Reviewer’s report

Title: SANRA – A Scale for the Quality Assessment of Narrative Review Articles

Version: 0 Date: 30 Dec 2018

Reviewer: Caroline Struthers

Reviewer’s report:

Although I work for the EQUATOR Network I would class myself as a consumer reviewer as I am not a trained researcher or statistician. This manuscript was of interest to me because it is aimed at evaluating a tool designed to improve conduct and reporting of research, in this case narrative reviews. I am an extremely infrequent reviewer, so my review reveals a lack of knowledge of much of the statistical and research jargon used. I think these comments are valid as the target reader and user of the knowledge (journal editors) may also not be familiar with the technical terms.

I have made comments roughly aligned with the main sections of the manuscript

1. Introduction

Could maybe mention that AMSTAR, CASP etc. are complex and you wanted to develop something simpler for narrative reviews which are more numerous. I don't see the logic of arguing that narrative reviews are "unappreciated" because they are the largest text type. Or that they are influential for the same reason. The assertion that narrative reviews are widely known to be unreliable points to a 1987 reference. It is widely known that all the medical research literature is unreliable.

It would be good perhaps to phrase the last sentence of the background to explain what you mean by item-total correlation, internal consistency and criterion validity. These may not be understood easily by the target readership? I had to work them out by reading and re-reading the article. I'm still not quite sure…

2 Methods

Some of the language in this article is quite complicated/wordy with a lot of jargon - almost seems to be over-compensating for the relative simplicity of the study! Eg. "The maximal sum score is 12." "The sum score of the scale is intended to measure the construct…"

In describing the statistics I find it that I am expected to know about things like Cronbach's alpha and ICC. I also needed to know what your pre-specified criteria for "success" in consistency and reliability scores, based on tests of comparable scales such as AMSTAR. I have no idea what criterion validity or pointbiserial correlation is.
It's great that you followed the GRRAS reporting guidelines. It would be good therefore to have a copy of the checklist submitted with the ms so reviewers (including me) could see where you have reported each of the items, and an explanation of why some items weren't appropriate.

3. Results

Surely the total number of ratings was 3 (number of raters) x 30 (number of mss) x 6 (number of items) = 560. It would be good to make the raw score data available in an excel spreadsheet as well as presenting summary statistics which could then be checked. In the methods you said you would present medians where appropriate. I think natural frequencies might be more appropriate than means for a 3 point scale? I didn't see any medians …were there any?

Disagreements most often occurred with item 1 and 4. Would be useful to know how often the disagreement was total (ie. 0 and 2) and how often 0 and 1 or 1 and 2.

Again the language is quite complicated…eg. "the lower level of the confidence interval (0.57) indicates that we cannot exclude a degree of reliability that would be difficult to accept in most settings” I have no idea what this sentence means. Where does the 0.57 come from? This doesn't look like a confidence interval

The result that all raters confirmed that completing the scale is feasible in everyday editorial work is unsurprising given that the raters were the editors developed the scale specifically for that purpose.

4. Discussion

It would be good to have included discussion of why there was more disagreement with items 1 and 4. They have since been edited along with 5 and 6, but you didn't report what changes have been made.

The conclusion that the internal consistency and item total correlation are sufficient and the inter-rater reliability is satisfactory seems based on the authors' personal judgement rather than evidence. As I mentioned before there is no comparison with reliability tests of other similar scales such as AMSTAR which I presume exist.

I read up on Cronbach's alpha here https://stats.idre.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/and it says that a reliability coefficient of .70 or higher is considered "acceptable" in most social science research situations. In this study the figure was 0.68. The earlier version of the scale with seven items had 0.80 and 0.84. I think the argument that the manuscripts which were included in the two earlier studies might have been easier to rate is weak.

You conclude that because the intraclass correlation is .77 SANRA can be used by different raters. Again it would be good to have a comparison with the correlation achieved by other scales. And what intraclass correlation actually means.
I think rather than recommending training, it would be better to test the further tweaked SANRA again pragmatically - perhaps in a randomised trial with a larger set of mss's (sampled from the published literature) and with a larger number of raters independent of the authors of this study.

Another idea would be to collaborate with others to develop a tool to help researchers plan, conduct and report narrative reviews. They are indeed a very important type of study which form the foundation of all primary research so it would be more constructive to work with other stakeholders (eg. the authors referenced 12-14 as having written editorials.), interested in improving both research conduct and reporting. I find it frustrating that a lot of research on improving conduct, reporting and quality appraisal is not joined up.

I think it's impractical to expect journals or anyone else to instigate rater training on the use of such a tool for editors and peer reviewers of one particular type of study however common it is. Having said that, I can see the value of using SANRA for training early-career researchers in how to conduct better quality narrative reviews in the first place whether for publication, or as foundation of their primary research.

I am totally in agreement that simplicity is key but I still think the simplicity should preclude the need for training which in the real world will not happen - at least not for one study design. More development and testing needs to be done, and if the tool is proved reliable and useful, it should be made easy to find and use for authors, editors and peer reviewers.

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

An article whose findings are important to those with closely related research interests

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

Needs some language corrections before being published

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