

February 15, 2022

University of California, Merced

Dear Dr. Horton, associate editor, and reviewers,

We appreciate the helpful and thoughtful comments, and thank you for the opportunity to revise and re-submit our manuscript, formerly titled “*Promoting analysis reproducibility with accessibility: An example in evolutionary biology*”, and now titled **“Approachable case studies support learning and reproducibility in data science: An example from evolutionary biology”** (please see response to reviewer 1).

We have thoroughly revised and restructured the manuscript. The comments of the editors and reviewers have greatly improved the manuscript. We focused on the contributions of our work related to teaching reproducibility, and connected our strategies to the pedagogical literature. We hope that you find it suitable for publication in the special issue about teaching reproducibility and responsible workflow.

We address individual comments line by line below.

Please receive our respectful regards,

The authors

22-Dec-2021

Dear Dr Luna Luisa Sanchez Reyes:

Your manuscript entitled “Promoting analysis reproducibility with accessibility: An example in evolutionary biology”, which you submitted to *Journal of Statistics and Data Science Education*, has been reviewed by myself, an associate editor, and two reviewers. Overall the reviews were quite positive about the submission, though some issues were noted that dampen enthusiasm. The reviewer comments are included at the bottom of this letter or as attachments.

I regret to inform you that the reviewers have raised serious concerns, and therefore your paper cannot be accepted for publication in *Journal of Statistics and Data Science Education*. However since the reviewers do find considerable merit in the paper, I would be willing to reconsider if you wish to undertake major revisions and re-submit, addressing the reviewers’ concerns. Please note that resubmitting your manuscript does not guarantee eventual acceptance, and that your resubmission will be subject to re-review before a decision is rendered.

Sincerely,

Dr Nicholas Horton

Editor, *Journal of Statistics and Data Science Education*

nhorton@amherst.edu, nicholasjhorton@gmail.com

*Comments from the Editors and Reviewers:*

*Editor*

*Comments to the Author:*

*It would be helpful to ground this research in broader initiatives (e.g., NASEM 2019 reproducibility report, NASEM 2018 data science for undergraduates definition of “data acumen”).*

These references were really helpful, thank you. We address the NASEM initiatives and its relation to the principles we propose in the third paragraph of the **Conclusions** section.

*In addition, a close edit would be indicated as there were a number of typographical errors throughout the initial submission.*

Thanks. We performed a close edit to eliminate as many typographical errors as possible. We apologize in advance for any errors that might have passed unnoticed and hence uncorrected.

Associate Editor

Comments to the Author:

*While the teaching materials developed by you have a potential to be useful for other lecturers and students of statistics and evolutionary biology, there are some deficiencies in the paper related to how you present your work. Especially sections 3 and 4 need some extensive revisions and improvement (see below). Your writing style could be improved in places too, as many sentences are short enumerations of steps you have undertaken (i.e., ‘We identified...’, We did this and We did that), without unpacking your rationale or any reflections on the process. So some sentences read like notes, not arguments in a scientific paper.*

Thank you for your comment. You accurately point out that the last sections of the ms were quite raw. We revised our ms extensively and worked especially on the sections that appeared as enumerations. **Section 3 - Results and Discussion** and **Section 4 - Conclusions** are now one single section **Conclusions**. We decided to eliminate the discussion section based on a suggestion of Reviewer 2: “This isn’t a study that requires the traditional Intro/Methods/Results/Discussion format. This is more of a description or commentary. I believe that this inevitably leads to the Results/Discussion section not making much sense, at least to me. I suggest replacing the section headers with something more appropriate and then after explaining your main thesis/criticism you then wrap up the article with a concluding paragraph [...]” We elaborate more on this below.

*Section 3 - Results and Discussion - please revise and expand the section. It is not clear how it discusses your results and results of what activity or a test are presented here. The section reads in many places like notes which need further explanation and evidence.*

We expanded and elaborated the ideas that were originally stated in the Section 3 of the original ms, and connected them to the barriers and principles identified in previous sections.

*Section 4 - Conclusions - this section reads as an unfinished draft. Please polish it in the light of the original contribution of your work related to teaching reproducibility. You should also link your contribution to the current debates on reproducibility, open scholarship and pedagogical approaches.*

We added a final paragraph addressing what we believe to be the main contribution of our work to teaching reproducibility.

*When revising the paper please remember that your manuscript was submitted to a special issue on ‘Teaching Reproducibility and Responsible Workflow’ (more details can be found here: <https://magazine.amstat.org/blog/2021/07/01/jsdse-call-for-papers/>).*

Thanks, we framed our analysis around the particular challenges related to teaching reproducible workflows. As suggested by the editor, we ground our principles around the “data acumen” concept defined in the “Data Science for Undergraduates” report (NASEM, 2018 doi:10.17226/25104.)

