Author’s response to reviews

Title: Validity and Reliability of Evidence Utilization in Policy-making Measurement Tool (EUPMT)

Authors:

Mohammad Hasan Imani-Nasab (imani_mh@yahoo.com)

Bahareh Yazdizadeh (baharehyazdizadeh@yahoo.com)

Masoud Salehi (salehi74@yahoo.com)

Hesam Seyedin (h.seyedin@iums.ac.ir)

Reza Majdzadeh (rezamajd@tums.ac.ir)

Version: 1 Date: 19 Jun 2017

Author’s response to reviews:

Reviewer #1

Many thanks for your overall evaluation about our manuscript, especially for your technical comments and kindly suggestions for its clarity. On your valuable advice it was reviewed by our statistical adviser.

BACKGROUND

Page 4, Line 3: Suggest editing to a slightly more conservative statement 'may lead to' (rather than leads to) The valuable comment is considered. Please check the blue colored modification on page 2, Line 3.

Page 4, line 4: Suggest change from 'encouraging … toward' to 'increasing ... health systems policy makers and managers awareness of ...' the valuable comment is considered. Please check the blue colored modification on page 2, Lines 4-5.

Page 4, line 8: Suggest change from "difficulty in changing policy-makers’ behaviour before and after the interventions" to "difficulty in measuring policy-makers' research-use actions before and after interventions ..." This will tie the argument more to the purpose of the paper. The valuable comment is considered. Please check the blue colored modification on page 2, Line 6.
Page 4, line 13: Suggest changing from: "has been the theoretical method predominantly employed" to "has been widely used to study health related ..."

Page 4, line 23: 'explain' rather than 'determine'

Page 4, line 40: I don't understand what is meant by 'and then their outcomes'.

Page 4, line 53-4: I don't think the statement about Boyko et al's test of their TBP measure is completely accurate. Boyko et al reported test-retest reliability and internal consistency, but stated that validity could not be assessed. Please check and correct to ensure consistent with what Boyko et al reported.

Page 4, last paragraph: This paragraph needs some editing to ensure it is clear whether the authors are referring to, i.e. is it Boyko et al's study or their own study? Suggest replacing with 'Boyko et al' for the former and 'our' for the latter.

Page 4, line 60: "The designed tool may be employed". Suggest editing this to state as an aim, rather than a conclusion (which pre-empts whether the tool is suitable for the measurement purposes listed here). ie. "Our aim was to design a tool that may be employed". Also, it would be helpful to state what is meant by 'determine the status quo'. Determine current beliefs and intentions about using research to inform policy?

Page 5, line 1: As above, edit "The results of
this study can effectively help" to reflect that this is your aim.

check the blue colored modification on page 3, Line 34-35.

METHODS

Page 5, lines 13-14 (and abstract): specify country.

Page 5, lines 33 and 36 (and abstract): Edit to clarify whether "number of eligible individuals" and "all of them were included in the study" refers only to the sampling frame/number of individuals invited to complete the survey ('included in the study' suggests you have data for all 373 and that the sample may have been larger than this).

Page 6, ethical considerations. Suggest including more information about the ethics/consent processes, specifically, consider reporting:

- an ethics approval number for the project if you have one
- whether participants were required to consent, if so describe the consent process
- whether respondents received information about the purpose of the survey and how data would be used, confidentially provisions etc.

Page 6, line 21-22: The process described is unclear to me. Should this read "57 questions were generated from qualitative data to develop a tool which was subsequently assessed for face and content ...'. In addition, if this data came from interviews/focus groups, perhaps state the source of the data (generated

Thanks for your attentions. Please check the blue colored added explanations on page 5, Lines 3 and 6-9.

Page 6, line 21-22: The process described is unclear to me. Should this read "57 questions were generated from qualitative data to develop a tool which was subsequently assessed for face and content ...'. In addition, if this data came from interviews/focus groups, perhaps state the source of the data (generated

Thanks for your attentions. Please check the blue colored modification on page 4, Lines 38-39 and page 5, lines 21-22.
from analysis of interviews …).

Page 6, line 29: perhaps change to "validities of the questions and rated them as 'completely undesirably' …." Then new sentence "Respondents were asked for their opinions on how the face validity of questions could be improved".

Page 6, line 32-3: please provide a reference for the methods involving calculation of "Content Validity Ratio (CVR) and Content Validity Index (CVI) indicators".

Page 6, line 39: I'm not sure what is meant by "The quantitative results were finalized" - "The questionnaire was revised by three panels of experts based on …"

Page 6 line 42: please be more specific about the skills held by those described as "renowned researchers"; i.e. "researchers with expertise in …"

Page 6, line 46-7: the description of construct validity is unclear. Maybe revise to: "Construct validity shows the extent to which observed scores on the tool are as predicted by the theories upon which the tool was based." Then "Accumulation of evidence about the face, divergent, convergent, and discriminant validities of a tool is used to provide evidence about construct validity".

