Fw: [External] Political Science Research and Methods - Decision on Manuscript ID PSRM-RE-2022-0155

Tai, Yuehong <yuehong-tai@uiowa.edu>

Fri 2023-05-19 11:36

To: Solt, Frederick <frederick-solt@uiowa.edu>;Yue Hu <yuehu@tsinghua.edu.cn>

🔗 1 attachments (74 KB)
Attached standard file: 2023-01 PSRM-review.pdf;

Hi Fred and Hu,

PSRM gives us an R&R. Although the review process lasts nearly 1 year, we get very positive feedback from two reviewers (one of them is Alexander Wuttke~) and the editor.
I just read it quickly. Both reviewers offered constructive suggestions and useful literature on how to solve the tension between "general terms" and "very specific content" in our manuscript.

I'll drive to Pitts for state politics conference tomorrow and will not be able to itemize these suggestions before next Monday. However, please feel free to open an issue for revision in our repo if you'd like to and have time to do it. Otherwise, I will initiate the issue before our Wed's meeting.

Also, the due day for resubmitting is 16-Aug.

-Cassandra

Yuehong Cassandra Tai
PhD Candidate
Department of Political Science
University of Iowa
yuehong-tai@uiowa.edu

From: Political Science Research and Methods <onbehalfof@manuscriptcentral.com>
Sent: Thursday, May 18, 2023 1:45 PM
To: Tai, Yuehong <yuehong-tai@uiowa.edu>
Subject: [External] Political Science Research and Methods - Decision on Manuscript ID PSRM-RE-2022-0155

18-May-2023

Dear Dr. Tai:

Manuscript ID PSRM-RE-2022-0155 entitled "On Data 'Janitor Work' in Political Science:The Case of Thermostatic Support for Democracy" which you submitted to Political Science Research and Methods, has been reviewed and I have read it carefully as well. The comments of the reviewers are included at the bottom of this letter.

The reviewers recognize the potential of your manuscript but they also suggest some major revisions. Therefore, I invite you to revise your manuscript.

In your revision, you should pay attention to two issues (besides the comments of the reviewers, of course):
First, I think the article will have more long-lasting impact if it reads less like a replication of one specific paper, but more like a general point about data entry errors (as well as the distinction between real "replication" versus "repetition" of an analysis). Second, I must admit I was somewhat confused when reading the replication part of the manuscript. It does not always seem obvious to me that the differences between your results and the replicated paper are due to "data entry errors". Some just look like simple coding mistakes or "case-by-case decision" that turn out to be questionable/consequential. I would help if you could clarify this in your revision.

To revise your manuscript, log into https://mc.manuscriptcentral.com/psrm and enter your Author Center, where you will find your manuscript under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number will be appended to denote a revision. You may also click this link to start your revision:

*** PLEASE NOTE: This is a two-step process. After clicking on the link, you will be directed to a webpage to confirm. ***

https://mc.manuscriptcentral.com/psrm?URL_MASK=23ed5b28668c4d36a381fa170cd87fd4

When submitting your revised manuscript, you will be able to respond to the comments made by the reviewer(s) in the space provided. Please use this space to document any changes you make to the original manuscript. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewer(s).

Because we are trying to facilitate timely publication of manuscripts submitted to the Political Science Research and Methods, your revised manuscript should be uploaded as soon as possible. We expect to receive your revision by 16-Aug-2023. If it is not possible for you to submit your revision by this date, please contact the Editorial Office to rearrange the due date. Otherwise we may have to consider your paper as a new submission.

Once again, thank you for submitting your manuscript to the Political Science Research and Methods and I look forward to receiving your revision.

Sincerely,
Daniel Stegmueller
Political Science Research and Methods
ds381@duke.edu

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author
Since the 1990s political scientists talk the talk of replication but ever so rarely walk the walk. Given the current incentive structure in academia which puts a high premium on novelty, it is exceedingly rare that scholars put substantial time and resources into checking existing studies. There is sufficient meta science research to believe that due to these structural problems the published literature is filled with

empirical studies with a seemingly strong evidence base that actually do not deserve our trust but no one has ever bothered to check. As a result, as a scholarly community we rarely separate the wheat from chaff and the bad apples (to go all in all bad metaphors) remain in our basket for future scholars to eat and perhaps even to build their dissertation or grant projects on. In other words, we should reward those scholars who actually put into practice our lofty ideals and it is laudable that the authors of this manuscript did the meticulous work of checking another person's work which the original authors can see as a badge of honor for the relevance of their work.

