Reproducibility of Research

Lucy Bowes, Juuso Repo (MSD)

Exploring AI-human triangulation for research reproducibility

Lucy Bowes & Juuso Repo

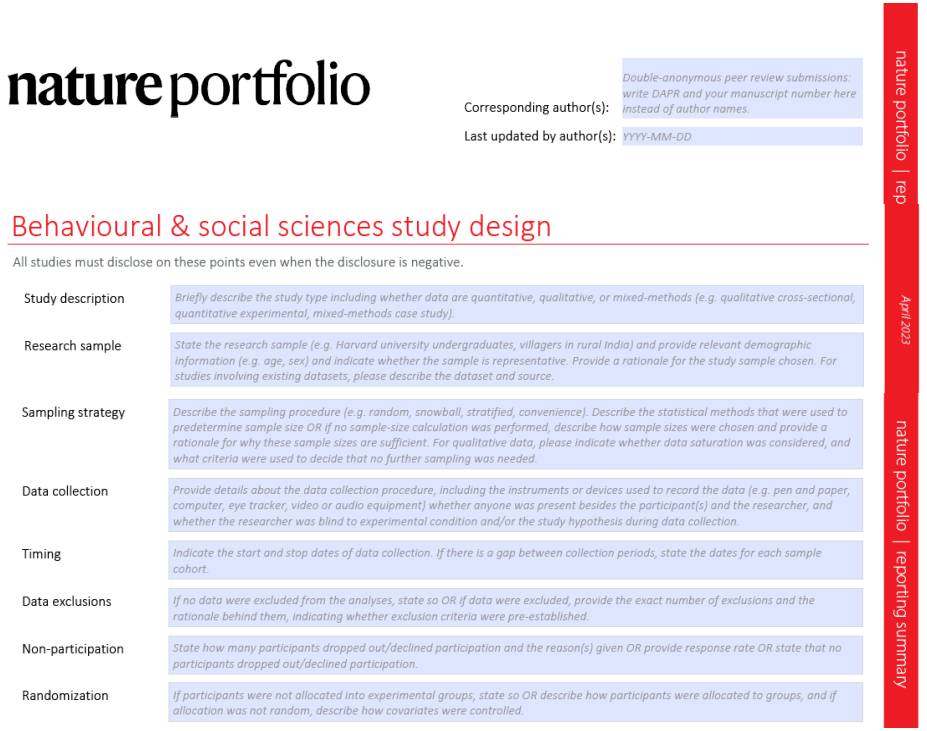
Team: Laurence Hunt, Nicholas Yeung, Xavier Laurent, Haseeb Ahmad, Mirela Zaneva, Tawa Edwards, Dominic Alonzi

Opening the research process - not only the outcomes

Procedural reproducibility

If I gave my paper to a colleague, would the text and supplementary materials enable them to replicate the methods, collect comparable data, process it similarly, and draw the same conclusions?

Reporting guidelines



nature portfolio

Corresponding author(s): *Double-anonymous peer review submissions: write DAPR and your manuscript number here instead of author names.*

Last updated by author(s): *YYYY-MM-DD*

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

nature portfolio | rep

April 2023

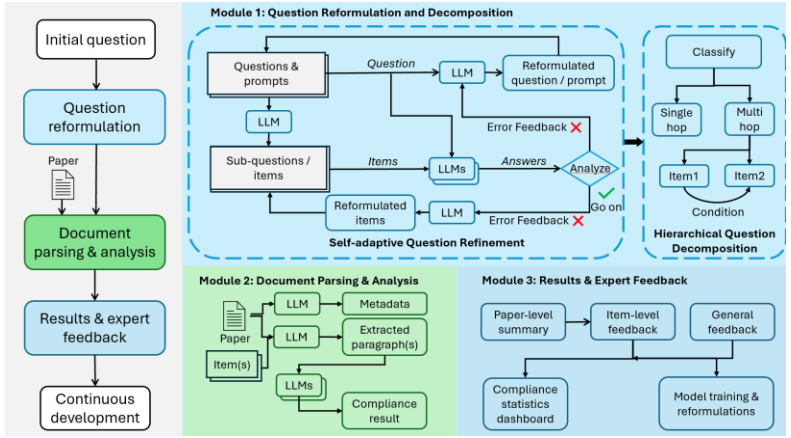
nature portfolio | reporting summary

Our example based on MDAR standards: https://osf.io/2k3va/

ReproAI Analyser

1. Insert PDF or DOI
2. Select checklist
3. Insert email

4. Start analysis



5. View results and the PDF annotated

Statistical analyses

Our sample calculation was based on previous research (reduction from 18% to 14% in rates of victimization) (Skivington et al., 2021), and a similar baseline victimization rate of 18% from a small UK based pre-post study (Edwards & McIntosh, 2019). Assuming 111 students in Years 3 to 5, an intracluster correlation coefficient (ICC) of 0.02, and allowing for one school dropout per arm, 10% student dropout due to either opt-out or loss to follow-up, an 18% rate of victimization, and a relative reduction of 22%, a trial involving 116 schools (58 per arm) would provide 90% power at a 5% significance level (a total of 12 828 students).

All analyses were intention to treat without imputation (a complete case analysis restricting to pupils with responses at both baseline and follow-up), with outcomes compared between KiVa and UP groups using three-level regression models (allowing for clustering between students within school, and between schools within sites). Analyses controlled for school level stratification variables (school size, proportion of students eligible for FSM), key student characteristics (age, sex) as well as baseline outcome measures (where collected).

