

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)	Do conceptualisations of health differ across social strata? A concept mapping study among lay people
AUTHORS	Stronks, Karien; Hoeymans, Nancy; Haverkamp, Beatrijs; den Hertog, Frank; van Bon-Martens, Marja; Galenkamp, Henrike; Verweij, Marcel; Van Oers, Hans

VERSION 1 – REVIEW

REVIEWER	Daniel Holman Department of Sociological Studies, University of Sheffield, UK
REVIEW RETURNED	14-Nov-2017

GENERAL COMMENTS	<p>This paper examines how those with three different levels of education conceptualise health. It is based on a concept mapping exercise which involved face-to-face sessions with each of the three groups. A number of similarities and differences in conceptualisations of health are identified across the groups, which it is argued are relevant to policies aimed at tackling health inequalities.</p> <p>Overall I think this is a well-written and useful contribution to the literature on health inequalities. I think the results on how different classes conceptualise health will be of interest to others working in this field. The methodology is appropriate to the research question. I think there are some weaknesses in the sampling which need to be better acknowledged. I would also like to have seen how this study fits more in terms of existing studies on this topic, perhaps by the same authors or others, and ideally using this concept mapping technique. At the moment, the introduction is very short (though I appreciate space is limited), so it is difficult to get a sense of how the paper is situated in terms of what is already known. I also think the concept map figures could be made clearer. Specific points are given below:</p> <p>In the abstract (and elsewhere if necessary), instead of referring to 'lower groups' and 'higher groups', which might be perceived by some as carrying value judgements, I think it is better to refer to 'lower education' and 'higher education' groups.</p> <p>The introduction states that 'the legitimacy of policies that aim at improving the health of people in lower socio-economic groups can be challenged if these policies do not reflect the conceptualisations</p>
-------------------------	---

of health that are valued in all strata.' I think this needs explaining a little more. Does it mean legitimate to the people who are the targets of health improvements, or that the policies themselves might not be legitimate? In either case, returning to this point in the discussion is warranted. In light of what you have found, how can the legitimacy of such policies be increased?

Perhaps to engage with theory is beyond the scope of a medical journal such as this but it struck me how useful Bourdieu's idea of dispositions would fit here. For example, the disposition to be focussed on the mind and inner health and reflection versus material circumstances, being functional, and 'getting by'. Relatedly, 'values they adhere to' in the second paragraph under the introduction suggests a conscious choice but reframed as a disposition this statement would allow for things like 'seeking immediate satisfaction' to be an orientation rooted in social class circumstances – which may or may not be conscious.

Why was education level chosen as the SES measure, and what implications does this have for the research? For example, differences in education are generally set earlier in the lifecourse, whilst occupation may be thought of as a more proximate measure. The former might measure e.g. literacy, understanding, curiosity, etc., whilst the latter e.g. life stressors, working conditions, resources, etc. This in turn probably has implications for how people conceptualise health which should be acknowledged in the paper.

Under step 2 it states: 'Double or overlapping statements were deleted in consultation with the group.' – does this mean the concept mapping group? How did this happen – was there a separate meeting to do this work, or was it electronic (or some other way?).

Under step 4 'enables to manually inspect' should be 'enables for manually inspecting'.

Under the results '16 in the higher and intermediate educational group' should be '16 in both the higher and intermediate educational groups' otherwise it suggests that there are only two and not three groups in total.

It is a limitation that the majority of participants were female – this should be acknowledged. Since we know that men and women experience health differently, it is likely that the results would have been different had the majority of participants been male.

I would remove 'apparently' from the paragraph under the heading interpretation of differences. This makes it sound like a surprise finding when it actually seems quite intuitive.

In the following paragraph, remove 'quality of a person' – sounds judgemental.