Page 6, line 50-1. Further, the definition of convergent and discriminant validity is not entirely accurate. It is true that factor analysis can be used to provide evidence of convergent and discriminant validity, but in itself this is rarely considered sufficient. Perhaps rephrase to "Evidence of convergent validity can be demonstrated by showing positive and significant correlation between scales scores …"
for constructs that are expected to be related.

Page 7, line 11: the authors mention comparison of observed values with "acceptable values". The latter should be specified and referenced (e.g. correlation coefficients of ..., ICC of ...).

Page 7, line 19: It was not clear to me which tests the following sentence refers to "The response rates in the two aforementioned phases". Please specify the response rate (RR), total number of respondents/number invited for both the factor analysis, and test-retest analysis.

Page 7, line 19 and line 40: in addition to reporting % RR, please specify the number of respondents/total number invited. Similarly, not clear what test "The response rate was 92.76%." refers to (RESULTS line 1).

RESULTS

The results of reliability testing (test-retest and internal consistency) appear not to be tabulated, or reported in the results section, yet they appear in the abstract, methods for assessment are described, and the results are interpreted in the discussion. I may have missed this information, otherwise please report these results in the results section.

Page 7, line 42-3: suggest changing to "27 individuals did not return their questionnaire".

Page 7, line 60: two decimal places is sufficient

Page 8, line 3-4: Suggested deleting

The results of reliability tests are presented in the results section with specific subheading (Reliability). Please check it on page 6, Line 26.

Page 7, line 11: This sentence removed on the advice of second reviewer.

Page 7, line 19: Thanks for your attentions. Please check the blue colored modifications on page 6, Line 11-13.

Page 7, line 19: Thanks for your attentions. Please check the blue colored added explanations on page 6, Line 35.

Page 7, line 42-3: The valuable comment is considered. Please check the blue colored modification on page 6, Lines 36-37.

Page 7, line 60: two decimal places is sufficient

Page 8, line 3-4: The valuable comment is considered. The
"Eventually, the face and content validity and reliability of 54 questions were confirmed."

Page 8, line 18-9: Not sure what this means "However, the correlation of 7 of the variables had no theoretical justification with the theory's constructs." Seven variables showed correlations not predicted by the theoretical model? Or correlations predicted by the theoretical model for seven variables were not observed?

Page 8, line 24-5: The authors have made some assumptions throughout the results about the interpretation of correlation coefficients, such as is described here "was greater than 0.5, hence, confirming their convergent validity". It would help to include a section in the methods which specifies the thresholds considered as demonstrating convergent and divergent validity, especially given the complexity of this interpretation (i.e. it is desirable that constructs are sufficiently correlated that they are shown to be related, but not so much that they cannot be discriminated from each other).

Overall comment. Testing a measurement instrument involves accumulating evidence that supports validity/reliability, and no single test "confirms" validity/reliability. For this reason, I suggest editing throughout the results and discussion to reflect this, for example change "discriminate validity was confirmed" to "providing evidence of discriminate validity".

Page 8, line 39-40: suggest changing "Eventually, the validity and reliability of the tool was confirmed with a total of 40 questions." to "After refinement, the final version of the tool had 40 questions."

The valuable comments are considered. Please check the blue colored added reference on page 5, Line 39, on page 5, Lines 35-37, on page 7, Line 10, and on page 8, lines 4 and 6.
DISCUSSION

Overall, some of the information in the discussion seems a better fit for the results section (e.g. the second paragraph on p9 that details results of the CFA). While I appreciate that it is difficult to describe findings from factor analysis in simple language, shifting technical information to the results might improve the discussion. It would help those wanting to use the tool to focus the discussion on the implications of the results (rather than describing results). For example, do the scales identified from factor analysis support the TPB model, does their content make sense (have face validity), are there implications of dropping items or including them in scales for which they were not originally intended?

Page 8, discussion. The first sentence needs revision/simplification. It is probably sufficient to say the validity and reliability of the instrument were examined using factor analysis and test-retest reliability. Here and in the second sentence, I think 'determined' should be changed to 'examined'; similarly in the third sentence 'used for the…' should read 'used to examine the construct validity.'

Page 8, line 58-9: I'm not sure where the 'meeting with experts' was described in the methods: "thus, it was improved upon setting up a meeting with experts." I couldn't see any other mention of a meeting - perhaps this was part of the process to determine face/content validity. If so, using consistent language would help the reader make this link.

