniel Banks

Comment 1: Are limitations of the work clearly stated? I should note that no limitations of the work are discussed.

Response: A sub heading ‘limitations’ has been added to the Discussion section to enable the limitations of this study to be more easily identified. Within this section each of the study limitations which were previously included in the body of the discussion have been identified and discussed in detail.

Action taken: The discussion section has been revised and now includes a ‘Limitations’ section.

Comment 2: On Table 4, the sum of comments for “Face to face contact with learners” should be 7 (not 8).

Response: We appreciate your time and diligence reviewing this manuscript. Based on your observations for Table 4 we have undertaken further review of the sum of comments for “Face to face contact with learners” and confirmed that there were eight comments provided by participants for this item. Therefore no further changes have been made to this table.

Action taken: No amendment undertaken.

Comment 3: On page 11, 27 of 36 responders yields 75%, not 79%.

Response: Our calculation for the response rate of 79% was based on the withdrawal of two participants from the Delphi survey, thus our calculation was based on a total of 34 participants. Further information to clarify this calculation has been added to Table 2 to include the total number of participants for each round (n). Additional information regarding the withdrawal of two participants has been added into the Results section.

Action taken: Previously read: The uptake rate for the Delphi survey was 34 per cent, with 36 out of the 105 potential participants accepting the invitation to participate (Table 2). Response rates across the four rounds were 100% (R1), 94% (R2), 97% (R3) and 97% in Round 4. A total of 27 participants responded to all four rounds, achieving an overall response rate of 79 per cent of the participants accepting the initial invitation to be involved in this study.

Now reads: The uptake rate for the Delphi survey was 34 per cent, with 36 out of the 105 potential participants accepting the invitation to participate (Table 2). Two participants withdrew over the course of the Delphi survey (one after Round 1 and Round 3), resulting in 34 participants for Round 4. Response rates across the four rounds were 100% (R1), 94% (R2), 97% (R3) and 97% in Round 4. A total of 27 out of the final 34 participants responded to all four rounds, achieving an overall response rate of 79 per cent.
Comment 4: In addition, there is no “asterisk” or “cross” for “What method used to decide content” in Table 3 despite this achieving content.
Response: For the item ‘What method used to decide content”, in Table 3 there is no asterisk or cross allocated to this item as consensus agreement was not achieved for this item even after the categories of importance were collapsed. Although this item did achieve a mean score of 7, 64 per cent of participants rated the importance of this item as >7. To achieve consensus agreement participant agreement of at least 80% was required and this was not achieved in this case. As this item did not achieve consensus agreement, the likelihood for inclusion in GREET was characterised as specified in the study protocol (likely to be included, consider for inclusion or unlikely to be included). In the final column for Table 3, (Include in reporting guideline) “What method used to decide content” was erroneously allocated a ‘Yes’. This has been amended to read ‘Likely’.
Action taken:
Previously read:

<table>
<thead>
<tr>
<th>Information item</th>
<th>n</th>
<th>Mean (SD)</th>
<th>Median</th>
<th>MAD-M</th>
<th>Frequency (%) per category of importance</th>
<th>Include in reporting guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>What method was used to decide content</td>
<td>25</td>
<td>7.1 (2.3)</td>
<td>7.0</td>
<td>1.7</td>
<td>32</td>
<td>Likely</td>
</tr>
</tbody>
</table>

Now reads:

<table>
<thead>
<tr>
<th>Information item</th>
<th>n</th>
<th>Mean (SD)</th>
<th>Median</th>
<th>MAD-M</th>
<th>Frequency (%) per category of importance</th>
<th>Include in reporting guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>What method was used to decide content</td>
<td>25</td>
<td>7.1 (2.3)</td>
<td>7.0</td>
<td>1.7</td>
<td>32</td>
<td>Likely</td>
</tr>
</tbody>
</table>

Comment 5: There are typographical errors in the Reference section.
Response: The reference list has been reviewed and corrected.
Action taken: Please refer to our response to the Editor, Comment 1.