For binary outcomes a logistic model was used, and the result presented as adjusted odds ratios (ORs) comparing the odds of an event in KiVa schools compared with UP schools. For continuous outcomes, we fitted a linear-regression model and presented results as difference in adjusted means (KiVa minus UP). Multilevel ordinal logistic regression model was used to compare TSDQ scales. Due to skewness in the TSDQ scales, data were categorized according to the clinical cut-offs (normal, borderline, abnormal). Box-cox transformations were applied to skewed data when necessary and Glass's delta standardized effect size calculated as the difference in means (KiVa – UP) divided by the standard

RESULTS

1.6 Data exclusions Partial

The manuscript partially complies with the requirements for reporting on data exclusions, mentioning primary and secondary outcomes and analysis methods but lacking specific details on exclusion criteria and handling of missing data.

1.6.2 Exclusion reasons Partial

If data were excluded, does the paper provide the exact number of exclusions and the rationale behind them?

The manuscript mentions the handling of missing data and the intention to treat analysis, which implies that data were excluded. However, it does not provide the exact number of exclusions or a detailed rationale behind them.

Key sentences: 150 151

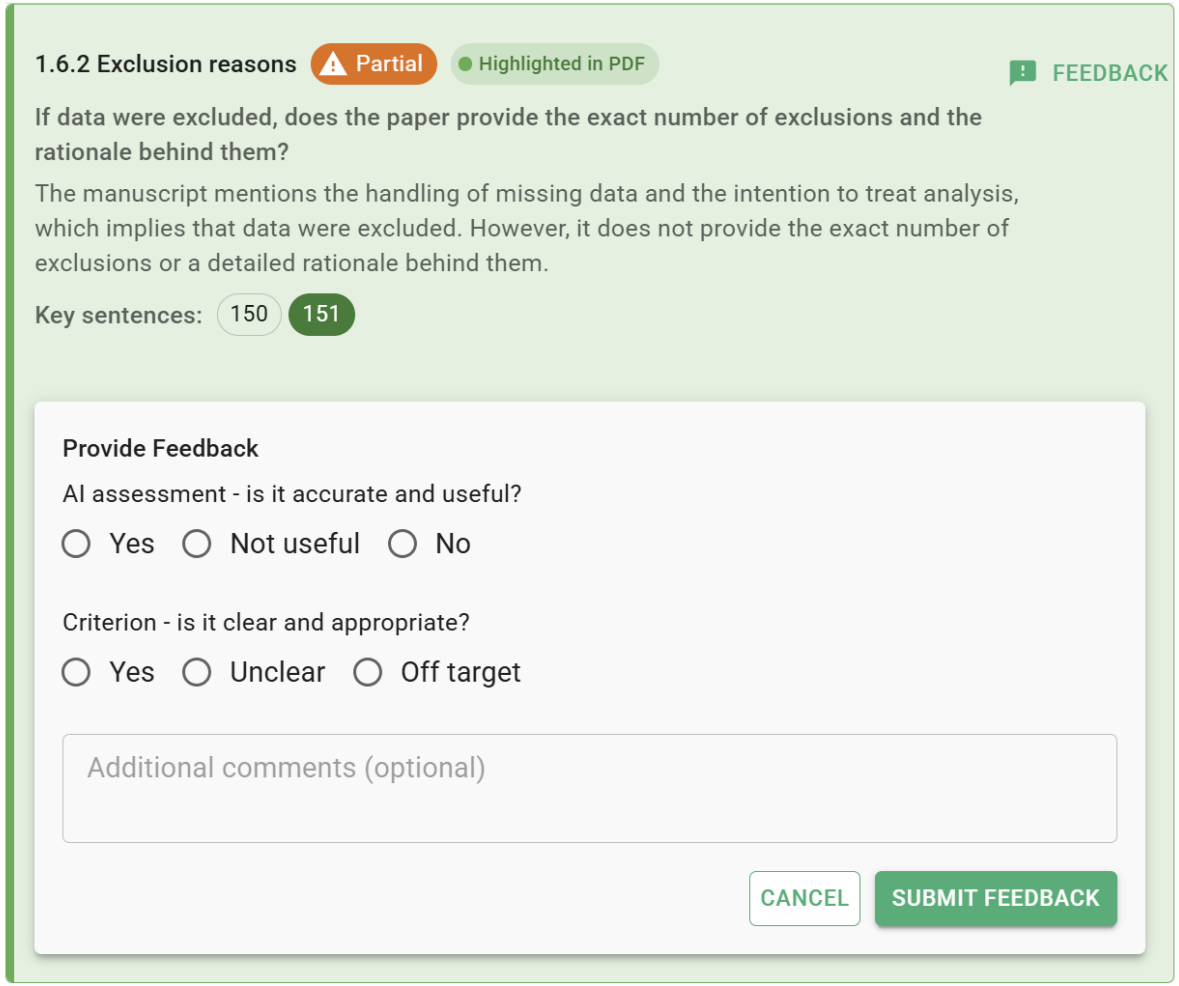
1.6.3 Exclusion criteria No

Does the paper indicate whether exclusion criteria were pre-established?

The manuscript does not indicate whether exclusion criteria were pre-established.

User validation & feedback

- User feedback gathered on item-level
- Both on AI-analysis and on criterion
- Stored with context: discipline and study design
- Future feature: What was the reason for not complying with this item?



Potential uses

1. **Students**: interactive learning about AI & open science
2. **Authors**: pre-submission support
3. **Publishers**: paper review, enhancing guidelines
4. **Research**: adherence and contextual variation