

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 is 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: | Setting Physical Activity Goals with a Virtual Coach: Vicarious Experiences, Personalization and Acceptance |
| Name(s) of researcher(s): | Nele Albers, Beyza Hizli, Bouke L. Scheltinga, Eline Meijer and Willem-Paul Brinkman |
| Name of the reviewer: | Bouke L. Scheltinga |
| 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? | x | |
| 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 dataset, all column names and possible cell values are explained. | | |
| 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. | x | |
| 4. All data has been anonymized in accordance to promises made in the Data Management Plan. | | |
| 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). | | |
| 6. All required input files for the analysis are available in the dataset. | | |

Commented [BS1]: I miss a description of headers of the data files. Some of them are self-explaining. But for example, what do the values mean in \Bayesian_Analyses\Data\df_acceptance.csv

Commented [BS2]: In \Prediction_Model, the required input files are missing, and the output file is incomplete (should add centers.csv, clusters.csv and \Figures\km_clustered.png)

Commented [BS3R2]: In \Participant_Characteristics_and_Cohens_Kappa the required input files and generated output files are missing.

Commented [BS4R2]: Also for \Bayesian_Comparison_Conditions\comparison.Rmd the required input and output files are missing

Commented [BS5R2]: Also for \Bayesian_Analyses\analysis.Rmd

Commented [BS6R2]: Also for Prediction_Model\collaborative_filtering.ipynb and Prediction_Model\preprocess_for_model.ipynb

Commented [BS7]: the files "data/prolific_data_AB.xlsx," "data/post_questionnaire_C.csv," and "data/questionnaire_data_AB.xlsx." need to be downloaded from another repository.

Commented [BS8R7]: Also for "thematic_analysis/coder1.csv" and "thematic_analysis/coder2.csv"

| Statement | Yes | No |
|--|-----|----|
| 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). | | |
| 8. The analysis script is clean and comprehensible in the sense that: <ul style="list-style-type: none"> 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. | X | |
| 9. The analysis script can be run successfully. | | |
| 10. All preprocessing steps are clearly described and traceable, especially when preprocessing code cannot be executed because raw data is not available. | X | |
| 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. | | |
| 12. All results reported in the manuscript accurately correspond to the output produced by the analysis script. | | |

Commented [BS9]: - In \Prediction_Model, the model.pdf is not created

Commented [BS10]: If on the C drive: yes. Maybe add this to the README.txt

Commented [BS11R10]:

Commented [BS12R10]: Error with running \Bayesian_Analyses\analysis.Rmd:

Error in library(BayesianFirstAid) : there is no package called 'BayesianFirstAid'

Commented [BS13]: For Participant_Characteristics_and_Cohens_Kappa\thematic_analysis.i pynb and participant_characteristics_example_people.ipynb the exact location in the paper is not given (table 4 and figure 6)

Commented [BS14R13]: For \Bayesian_Comparison_Conditions\comparison.Rmd, it is not stated where the results of the code can be found back in the paper.

Commented [BS15]: Data in Table 2 is different. E.g. Number of people in generic condition: 19 Number of people in personalized condition: 18

Instead of 20 and 19 as seen in the paper.

All data is different for table 2.

Commented [BS16R15]: Outcomes of Bayesian_Comparison_Conditions\comparison.Rmd are different compared to the given .pdf, also if the .pdf is recreated

III. Additional comments by reviewer

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

Nice work, most of the files are doing what's needed and the code is clean to read.

The files "prolific_data_AB.xlsx", "post_questionnaire_C.csv", "questionnaire_data_AB.xlsx", "thematic_analysis/coder1.csv" and "thematic_analysis/coder2.csv" are not in the data folders already, but need to be downloaded separately. It might be nicer to include these files in this repository as well.

Also it is seen that everything is executed in a docker environment except for the preprocess_data.py. Don't know whether this is desired, maybe for consistency also put this into a docker environment.

IV. Review log

| Round | Date | Paper Status | Checklist Items | Signature Reviewer | Signature Researcher |
|-------|------------|---|-----------------|--------------------|----------------------|
| 1 | 21-09-2022 | First draft sent to co-authors for feedback | 1-3, 5-12 | B.L. Scheltinga | Nele Albers |
| | | | | | |
| | | | | | |
| | | | | | |