Comment 6: 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.

**Response:** How to appropriately determine consensus for our Delphi survey was the result of considerable discussion and debate for our team. In the absence of any gold standard for determining consensus, we decided to base the level of agreement on that used in previous Delphi surveys undertaken in the development for reporting guidelines and the recommendations by the RAND Corporation, the founders of the Delphi process. Khodyakov et al. (2011, page 3) describe the MAD-M as “the preferred measure of disagreement in expert panels that has been widely used since 1980s when the RAND/UCLA Appropriateness Method was originally created. It is a good measure of disagreement because it is not affected by extreme observations and measures deviation from the median, a measure of central tendency typically used in consensus development …”

Reviewer 1 has raised an interesting alternative for determining consensus and we have undertaken Wilcoxon signed rank analysis for the 39 intervention items in the Delphi survey. The results for these analyses and our recommendations for the inclusion of items in GREET appear very similar.

Using the calculations and degrees of consensus defined by Banks et al. (2009), there were 34 items which achieved agreement ratings between ‘somewhat’ and ‘very good’; this was slightly lower than the number of items (n=37) categorised as either warranting inclusion (yes rating, n=19) or consideration (n=18) for inclusion in GREET.

With respect to the items which were not considered to have good agreement or to warrant inclusion in GREET, there were more items which were considered to have agreement levels of none (IQR ≥4 or \( p ≥ 0.05 \)) (n=5), when compared to the items which were categorised as unlikely to be essential for inclusion, or not for inclusion in GREET (n=2).

If we had used the method recommended by Banks et al. (2009), the recommendations would have altered from could be considered for inclusion, to unlikely to be essential for inclusion, for three items: Relation of instructor to learner / program, Who was involved in designing the content and Whether the educational intervention was endorsed by an academic, educational or professional institution.

However as discussed in the next response (Comment 7), the final decision to include or exclude items in GREET were not based on the results of the Delphi survey alone. The levels of agreement for recommendations to include or not include were noted as described but all items provided by Delphi participants, irrespective of the level of agreement, were considered and discussed in the next stage of the reporting guideline development (consensus meetings).

**Action taken:** A paragraph has been added to the Discussion section (pg. 16) outlining the possible outcome of the Delphi survey with the alternative method for determining consensus based on probabilities (Banks et al. 2009). The text reads:

*There is an alternative scientific method to address agreement which is based on probabilities (p values) using the Wilcoxon signed rank test and interquartile range (IQR)* [28]. Based on the outputs from Wilcoxon signed rank analyses, the recommendations for
the inclusion of items in GREET are very similar to the analysis we employed. The final recommendations for three items (Relation of instructor to learner / program, Who was involved in designing the content and Whether the educational intervention was endorsed by an academic, educational or professional institution) would change from consider for inclusion, to unlikely to be included in GREET. The outcome of the Delphi survey would have been similar, regardless of whether the current or alternative method was applied to determine consensus.

Comment 7: I ask that the authors address these two outliers and the validity of their “empiric” categories.

Response: The two outliers are “Timing of intervention” and “Extent of peer interaction”) which did not achieve consensus but have comparable mean, median and MAD-M scores to items which have achieved consensus; we agree that the mean scores are comparable with other items which have achieved consensus.

The main difference for these two items, when compared to the items which achieved consensus agreement, is the spread of participant ratings, which we have inferred to indicate less participant agreement. The SD scores of 2.3, for “Timing of intervention” and “Extent of peer interaction” was greater than the SD which ranged from 1.4 to 1.9 for the other items which did achieve consensus.

Admittedly “Timing of intervention” and “Extent of peer interaction” were close to achieving consensus, with 76% of participants rating the importance of these items as >7. However, this was lower than the participant agreement of 88 per cent achieved for the other items.

