PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Building capacity to use and undertake research in health organisations: survey of training needs and priorities amongst staff
AUTHORS	Barratt, Helen; Fulop, Naomi

VERSION 1 - REVIEW

REVIEWER	Bev Holmes Michael Smith Foundation for Health Research/Simon Fraser University/University of BC/ Vancouver Canada
REVIEW RETURNED	30-May-2016

GENERAL COMMENTS	Review of Manuscript ID bmjopen-2016-012557, "Building capacity to use and undertake research in health organisations: survey of training needs and priorities amongst staff." This is a well-written paper that tackles an important area of need: training for health care staff to do and use research. I have a few comments that I think should be addressed before the paper is ready for publication. In no particular order, they are: - The authors point out at the beginning of the piece and near the end that training is only one part of a much bigger picture of what is needed, but it would be good to reiterate this at appropriate points in the article. For instance “lack of time and resources” and “lack of institutional support” are huge barriers to the production and use of research. The latter especially needs a bit of unpacking – what is it? Is it only for the training, or is it for the production and use of research itself? Infrastructure, buy-in, appropriate wording in job descriptions – these are all part of the picture. I realize the survey was not meant to get at those, but a better sense of the context in which the training would ideally be implemented would be welcome. - The survey was cascaded to “relevant staff” – what does that mean? How did the initial contacts know who should fill the survey out, and what instruction was given to them? It would help the reader to know that in order to be confident in the sample. - There is one mention of the fact that “importance to the role” is “self-assessed,” but otherwise it stand alone – including in a title – and could be confusing. I think it’s really important that readers understand that the skills’ importance to the role are not necessarily due to their being in a job description (ie demanded by employers) but by the respondent him/herself. Perhaps the authors could comment on the implications of the difference, even if they think there are none?

	- What was the process by which you picked “additional tasks” – did you just ask the seven people (page 9/10) what they thought? Was the process more robust than that? - Similarly, how was the qualitative data analysis done? You got a lot of rich information – it would be good to know how it was categorized, what the process was – and who did it etc. - Could we have a bit more information about the pilot test? Did the survey have to change much? These surveys are notoriously hard to do, since there are so many interpretations of questions, so feedback from the pilot would be helpful. - I’m intrigued by what level the respondents were in the organization, and it would seem to matter. Are they are all fairly autonomous? It would help us understand the results a bit more if we understand the level of authority and autonomy these people had. I was really interested by the task “evaluating your organization’s performance” and agreement with same. There are many staff interested in the production and use of research that would not be thinking of the organization level per se – so would these people likely have some organizational responsibility? - The task of “learning about new research developments” sounds communications related to me, but here it seems to have been conflated with implementing evidence informed practice. Could you explain why “learning about new research developments” was categorized with “using and conducting research?” And perhaps provide some assurance that it didn't come across to respondents as simply being interested to know what the latest is? That’s it from me. Thanks for the opportunity to review this paper.

REVIEWER	Leider, Jonathon de Beaumont Foundation, US
REVIEW RETURNED	04-Jun-2016

GENERAL COMMENTS	This manuscript is well-written, and focuses on an important topic. The methods seem reasonable, although the pilot/pre-test appears quite small. It would be impossible, I think, to run any meaningful psychometrics with so few pretesters. The one concern I have relates to the analysis and presentation of your Likert-type responses. I concur wholly with the concerns you have raised regarding displaying and comparing these questions. So I was quite surprised to see you did not handle this in a more appropriate way - namely dichotomizing the responses around a particular point (e.g., positive versus all others). Even with a 7 point LT, this would be a way of handling the issue you mention. Was there significant skewing in the responses? Difficult to tell about the distribution with what is presented.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

This is a well-written paper that tackles an important area of need: training for health care staff to do

and use research. I have a few comments that I think should be addressed before the paper is ready for publication. In no particular order, they are:

1. The authors point out at the beginning of the piece and near the end that training is only one part of a much bigger picture of what is needed, but it would be good to reiterate this at appropriate points in the article. For instance “lack of time and resources” and “lack of institutional support” are huge barriers to the production and use of research. The latter especially needs a bit of unpacking – what is it? Is it only for the training, or is it for the production and use of research itself? Infrastructure, buy-in, appropriate wording in job descriptions – these are all part of the picture. I realize the survey was not meant to get at those, but a better sense of the context in which the training would ideally be implemented would be welcome.

We have made a number of changes to address this point. This includes, clarifying in both the results (page 23, lines 9-11) and discussion section (page 24, lines 10-12) that training is only one of a range of activities that may be used to increase use of and participation in research. We have also explained in the abstract (page 2, line 26) and in the article that respondents’ comments about ‘lack of institutional support’ specifically referred to their participation in research. Finally, we have added text to the discussion noting the importance of context and the range of factors that should be considered when implementing training (page 27, line 22 – page 28, line 2).

