Interactive Intelligence
Checklist for Review of Dataset
(Version 1)

We recommend that students or employees wishing to publish on their data and results for a given research project in the form of a dataset asks a fellow student or colleague to review this dataset with regard to the points in this checklist. The purpose of the checklist ist to ensure that all data that can be made available is made available, that all analyses were conducted conscientiously by the researchers, that all results are reported accurately, and that all methods are transparent and sufficiently clear to be reproducible.

If you choose to have your code reviewed according to this checklist, we advise you to upload this document together with your dataset to the research data repository of your choice (e.g. 4TU Research Data) upon publication of your work.

I. Basic Data

Paper title:	**Tailoring for Motivation: Reproducible, Sustainable Generation of Tailored Motivational Messages**
Name(s) of researcher(s):	Ramya P. Ghantasala, Nele Albers, Kristell M. Penfornis, Milon van Vliet and Willem-Paul Brinkman
Name of the reviewer:	Milon van Vliet
Data repository platform (e.g. 4TU Centre for Research Data):	4TU Centre for Research Data

II. Checklist

Statement	**Yes**	**No**
1. The dataset contains a README file that fulfils the requirements of the data repository platform that the researcher wishes to use. If no such requirements can be found, the dataset nonetheless contains a README file that clearly explains the contents of the dataset?	Yes, it contains a README file. Some things that could be improved (not sure if all of these points are necessary): - In the guidelines of 4TU Research Data they recommend to include your contact details. - The same README file is located in the 'Dataset_paper' folder and in the 'Analysis' folder, which was a bit confusing for me (especially because you need to provide the current file path in the code).	
2. Either within the README file or within an extra, easily findable file, the researchers have explained their data. This means that, for example, for every column of a tabular	The 'Columns explanation' file gives a clear overview of the data that can be	

Statement	Yes	No
dataset, all column names and possible cell values are explained.	found in the 'experiment_data' file.	
3. data is in readily readable file formats. If this should not be the case, the README (or similar) clearly explains the file format and which software can be used to access the contents.	The data in the 'experiment_data' file is readable, but due to the format (all data of one participant is displayed in one cell) it is not very clear.	
4. All data has been anonymized in accordance to promises made in the Data Management Plan.	Participant ID is anonymised (not sure what promises were made).	
5. The analysis file or files contain a header with meta-data (name of author, date of writing, required input files and generated output files).	Yes	
6. All required input files for the analysis are available in the dataset.	Yes	
7. There is an output file that is generated by the analysis script that neatly combines code and commentary (e.g. markdown output file). This output file is in a readily readable file format (e.g. pdf).	Yes	
8. The analysis script is clean and comprehensible in the sense that: • There is sufficient, useful, and clearly written commentary • Irrelevant code (such as old analyses) has been removed • The details of analyses that are not reported in the paper (e.g. assumption checks) are proportional to those that are reported in the paper. This means that unreported analyses should not clutter up the script, making it long and unreadable.		I experienced some difficulties when following the README file instructions. See my comments below ('Additional comments reviewer')
9. The analysis script can be run successfully.		I experienced some difficulties when following the README file instructions. See my comments below ('Additional comments reviewer')
10. All preprocessing steps are clearly described and traceable, especially when preprocessing code cannot be executed because raw data is not available.	I am not sure what exactly falls under preprocessing. Any troubles I ran into are described at 'Additional comments reviewer')	
11. The analyses and results reported in the manuscript can be found back in the analysis script with labels according to where they appear in the manuscript.		Some points to improve clarity: - I think you can save the researcher who will reproduce your analyses and results time by indicating more clear which code / results in the PDF files relates to which

Statement	Yes	No
		text and results in your manuscript. - At the end of the README file an overview of 'Tables/figures and corresponding files' is provided. As 'Motivating factors.png' and 'Demotivating factors.png' are also provided in this overview, does this mean that these results can also be reproduced with your instructions / code? If so, I do not understand how I can reproduce these results. If not, I would make it more clear that this is not part of the results that can be reproduced.
12. All results reported in the manuscript accurately correspond to the output produced by the analysis script.		- In the manuscript, paragraph 3.1 'H1: Motivational impact of tailored vs. generic messages' the HDPI ranges are reported. If I understand it correctly, this corresponds the data of PDF file 'Statistical-Analysis', page 2 'Posterior Probability', values provided in row 'b'. However, the mean, sd and lower range match the data reported in the manuscript, but the upper range differs (1.27 in the PDF results and 2.37 reported in the manuscript). - In the manuscript you also report Cohen's Kappa values to indicate the agreement between coders. If I am correct, these results cannot be reproduced with the instructions / code. Is this a problem?

III. Additional comments by reviewer

Please state any additional things you noticed in reviewing the dataset or possible points of improvement for the reviewer.

- For someone who does not have a background in coding / data science, it would be useful to provide a bit more guidance in writing the codes and reproducing the results. For instance, something simple like mentioning that you should open your Windows command prompt by typing 'CMD' in your Windows search field, and explaining that <…> in the code should be filled in and after filling it in <> should be removed. So when running the code to run the Docker container, it can help to explain that <PATH_TO_CURRENT_FOLDER> should be replaced by the folder path in which the README file is located and that it is not allowed to use capital letters (so that the capital letters need to be replaced lower case).
- In addition, because both the 'Dataset_paper' folder and in the 'Analysis' folder contain a (similar) README file, it was not clear to me what the correct <PATH_TO_CURRENT_FOLDER> was (the one in the 'Dataset_paper' folder or the 'Analysis' folder). As I first filled in the path to the 'Dataset_paper' folder I ran into an error.
- Navigating to https://localhost:8787 only worked if I navigated to localhost:8787 (without https://).
- When I unzipped the 'Dataset_paper' folder, at first I was not sure which folder to use, as the main folder consists of the 'Dataset_paper' folder and the '_MACOSX' folder, and both contain README files.
- It was not clear to me if I needed to follow the steps under 'Knitting from the command line' AND 'Knitting from RStudio', or that I could chose to follow the steps for the command line OR RStudio.

IV. Review log

Round	Date	Paper Status	Checklist Items	Signature Reviewer	Signature Researcher
	21-07-2022	Adjusting manuscript after feedback co-authors		MV 	Ramya Praneetha Ghantasala