The allocation of the four importance categories (low 0 to 4, moderate 5 to 6, high 7 to 8 and very high >8) was considered by the research team to be a practical and realistic decision. We are not able to provide validation for our empiric categories and acknowledge that in future Delphi surveys, using a nine point Likert scale with three, three point categories as recommended in the recently released RAND online resource (2013) (http://www.rand.org/pubs/monograph_reports/MR1269.html), is ideal. We have acknowledged this limitation and included the recommendation for using a nine point Likert rating scale for future Delhi surveys.

Despite this, the decision to include or exclude items within GREET was not intended to be based on the outcome of the Delphi survey alone. It was planned that all intervention items arising from the Delphi survey would undergo further discussion and review during the consensus meetings to determine the final inclusion and exclusion of items in GREET. The intent for calculating participant agreement for the Delphi items was to provide a guide for these consensus meetings regarding the relative importance for each of the items determined in the Delphi survey.
We acknowledge that our determination of consensus and the allocation of categories of importance used in the rating for the Delphi information items is a limitation for this study. Action taken: Added to limitations: Secondly, there are no current recommendations for determining the threshold for consensus agreement. In the absence of a gold standard method for determining consensus, the a priori level of agreement used in this study was based on previous Delphi surveys undertaken in the development for reporting guidelines and the recommendations from the RAND Corporation [27]. Despite excellent agreement for many items (evidenced by low MAD-M scores), a priori consensus agreement was only achieved for two items. This stringent level required for consensus may account for the small number of items achieving consensus agreement after four rounds. Furthermore, the allocation of four categories of importance resulted in narrow groupings for the Likert ratings. With the exception of low importance, each category spanned two Likert ratings (low 0-4, moderate 5-6, high 7-8, very high 9-10). Merging the importance categories into two (<7 and ≥7) resulted in an eight-fold increase in the number of intervention items achieving consensus agreement (from two items to 18 items). On reflection, a 9 point Likert rating scale as recommended in the recently released RAND online resource, with three, three point rating categories may have been more appropriate.

Comment 8: 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.

Response: The abstract has been modified to reflect the two ways that consensus was able to be reached.

Action taken:
Previously read: Over three subsequent rounds participants were invited to rate the importance of each of the Delphi items using an 11 point Likert rating scale. Consensus agreement was set a priori as at least 80 per cent participant agreement.

Now reads: Over three subsequent rounds participants were invited to rate the importance of each of the Delphi items using an 11 point Likert rating scale (low 0 to 4, moderate 5 to 6, high 7 to 8 and very high >8). Consensus agreement was set a priori as at least 80 per cent participant agreement. Consensus agreement was initially calculated within the four categories of importance (low to very high), prior to these four categories being merged into two (<7 and ≥7).

Comment 9: 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?

**Response:** The structure of the identified sentence has been amended for clarity. The intent of the sentence was to convey that for items which did not achieve the level of participant agreement required for consensus, their final rating of importance would be determined based on the category which achieved the greatest participant agreement. This allocation for the items which did not achieve consensus agreement was determined during the prospective planning for the Delphi survey and included in the published study protocol. The intent for this allocation of the items not achieving consensus was to ensure that a contingency plan was in place to provide guidance in the instance where consensus agreement was not achieved.

However, as the Delphi survey progressed and it became clear that despite high levels of participant agreement, the a priori levels of participant agreement required for consensus were not likely to be achieved. The decision was made to merge the four categories of importance into two. As stated in the study protocol, the importance for the Delphi information items which did not achieve consensus agreement, were characterised according to the category with the greatest participant agreement.

**Action taken:** Previously read: Items with the greatest participant agreement in the low importance category (Likert scores 0 - 4) were deemed unlikely to be essential in reporting; items with the greatest participant agreement in the moderate to high importance category (Likert scores 5 – 8) could be considered for inclusion; items with the greatest participant agreement in the very high importance category (Likert scores >8), should be included in reporting.