I think the concept maps need to be made easier to understand. They would benefit from a little more explanation, or there needs to

	<p>be more labelling of the maps. Why is the number of layers needed to represent the ranked importance when the numbering already seems to do that? What do the shapes/their sizes represent?</p> <p>It would be good to know more about the following process: ‘Within this group, it was discussed and decided which cluster solution made most sense, given the content of the statements within the clusters.’ Presumably, if each participant did the clustering process, then for e.g. the lower education group there were 14 different clustering solutions. So for these 14, did you select discrete clusters across all participants, or was there some merging involved? I ask because for example, under the low SES group, if ‘feeling safe in the home’ and ‘enjoying work’ were removed (which seem like distinct things), then this cluster would be very well described as ‘being sociable’ or something similar. So, was this option the best available, does it come from one participant, and how did you handle the same statement belonging to different clusters among different participants? Some more clarity around this process would be very helpful for the reader.</p> <p>Lastly, I note there is no discussion of ethical issues/review. As per the SPQR guidelines this should be in place for primary qualitative research. Given this omission, as well as more attention needed to study limitations, I have currently marked these items as ‘No’ on the Review Checklist.</p> <p>In sum, I think with a little more work based on the above comments would bring the paper to a publishable standard.</p>
--	---

REVIEWER	Gerald Elsworth Deakin University, Australia
REVIEW RETURNED	28-Dec-2017

GENERAL COMMENTS	<p>Review of paper titled: “Conceptualisations of health across social strata: a concept mapping study.”</p> <p>This is a succinctly written report of a concept mapping study designed to explore possible differing conceptions of health and being healthy across social strata. The study utilises a typical ‘structured concept mapping’ approach as originally described by WMK Trochim and colleagues and currently implemented in an on-line computer package, the Concept System Global Max. In suggesting that the paper is succinct, it is, for this reviewer, perhaps too succinct, particularly in the detail provided about the method used and the analysis and description of results. I elaborate below some of the more important instances, largely in the Methods and Results sections of the paper, where I believe more information and discussion could be provided to facilitate a better understanding and evaluation of the work.</p> <p>First, a brief comment on the theoretical stance of the paper as developed in the Introduction. Here, ‘health’ is designated as a ‘thick’ concept i.e. a concept with a clearly evaluative connotation but one that also contains useful descriptive content. From this designation, given the evaluative component, the authors anticipate that conceptions of health will vary across socio-economic strata. The reference given for designating health as a thick concept is a</p>
-------------------------	--

'forthcoming' paper in the Journal of Medicine and Philosophy by Haverkamp et al. As this paper does not appear to be currently available in any form, I believe it would be helpful for the argument in the present paper if the authors were to elaborate more fully on both the designation of 'health' as a thick concept – in particular, on the evaluative component of 'health' as a thick concept – and on the derivation of the expectation that, as such, conceptions of health might be expected to vary across social strata. Certainly, some research is cited to support this latter claim and numerous important epidemiological indicators of health status are well known to vary along the social gradient, but what are the grounds for arguing that, given the evaluative component of the concept, individuals' conceptions of health might be expected to similarly vary along the social gradient? Also, possible issues in using educational level as a proxy for differentiation of participants into social strata in this study should be discussed and references cited – it could reasonably be argued that the differences in conceptions of health observed in the study are the direct result of differences in the amount of education received by the participants (and, relatedly, their age, as acknowledged by the authors) largely independent of other aspects of socio-economic status.

Secondly, there are aspects in the Methods section of the paper that I believe require elaboration and/or justification. I mention a number in the points below; some may be quite simply addressed, others may require more detailed discussion/justification, all revolve around the general method of structured concept mapping as originally devised by Trochim and colleagues, and the specific features of the Concept System Global Max program that the present authors used in their study.

1. In Step 2 of the procedure the moderator of the concept mapping groups "also mentioned aspects of health emerging from earlier studies in Netherlands on health, in case they had not been mentioned by the participants." (p. 5). This reviewer has always understood structured concept mapping as a procedure that is firmly grounded in the conceptions of the participants. As such, one of the strengths of the approach (similar to the nominal group technique) is both this grounding and that participants work alone at critical stages of the procedure, thus potentially achieving a more balanced representation of their ideas. To my mind introducing research-based ideas from outside those of the participants disturbs this balance and thus moves the method critically away from a secure focus on the perceptions and values of the participants.

2. In Step 3 it is indicated that the participants rated the statements along "... a 5-point Likert-type response scale ... of increasing importance for understanding what health is." What were the response options for this scale of importance?

