Reviewer’s report

Title: Quantifying Professionalism in Peer Review

Version: 0 Date: 18 May 2020

Reviewer: Noah Haber

Reviewer's report:

The issues addressed by the authors are extremely important in modern research, and vastly underappreciated. Abusive or inappropriate review is a huge and serious problem in academia, particularly with regard to ECIs. In broad strokes, a systematic review of inappropriate review comments is an excellent way of bringing these issues to light.

However, while the issue and broad idea is generally excellent, I have major concerns about the design, execution, and reporting. I was unable to follow the systematic process of the methods, nor was I sure about the justification of many of the key methodological decisions. It appears that this was a qualitative or mixed quantitative / qualitatively coded study, but neither the qualitative nor the quantitative frameworks or methods were described, if any such frameworks were utilized. There also appears to be writing issues that made it extremely difficult to follow and understand what was done, and what the results were. Finally, I have ethical concerns about the conduct of the case studies, which may be in violation of human subjects research requirements.

This study clearly took a lot of work, and I sincerely hope that this work can be shine in a future version of this manuscript so that we may use it to best tackle these issues. I also hope that the authors find that these review comments to be free from the abuses that are detailed in their collected review data, which would be terribly ironic. Were this to be published, I believe it requires a major overhaul, if not a complete rewrite, of the methods, results, discussion, and conclusions sections. As such, I have limited my comments largely to those sections.

I have focused disproportionately on the criticisms below, largely because I hope that they can be used to improve this paper so that the world can see this work.

Methods:
To review this section, the criteria I am using is that I - an experienced but otherwise naive researcher - should be able to understand well enough to at least roughly replicate what the authors did. Unfortunately, after several re-readings, I was unable understand the process of how the data were acquired and organized. This is in part due to inconsistent writing and language, but also in part due to incomplete information. Using headings and subheadings, and text that explicitly said that these were independent searches would go a long way, as would diagrams explaining the process.
I would strongly suggest flow charts or similar. In the results section, diagram types that might be found in systematic reviews, such as PRISMA diagrams, would go a long way to help guide the reader through the process.

As best I can tell (but I am still honestly not sure that this is the case), there were two completely independent sources of review data: "author case studies" and Publons reviews. If that is the case, this should be spelled out explicitly. A full paragraph or similar laying out the data collection process would help greatly.

Author case studies: The case studies seem out of place, and I am unsure they add to the study. I am concerned that this will produce a highly selected group of individuals who experienced particularly brutal reviews (people who have worse reviewers will be more likely to volunteer, heavily biasing the sample). As far as I am aware, this is not discussed in either the methods or discussion sections.

In general, I am not sure why these were included at all, given that you have a systematically selected sample elsewhere. One suggestion for this section would be to clearly lay out what you are trying to accomplish with the early career author case studies, and structure the results around these justifications.

How were the original 327 researchers selected to be e-mailed? What was in the e-mail that was sent to them?

I am also concerned about ethical consent, which is not at all addressed. Was consent obtained from these individuals? What about the reviewers of the articles whose reviews (abysmal though they may be) were written under the assumption that they would not be distributed. These are non-public data of a sensitive nature that are being used for research purposes in a potentially vulnerable population, which at minimum necessitates specific justification for why it would be exempt from human subjects review.

Publons: Why were these three journal types selected? This seems simultaneously like an arbitrary but also very specific set of sections. If you were looking for diversity, why not a random selection from all of Publons' subject areas?

There are also quite a few results in this section. The methods should contain the search strategy, while the results should contain the counts of articles contained at each stage of the process (preferably with a PRISMA-like diagram).

Reviewer comment sets: The definition for this appears after it is used several times, which is very hard to follow. Please integrate this into the text. You also describe an "interaction" but do not define what this means. It would help a lot if you gave a basic description of what Publons is, and the data therein.

