Author's response to reviews

Title: Protocol for the development of the guideline for reporting evidence based practice educational interventions and teaching (GREET) statement.

Authors:

Anna C Phillips (phiac002@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 Tilson (tilson@usc.edu)
Marie T Williams (marie.williams@unisa.edu.au)

Version: 2 Date: 5 December 2012

Author's response to reviews:

Title: Protocol for development of the guideline for reporting evidence based practice educational interventions and teaching (GREET) statement.

Reviewer 1: Patricia McInerney

Level of interest: An article of importance in its field

Comment 1 - pg 9 - 2nd sentence at top: do you mean that reference lists of included studies will be reviewed to identify further studies? As it stands, it is not clear.

Response: Yes, reference lists of included studies will also be reviewed to identify any further relevant studies. This sentence has been revised to clarify the point.

Action: Previously read: “Studies will be identified by searching electronic databases and reviewing the reference lists of included studies.”

Revised to read: “Studies will be identified by searching electronic databases and reviewing the reference lists of included studies to identify further studies.”

Comment 2 - pg 9 - is any time span being given to the inclusion criteria?

Response: We have elected to include all eligible publications irrespective of publication date.

All databases will be searched from inception to ensure comprehensive
coverage.

Action: Previously read: “Studies will be eligible for inclusion in this review if they are:

• Published in peer reviewed journals.
• Published in any language…….”

Revised to read: “Studies will be eligible for inclusion in this review if they are:

• Published in peer reviewed journals irrespective of language of publication or publication date.”

Comment 3 - pg 11 Under data extraction, first paragraph - should this not be Table 2

Response: The reviewer has identified a formatting error. This should read Table 2 not Table 1 and has been amended.

Action: Table number corrected.

Comment 4: How easy will it be to extract professional discipline?

Response: The ease and usefulness of extracting data on professional discipline has been topic of discussion amongst the team. Given the number of reviews where statements have been made concerning the paucity of studies involving professions other than medicine, we have proposed retrieving a small number of demographic items for authors one of which is professional discipline. The difficulty posed with this data item is that, short of contacting authors, it can only be extracted if the authors have included their qualifications either in the author list in the study title or as part of their contact information. In the process used to assess the reliability of raters in using the data extraction template, of the 15 randomly selected articles eligible for inclusion within the review, data on profession was able to be extracted in 8 articles. We believe that this could be useful descriptive information.

Action: No amendment undertaken

Reviewer 2: Note from authors: The difficulty with responding to Reviewer 2 specific comments is that while there are wide ranging philosophical issues raised, there are very few identifiable specific comments. As such we have elected to reproduce the report provided by Reviewer 2 and respond at the end of the report.

Reviewer 2: W. Dale D. Dauphinee

Comments This submission is a protocol (either proposed or partly in process) with the goal of developing guidelines for reporting evidence based practice in
educational interventions and teaching. It appears to have arisen from the 2011 Sicily Statement on the classification and development of evidence based learning assessment tools. In turn, that statement was an outgrowth of the 2005 Sicily Statement on evidence based practice which is based on the epidemiologically anchored Cochrane approach to evidence based practice. In fact the whole protocol makes frequent references to this approach to the exclusion of all other views of ‘evidence’. The authors are seemingly unaware of the real problems in applying the Cochrane approach to the existing literature on many aspects of teaching and evidence versus defining evidence around clinical decisions made in medical practice. Cochrane relies heavily on hierarchies of evidence of which RCTs are, of course, considered ideal. It is of note that one of the authors of this proposal is associated with the Best Evidence Medical Education Collaborative which has led to many systematic reviews of evidence in medical education and teaching. After many of the reports emerging from BEME, it is apparent that, with the exception of learning associated with very disease focused and clinically oriented training situations, such as line insertions, BEME reports regularly conclude that there is not strong evidence one way or another and that more studies are needed. Why have the results from BEME reviews been so inconclusive? There are many reasons. Poor librarian and informatics support, poor choice of search words, and very limited scope of search into other disciplines. But the most important reason is that RTC and other so called higher levels of evidence are seldom carried out in education for reasons that are oft cited, including cross contamination of groups, studies are not randomized, or are carried out at only one of two sites. Thus generalizability and extrapolation of results to the specific end points and specific target groups behaviours is difficult. Kane’s discussion of types of validity” discussions needed when interpreting the results of the studies is seldom appreciated in such studies. Thus validity (which what I presume the authors mean by external validity) is critical in this field – especially if performance is an outcome measures. In short, validity has to be established with each use of a tool or instrument. Hence I would argue that the Cochrane approach has yet to be shown to validly address and interpret existing evidence in the proposed field of study to the exclusion of other valid forms of evidence.