Page 10. The authors note "One of the strengths of this study was its high response

The valuable comment is considered. The second and third paragraphs of the discussion section were transferred to the results section (page 7, lines 11-41). Please check the blue colored added explanations in the discussion section.

Page 8, discussion. The first sentence needs revision/simplification. It is probably sufficient to say the validity and reliability of the instrument were examined using factor analysis and test-retest reliability. Here and in the second sentence, I think 'determined' should be changed to 'examined'; similarly in the third sentence 'used for the…' should read 'used to examine the construct validity.'

Page 8, line 58-9: I'm not sure where the 'meeting with experts' was described in the methods: "thus, it was improved upon setting up a meeting with experts." I couldn't see any other mention of a meeting - perhaps this was part of the process to determine face/content validity. If so, using consistent language would help the reader make this link.

Page 10. The authors note "One of the strengths of this study was its high response

The valuable comment is considered. Please check the blue colored modification on page 8, Lines 16-17.

Page 8, line 58-9: I'm not sure where the 'meeting with experts' was described in the methods: "thus, it was improved upon setting up a meeting with experts." I couldn't see any other mention of a meeting - perhaps this was part of the process to determine face/content validity. If so, using consistent language would help the reader make this link.

Page 10. The authors note "One of the strengths of this study was its high response

The valuable comment is considered. Please check the blue colored modification on page 8, Line 18 and on page 5, line 29.

Page 10. The authors note "One of the strengths of this study was its high response

Thanks for your attention. Please check the blue colored added explanations on page 8,
rate." I agree that the response rate is pleasingly high for a study/participants of this type, and wonder if the authors might be able to comment on factors they think may have motivated participants to respond?

Page 11. The manuscript would benefit from some discussion of the limitations of the study, and more description of testing that could be done to generate evidence of the instrument's measurement properties (e.g. examining how scale scores relate to scores on other instruments to which this new measures is expected to relate).

Thanks again

Warm regards

Reviewer #2

Thanks for your overall evaluation about our manuscript, especially for your technical comments. On your valuable advice the discussion of findings improved by links to past measures of research and the whole manuscript was edited by an English expert.

Abstract

Comment

The abstract says nothing about what exactly this tool measures - there are no insights into what is unique about this measure compared to others. It does not acknowledge the fact that there are quite a few measures out there, and some have indeed been validate.

Answer

Thanks for your attention. Please check the green colored added explanations on the abstract.

BACKGROUND

Introduction - line 3 - the quantity of citizen's

Your valuable comment is considered. Please
lives is not the best way to describe this, rather "Reduced quality of life and life expectancy" or something similar.

You need to explain what EXACTLY is a tool that measures research use based on the theory of planned behavior - what kinds of items are in such a scale? More detail about what the TPB is needed - since this is the theory with which you are basing you confirmatory factor analysis on - what are the key constructs in it, and how does your tool map onto these constructs. The difference between a DIRECT and INDIRECT approach to measurement is not clear at all - what exactly does it mean in concrete measurement terms - what is a direct measurement approach, what is an indirect approach? - Please provide examples, the advantages and disadvantages of each, and examples of measures that assess research use in both ways. More detail about these "protocols" (line 43) that measure the factors affecting evidence utilization using the TPB is needed. And why is a TPB-based approach to measurement needed? Why this theory? There are many other frameworks and theories that try to account for and explain why research is or is not used in policy - why have you decided to focus on the TPB, what is its utility?

You have NOT discussed many of the currently available and published measures of research use that are currently available. This article provides a good summary of the available measures: https://academic.oup.com/rev/article/25/3/315/2364577/The-development-of-SAGE-A-tool-to-evaluate-how Many measures of research use have been ignored in the paper - you need to discuss these, their strengths and weaknesses - and how your measure improves upon these, Your valuable comments are considered. Please check the green colored added explanations on page 2, Lines 10-18 and 24-27. Please check the blue colored added explanations on page 4, Lines 20-43. Please check the green colored added explanations on page 3, Lines 2-5.
what makes it different? What does it target that these previous measures do not?

Furthermore, there are tools that have been validated, for example: http://journals.sagepub.com/doi/abs/10.1177/10049731514560413

METHODS

What country are the sample from?

Thanks for your attention. Please check the blue colored added explanation on page 4, Line 3.

There are indeed guidelines about sample sizes for confirmatory factor analysis, based on simulation studies: http://people.musc.edu/~elg26/teaching/psstats1.2006/maccallumetal.pdf. Although your strength is that you have targeted a large sample.

Designing the questions - line 38 onwards

You talk about a qualitative study - you do not provide any details about what this equal study found and how you are using the results from that study to inform your development of this current questionnaire.