Now reads: Items with the greatest participant agreement in the low importance category (Likert scores 0 to 4) were deemed unlikely to be included in GREET; items with the greatest participant agreement in the moderate importance category (Likert scores 5 to 6) were characterised as could be considered for inclusion in GREET and items with the greatest participant agreement in the high to very high importance category (Likert scores ≥7), were characterised as likely to be included in GREET.

**Comment 10:** 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.

**Response:** With respect to the item “Supporting structures in organisation to maintain behaviours targeted by the organisation,” the reason this item was recommended for inclusion in GREET was that consensus agreement was achieved after the categories of importance were merged. Eighty-six per cent of participants rated the importance for “Supporting structures in organisation to maintain behaviours targeted by the organisation” as ≥7.
Action taken: No amendment undertaken.
Reviewer 2 Misa Mi

Comment 1: Second paragraph, page 12, and first paragraph, page 13. Is it an error to refer to Table 2? Table 2 shows demographic information on participants. Do the authors mean Table 3?
Response: Thank you for your diligent review. Your observation is correct. It was an error to refer to Table 2 in these instances as we were discussing the results from Table 3.
Action taken: These errors have been corrected in the manuscript.
Previously read: …achieved consensus agreement (Table 2). When the four categories of importance were merged in to the two categories of low to moderate importance (0 - 6) and high to very high importance (>7), a further 16 items achieved consensus agreement (Table 2).
Now reads: …achieved consensus agreement (Table 3). When the four categories of importance were merged in to the two categories of low to moderate importance (0 - 6) and high to very high importance (>7), a further 16 items achieved consensus agreement (Table 3).

Comment 2: Table 3 on page 28. Several sets of Delphi intervention items share the same domain or overlap in terms of the domain covered. For instance, the following 2 sets of items are all related to one another and can be combined:
1. Contact time
   - Number of face to face teaching/learning sessions
   - Face to face contact time with learners
   - Student time NOT covered by face to face contact
   - Non-face to face contact time with learners
2. Post-training activities
   - Any post-intervention activities required
   - Supporting structures in organization to maintain behaviours targeted by intervention
   - What post-training support was provided
   - Whether follow-up sessions planned
Response: We completely agree with you that there is considerable overlap across many of the information items volunteered by the Delphi participants.
This was an issue that generated significant discussion within our research team to determine how to group the Delphi items and whether it would be more efficient to combine the items. However, the decision was made, in keeping with the Delphi process and to minimise the potential for investigator bias, not to impose our interpretation on the meaning for the items or to modify / combine any of the information items volunteered by the Delphi participants.
Action taken: The relationship and overlap across many of the Delphi information items has been acknowledged in the limitations section in the discussion.
Almost half of the Delphi items (n=37, 49%) did not relate to describing the intervention, which was the primary question posed by the Delphi survey. Furthermore, there was considerable overlap between many of the information items volunteered by the Delphi participants. However, in keeping with the intent of the Delphi process, irrespective of their interpretation, no items were discarded or modified by the researchers.

Comment 3: Method: First paragraph, page 6. Can you provide the number of studies included in the systematic review from which the corresponding authors were contacted for the Delphi survey?
Response: To assist with the context for the Delphi survey a copy of the companion systematic review manuscript has been provided. All corresponding authors from the 61 included studies were contacted via email and invited to participate in the Delphi survey. This information has been added to the participants section of the manuscript.
Action taken: Manuscript amended.
Previously read: Invitations for the Delphi survey were sent to corresponding authors of studies included in a recent systematic review [7] and to the editors of the journals in which these studies were published.

Now reads: Invitations for the Delphi survey were sent to corresponding authors of the 61 studies included in a recent systematic review [7] and to the editors of the 44 journals in which these studies were published.