---

### **Reviewer: 1**

#### *Comments to the Author*

*This article emphasizes the importance of accessibility of computational tools and datasets, as opposed to their mere availability. Accessibility is discussed using the example of a tutorial given as part of a workshop on using the “Open Tree of Life”, an online database front-end meant to facilitate access to public phylogenetic data. The tutorial focuses on the use of the R language for accessing the Open Tree of Life, assuming some prior familiarity with this language. The support material for the tutorial is publicly available, both for students who wish to re-visit it, and for educators as a source of inspiration or reuse.*

*Accessibility is an important and often overlooked aspect in the ongoing transition to more open, transparent, and reproducible research practices. Following the lead of the Open Source movement for computer software, Open Science activists tend to focus on publication and citation of digital assets, claiming that “if it’s available, people can inspect and reuse it”. The emphasis on accessibility is an important complement contributed by this work. Accessibility can be improved in essentially two ways: writing code and documentation in such a way to be usable with less technical knowledge, and educating scientists in the techniques required to work with published digital assets. The article discusses both aspects, focusing on the role of documentation. Another aspect that the article emphasizes is the need to teach not only the correct use of software, but also how to deal with the inevitable errors.*

Thank you for your comments!

That is exactly what we were trying to convey: the fact that something is available does not ensure that people can actually reach that something and interact with it in the way it is intended.

While our main thesis remains unchanged, we decided to replace the term **accessibility** for **approachability**. After sharing a reworked version of the ms with colleagues for feedback, several of them pointed out that the term accessibility is historically and practically tied to addressing barriers faced by people with different physical abilities

(visual, cognitive, motor, etc.). As we were not referring to that type of barriers in our ms, they suggested that we should elaborate more on the way we were applying the term **accessibility** for our thesis. This inevitably lead to somewhat of a dark rabbit-hole of jargon. Trying to be consistent with our own principles (avoid use of specialized language), we decided to use a word that conveyed the idea without becoming a coined concept of its own. Approachable appeared as the most friendly candidate, so we decided to reframe the whole ms in terms of **approachability** and move away from accessibility. Nevertheless, the problem that we identify and try to tackle with this work remains the same.

*There are two minor points that need to be addressed in a revision:*

*Page 6: “following the carpentries” This comment deserves a reference, and maybe a few words about the carpentries’ approach, unless this can be considered generally known in the intended audience.*

We introduced “the Carpentries” and added references to the materials and research papers that support the theoretical and practical backbone of our paper. We also added “the Carpentries” to acknowledgements, as the resources released by the organization were truly essential for the existence of this work.

*Page 7: “Discuss: why address accessibility and not other aspects of reproducibility?” This looks like an unhandled reminder to the authors themselves.*

Indeed, thanks for pointing this out. We removed this sentence. As we no longer address the issue of reproducibility from the viewpoint of accessibility in particular, but from a broader concept of approachability, it is no longer needed to elaborate on this reminder.

---

## **Reviewer: 2**

*Comments to the Author*

*General comment*

*This manuscript touches on some really important facets of the replication/reproducibility problem, namely the link between producing and publishing code and the end user’s ability to use the code. The problem is often one of communication or what Steven Pinker calls “the curse of knowledge”, the speaker knows what they are talking about so well that they leave out important details that the naïve listener needs to make sense of the statement.*

Thanks!

*My main comment here is that I do not think the format of the manuscript is appropriate. This isn't a study that requires the traditional Intro/Methods/Results/Discussion format. This is more of a description or commentary. I believe that this inevitably leads to the Results/Discussion section not making much sense, at least to me. I would suggest replacing the section headers with something more appropriate and then after explaining your main thesis/criticism you then wrap up the article with a concluding paragraph highlighting what authors of future vignettes and tutorials need to provide for everyone to be able to use the code.*

We agree. We reorganized the ms and removed the results/discussion section. We elaborated on the sections where we identify the barriers and describe the principles and its application. We wrap up the ms with a **conclusion** section.

#### *Specific comments*

*Page 3, Line 31: Replace “in” with “on” - Done*

*Page 3, Line 45: Accessibility of what specifically? - Accessibility turned out to be a somewhat confusing term (see response to reviewer 1). What we were trying to convey is that most software documentation is difficult to be approached by a wider audience, which renders analysis workflows that rely on that software harder to reproduce by the majority of the potential user base. We replaced the term accessibility for approachability across the ms.*

*Page 3, Line 52: Remove a closing bracket. - Done*

*Page 4, Line 15: Replace “utilized” with “used” - Done*

*Page 4, Line 24: required baseline, not baseline required - Done*

*Page 4, Line 26: The required computational skills are likely... - Done*

*Page 4, Line 35: Replace “for reproducibility research...” with “...of reproducing published research” - Done*

*Page 4, Line 36: What does “the utility of data resources...” mean? Please clarify. We removed this sentence, it was no longer necessary. But it referred to the usability or usefulness of data resources.*

*Page 4, Line 38: What does “that gap” refer to? Please clarify. - We removed that statement, it was no longer necessary.*

*Page 4, Line 42-43: Replace “is written using” with “communicates in a”. - Done*

*Page 5, Line 19: By “external” do you mean “naive”? - We mean “the non-specialized users” that we now refer to as the “wider audience”.*

*Page 7, Line 35-36: Should be: “the individual user” or “individual users”. - Done*

*Page 7, Line 49-50: Should be: “...users to return to to at any time, and to be...” - Done*

*Page 7, Line 52: What are carpentries? This article aims to make data and code clearer for users yet you drop lingo like this into it; make sure you follow your own advice. - You are right. We define the carpentries and reference it when relevant.*

*Page 7, Line 54: Should be: “...presented in workshops...” - Done*

*Page 8: Results and Discussion, and Conclusions. Perhaps transform these into single summary paragraph - Agreed. All those are now a single section named “**Conclusions**”.*