.. As is now obvious, I have cited BEME’s challenges because it seems to me that these authors may not have appreciated the magnitude of the challenge of finding evidence in education and teaching - especially in the health sciences, if one plays exclusively by Cochrane rules. This is reflected in the complete bias of their literature review which should always be about what is known in the field. The search must be cast widely as much of the relevant literature is not found in the mainline educational or clinical journals or even within medical and health sciences education journals. Thus if search terms are limited, much is missed. BEME reviews have demonstrated this time and again. Expertise in search strategies is essential and must be executed by individuals who are expert in the field and who realize that many of the key studies in educational interventions demands a very wide casting of search terms. If not, the search can be a potentially inefficient process as the field is vast and terminology
Even more disconcerting in BEME’s experiences, and hotly debated is: is the Cochrane approach in appropriate for education? What of the Campbell Collaborative and the nature of evidence in work in the Social Sciences and policy fields? It may be of great interest that at the BEME Congress in Lyon in September 2012, the very anchors of the BEME protocol strategies and methodology currently under use is being seriously questioned. At that meeting, it was agreed that to date BEME had yielded little. It was pointed out that a hard look at strategic assumptions and its ready adoption of the Cochrane approach may have been its prime failing. These same points had been raised by others as early as 2002 when the BEME movement began – not because the idea was wrong – but because the assumptions and proposed methodology and rubric was not compatible with much of the science and non-epidemiological evidence in the educational and teaching literature (see Norman and Eva and Dauphinee and Wood Dauphinee).

To summarize, I am concerned that the authors are on a journey that maybe ill-fated due to lack of preparation and lack of empirically based pre-testing of assumptions. The Scott-Amundsen race to the South Pole illustrates my point perfectly. One had done his homework and had validated his assumptions and pathway and thus had anticipated problems by having redundancy of back-up supplies, etc. Scott did not. Thus, my challenge to the authors will focus on their assumptions and their implied native view that ‘validated’ (there is not such thing as each tool must be validated with each use – just like one’s bank account)! Last week’s balance will not be fine if one decides to buy a new automobile this week. That is the essence of validity – how does one interpret the findings in terms of proposed uses.

In reality, I would have preferred to see a staged approach wherein each proposed step and all assumptions were evaluated before this massive undertaking begins. Here is what I would suggest. Cut the submission in half and tell the readers how you will establish the validity of the proposed approach and all its tools – for this purpose.

Response: Reviewer 2 raises a number of philosophic concerns; the “Cochrane approach and or the Best Evidence Medical education collaboration (BEME)”, the nature of evidence, the appropriateness of higher level research designs in educational interventions and clinical setting, and the validity of systematic review processes and outcome assessment. These concerns are broad and we do not necessarily agree or disagree with many of these conceptual issues.

We agree that there are knowledge frameworks other than EBP, but this decision making paradigm is rapidly becoming part of accreditation requirements for both entry-level and post graduate health programs in educational institutions and various health care registration bodies. There is a rapidly growing body of research concerning educational practices used to teach and facilitate knowledge and skills in EBP, hence the consensus papers concerning what should be taught in EBP (Sicily 1) and the development of specific outcome instruments for
use in assessing EBP foundation knowledge, skills, behaviors and attitudes (Sicily 2). As Reviewer 2 correctly indicates, intervention studies (of all designs/approaches) in education have a number of challenges and amongst the difficulties in interpreting and contrasting such studies is the variety of ways in which the actual educational intervention is reported.