Nonetheless, to the extent that we do stick to the problematic habit of rewarding novelty, it is worth pointing out that the authors present not just any kind of generic replication study but they point to manual data entry as a source of non-replicability that has not received much attention. It really is an intriguing idea that the whole notion of reproducibility as an indicator of credible is insufficient and perhaps even misleading because errors may linger in the data generating process (and not in data wrangling or analysis).

Given these merits, the study deserves publication in a reputable journal such as PSRM. I assume the replicated author will be among the reviewers and am confident that he will competently engage with the analytical/ empirical aspects of the study. In the remainder I want to focus on the gap between what is promised and given and on the study. Altogether, my recommendation is that the authors should exhibit more caution in generalizing their findings as the heart of their study is one reproduction attempt on one study.

** Manual vs automated data entry.
The introduction and conclusion seems based on the premise that it is always preferable to automate manual tasks. But this seems questionable. The claim is unproven and deserves more systematic discussion. For instance, manual tasks indeed run the risk of unconscious, idiosyncratic errors. But automation runs the risk of systematic errors. For an overall judgment of what is preferable a balanced discussion of all potential error types would be necessary. It is also not clear that the presented evidence gives compelling reason to believe the superiority of automated data entry even in this case because we do not know whether the automated data are valid.

** Framing, contribution, implications
A related publication has been recently published in the APSR (https://doi.org/10.1017/S0003055422000429). Although that study has a slightly different focus (on the role of measurement uncertainty), the authors should transparently discuss this aspect to rule out any doubts that a motivation would have been to slice one research project into minimally publishable units. Perhaps the fact that some of the empirical evidence has already been presented elsewhere gave rise to the manuscript's framing where I perceive a conflict between the manuscript ambition and the manuscript's content.

Specifically, the tension concerns the article's framing on manual vs automated data entry in general terms and the article's very specific content (reporting a reproduction of one article in empirical democracy research). The actual content is much narrower and more specific than the framing implies. The framing makes the reader expect a comprehensive treaty of the pros and cons of manual vs automated data entry, perhaps a systematic review of available literature etc but all of this is missing. The study makes it seem as if automated data entry is always preferable to manual entry but this conclusion seems due to the superficiality of the discussion. I believe a more balanced, comprehensive treaty would reveal trade-offs and contexts where the one works better than the other because, to name one point, automated processes come with lower levels of control and lower levels of researcher

awareness of what is actually happening under the hood of the automated processes.

In a similar vein, the recommendations given in the discussion section sound reasonable but stand on thin ground as the empirical evidence for the superiority of each suggestion over its alternatives is weak. Rather than putting us in a position to make general suggestions, I would interpret the one case study presented in this manuscript as giving rise to an interesting and innovative research agenda on the (presumed) importance of manual data entry. As part of that research agenda, each of these suggestions could be empirically tested on a greater set of studies.

Altogether, one way forward to alleviate the described problems could be to take the issue manual data entry seriously in theoretical respect. But expanding these discussions could run the risk of artificially expanding the manuscript. Another path could be to be more modest in the framing and conclusion and treating this study as an exploratory case study with the inductive value of giving us reason to think more systematically about data vs automated data entry and test these questions in an future research agenda.

Minor

There is a new OA political science book out this month which may be a useful reference in the introduction on the importance of data janitor work but feel free to reject this suggestion if not helpful: https://www.cambridge.org/core/books/data-management-for-social-scientists/33356BF6DE034B25239DCF7C436CBFCD

I hope these comments were useful to the authors and I wish all the best for this interesting project.
Alexander Wuttke

Reviewer: 2

Comments to the Author
(There are no comments.)