Comment 4: Method: Third paragraph, page 7. Who was the four staff participating in the pilot testing of the first round survey? Were they experts in EBP? Did they ever teach EBP? Did they comment on the appropriateness of the example in Box 1 during the pilot testing of the survey?
Response: The four staff who participated in the pilot testing of the first round survey were experts in EBP, all with extensive expertise in research and teaching EBP at the tertiary level.
Action taken: Further detail regarding their experience in teaching EBP has been added to clarify this point in the manuscript.
Previously read: The initial draft survey was pilot tested by four staff members of the International Centre for Allied Health Evidence (iCAHE) at the University of South Australia, who were not involved in the Delphi proper [24].

Now reads: The initial draft survey was pilot tested by four staff members with expertise in EBP research and the practice and teaching of EBP from the International Centre for Allied Health Evidence (iCAHE) at the University of South Australia, who were not involved in the Delphi proper [25].
Comment 5: Can the editors of the journals in which the studies selected for the systematic review were published be considered as experts in EBP to provide information items for EBP educational interventions in the Delphi study? What was the rationale for selecting journal editors as part of the expert panel?

Response: The rationale for the inclusion of journal editors in the Delphi survey is based on the recommendations by Moher et al. (2010) in their guidance statement for developers of reporting guidelines. Moher et al. (2010) recommend inviting an international multidisciplinary group to participate in the Delphi survey, which includes journal editors. For example, the development process for the CONSORT statement (Schulz et al. 2010) included participants who were journal editors, health care professionals, methodologists, clinical trialists and others with expertise in the reporting of randomised trials.

As the focus of our reporting guideline (GREET) is for the reporting of the intervention component of EBP educational interventions for teaching foundation skills and knowledge in EBP, the journal editors from the journals which published the EBP educational interventions identified and included in our systematic review were considered to be well placed to provide information items recommended for the reporting of EBP educational interventions.

Action taken: No amendment undertaken.

Comment 6: Table 2, page 27. Only three medical doctors or 4 clinicians participated in the Delphi study. It is assumed that the majority of participants (22) had PhD or a doctorate degree, and that those participants may never be involved in practicing EBP. EBP is a health care practice or a different way to practice medicine. Could the composition of the expert panel affect the resulted information items from the four rounds of the Delphi study?

Response: This point was also raised by the editor Dr Lauren Maggio. The demographic information reported for the Delphi participants was separated into the Participants’ sex, professional role, professional discipline, their highest qualifications, experience in their current role and the country of work. The number of medical professionals included in the Delphi survey was 14 (39%) (Table 2, row four). As several of these medical doctors also had PhD qualification, the medical doctors with PhD’s were listed as PhD, rather than medical doctors may have resulted in some confusion. We have amended the discussion section to reflect the proportion of medical practitioners in the Delphi survey.

Action taken: Please refer to our response to the Editor, Comment 6.

Comment 7: Box 1, Page 8. How was the example of an EBP education intervention in Box 1 be selected? It does not seem to be a well-designed and described educational intervention. For example, it does not mention anything related to evaluation or learning assessment. I can't help but wonder if the study participants may have taken the cue from the example in providing information items during the first round of the Delphi study.
Could the information provided in Box 1 may also have led participants to offer 37 items unrelated to educational interventions that mentioned in the last paragraph on page 11?

Response: The purpose of the reporting guideline we are developing (GREET) is to provide guidance for the reporting of the EBP educational interventions. GREET is intended to be used in conjunction with the reporting guideline appropriate for the study design to ensure that all aspects of the study outside of the intervention are reported adequately.

We recognised that this may be a difficult concept for participants to separate out the items relating to the description of the intervention from the items relating to the reporting for the rest of the study such as the participants, methodology and outcomes. In an attempt to clarify that the information we were asking them to volunteer was for the intervention only, the information we included in Box 1 was an example of the description for an EBP educational from a published study that provided only limited detail of the intervention. The purpose of using this example was to prompt participants to think about what information was missing to describe the intervention.

To help clarify the reasoning behind the inclusion of the information in Box 1, further information has been included in the manuscript.