We would like to clarify and reiterate the aim of the intended protocol.

We aim to produce a reporting guideline for use in describing/reporting educational interventions used to facilitate foundation skills and knowledge in evidence based practice. The focus of this guideline is how educational interventions are reported rather than the specific study design or approach (e.g. RCT, cohort, case study, action research etc.) for which there are already pre-existing reporting requirements. These new reporting guidelines are intended to be specific for the educational intervention and would be used in conjunction with, rather than a replacement for, existing reporting guidelines for, specific research designs (e.g. CONSORT, TREND and STROBE).

We are not seeking to determine which research methods or approaches should be used to evaluate educational interventions, best teaching practice, the impact of educational interventions or to analyse learning outcomes or generalize the educational interventions.

The reviewer makes relevant comments regarding the potential pitfalls of systematic reviews of RCT’s in education, especially when the question concerns the effectiveness of interventions. However, the systematic review proposed in our protocol does not pose a question of effectiveness.

The primary aim of the systematic review (Phase 1) is to sample studies which investigate an educational intervention used for teaching basic skills in EBP, in order to describe the manner in which EBP educational interventions have been reported. (i.e. determine what information items have authors reported regarding the intervention). We have restricted the eligibility to “higher level” designs as a way of sampling the literature.

During the development phase of this search strategy, we included a larger range of original information (all publically accessible peer reviewed materials, books, conference presentations etc.) that might report or describe educational interventions. As might be expected, the number of materials and citations eligible was large with 17577 citations identified. Consequently we have elected to only include the “higher level” designs to provide a sample of documents for their descriptions/reporting of educational interventions. At this stage, we have reviewed 504 citations in order to identify 68 higher level designs (RCT’s and controlled trials).

Reviewer 2’s comments infer the possibility that different source documents (designs/research approaches/descriptions) may provide different or more detailed descriptions of the educational strategy used to teach EBP foundation
skills. As part of the assessment of inclusion / exclusion criteria for this search, we identified the nature of the information source (i.e. RCT, controlled trial, pre-post, narrative review etc.). To explore the possibility that different designs/sources might result in different descriptions of the educational strategy, we compared randomly selected articles of different types (RCT/controlled trials (n=15), pre-post designs (no control group n=5) and narrative reviews of EBP education (n=5), using the data extraction template developed for this review. There was little difference in the specific reporting of the educational interventions in terms of the details provided re the participants, the instructors, educational philosophy, prior exposure, content of the intervention, mode of delivery, frequency, duration, assessment or confounding variables (and the higher level designs intended for inclusion within this review provided slightly greater detail for a number of items). This was a very useful process to explore the possibility and we have revised the protocol to include this.

To use Reviewer 2’s metaphor of the ill-fated Scott-Amundsen race to the South Pole.

Our team is not intending to undertake an expedition to the South Pole. We are interested in how people report such expeditions. Using a staged approach, we plan to sample a variety of source documents that report the expedition and collate items. We will then seek the perspectives of authors who have written documents covering the expedition and ask their views about which items should be covered in reporting the expedition. We will supplement these views with any items uncovered in the document search but not volunteered by authors. We will then propose draft guidelines for authors planning to describe or document South Pole, dog led expeditions and seek comment.

We accept that there are philosophical differences in terms of which frameworks are appropriate for exploring educational interventions and definitions of what evidence should be considered. Regardless of which framework is proposed or accepted, there is value in having a guideline for use when reporting the educational intervention. This guideline would allow consistency in describing the educational intervention irrespective of the study design / methodological approach used by researchers.

Action:

Revision 1: Clarify that the intent of this new guideline is that it will be used in conjunction with, rather than a replacement for existing reporting guidelines. The design used for each study will determine which reporting guideline would be used to report a study of that particular study design.

Previously read: “Developing the GREET statement will provide a reporting guideline for EBP educational interventions which will potentially benefit all stakeholders in EBP education; researchers, educators, editors, reviewers and students.”
Revised to read: “Developing the GREET statement will provide a reporting guideline for EBP educational interventions that can be used in conjunction with existing guidelines to potentially benefit all stakeholders in EBP education; researchers, educators, editors, reviewers and students.”

