

January 20, 2018

Dear Professor Hubbard,

I have now received reports from three reviewers regarding your manuscript, “A proposal for replicating Evanschitzky, Baumgarth, Hubbard, and Armstrong’s ‘Replication research’s disturbing trend’ (Journal of Business Research, 2007)” (*Manuscript Number 2341; Discussion Paper Number 2017-75*).

Based on the feedback I have received, I am inviting you to submit a revision. Your revision should address the individual points raised by each of the reviewers. In addition, you should include a separate document that provides a point-by-point response to each of the reviewer’s comments. You are not required to do everything the reviewers say, but if you disagree with a recommendation, you should state your reason. A successful revision will include addressing the following:

1) A revision should be clearer about the “type” of replication you are doing. You describe your replication as an “exact replication.” Yet, your use of different journals and time periods than those used by EBHA runs contrary to how many would understand this term. To better help the reader, it would be helpful if you mapped your replication to the Tsang and Kwan (1999) taxonomy of replication types that you present in your paper. My interpretation of what you are doing is that your replication maps onto TK types (4), (5) and (6). You may disagree. In any case, your revision should be clearer about this.

2) Relatedly, Reviewer 3 asks “what is the difference between a replication (especially with regard to the replication type 6) and a follow-up study?” As this relates to the purview of the Special Issue, it would be valuable to readers to hear your thoughts on this subject.

3) On page 4, you state that “to qualify as a ‘replication,’ an article has to contain an explicit citation of the original work. This does not mean that the replicating author(s) must identify their own research as such. This will be our responsibility.” Actually, if this is to be a replication, then it is important that you use the same criteria as employed by EBHA. Those criteria are not specified in your paper. Is that because EBHA were not precise in how they determined what constituted a “replication?” Or perhaps their criteria were not explicitly stated and now cannot be recalled? Whatever the case, you need to try and match EBHA’s criteria as much as possible, and to state those criteria as clearly as possible in your paper. Ideally, your description of what constitutes a replication should be sufficiently precise that another researcher, armed only with your description, would identify the same studies that you identify, and be able to reproduce your research. You may find it useful to give a “rule” for how one can identify a replication, and then give examples of one or two papers that sit closely on one side of the line or the other, so that the reader can better understand how you are defining a “replication.”

4) Another issue concerns the criteria for determining whether the set of replication studies you identify in your replication “successfully replicated” their original studies. Again, since you are proposing a replication of EBHA, it is important that you use the same measure of “success” as EBHA. If they used p-values, then you must use p-values. While it might be “better” on some level to use a different measure of success, such as overlapping confidence intervals, this would no longer make it a replication, because you would not be following EBHA.

5) You state on page 5, “Since the data in EBHA and my replication constitute entire populations, no statistical testing will be involved. Therefore, the results will be analyzed by ‘eyeballing’ them.” However, as I read what you have written, you are studying samples, not populations. On page 4, you define the population as “highly regarded marketing journals.” Note that the set of 6 journals that you propose to study in your replication plan is larger than the set of 3 journals included in EBHA. That immediately implies that EBHA used a sample. And if the set of “highly regarded marketing journals” is more than the 6 identified in your replication plan, then your “census” is also a sample. Even if “highly regarded marketing journals” exactly maps onto your set of 6 journals, one could still consider the associated set of articles a “sample,” say, by assuming that the years 2005-2020 are a sample from a larger set of years. Alternatively, one could imagine that the 6 journals are themselves a “realization” from an unobserved, larger population of possible “highly regarded marketing journals” that could have existed; say, the same journals with different editors, or the same journals/editors with different researchers submitting their work to those journals. As the distinction between population and sample is key to how you choose to compare the results of your replication, this issue needs to be discussed.

6) If one takes the view that both EBHA and your replication analyze “samples” and not “populations,” which I think is most appropriate, then how should one determine whether the results are “different.” Your paper promotes the criterion of “overlapping confidence intervals.” However, it is not clear what the statistical basis for this criterion is, as “overlapping confidence intervals” is not the same thing as difference in sample means (see <https://replicationnetwork.com/2017/11/11/parasurama-why-overlapping-confidence-intervals-mean-nothing-about-statistical-significance/>). You either need to provide a statistical argument for using overlapping confidence intervals, or provide an alternative criterion for determining whether results are “different.”

7) Related to the previous point, how would you consider the following issue: Suppose EBHA reported a replication rate of 2 percent, with a confidence interval of [1%,3%]. And suppose your replication estimated a larger replication rate of 5%, with a corresponding confidence interval of [2%,8%], so that there was no statistically significant difference between the replication rates? How would you interpret these findings? Researchers who have problems with null hypothesis statistical testing would benefit greatly from your view on this.

8) Finally, you need to be clearer about how to interpret the results from your replication. As Reviewer 1 points out, what would it mean if your replication produced different results than EBHA? If you really were doing an “exact replication”, it seems that the appropriate conclusion would be that EBHA’s findings were not reliable. Alternatively, and this seems to be closer to what you have in mind, one could interpret this difference as indicating that professional practice was different in 2005-2020 compared to 1990-2004. This all goes back to being clear about the “type” of replication you are doing, as different types require their results to be interpreted differently.

9) Minor: Drop the discussion of the rationale/benefits of using a PhD student on page 4. This is too much detail, and is not germane to the purview of the Special Issue.

10) Minor: I suggest the discussion on page 6 about power be dropped for two reasons. First, as noted above, you need to use the same measure of “replication success” adopted by EBHA

in order for his study to be a replication. If EBHA did not use power, neither can you if you want to keep things comparable. Secondly, it is unclear how one would measure the power of completed studies. Note that any calculation of “post-hoc power” is invalid (see <https://replicationnetwork.com/2017/05/23/reed-post-hoc-power-analyses-good-for-nothing/>)

A revision that satisfactorily addresses the points above, as well as responding to the individual reviewers’ concerns, is likely to result in a favourable publication decision.

Thank you for submitting your manuscript to *Economics: The Open-Access, Open-Assessment E-Journal*. I look forward to receiving your revision.

Sincerely,

W. Robert Reed  
Co-Editor, *Economics E-Journal*