3. Steps 4 and 5 – Representation and Interpretation. As I read the description of the method used, the input from the participants ceased following the card sorting and rating phases. Representation of the concept maps was achieved by the researchers' use of the program and its output over two meetings. Thus, important decisions including the number of clusters that best represented the structure of the groups' conceptions of health (decided by the researchers) and the names of these clusters (recommended, as I read the account, by the software on the basis of an item being closest to the centroid of the cluster) received no input or discussion from the participating groups. This, to my mind, clearly violates the spirit of the 'grounded' nature of structured concept mapping. Certainly, in many early implementations of the process, these steps were undertaken as part of the group process – even often extending to

	<p>the group deciding to move a small number of statements from their 'program designated' cluster to clarify the meaning of a cluster, and discussing the nature (e.g. causal) of possible associations between clusters. I believe the issues raised in 1 and 3 above are a serious threat to the validity (however conceptualized) of the study and should be addressed by the authors by argument and/or citation.</p> <p>4. Some valuable aspects of the process were not used, including interpreting the so-called Point Map. Here, prior to examination of the Cluster Map the group can be encouraged to interpret the relationships between individual statements and the arrangement of points radiating from the center to the periphery of the map as an important step in the process. Indeed, Trochim emphasized that, given what he viewed as the stronger mathematical basis of multi-dimensional scaling over cluster analysis, the arrangement of the statements in the Point Map may well be more reliable than their subsequent arrangement into clusters. Additionally, the axes derived from the multi-dimensional scaling analysis or axes placed by agreement among the participants can be used to add additional organization to the Point Map. Notable benefits from working with the Point Map include uncovering inconsistencies in the Cluster Map and resolution of these to achieve more clearly delineated clusters together with an enriched understanding of the group's overall conception of the target concept. From this perspective, I believe examination of the Point Maps in addition to the Cluster Maps would have strengthened the conclusions from this study.</p> <p>5. Similarly, the authors make no use of bridging values, either for individual statements (as portrayed in the Point Bridging Map of the Concept System Global Max program) or for the clusters (similarly portrayed in the Cluster Bridging Map). Interpretation of bridging values can be a very useful adjunct/foil to the algorithmic clustering of items and the naming of clusters provided by the Concept System program as used by the authors in this paper. Finally, aside from the call in the Methodological Limitations section of the paper for replication of the study in different regions and with different groups very limited attention is given in the paper to the reliability and validity of the results of the study. I believe this is of particular concern given those aspects of the method used that I highlighted in points 1 and 3 above. While I'm hesitant to use the idea of bias in a mixed-method study such as this, could the authors indicate the steps (standard in qualitative research practice) that were taken in the Representation and Interpretation phases of the study to avoid the potential influence of their prior reading, discussion and theorizing on, for example, the decisions made about the number of clusters and the interpretation of their meaning? The transfers of decision-making responsibility from the participants to the researchers at critical phases of this study raise serious issues for me around the objectivity of the findings and interpretations offered. For me, the authors' use of the automatic features of the Concept System program do not alleviate this concern.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Editorial Requests:

- Please revise your title to indicate the research question, study design, and setting. This is the preferred format of the journal.
- >> We have changed the title accordingly.

- Please ensure that the bullet points in the 'Strengths and Limitations' section relate to the design and methods of your study. The first point is a summary of the study rather than a specific strength or limitation.

>> We have now changed the first bullet point so that it relates to the strength of the methodology employed in our study.

- Please briefly explain in the methods section why this study did not require approval from a local ethics committee.

>> We have now explained this in the method section (p. 7).

- Please add a statement to methods section confirming that you obtained written informed consent from participants.

>> We have now added this to the method section (p. 7).

Reviewers' Comments to Author:

Reviewer: 1

Reviewer Name: Daniel Holman

Institution and Country: Department of Sociological Studies, University of Sheffield, UK

Competing Interests: None declared

This paper examines how those with three different levels of education conceptualise health. It is based on a concept mapping exercise which involved face-to-face sessions with each of the three groups. A number of similarities and differences in conceptualisations of health are identified across the groups, which it is argued are relevant to policies aimed at tackling health inequalities. Overall I think this is a well-written and useful contribution to the literature on health inequalities. I think the results on how different classes conceptualise health will be of interest to others working in this field. The methodology is appropriate to the research question.