Revision to discussion: paragraph added.

There are knowledge frameworks other than EBP, but this decision making paradigm is rapidly becoming part of accreditation requirements for entry and post graduate health programs in educational institutions and a various health care registration bodies. There is a rapidly growing body of research concerning educational practices used to teach and facilitate knowledge and skills in EBP, hence the consensus papers concerning what should be taught in EBP (Sicily 1) and the development of specific outcome instruments for use in assessing EBP foundation knowledge, skills, behaviours and attitudes (Sicily 2). There are philosophical differences in terms of which frameworks are appropriate for exploring educational interventions and definitions of what evidence should be considered. Regardless of which framework is proposed or accepted, there is value in having a guideline for use when reporting the educational intervention. This guideline would allow consistency in describing the educational intervention irrespective of the study design / methodological approach used by researchers.

Revision 2: To clarify that the included studies are only providing a sample of reporting in studies reporting and EBP educational intervention in recognition that this sample may not be representative of all study types. We have now added to systematic review analysis:

It is possible that lower level designs (i.e. not RCT/controlled trials) may provide different descriptions of the educational intervention and report different items. To explore this possibility, data will be extracted from a random selection of excluded studies with lower level designs (pre-post without control and narrative reviews) and compared with the higher level studies (RCT/controlled trials) used during the reliability process (n=15).

Reviewer 3 Daniel E Banks

Level of interest: An article of outstanding merit and interest in its field.

Comment 1.My group introduced our “solutions” to some of the same concepts that you consider. First, we wanted some early success so we developed some questions that would likely engender agreement in the first round. We developed a statistical test for agreement that allowed us to complete some questions in the first or early rounds and make the burden on the participant less. You report “consensus will be deemed to be achieved for each item of >80 per cent agreement.” Does this mean that more than 80% of the responses to each question will need to have the same Likert scale score to be considered for agreement.
Authors’ response

The plan for the Delphi procedure used in this study is based upon the recent recommendations for conducting and reporting Delphi procedures by Sinha et al. (2011). We believe these guidelines for Delphi procedures were not available at the time that your paper was published. Based upon the results of a recent systematic review of the reporting of Delphi surveys Sinha et al. (2011) recommend that where possible it is desirable to commence the Delphi survey with an open ended question. This will minimise the introduction of bias associated with selecting questions or items for the participants to comment on. Commencing with an open ended question seeks the opinion of all participants and assigns equal weight to all answers.

Consensus for this study is intended to be when 80 percent or more respondents rate an item within a pre-defined range of agreement. Specifically the four categories according to the Likert score (0-10) are:

- 0-4: Low importance. Item not included.
- 5-6: Moderate importance. Possible consideration for inclusion.
- 7-8: High importance. Likely to require inclusion.
- >8: Very high importance. Essential for inclusion.

Descriptive statistics (percent agreement, mean, score, SD, range) for each item and each category will be computed. For an item to achieve consensus, 80 per cent or more of respondents must have rated the item into one category (low importance, moderate importance, high importance, very high importance). This does not mean that the respondents must rate the item with exactly the same score (e.g. 6) but rather within the pre-defined range.

Action:

Previously read: “For an item to achieve consensus, 80 per cent or more of respondents must have rated the item into one category (low importance, moderate importance, high importance, very high importance).”

Revised to read: “For an item to achieve consensus, 80 per cent or more of respondents must have rated the item into one category. To be considered to have met the consensus criterion, 80 per cent or more respondents will need to rate an item’s importance within in the range for one category (low importance 0-4, moderate importance 5-6, high importance 7-8, very high importance>8).”

