Geosci. Commun. Discuss.,
https://doi.org/10.5194/gc-2019-20-RC4, 2019

Interactive comment on "The benefits to climate science of including Early Career Scientists as reviewers" *by* Mathieu Casado et al.

Lonni Besancon (Referee)

lonni.besancon@gmail.com

Received and published: 28 December 2019

Disclaimer: I have not been involved before with the IPCC report reviewing (although I have read some chapters for curiosity or facts checking) and I don't work in climate science at all. I have, however, been involved before on open science and open review efforts. I am also myself an ECR (postdoctoral fellow, PhD + 2 years).

In this submission, the authors present the results of a study on the impact of including ECSs in the reviewing of IPCC reports. The paper is well written and very interesting. It presents interesting data and effectively argues for a change in the recruitment of reviewers.

I nonetheless have some issues with the current submission that I will list below.

First of all, I find the name ECS to be confusing. In many scientific communities I have been seeing ECRs used instead and I would argue that the authors should perhaps switch to using ECRs instead. Both are fine eventually, I just want the authors to know that in my field of expertise and the other fields I have contacts in, this is not a commonly used term.

"Each chapter was distributed by the project leaders to the participants. Depending on the number of participants for each chapter, 10 to 20 pages were assigned to each participant. We attempted to assign whole sections as much as possible. We also attempted to balance the workload and in some instances, reviewers were assigned non-contiguous sections to even out the number pages they were responsible for." I am particularly skeptical about the consequences of that decision. I am not learned in how IPCC reviewing usually works (although I have read some chapters from IPCC reports before), I would argue that a thorough reviewing process probably cannot be completed if a whole chapter is not assigned to reviewers. In particular, I would personally feel extremely uneasy about having to review only parts of a scientific text. While it is true that, at times, I find myself reviewing only the parts of a scientific communication that I feel I have the expertise to review, I nonetheless read through the entire communication in order to understand the context/application/goal better. I would therefore like the authors to clarify what the impact of this decision to review parts of a chapter could be on the results they obtained. In particular, my concern is even more important if we consider the text two paragraph later, stating "Though participants themselves chose the chapter which they would have to review, a significant number of concerns were raised from participants that felt that the section they were assigned did not correspond to their particular expertise." I think this should be really clarified in the submission and potentially highlighted as a clear limitation of the work produced by the authors.

"The total workload of the participants was less than the project leaders, who spent an estimated 40 hours to prepare the project, participate in the webinars, read the different chapters in which they were involved, and sort all the comments." How is the workload of participants estimated? I would expect the workload of participants to be vastly heterogeneous, and in particular students to be, potentially, extremely rigorous in their reviewing assignment and therefore take quite some time to complete their review.

Line 175 “These three figures are not significantly different.” Are the authors talking about statistical significance? If yes how was that determined? Dichotomous reports of statistical results have been shown to be particularly harmful by many different scientists (see references [A–F] for instance, although many others are available and usually cited in the the ones I provided hre) and I would like to advise against such reports (in particular if no exact numerical results are reported).

Line 185: It is unclear what significance threshold the authors are using. Alpha = 0.05 is often used, yet the p-value obtained by the authors is greater than this. It is absolutely fine not to use a specific cut-off for significance but in this case I would argue that the authors should not use the word “significant” to describe their results. I also agree with the authors’ results interpretation that this result seems interesting and definitely noteworthy, but the authors could simply interpret it in terms of strength of evidence obtained instead of using “significant”.

I would argue that figure 1a is not really appropriate. While having a map is always informative, the high level of clutter in Europe makes it very hard to read. A simple bar chart would have been more helpful in this case, and would have made comparisons much easier. Similarly, the donuts charts in figure 1c are very hard to compare between different categories of researchers. Also it would seem that the time to get training on reviewing for the study presented was not included in the workload of participants, which can be problematic.

Line 210: “The PhD students provided as many substantive comments as the more experienced participants of the group review (i.e. Post-docs and Early Career Academics), thus the length of the academic career was ruled out as a factor in the ability to effectively produce reviews.” seems like an overstatement. The number of sub-

stantive comments itselfs is not enough to rule out the length of academic career as a factor in the ability to produce good reviews. While we can trust the categorization made by project leaders, this number alone is not enough to make the conclusion that the authors are making. They are still many things that need to be considered: how critical are the comments made by the reviewers, do the comments identify critical bottlenecks/flaws... This specific conclusion should therefore be rephrased and take into account the limitation of this single metric as a measure of ability to review.