>> Thank you for acknowledging the contribution of our paper to the literature.

I think there are some weaknesses in the sampling which need to be better acknowledged.

>> We agree that the sampling suffers from weaknesses which might affect the generalisability of the results. We have acknowledged this in the discussion section (p. 12 and 13). From the reviewer's comment, we do not understand how to elaborate on this. If the reviewer could further explain his comment, we are of course to further elaborate on this.

I would also like to have seen how this study fits more in terms of existing studies on this topic, perhaps by the same authors or others, and ideally using this concept mapping technique. At the moment, the introduction is very short (though I appreciate space is limited), so it is difficult to get a sense of how the paper is situated in terms of what is already known.

>> To the best of our knowledge, we are the first to apply the concept mapping technique to explore conceptualisations of health in different social strata. Given the explorative character of our analyses, we prefer to situate the results of our study in the context of previous papers in the discussion section rather than the introduction. This gives us the opportunity to discuss the results of previous studies in more detail than we could do in the introduction section.

I also think the concept map figures could be made clearer.

>> Thank you for this suggestion. We have now improved the lay-out of the figures, and we have added the points indicating statements.

Specific points are given below:

In the abstract (and elsewhere if necessary), instead of referring to 'lower groups' and 'higher groups', which might be perceived by some as carrying value judgements, I think it is better to refer to 'lower education' and 'higher education' groups.

>> Thank you for bringing this to our attention. We have changed this accordingly throughout the manuscript.

The introduction states that 'the legitimacy of policies that aim at improving the health of people in lower socio-economic groups can be challenged if these policies do not reflect the conceptualisations of health that are valued in all strata.' I think this needs explaining a little more. Does it mean legitimate to the people who are the targets of health improvements, or that the policies themselves might not be legitimate? In either case, returning to this point in the discussion is warranted. In light of what you have found, how can the legitimacy of such policies be increased?

>> We have now added an example of what we meant with the statement on legitimacy of policies (p. 4, first paragraph). In addition, we elaborated on the potential implications of what we found for the legitimacy of health policies that aim to tackle health inequalities (p.15)

Perhaps to engage with theory is beyond the scope of a medical journal such as this but it struck me how useful Bourdieu's idea of dispositions would fit here. For example, the disposition to be focussed on the mind and inner health and reflection versus material circumstances, being functional, and 'getting by'. Relatedly, 'values they adhere to' in the second paragraph under the introduction suggests a conscious choice but reframed as a disposition this statement would allow for things like 'seeking immediate satisfaction' to be an orientation rooted in social class circumstances – which may or may not be conscious.

>> Thank you for your suggestion. We have now reframed the second paragraph in terms of Bourdieu's idea of being disposed to certain practices (p. 4).

Why was education level chosen as the SES measure, and what implications does this have for the research? For example, differences in education are generally set earlier in the lifecourse, whilst occupation may be thought of as a more proximate measure. The former might measure e.g. literacy, understanding, curiosity, etc., whilst the latter e.g. life stressors, working conditions, resources, etc. This in turn probably has implications for how people conceptualise health which should be acknowledged in the paper.

>> In the Netherlands, as well as in many other countries, education is considered a good indicator of adult socio-economic status. It is strongly associated with occupational level and income. In comparison with occupational and income level, education has the advantage of being available for each individual (e.g. also for those without a paid job). We agree that different indicators for socio-economic status might indicate different aspects of someone's position (e.g. educational level-literacy versus occupational level-working conditions). We do not expect our results to be biased by the choice for this indicator, however, as the conceptualisation of the concept mapping groups not only reflect aspects closely related to educational level (such as autonomy in the higher education group), but also aspects more related to income and occupational level (e.g. having access to health care or living in a healthy environment in the lower education group). We have now added this reflection in the limitation section (p. 14).

Under step 2 it states: 'Double or overlapping statements were deleted in consultation with the group.' – does this mean the concept mapping group? How did this happen – was there a separate meeting to do this work, or was it electronic (or some other way?).

>> Indeed, we meant to say: the participants in the concept mapping group. This has been done during the sessions. We have now clarified this in the text (p. 5).

Under step 4 'enables to manually inspect' should be 'enables for manually inspecting'.