Comment 2. We began by planning for 4 rounds, but with time found that there would not be agreement on some questions (no matter how many rounds). We used the same statistical approach for agreement to also define disagreement. When there was no agreement after three rounds, we recognized that it was not
Authors'Comment: The decision to use up to four rounds was based upon the recommendations by Sinha et al. (2011). Four rounds will also enable the Delphi participant’s two rounds of review for any items introduced from the systematic review in Round 2. In our Delphi survey round 1 will serve as an opportunity to gather items which the participants consider important for reporting (which will be rated using Likert scales in Round 2). This differs from your Delphi survey where Round 1 was used to ask participants to agree or disagree with a list of predetermined items. In this case participants were able to rate their agreement or disagreement on three occasions by the end of round three. In our Delphi survey participants will require four rounds to enable the same opportunity to rate their agreement and disagreement (as the first round is an open ended question and item generation). If consensus has been achieved after three rounds, the fourth round will no longer be required and would not be undertaken. Where there are any items unable to reach consensus after the fourth round, these will be subject of discussion with the expert panel in Stage 3.

Action: The descriptive information regarding the intended Delphi procedure is provided on page 14 of the manuscript. No further action required at this stage.

Comment 3. We set up a web site to show all participants work in progress. For each round, we included all comments and manuscript citations as well as Likert scale outcomes (http://www.sh.lsuhs.edu/medicine/delphi/) from round to round. The information provided by each expert was anonymous. In this manner, information was available to all participants on a cumulative basis. The experts could review questions and answers over the “rounds” and change their decisions as appropriate.

Authors’ response: Setting up a website to show all participants the results of the survey as it progresses is an excellent means of providing participants information and feedback about the Delphi survey. While we are not intending to establish a website to enable participants to be able to access the results of the Delphi, we will provide comprehensive feedback to all participants after each round of the survey. This feedback will include all items listed by participants and which items had reached consensus. Any items which have reached consensus will not require further comment or rating.

Comment 4: The reason for the number of “rounds” in the Delphi study is to expose each of the experts to the answers of their peers in order to “drive the experts towards convergence.” At the end of the instructions for round one you ask 2 questions: first, are there any other items that you believe should be reported when describing EBP educational interventions or teaching? And second, are there any comments you would like to add? Nowhere do you suggest that your experts make persuasive arguments for their choices. You are not just asking these scientists for their opinions, rather your goal is to “get them to agree.” I ask that you consider asking the experts to justify their statements for
each Likert score that they put down. For example, if the expert were to score a “9”, he/she would be expected to provide a reference which would justify their view and prove their point to their peers. Similarly, you have expressed no such desire for expert advocacy in rounds two or three. Without this encouragement to make these experts “prove their points”, how can you expect the group to change.

Authors’ response: This is an excellent point and something that has generated a lot of discussion within our team. We agree that it would be very useful to ask participants for a justification for their scores (either agreement or disagreement) to prove their point during the delphi survey. We will include a question asking participants to provide justification for their views in the second round of the Delphi as this will give the experts two rounds to consider the list of items and their response to this list.

Action: Previously read: “After completing the rating exercise, participants will have the opportunity to provide further items by responding to the following questions:

• Are there any other items that you believe should be reported when describing EBP educational interventions or teaching?

• Are there any comments you would like to add?

Revised to read: “After completing the rating exercise from Round 2, participants will have the opportunity to explain or justify their rating choice and provide further items by responding to the following questions:

• If you rated any items <4 or >8, please provide a brief justification or reference to support your choice?

• Are there any other items that you believe should be reported when describing EBP educational interventions or teaching?

• Are there any comments you would like to add?

Comment: 5. Many of your experts will be knowledgeable and spend a lot of time on this. You might ask them if they wish to be acknowledged at the end of the manuscript.

Authors’ response: This is a valuable insight and one that we had included but did not explain in the manuscript. As part of the participant’s demographic information collected in the first round of the Delphi survey, participants are invited to provide their name and consent to be acknowledged as a member of the Delphi panel in presentation / publications arising from this research.

Action: Previously read: “Demographic data regarding the participant’s profession, qualifications and contact details will be recorded.”
Revised to read: “Demographic data regarding the participant’s profession, qualifications and contact details will be recorded. Participants are invited to provide their name and consent to be acknowledged as a member of the Delphi panel in presentation / publications arising from this research.”

Comment 6. As time goes on, some of the experts will drop out. It is likely that this will not occur in a “symmetric distribution” but rather randomly. That is, if five experts decline to answer in the third round it may be that these experts had all scored the question a “9.” In their absence the Likert score for this specific question dramatically changed. How will this be addressed?