Line 218: “This latter aspect is particularly relevant for the climate change community considering the need for transparency in the peer-review process (Edwards and Schneider, 2001).” this comment is not particularly clear. The clear need for transparency in the PR process for climate science should first be made clear and then tied back to the claims of the authors.

Line 226: “The relatively reduced time commitment might make participating in future IPCC reviews more appealing to ECS.” My comment here might be a bit outside of the scope of that submission and review. I will make it nonetheless. Reviewing can be appealing to ECRs if it is somehow also relevant for their careers. While open and non-anonymous reviews seem to be on the rise in many scientific communities, ECRs get a lot of pressure from the publish or perish system (leading to multiple calls for a change of system, see for instance [G]), from their supervisors and from peers which are incompatible with an involvement or a rigorous participation in the peer review system. While the authors then highlight the benefits that ECRs self-report they would get from participating in the reviewing process, this has to be contrasted with the limitations I have just highlighted. The figures provided by the authors in the next paragraphs are, still, very important and very nice indication that ECRs are willing to participate in the reviewing system and see value in it.

Line 254: “There is no evidence that this attribution process influenced the quality of their comments.” This statement is quite bold and unsupported and ties back to one of my first comments at the beginning of this review. The manuscript should clearly state how the authors came up with this conclusion?

Figure 2a, I find it really interesting that the number differ from group peer review to individual review. Indeed, it would seem that the participants, according to the authors, did not communicate much within groups. I wonder if the authors can think of an explanation for that result somehow. Do they have some data on this?

Line 271: I would particularly like to highlight two things in the argument made by the authors here. First, point number 1 is not a direct benefit for an ECR, but a benefit for scientific communities, for the report at hand, and/or journals/publishers instead, and I would therefore argue that it should not be listed there. Second, I would argue that point 3 is not always true. Many fields and journals still have double/single blind reviewing processes and this argument simply does not apply in this case. This should be highly contrasted by the authors, should it stay in the final version of this manuscript. Instead of these benefits, I would mention that conduction reviews offers other advantages: 1/ better understanding how one's writing can influence how a paper is perceived and therefore improving on one's own writing skills, 2/ keeping up to date with recent literature, 3/ participating and understanding the reviewing process better and therefore getting an understanding of how one's paper is reviewed by others. Other benefits can be found, (in particular in open review processes [H,I,J,K]) and I wonder whether the authors should focus on these points instead.

Line 295: this echoes previous findings from the literature (e.g., [J]).

Is the data going to be made available in some way by the authors? There are a few things that do not appear clearly in the manuscript. For instance how did project leader categorize comments? Was some form of agreement score computed? Or did all comments only get reviewed by one project leader? This is not very clear in the current manuscript. Making the data available would help to clear that but this should also be clearer in the manuscript too.

Update after having read the other reviews: I agree with RC2 that, should this work be

a close replication of previous work, it should be clearly mentioned in the submission. I don't think this hinders the work conducted by the authors in any way, but it should simply be made clearer. Should it not be a close replication of previous work, then the authors should clearly highlight what the differences are. I disagree with RC1 that figure 1 is excellent. The data is interesting, the visualization is not well chosen though. I agree with RC2 that the guidelines given to ECRs should be made available. This will, in particular help future replications of this work.

Overall, I am very positive about the work done by the authors and I simply think that some rewriting is needed to make some point clearer or more contrasted. I would personally argue that the revisions needed to accept this submission are minor and that the authors should be able to address all concerns raised so far (by myself and the other reviews I have read) in a short revision time.

REFs: [A] https://www.nature.com/articles/d41586-019-00857-9

[B] https://peerj.com/preprints/2921v2/

[C] https://journals.sagepub.com/doi/10.1177/0956797613504966

[D] https://link.springer.com/chapter/10.1007%2F978-3-319-26633-6_13

[E] https://hal.inria.fr/hal-01980268/file/alt.chi_2019_CPDI-authors.pdf

[F] https://discourse.datamethods.org/t/language-for-communicating-frequentist-results-about-treatment-effects/934

[G] https://opensciencemooc.eu/evaluation/2019/10/15/solve-research-evaluation/

[H] https://academic.oup.com/femsle/article/365/19/fny204/5078345

[I] https://doi.org/10.18243/eon/2017.10.8.1

[J] http://dx.doi.org/10.20944/preprints201905.0098.v2

Interactive comment on Geosci. Commun. Discuss., https://doi.org/10.5194/gc-2019-20, 2019.