January 10, 2022

University of California, Merced

Dear Dr. Horton, associate editor, and reviewers,

We appreciate the helpful and thoughtful comments made to our revised manuscript. We are excited to learn that the manuscript is now suitable for publication, subject to minor revisions.

Please receive a newly revised version of 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"**.

We have incorporated into the revised manuscript all comments and suggestions from you, associate editor and 2 reviewers. We believe that all of them have greatly improved the manuscript. We hope that you find it ready 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

Comments from the Editors and Reviewers:

Editor Comments to the Author

In addition to the comments from the reviewers and associate editor, I had the following additional points:

1. The abstract (in the metadata associated with the paper) is missing spacing in many places (e.g., "reproducibilityare". This is likely due to cut and paste issues. Can you please review and fix as this will promulgate through the production process.

Thanks. We reviewed and fixed missing spaces across (hopefully) the entire manuscript.

2. It might be appropriate to consider citing this in-process work (https://arxiv.org/abs/2202.09504)

We agree. We have cited this work at the beginning of the second paragraph of the Conclusion section: "These approaches not only teach reproducibility, but also make the teaching process itself reproducible (Dogucu & Cetinkaya-Rundel 2022).

3. page 2, end of third paragraph: would it be useful to cite the NASEM 2018 (Data Science for Undergraduates) report, which includes reproducibility and workflow as important components of "data acumen"? In addition, would citation of the NASEM 2019 Reproducibility report be appropriate?

Yes, after pondering your question, we concluded that the NASEM 2018 report should also be cited in this section.

The NASEM 2019 report is very interesting and addresses many important points relevant for this work. We cite it at the end of first paragraph, Page 5: "As computers continue to play a larger role in most scientific disciplines (Piccolo & Frampton 2016), higher baseline computational skills are required across all natural sciences not only to develop an original research workflow, but to be able to follow and reproduce research workflows from other researchers (NASEM, 2019)"

4. Note style for journal (Taylor & Francis) is to capitalize article and book titles

Thanks. We went over the entire bib file and capitalized article and book titles. Hopefully they are all appropriately styled now.

Associate Editor Comments to the Author: Thank you for reviewing the manuscript according to reviewers feedback. One issue which is pointed by one reviewer is mentioning package version - which is your first section in the online resource you discuss, so it will be easy to incorporate.

Thank you! We addressed that on the second paragraph of the Conclusion section.

In the Conclusions when you provide examples of some institutions adopting reproducibility as a main teaching principle, you could also mention University of Glasgow School of Psychology and Neuroscience - https://psyteachr.github.io/

Thanks for pointing us out to this resource. We thought it was appropriate, so we added psyteachR to the examples.

Reviewer: 1

Comments to the Author The authors have addressed all the issues I had raised about the initial submission.

We really appreciate all comments and suggestions made on the initial review!

Reviewer: 2

Comments to the Author A significant barrier not discussed here is the deprecation of functions in open software packages like R and Python. It often requires considerable programming savvy to repair code that has fallen victim to deprecation.

Thank you, the point that you raise is very important. Associate Editor highlighted your comment and points out that this can be addressed by providing information on package version used; we agree. In general, old package versions remain accessible, and persist in time, so they can be used to reproduce an analysis, even long time after a function had been deprecated. The teaching materials that we use as example include a section that addresses documenting package versions used for an analysis.

We now discuss this barrier and suggest ways to address it at the end of the second paragraph of the Conclusion section.

Line 38: considerably - Fixed

Line 54: "... as a model..." - Fixed

Line 30: Define "live programming" - Added a sentence defining it.

Line 37: This is called Carpentries here but Software Carpentries in Figure 3's caption. Are these the same thing? - Software Carpentries is a component of the Carpentries, we clarified this in the Figure caption.

Line 53 and elsewhere: "learner's" and "user's". - Fixed

Line 12: Warner et al. 2021 do not make this claim in their paper. - You are right, we changed the sentence accordingly to: "[The principles described here] can also help close the academic gap that is generated by uneven access to computational resources across students belonging to different groups (KewalRamani et al. 2018), where the most affected learners usually belong to underrepresented minorities and rural areas (Warner et al. 2021)."