>> Thank you. We have changed this accordingly.

Under the results '16 in the higher and intermediate educational group' should be '16 in both the higher and intermediate educational groups' otherwise it suggests that there are only two and not three groups in total.

>> We have changed this accordingly.

It is a limitation that the majority of participants were female – this should be acknowledged. Since we know that men and women experience health differently, it is likely that the results would have been different had the majority of participants been male.

>> We agree that the concept maps could have been different had the majority of the participants been male. We now acknowledge this in the discussion section when we discuss the external validity of the maps (p. 14). However, as our research questions focuses on differences between educational groups (with comparable sex distributions), we do not expect this sex distribution to bias our results.

I would remove 'apparently' from the paragraph under the heading interpretation of differences. This makes it sound like a surprise finding when it actually seems quite intuitive. In the following paragraph, remove 'quality of a person' – sounds judgemental.

>> We have changed this accordingly.

I think the concept maps need to be made easier to understand. They would benefit from a little more explanation, or there needs to be more labelling of the maps. Why is the number of layers needed to represent the ranked importance when the numbering already seems to do that? What do the shapes/their sizes represent?

>> As said, we have now improved the lay-out of the figures, and we have added the points indicating statements, so that it becomes clear what the shapes represent.

It would be good to know more about the following process: 'Within this group, it was discussed and decided which cluster solution made most sense, given the content of the statements within the clusters.' Presumably, if each participant did the clustering process, then for e.g. the lower education group there were 14 different clustering solutions. So for these 14, did you select discrete clusters across all participants, or was there some merging involved? I ask because for example, under the low SES group, if 'feeling safe in the home' and 'enjoying work' were removed (which seem like distinct things), then this cluster would be very well described as 'being sociable' or something similar. So, was this option the best available, does it come from one participant, and how did you handle the same statement belonging to different clusters among different participants? Some more clarity around this process would be very helpful for the reader.

>> We have now clarified this step in the methods section (p. 6/7). We would also like to refer to our reply to the second reviewer (point 4 and 5).

Lastly, I note there is no discussion of ethical issues/review. As per the SPQR guidelines this should be in place for primary qualitative research. Given this omission, as well as more attention needed to study limitations, I have currently marked these items as 'No' on the Review Checklist.

>> We have now added a paragraph an ethical approval in the methods section (p. 7).

In sum, I think with a little more work based on the above comments would bring the paper to a publishable standard.

>> Thank you. We hope that the revised version meets your expectations.

Reviewer: 2

Reviewer Name: Gerald Elsworth

Institution and Country: Deakin University, Australia

Competing Interests: None declared

Review of paper titled: “Conceptualisations of health across social strata: a concept mapping study.” This is a succinctly written report of a concept mapping study designed to explore possible differing conceptions of health and being healthy across social strata. The study utilises a typical ‘structured concept mapping’ approach as originally described by WMK Trochim and colleagues and currently implemented in an on-line computer package, the Concept System Global Max. In suggesting that the paper is succinct, it is, for this reviewer, perhaps too succinct, particularly in the detail provided about the method used and the analysis and description of results. I elaborate below some of the more important instances, largely in the Methods and Results sections of the paper, where I believe more information and discussion could be provided to facilitate a better understanding and evaluation of the work.

First, a brief comment on the theoretical stance of the paper as developed in the Introduction. Here, ‘health’ is designated as a ‘thick’ concept i.e. a concept with a clearly evaluative connotation but one that also contains useful descriptive content. From this designation, given the evaluative component, the authors anticipate that conceptions of health will vary across socio-economic strata. The reference given for designating health as a thick concept is a ‘forthcoming’ paper in the Journal of Medicine and Philosophy by Haverkamp et al. As this paper does not appear to be currently available in any form, I believe it would be helpful for the argument in the present paper if the authors were to elaborate more fully on both the designation of ‘health’ as a thick concept – in particular, on the evaluative component of ‘health’ as a thick concept – and on the derivation of the expectation that, as such, conceptions of health might be expected to vary across social strata. Certainly, some research is cited to support this latter claim and numerous important epidemiological indicators of health status are well known to vary along the social gradient, but what are the grounds for arguing that, given the evaluative component of the concept, individuals’ conceptions of health might be expected to similarly vary along the social gradient?