Your valuable comments are considered. Please check the green colored added explanations on page 4, Line 38-41. Please check the blue colored added explanations on page 4, Lines 20-37. The initial questionnaire was provided and presented on table 2. The variables are not same questions. As is mentioned in the methods section for each identified variable in the qualitative study, two questions were generated. The constructs of TPB, factors loading of variables and other helpful information is bring in figure 2.

Also, you do not provide the actual questionnaire anywhere - so your discussion about variables, and constructs is not clear - what constructs are you measuring? What do
you mean by variables? Are variables the questionnaire items? What are the items? Where did you get these items from? What is a DIRECT question? What is an INDIRECT question? Explanation of concepts, and providing examples is essential. List all the questions somewhere. These could be displayed in a table of results, where you do the factor analysis, a table with all the items and their loadings on the identified factors from the TPB.

Tool validity - line 16 onwards

Explain content and face validity - these are not common knowledge terms.

Where did you draw your techniques for evaluating content and face validity - i.e., where did you get the idea to use a "content validity ratio".

Also, content validity is not about individual questions - it is about whether the RANGE of questions measures the full breadth of the construct you are interested in measuring. SO your content validity ratio is not actually measuring content validity but rather face validity. The content validity index also is more so measuring face validity not content validity. Content validity needs to measure whether respondents think the questions AS A WHOLE measure the full extent of the construct, and your current indices do not capture this.

As is mentioned in added explanations the content validity ratio can be calculated for both each question and a range of questions.

What exactly do you mean by "individuals who believe the content of the question is good" - Thanks for your attention. Please check the green colored added explanation on page 5, Line 15.

Your valuable comments are considered. Please check the green colored added explanations and references on page 5, Lines 12-15. Please check the presented references on page 5, line 15.
What is "good"?

Tool reliability

Line 7 I have never heard of external consistency. This should be changed to what you are actually measuring which is test-retest reliability. I have also never heard of the term construct reliability. Please remove these terms and say that you are measuring: test retest reliability and internal consistency.

When you refer to goodness of fit tests - what fit are you measuring here? Whether the model in the confirmatory factor analysis has good fit with the hypothesized factor structure? If so, this is NOT about tool reliability, but about validity. Why do you not have a section in your method that talks about the methods you used to test validity?

Why was the reliability of data entry examined? If this is a questionnaire, then data entry should be 100% reliable, because you are simply entering the choices of respondents - it is not an externally rated measure.

It is desirable that data entry must be 100% accurate, but in reality may be many errors; So it is necessary that we check its accuracy. As we mentioned in the method section we randomly extract 10% of questionnaires and entered their data again. After that we checked our accuracy in the data entry.

RESULTS

In the results - line 39

Have separate sections or subheadings - one for reliability and one for validity.

Your discussion of which items loaded on which factor is meaningless until you actually provide a list of all the items.

When you talk about the goodness of fit, please

Your valuable comments are considered. Please check the green colored modifications on page 6.
DISCUSSION

The sentence - "based on our results... impact of knowledge translation intervention in health policymaking" (line 8-10) - you have actually NOT demonstrated this. This requires you to demonstrate that your measure has sensitivity to detect changes as a result of knowledge translation interventions. And the use of the term "impact" is very general - what kind of impact? Impact on research use? Engagement with research?"

I am uncertain as to why you discuss exploratory factor analysis and rotation in the discussion - this should be in the results. Furthermore, why did you do EFA if you already did CFA and already demonstrated a good fit to the hypothesized model? The discussion of loadings is also meaningless unless you provide a table showing all items and their relative loadings on all the individual factors.

The variables were extracted from open interview, so we first conduct EFA and then EFA.

The discussion reads like a results section and does not make any reference to the implications of your findings for improving or strengthening evidence based policy. This needs to be a key focus in the discussion, and how your tool can be used to improve the use of research in policy.

Your reference to specific constructs from the TPB in the discussion, again, is difficult to follow and lacks meaning since you do not

This is a very strong conclusion that is not supported by your evidence. Thanks for your attention. The mentioned sentence is omitted

The valuable comment is considered. The second and third paragraphs of the discussion section were transferred to the results section. Please check on page 7, lines 8-40.

As is mentioned previously, the initial questionnaire was presented on table 2.

Your valuable comments are considered. Please check revised version of discussion on pages 8-9. Please check the green colored added implications on page 9, Line 36-40.

Thanks for remembering again. The necessary explanation of the TPB constructs was added on page 2, Lines 10-18 and the initial
explain the theory adequately in the beginning, as questionnaire was presented on table 2. and how your measurement tool aligns with the TPB, and how it measures each of the key constructs in the TPB.

Thanks again

Warm regards