Authors’ response: As you are aware, participant drop out is a commonly reported issue in studies using the Delphi method. Many studies report response rates of 50% or less and the response rates from your Delphi ranged from 33 per cent (round 3) to 48% (round 1).

In terms of participant drop out after stage 1 of the Delphi process, we would have data on their last round ranking of items. If participants dropped out after round 1, but before round 2 there would be no ranking data; if they dropped out after Round 2 but before Round 3 we would have Round 2 rankings; if they dropped out after Round 3 but before round 4 we would have round 3 ranking data. Depending on the number of “drop outs’ and last available round scored, we may, at least, be able to identify whether ranking for an item were consistent or random and predict the likelihood of the impact on the attrition.

As this project is a three stage project, the Delphi survey (stage 2) will be followed by further expert review and discussion in Stage 3. This means that all items arising from the Delphi will go forward for further discussion and review by the expert panel in stage 3. The rating scores for each item will be provided to the expert panel, however these on their own will not necessarily determine the items inclusion or exclusion from the guideline. The final decision to include or exclude each item will be determined by the expert panel in stage 3. In addition, the draft reporting guidelines will be “tested” by participants of the Delphi (see stage 3 information in the manuscript).

Comment: 7. Finally, we used only questions that could be addressed on the Likert scale. Inclusion of open-ended questions will require a plan on what the frequency of answers will need to be to add these questions to your Delphi program using Likert scale analysis

Authors’ response: As you are aware, each of the stages is based on the responses from the previous round. Round 1 is the open ended question. The data from round one will be grouped according to commonality and a Likert scale used in Round 2 for respondents to rate. In each of rounds 2 and 3, items will predominantly be Likert scales but we will need to allow participants to provide any further items that have not yet been volunteered. Where there are additional items suggested by participants for inclusion it is intended that all items
irrespective of their frequency will be included for review in the subsequent round (i.e. converted to a Likert scale question).

Comment 8: There are no power calculations. I ask that you consider the number of experts in your proposal. You wish to include a maximum of 250 experts. Is this large sample necessary?

Authors' response: As there are no current guidelines for how many experts should be included in a Delphi survey, we based our initial sample size prediction of 250 upon early estimates of how many studies would likely be included (i.e. inviting the corresponding authors of articles included in the systematic review, stage 1). As the systematic review process has evolved it has become evident that we have overestimated the sample size. At this stage we have 68 included studies in the systematic review published in 42 journals. After inviting 105 participants (65 first authors and 40 journal editors) a total of 36 experts have agreed to participate in the Delphi survey. The sample of 36 experts is similar to the number of experts you included in the first round of your paper. Furthermore, 36 is within the range used in previous reporting guideline development which ranged from 11 (CONSORT Begg et al. 1996) to more than 50 (SQUIRE Davidoff et al 2008), with an average of 24 participants.

Action:
Previously read: “All corresponding authors of studies included in the systematic review and the editors of the journals in which these studies were published (stage 1) will be invited to participate in the Delphi survey (likely maximum number 250).”

Revised to read: “All corresponding authors of studies included in the systematic review and the editors of the journals in which these studies were published (stage 1) will be invited to participate in the Delphi survey (likely maximum number 110).”

Comment 9. You wish to include the editors of the journals the manuscripts were published. Each journal masthead includes many editors – which one(s) will you choose? Will the editors have specific knowledge to be able to fairly answer the questions that you pose? They may or may not. Yet, although they are “experts”, you cannot be certain that they will be able to appropriately answer the questions that you pose.

Authors’ response: The chief journal editors for each journal will be invited in the first instance and if they do not wish / are not able to participate they will be asked to nominate another journal editor who may wish to participate. While there is the potential that the journal editors are not experts in the field of EBP education, our team feels that journal editors are ideally placed to provide an expert opinion regarding the type of information that should be reported by studies reporting an intervention. The importance here and expertise that they
bring is regarding the detailed reporting requirements for an intervention irrespective of the type of intervention and whether or not they are expert in the field of EBP/ education.