Evaluation questions:
This is the first mention of "coding" in the paper, which is typically a qualitative concept and framework. However, no mention is made on the qualitative methods for how the codes were
determined. What qualitative frameworks were used, if any? I am concerned that the lack of mention here suggests that no such frameworks were used or considered, or that a qualitative research expert contributed to the methods design and execution.

Did TGG and JAR have an agreed-upon coding language beforehand? Was there any cross-validation of the coding?

These appear to be categories which are extremely subjective, and some appear to take quite a bit of effort and knowledge from the author coders. For example, what is an "unprofessional" comment vs an acceptably professional comment? I am not aware of any standard definition here, and this is the key argument of the paper. How was this determination actually made? Subjectivity is necessary, and even desirable here, but I have no confidence that I would be able to replicate this coding were I to attempt it.

I see no information on how the statistics were generated. Elsewhere in the paper, there are ranges of proportions which I assume (I did not find a label) are 95% confidence intervals generated somehow, but how?

Results:
The results are extremely minimal and lacking in detail, and missing critical information. For starters, we need some basic descriptive statistics here about the sample. How many total comments were coded, and from what sources? The total number appears in the Discussion, but needs to be in the results.

It also appears as though the authors pooled the case studies and the Publons reviews together. These should be treated separated (or alternatively the "case studies" removed from the paper altogether), unless there is a good justification for it.

I am very confused about what (I assume) are the 95% confidence intervals here. These appear to be extremely wide for a sample size of 1500. However, I have no way of verifying that this is a problem, as I am not sure what these numbers represent. What does this range represent?

I would strongly suggest a paragraph for each of your seven sections, including quotes. I suspect that the quotes themselves are likely the most valuable part of this paper, rather than the counts, and it would be great to see them. Much of these data exist, but in the discussion section and not in the results section.

I would strongly suggest modelling this section after common structures in the qualitative/mixed methods literature.

Discussion and conclusions:
Given my concerns about the methods and results sections, I am going to limit my critique of the discussion and conclusions to broad critiques, under the assumption that they undergo a major rewrite and reorganization. I have also combined my comments regarding the discussion and conclusion section, the critiques below are applicable to both.
A huge amount of the discussion section are results that should be in the results section. For example, the quotes, the statistics, etc., are much better suited to the results section, as these are the product of the methods. This partially explains why the results section is lacking.

Much of the discussion section appears to make broad and wide sweeping arguments regarding the publication system. While I believe they are well-intended, the paper would benefit from a much narrower focus.

For example, the authors describe "post truth" as a relevant concept, which is tangential to the topic at hand, adjacent at best. I was unable to follow the line of logic that lead us from abusive comments to post-truth society at large. The authors are, in my opinion, attempting to do vastly too much here, and not all of it relevant to the actual study performed.

For reference, there are 6 pages of discussion section for one paragraph of results. To me, this alone suggests that the authors should reconsider their broader argument, and keep the discussion relatively narrowly to the paper at hand.

Ironically, I found much the language in this section to, itself, be potentially unprofessional. For example, much of the language is heavily tinged with moral arguments and the authors' beliefs. While I don't disagree with any of them off hand, opinions and preferences should be clearly marked with "we believe," "we suggest" or similar. For example, there are large sections describing what reviewers "should" do. In general, I agree, but it is best to state these as the beliefs of the authors.

Similarly, I found the dismissal of key limitations as "obvious criticisms" rather than key limitations to be poorly worded. It may not be "obvious" to many, nor are they necessarily criticisms, but rather limitations.

Many key limitations were missing from this section, including the selection bias induced by self selection of the "case studies," lack of comparison with non-ECR-generated papers, subjectivity of the assessment of the questions, ethical concerns, etc. A more thorough assessment of the limitations of this paper would help greatly. I was also not fully able to assess these limitations, in part because I didn't fully understand the methods.

In sum: there is some really great thoughts here, and this is important work. I believe that this paper can be revised to generate a study which is both scientifically and socially valuable. I hope to see a version of this to further public discussion on these issues soon.
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:

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