Although we are not entirely sure why participants offered 37 items that were not specific to the description of the intervention, we believe the most likely reason relates to the difficulty in separating information pertaining to the description of the intervention from the other essential information items that should be reported when describing an EBP educational study.

Action taken: Manuscript revised.

Previously read: The initial survey comprised three sections: a brief overview of the Delphi process, demographic information, and one open ended question asking participants which items should be included when describing an EBP educational intervention (Box 1).

Now reads: The initial survey comprised three sections: a brief overview of the Delphi process, demographic information, and one open ended question asking participants which items should be included when describing an EBP educational intervention. An example from a study which provided limited detail in the reporting of the EBP educational intervention was provided as a prompt for participants to help identify information relevant for the reporting of the intervention (Box 1).

Comment 8: The 19 intervention items for consideration within the reporting guideline presented in Table 3 do not include any items specifically related to learner needs assessment, the sequence of content addressed, or learning and learning outcomes assessment. Learning assessment is a very important component for an effective EBP or any educational intervention. One would question the value and utilization of such a consensus development report or guideline that does not include any items on evaluation/learning assessment.
Response: We agree that information related to learner needs assessment, the sequence of content addressed, or learning and learning outcomes assessment are all very important components in reporting a study for an educational intervention. Just to reiterate, the purpose of GREET is to provide a guide for the reporting of information relating to the intervention, rather than reporting all aspects of the study design. GREET is intended to be used in conjunction with a reporting guideline specific to the study design used. For example when reporting an EBP educational intervention with a randomised controlled trial design, GREET would be used for the reporting of the intervention and the CONSORT Statement (Schulz et al. 2010) would be used for the reporting of the other aspects of the study. Hence, GREET does not include items regarding evaluation or leaning assessment, as these items are not specific to the description of the intervention. For a randomised controlled trial design, the information relating to the evaluation and leaning assessment information would be covered in items 2a Background and objectives, 6a Outcomes, 13a Participant flow, 17a Outcomes and estimation in the CONSORT statement. To help clarify the intent for GREET for reporting of the intervention only and the need for GREET to be used in conjunction with existing reporting guidelines, further information has been added to the introduction for the manuscript.

Previously read: To enable the standardised and transparent reporting of educational interventions in EBP, we have proposed the development of a reporting guideline [19]. Rather than replicate reporting guidelines for study designs, the intent of this reporting guideline is that it should be used in conjunction with existing reporting guidelines for study design, and provide detailed information for describing the intervention only.

Now reads: To enable the standardised and transparent reporting of educational interventions in EBP, we have proposed the development of a reporting guideline, the guideline for reporting evidence-based educational interventions and teaching (GREET) [19]. Rather than replicate reporting guidelines for study designs, the intent of GREET is that it should provide detailed information for describing the intervention only and should be used in conjunction with existing reporting guidelines for study design.
Reviewer 3 Fares Alahdab

Comment 1: The aim of the study is well defined. The methods seem to be sound and the results are interesting. There is only a minor punctuation error.
Response: All punctuation errors have been corrected in the manuscript.
Action taken: No further action required.

Comment 2: Discretionary revisions: The length of the methods, results, and discussion sections is a little too much. I think it would be better if they are shortened.
Response: We agree that this manuscript is lengthy. However in keeping with the recommendations for the reporting of Delphi surveys (Sinha et al. 2011) we have attempted to report the methodology and results as transparently as possible. Finding a balance between keeping the manuscript succinct and providing sufficient detail in the reporting is difficult. We felt that the level of detail in the reporting for the methods, results, and discussion sections was necessary to enable the readers to understand the entire Delphi process. We appreciate the benefits to the reader of making the manuscript more concise, however we were not able to shorten the manuscript in light of the requests made by the other reviewers who requested further information and clarification of points within the manuscript.
Action taken: Minor amendments and editing has been undertaken in an effort to make the manuscript more concise.