>> We have now further explained what we mean by health being a ‘thick’ concept, and also elaborated our argument that the evaluative element of the health concept might vary along the social gradient (p. 4, second paragraph).

Also, possible issues in using educational level as a proxy for differentiation of participants into social strata in this study should be discussed and references cited – it could reasonably be argued that the differences in conceptions of health observed in the study are the direct result of differences in the amount of education received by the participants (and, relatedly, their age, as acknowledged by the authors) largely independent of other aspects of socio-economic status.

>> On this point, we would like to refer to our answer to the first reviewer.

Secondly, there are aspects in the Methods section of the paper that I believe require elaboration and/or justification. I mention a number in the points below; some may be quite simply addressed, others may require more detailed discussion/justification, all revolve around the general method of structured concept mapping as originally devised by Trochim and colleagues, and the specific features of the Concept System Global Max program that the present authors used in their study.

>> Thank you for your comments and suggestions, which were very helpful in improving the methods section of our paper.

1. In Step 2 of the procedure the moderator of the concept mapping groups “also mentioned aspects of health emerging from earlier studies in Netherlands on health, in case they had not been mentioned by the participants.” (p. 5). This reviewer has always understood structured concept mapping as a procedure that is firmly grounded in the conceptions of the participants. As such, one of the strengths of the approach (similar to the nominal group technique) is both this grounding and that participants work alone at critical stages of the procedure, thus potentially achieving a more balanced representation of their ideas. To my mind introducing research-based ideas from outside those of the

participants disturbs this balance and thus moves the method critically away from a secure focus on the perceptions and values of the participants.

>> We completely agree with the reviewer that the perceptions and values of the participants were the primary and only focus of the concept mapping groups. Therefore, the moderator only mentioned additional aspects after no additional items were mentioned by the group. These included a limited number only (maximum of 3 – this information was added to the text, at p. 5). Only if the item proposed by the moderator was recognized by at least one member of the group, it had been added to the list. This is in accordance with the recommendations by Trochim, as he mentioned also the possibility of including statement that were derived from the literature, under the condition that the group participants had the opportunity to discuss those statements, and make a decision on whether to include it in the list (Trochim 1989). In addition, as we asked the participants to rank the items by importance to them, they had the opportunity to say that the items added in second instance to the list were less important for them. For that reason, we believe that the introduction of research-based items from the side of the moderator did not disturb a balanced representation of the ideas of the group.

2. In Step 3 it is indicated that the participants rated the statements along “... a 5-point Likert-type response scale ... of increasing importance for understanding what health is.” What were the response options for this scale of importance?

>> We did not give response options, but instead asked the respondents to divide the index cards into five equal piles of increasing importance for understanding what health is, ranging from 1 (least important) to 5 (most important). We have now further clarified this (p. 6).

3. Steps 4 and 5 – Representation and Interpretation. As I read the description of the method used, the input from the participants ceased following the card sorting and rating phases. Representation of the concept maps was achieved by the researchers' use of the program and its output over two meetings. Thus, important decisions including the number of clusters that best represented the structure of the groups' conceptions of health (decided by the researchers) and the names of these clusters (recommended, as I read the account, by the software on the basis of an item being closest to the centroid of the cluster) received no input or discussion from the participating groups. This, to my mind, clearly violates the spirit of the 'grounded' nature of structured concept mapping. Certainly, in many early implementations of the process. these steps were undertaken as part of the group process – even often extending to the group deciding to move a small number of statements from their 'program designated' cluster to clarify the meaning of a cluster, and discussing the nature (e.g. causal) of possible associations between clusters. I believe the issues raised in 1 and 3 above are a serious threat to the validity (however conceptualized) of the study and should be addressed by the authors by argument and/or citation.