2. The survey was cascaded to “relevant staff” – what does that mean? How did the initial contacts know who should fill the survey out, and what instruction was given to them? It would help the reader to know that in order to be confident in the sample.

To address this point, we have expanded the section on questionnaire administration (page 9, lines 4-11). In this, we now explain that an email was sent to key CLAHRC contacts in partner organisations, with a request that they disseminate the questionnaire link to staff electronically. The email explained that the survey was to inform the design of training opportunities for health care and public health staff, to increase their skills in using research evidence. It stated that we were keen to receive responses from staff with a range of backgrounds and experience, at all levels, from a range of groups including but not limited to clinicians; nurses and midwives; allied health professionals; managers; and technical staff, such as laboratory workers.

3. There is one mention of the fact that “importance to the role” is “self-assessed,” but otherwise it stand alone – including in a title – and could be confusing. I think it’s really important that readers understand that the skills’ importance to the role are not necessarily due to their being in a job description (ie demanded by employers) but by the respondent him/herself. Perhaps the authors could comment on the implications of the difference, even if they think there are none?

To address this point, we have noted in several places in the methods and results sections that both the importance and performance ratings were self-assessed. See for example page 24, lines 15-88. We have also added text to the limitations section (page 26, lines 7-11) reiterating this point and explaining that there may be a discrepancy between what participants consider to be important, versus what their employers require of them.

4. What was the process by which you picked “additional tasks” – did you just ask the seven people (page 9/10) what they thought? Was the process more robust than that?

The modification process involved two stages. First, we identified possible additional tasks from the literature on research use and participation by health care and public health staff, and through one to one interviews with staff from a range of backgrounds (n=7). We did this iteratively by sense checking new suggestions with subsequent interviewees. Second, we presented the proposed modifications to

the seven individuals who piloted the questionnaire, and asked them to comment on whether the changes to the list of tasks appeared valid and realistic in the context of their current role. We have amended the methods section to reflect this (page 10, lines 3-11).

5. Similarly, how was the qualitative data analysis done? You got a lot of rich information – it would be good to know how it was categorized, what the process was – and who did it etc.

As we have outlined, we conducted qualitative content analysis of the free text data to identify research training priorities, using the systematic method set out by Mayring and others (Forum Qual Sozialforschung Forum Qual Soc Res 2000; 1.) Categories were derived iteratively using Mayring's step model of inductive category development. Within this, the researcher (HB) reviewed all the free text data in light of the research questions. Free text comments relating to similar topics (e.g. training in research methods; using research to inform practice) were grouped together. From this, provisional categories were deduced and revised, with constant reference to the data. The research team then checked the reliaibility of the final categories, before HB conducted quantitative aspects of the analysis (e.g. frequency of the coded categories). Priorities were first identified for the whole sample, and then compared using the same categories to examine potential differences between professional groups. We have amended the methods section to reflect this (see page 16 line 17 – page 17 line 2).

6. Could we have a bit more information about the pilot test? Did the survey have to change much? These surveys are notoriously hard to do, since there are so many interpretations of questions, so feedback from the pilot would be helpful.

Pilot testers were provided with a copy of draft questionnaire. As well as asking them about the proposed modifications to the list of tasks, we asked them whether the text of the questionnaire was clear and how realistic it was in the context of their current role. Overall, the tool was considered to be good, with clear instructions, and of appropriate length. Through this process we made minor modifications to the tool including: adding in a definition of research; providing additional job categories in the section on demographic data; and clarifying the instructions for the importance/ performance rating exercise. Pilot testers considered the list of tasks to be clear and comprehensible, acknowledging the challenges of compiling a list that would be of broad relevance across a range of different types of health organisation. Two testers suggested that we group together similar tasks. However, the questionnaire developers intended that the list should be organised randomly, so we opted to keep this approach to maintain the integrity of the tool. We have amended the methods section to reflect this (page 10 line 13 – page 11 line 2)

7. I'm intrigued by what level the respondents were in the organization, and it would seem to matter. Are they are all fairly autonomous? It would help us understand the results a bit more if we understand the level of authority and autonomy these people had. I was really interested by the task "evaluating your organization's performance" and agreement with same. There are many staff interested in the production and use of research that would not be thinking of the organization level per se – so would these people likely have some organizational responsibility?

We did not collect detailed information on participants' seniority within their organisation for two reasons: 1) this information would have been difficult to assess and compare across a large number of organisations of different types and sizes, and 2) we sought to ensure that individuals could complete the questionnaire without risk of being identifiable. However, in light of this comment, we have added text to the limitations section noting that this may have affected participants responses, for example to the evaluation question (page 25, 23-25).

