Reviewer's report

Title: Researchers’ perceptions of research misbehaviours: A mixed methods study among academic researchers in Amsterdam

Version: 0 Date: 07 Jun 2019

Reviewer: Alison Antes

Reviewer's report:

In this study, the authors report on the results of a survey and follow-up focus groups to understand how researchers at universities in Amsterdam in different disciplines and career stages view research misbehaviors. While the findings may offer some contributions to the field, and I appreciate the work that went into collecting these data, my view is that the current formulation of the manuscript lacks details to ensure the clarity and rigor of the study purpose, approach, findings, and conclusions. I would recommend the following edits, details, and rationale to be addressed in each section:

Title:
(1) I recommend editing the title, as it is not an accurate reflection of the study. The study does not examine whether research misbehaviors differ by field. Maybe edit to be "perceptions of research misbehaviors," or something like this.

Introduction:
(1) The introduction indicates that the study examines what researchers view at the most important research misbehaviors. Later we find that "importance" was computed through scores given regarding the frequency and impact of the behaviors. While this method for computing "importance" may be justified, I'm not sure it represents a measure of "what the researchers view as the 'most important' research misbehaviors." More justification of this metric is needed.

(2) Furthermore, more conceptual clarity of what an "important research misbehavior" means is necessary, as it seems confusing to refer to a misbehavior as important. What is meant by "important"? Please be specific.

(3) Similarly, in relation to your measurement and methods, if the most frequent are also not the most impactful—and the introduction specifically mentions that misbehaviors that are more frequent could be impactful because they are frequent—why multiply the ratings of frequency and impact together to create a new variable? What precisely does this new variable offer us in
interpreting the findings? I'm not saying this approach can't be potentially justified, just that the rationale isn't clear enough.

Methods and Results:

(4) Several additional sentences should be added to explain and justify the planned missing data design, and how the assignment of 20 items to each person ultimately yield satisfactory sample sizes across groups (discipline; career stage) and items. Also, in this type of design, how does the research team know if the participants’ response patterns might be affected by the 20 items they see? If you see 20 items that vary greatly from the 20 items another participant sees, could this affect responding? Additionally, please justify the sample size for individual items, especially in some disciplines where the n size is quite small. This to me seems like a design concern that has to be discussed. How can we know that 40 responses, for example, on a particular misbehavior represents how researchers in humanities view that behavior?

(5) Even though Bouter is referenced, and the items shown in an additional file, I would advise a paragraph describing the development and rigor of the survey items used in the survey. How were they developed? Are they reliable? Are the response scales used in this study from Bouter, or adapted for the current study? More details about the survey are needed. Also, were they translated? Or, was the study performed only in English?

(6) Page 7 indicates that scores range from 0 to 15; elsewhere it is indicated they ranged from 1 to 15. Please correct.

(7) Two additional questions about the computed variable: (a) If the computed variable has one score with a scale of 1 to 3 and the other score a scale of 1 to 5, is the 1 to 5 scale given more weight in this computation? If so, is this intended, and is this conceptually the desired meaning of these scores? (b) Are the standard deviations in Table 2 correct? They seem rather large. Please check, and please provide a brief interpretation of these large SDs.

(8) Please provide a rationale for examining only the top 5 behaviors. Why 5? And, how are these interpretations of the data as showing that fields are similar or different being made? Are group comparison tests being performed? Are you adequately powered to make these comparisons given sample sizes? Even if the top 5 behaviors across disciplines sort out to the same, or similar, order, are the mean scores statistically different, if so, what does this imply? Please explain. Additionally, was an analysis done to conclude that career stage didn't influence perceptions, and therefore aggregation of career stage within discipline was appropriate?

(9) The survey response information suggests over 1000 people quit before completing the 20 survey items. This seems like an unusually high partial-completion rate. Do the authors have any idea of why this partial-response rate was so high? How long was the survey—the entire study, as I understand that this was part of a longer study? (And, in general, it would be helpful to make it clearer how the existing report is a subset of the larger project. It is hard to fully
understand this from the manuscript. I read the supplemental materials, which helped to some degree, but also created information overload.)

(10) The link between the survey portion of the study and the focus groups is confusing, and I think largely due to lack of details in the manuscript. As 'sequential explanatory design' would suggest, the focus groups would follow-up on quantitative survey responses, but the results section focuses mostly on describing the new misbehaviors identified. So, was the purpose to understand the survey findings, identify new misbehaviors, or both? Page 7 states, "To better understand the relevance of the aggregated impact top 5 research misbehaviors established in the survey, we organized focus groups with researchers from three academic ranks and four disciplinary fields," however, it is confusing how the themes reported on from the qualitative interview were identified? Was this discussion of the top 5 research misbehaviors, the addition of new behaviors, or both? Also, it isn't clear what the participants were specifically asked to do in the Participative Ranking Methodology. I think this process involved reviewing all 60 behaviors, plus new ones, and ranking them. It needs to be clearer, as well, what photos were taken of. What exactly do the results of the focus groups tell us? Do they clarify why the top 5 from the survey are problematic? Do they add new behaviors to be concerns with? Also, the themes from the disciplines are fairly overlapping? Are their disciplinary differences that need to be clarified, or were the differences negligible? This isn't clear. Similarly, were there differences of note across career stage?

(11) More needs to be said in the methods about the coding approach and generation of themes. What made a theme a theme? Did it have to occur a certain number of times, etc.?

(12) It would be good to indicate if focus group participants could have also participated in the survey, or if this was not the case.

Discussion:

(13) I am left without a clear understanding of the key take-home points and contributions of this study in the discussion. Also, as noted above, did career stage make a difference? A brief comment might be worthwhile in the discussion.

(14) The discussion should connect the findings clearly to the research questions of interest, and engage a more lengthy discussion from how they connect or diverge from past findings. For example, on page 19, mentoring references should be provided.

(15) If the authors think it worth mentioning, do you think that the academic environment in Amsterdam would influence responding and results? Or, are these findings potentially universal? How far might they generalize? This might be worth noting, as your readership is likely to be international in scope.
The limitations note response bias is not a problem if the respondents are representative. Do you think your respondents were representative? Do you have any evidence or way to confirm this? The comment is correct, but are you saying you think this point holds in your study? This is unclear as written.

The discussion seems to bring in a new issue, namely whether the items as written in the survey are clear to participants. This raises concerns about the survey itself, and while may be worth mentioning, should not be a key focus in the discussion.

Overall, I feel most of my comments relate to the need for greater rationale and detail in the manuscript. The issue of greatest concern to me is the measurement of "importance"—how this variable was constructed and what it means conceptually, along with the sample sizes for individual behaviors. It is important to strengthen this information in particular. In terms of manuscript writing, take your reader on a journey through each section, bearing in mind that we have no background context for your design, analysis, and interpretation choices. I appreciate all the additional supplements, but I do not advise leaving your reader to do too much work to understand your approach. Best wishes in this and your future work.

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:

Acceptable

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 organization 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 organization 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