>> Thank you for your comments. Indeed, decisions regarding the clustering (including names) received no discussion from the participants in the concept mapping groups, although they were based on the names of the clusters that were used by the participants themselves (which means that they are not based on the items, as suggested by the reviewer). We believe further discussion with the participants could have improved our understanding of each of the clusters, as we already mentioned in the limitation section (p. 14, first paragraph). On the other hand, we were not unique in our choice for this strategy, as other concept mapping studies, such as those in the Netherlands, also do not always involve participants in this stage (van Bon-Martens et al. 2017 –ref. 17). This is in line with Trochim's recommendations, that state that different strategies can be applied, depending on the aim of the concept map. The advantage of the fact that we as researchers were responsible for the interpretation of the maps, is the comparability of the maps across education groups. We have now explained this in the limitation section in the discussion (p. 14).

4. Some valuable aspects of the process were not used, including interpreting the so-called Point Map. Here, prior to examination of the Cluster Map the group can be encouraged to interpret the

relationships between individual statements and the arrangement of points radiating from the center to the periphery of the map as an important step in the process. Indeed, Trochim emphasized that, given what he viewed as the stronger mathematical basis of multi-dimensional scaling over cluster analysis, the arrangement of the statements in the Point Map may well be more reliable than their subsequent arrangement into clusters. Additionally, the axes derived from the multi-dimensional scaling analysis or axes placed by agreement among the participants can be used to add additional organization to the Point Map. Notable benefits from working with the Point Map include uncovering inconsistencies in the Cluster Map and resolution of these to achieve more clearly delineated clusters together with an enriched understanding of the group's overall conception of the target concept. From this perspective, I believe examination of the Point Maps in addition to the Cluster Maps would have strengthened the conclusions from this study.

5. Similarly, the authors make no use of bridging values, either for individual statements (as portrayed in the Point Bridging Map of the Concept System Global Max program) or for the clusters (similarly portrayed in the Cluster Bridging Map). Interpretation of bridging values can be a very useful adjunct/foil to the algorithmic clustering of items and the naming of clusters provided by the Concept System program as used by the authors in this paper.

>> Thank you for pointing this out. As recommended by Trochim, in an iterative process, we carefully looked at the position of the statements (points) on the map in relation to that specific cluster, and the relationship with other clusters. We actually also used bridging values to support the final cluster solution. In an attempt to write as succinctly as possible, we did not mention this in the method section in the previous version of the manuscript. We have now extended the interpretation section with this information (p. 6). In general, we would like to mention that we based our decisions more on the content of the statements and clusters rather than the automatic features of the software.

Finally, aside from the call in the Methodological Limitations section of the paper for replication of the study in different regions and with different groups very limited attention is given in the paper to the reliability and validity of the results of the study. I believe this is of particular concern given those aspects of the method used that I highlighted in points 1 and 3 above. While I'm hesitant to use the idea of bias in a mixed-method study such as this, could the authors indicate the steps (standard in qualitative research practice) that were taken in the Representation and Interpretation phases of the study to avoid the potential influence of their prior reading, discussion and theorizing on, for example, the decisions made about the number of clusters and the interpretation of their meaning? The transfers of decision-making responsibility from the participants to the researchers at critical phases of this study raise serious issues for me around the objectivity of the findings and interpretations offered. For me, the authors' use of the automatic features of the Concept System program do not alleviate this concern.

>> When rereading the methods section, we indeed realised that we gave little information as to how we have come to an interpretation of the clusters in the concept maps for each of the educational group, and how we have promoted an open discussion and preclude bias. Thank you for pointing this out. We have now extended the description of this process at p. 6, and discussed this in the limitation section (p. 13).

VERSION 2 – REVIEW

REVIEWER	Daniel Holman Department of Sociological Studies, University of Sheffield, UK.
REVIEW RETURNED	13-Feb-2018
GENERAL COMMENTS	Thank you for the opportunity to review a revised version of the manuscript. I have read the authors' responses and revised manuscript carefully, and I am satisfied that they have responded to the issues raised appropriately. One minor point: the authors

	questioned exactly what I was referring to when I mentioned weaknesses in the sampling. My concern regarded the gender balance, but this has now been acknowledged as a limitation.
REVIEWER	Gerald Elsworth Deakin University, Australia
REVIEW RETURNED	28-Feb-2018
GENERAL COMMENTS	I thank the authors of this paper for their careful consideration of my comments on the original MS. I have carefully reviewed the revised paper and the authors' responses to the critical points I raised. I am very happy to recommend the revised MS for publication.