8. The task of "learning about new research developments" sounds communications related to me, but here it seems to have been conflated with implementing evidence informed practice. Could you

explain why “learning about new research developments” was categorized with “using and conducting research?” And perhaps provide some assurance that it didn't come across to respondents as simply being interested to know what the latest is?

We included ‘learning about new research developments’ in the questionnaire because, in order to implement new evidence, health care and public health staff need to first learn about the existence of that evidence. This is a central part of the ‘activities used to link research to action’ described in the Research Knowledge Infrastructure by Ellen et al (Implementation Science 2011; 6:60), which we have used as a basis for this paper. We cannot know how individual respondents interpreted this. However, this task was not highlighted by our pilot testers, all of whom regarded the list of tasks as valid and realistic in the context of their current role.

Reviewer 2

1. The methods seem reasonable, although the pilot/pre-test appears quite small. It would be impossible, I think, to run any meaningful psychometrics with so few pretesters. The one concern I have relates to the analysis and presentation of your Likert-type responses. I concur wholly with the concerns you have raised regarding displaying and comparing these questions. So I was quite surprised to see you did not handle this in a more appropriate way - namely dichotomizing the responses around a particular point (e.g., positive versus all others). Even with a 7 point LT, this would be a way of handling the issue you mention.

Drawing on advice from our statistician, we did not dichotomise the responses for two reasons: 1) the Hennesy-Hicks questionnaire is a validated tool, designed so that training needs are assessed by comparing the mean importance and performance scores given to each task. In dichotomising the data, we would have been making a significant change to the overall design, and would also have had to make a post-hoc decision about where to set the cut point. 2) By dichotomising variables, information is inevitably lost. Our goal was to assess research training needs and priorities, to inform a programme of training. This included assessing the relative importance of different needs. If we had dichotomised the data, we would only have been able to say, for example, that tasks were considered either highly or less important. This would not have helped us identify those tasks considered particularly important, to prioritize for training.

2. Was there significant skewing in the responses? Difficult to tell about the distribution with what is presented.

To address this point and provide information about the distribution of responses, we have provided a supplementary file, in which we present data about the overall median and inter-quartile range of responses for each question. We have referenced this in the manuscript (page 19, lines 2-3).

VERSION 2 – REVIEW

REVIEWER	Leider, Jonathon JHU USA
REVIEW RETURNED	01-Aug-2016

GENERAL COMMENTS	This is an interesting article on an important topic. I do wonder about comparing median scores - there is significant discussion around how to handle Likert type responses. A recent workforce study (PH WINS) in the US collapsed the categories into positive/negative instead. I wonder if this would be a more appropriate statistical approach. The background and discussion are very sharp.

	Minor typo early on ("wasclear" missing a space)

VERSION 2 – AUTHOR RESPONSE

Thank you for the opportunity to respond to the second set of review comments on the above paper. We describe below the changes that we have made. Line numbers relate to the revised version of the paper with changes tracked.

Reviewer 2

1. This is an interesting article on an important topic. I do wonder about comparing median scores - there is significant discussion around how to handle Likert type responses. A recent workforce study (PH WINS) in the US collapsed the categories into positive/negative instead. I wonder if this would be a more appropriate statistical approach. The background and discussion are very sharp.

Response: We note the findings of the PH WINS study with interest. As we outlined in our previous response to reviewer comments, drawing on advice from our statistician, we did not collapse the responses from our survey into two categories for two reasons: 1) the Hennesy-Hicks questionnaire is a validated tool, designed so that training needs are assessed by comparing the mean importance and performance scores given to each task. In dichotomising the data, we would have been making a significant change to the overall design, and would also have had to make a post-hoc decision about where to set the cut point. 2) By dichotomising variables, information is inevitably lost. Our goal was to assess research training needs and priorities, to inform a programme of training. This included assessing the relative importance of different needs. If we had dichotomised the data, we would only have been able to say, for example, that tasks were considered either highly or less important. This would not have helped us identify those tasks considered particularly important, to prioritize for training.

2. Minor typo early on ("wasclear" missing a space)

Response: We have corrected this error (Page 3, Line 20)

VERSION 3 – REVIEW

REVIEWER	Jonathon Leider JHU USA
REVIEW RETURNED	10-Oct-2016

GENERAL COMMENTS	I appreciate the authors discussion of my previous objection to the analytic approach they used. While I still do not agree that the current approach allows for the optimal comparisons, I think we're at a place where reasonable people can reasonably disagree, and would recommend accepting the manuscript at this point.