**Reviewer’s report**

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

**Version:** 0 **Date:** 30 Mar 2017

**Reviewer:** Steve Makkar

**Reviewer's report:**

The article discusses the development and validation of a measure of research use, which is an area of growing interest and importance, particularly with the need to improve the use of research in policy and practice. Although the present paper used strong methods, the narrative, discussion of findings, implications, and link to past measures of research use are extremely weak. There are also problems with the quality of written English throughout, with certain sentences and word choices being unclear. This needs major revision.

Some comments:

- 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 validated.

Introduction - line 3 - the quantity of citizen's 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, 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/10.1177/1049731514560413

METHODS

What country are the sample from?

Also, 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 qual study found and how you are using the results from that study to inform your development of this current questionnaire.

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.

Also what exactly do you mean by "individuals who believe the content of the question is good" - line 36. 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 hypothesised 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?

Furthermore, 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.

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 make explicit reference to table 3.

Discussion - line 48 onwards.
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"? This is a very strong conclusion that is not supported by your evidence.

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 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 explain the theory adequately in the beginning, and how your measurement tool aligns with the TPB, and how it measures each of the key constructs in the TPB.

**Level of interest**
Please indicate how interesting you found the manuscript:

An article of importance in its field

**Quality of written English**
Please indicate the quality of language in the manuscript:

Not suitable for publication unless extensively edited

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?
4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal