



University
of Glasgow



Evaluating Psychology Students' Data Skills For Reproducible Research

Dale Barr, **James Bartlett**, Heather Cleland Woods, Gaby Mahrholz, Ashley Robertson, Holly Scott, Wilhelmiina Toivo

james.bartlett@glasgow.ac.uk
[bartlettje.bsky.social](https://bsky.app/profile/bartlettje.bsky.social)

#PSY
TEACHR

Background

- Research methods challenges focus on itemising an APA-formatted report (e.g., Gurung & Stoa, 2020) or decision trees of statistical tests (e.g., Allen et al., 2021) but preparing data for analysis is often overlooked.
- Preparing raw data requires computational thinking to reverse-engineer the smaller steps required to reach an analysis end goal (Sundin et al., 2021).
- Thibault et al. (2024) helped update the 2024 BPS programme standards, where students should now learn how to “**clean and manage data**”.
- However, we are not aware of any studies exploring psychology students’ ability to clean/process data as part of the analysis process.

Research questions

Exploratory mixed methods study given the lack of previous research with comparable designs:

- Is there a difference in how psychology students can analyse a processed vs unprocessed data set?
- What elements of the data analysis process do students find easy or challenging?
- What sources of support do students use in the process of data analysis?

Method: Participants

- Two key recruitment strategies from July 2024 to July 2025:
 1. Online advert circulated via social media and university contacts.
 2. Our student RAs ran a challenge/employability event.
- 200 people accessed the study...
- 132 started the study...
- 53 completed...
- 42 met all inclusion criteria and we paid them £25.
- Modal demographics were UoGlasgow (N = 16), undergraduates (N = 22), who used R (N = 24) to complete the study.

Method: Materials

- Participants analysed data from a (simulated) experiment on the Simon task (congruent and incongruent trials) where “participants” completed it in a distraction or no distraction group.
- Everyone received the same instructions to prepare and analyse the data to answer the research question: “Is there a difference in the Simon effect under a distraction compared to no distraction?”
- We randomly allocated each participant to receive one of two versions of the data set, and quantitative data included summary statistics they calculated.

Data Set 1: Processed

- In the processed group (N = 25), there was one .csv file containing participant-level data with one pre-processing step to calculate the Simon effect:

id	distraction_group	congruent	incongruent
S01	distraction	562.2	982.7
S02	nodistraction	576.7	692.5
S03	nodistraction	440.0	606.5
S04	distraction	477.1	597.4

Data Set 2: Unprocessed

- In the unprocessed group (N = 17), there were two .csv files containing participant and trial-level data, requiring several pre-processing steps prior to analysis:

id	RT	simon_task_condition	trial_number
S01	1124	congruent	1
S01	579	congruent	2
S01	679	congruent	3
S01	457	congruent	4

id	distraction_group
S01	distraction
S02	nodistraction
S03	nodistraction
S04	distraction

Method: Free-text responses

For each step in the analysis process...

1. Planning and/or Data Wrangling.
2. Computing descriptive statistics and performing the statistical test.
3. Interpreting the output.

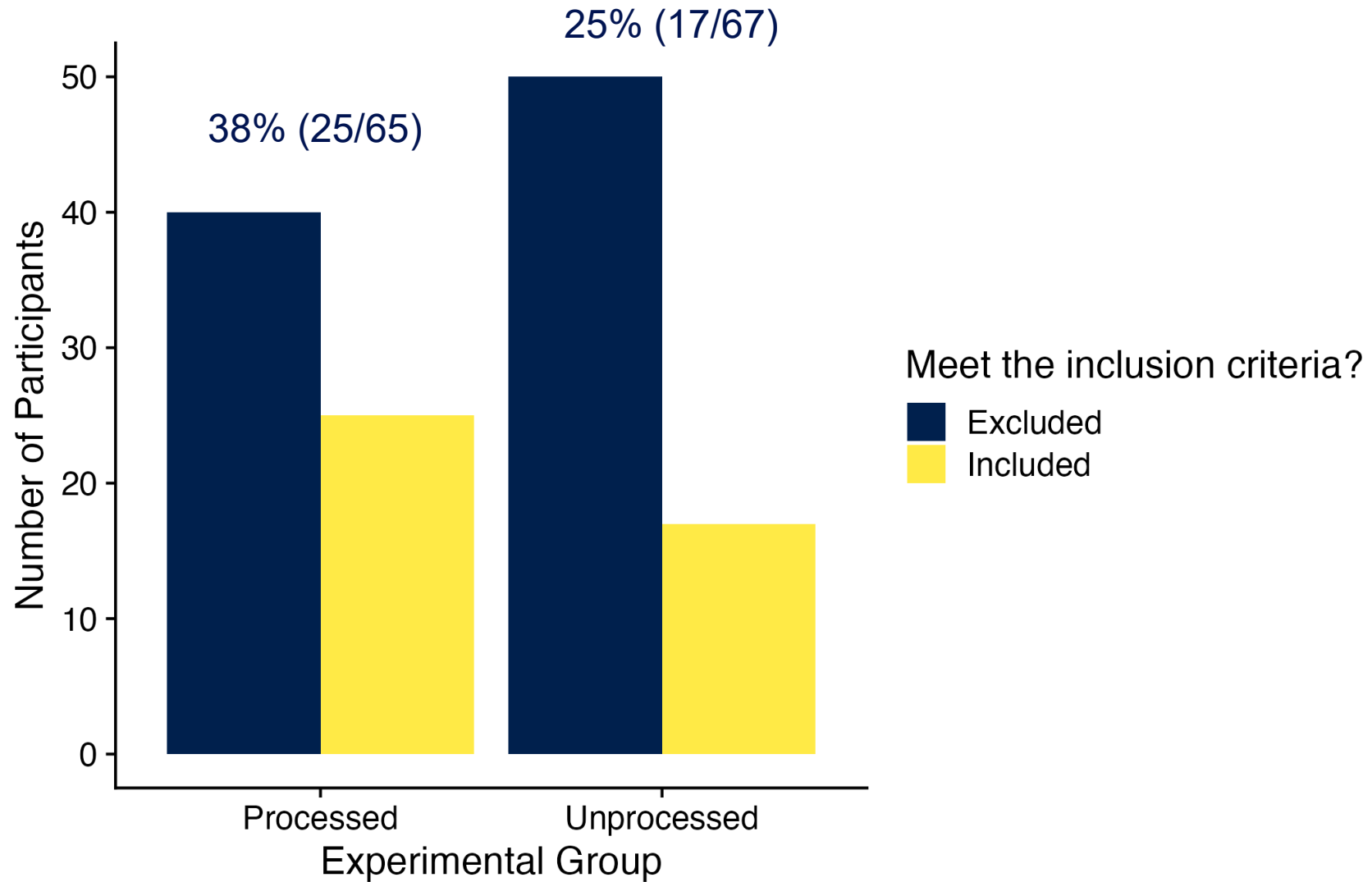
...we asked participants to reflect on their experiences:

1. Thinking about [1/2/3], how did you find this part of the process? What exactly did you find easy or challenging?
2. For anything that you found challenging, why was this?
3. For anything that you found challenging, how did you overcome this?

One final reflection question: Overall, how did you feel about doing this task?



Quantitative Results: Attrition



Qualitative results: Content analysis

- After finalising the coding scheme, two members coded the data independently, with inter-rater agreement ranging between 76% and 100% per category ($M = 95.8\%$, $SD = 0.05\%$).
- There were 14 broad categories for the nine process questions (e.g., using resources for information, worry about getting it right, uncertainty about how to complete the task).
- There were 7 broad categories for the final reflection question (e.g., the outcome of the task was rewarding, the process of the task was rewarding, self-reflection).

Qualitative results: Notable categories

Worry about getting it right (coded in 33/79% of participants)

- “The most challenging part was the data wrangling, especially since I hadn't done it in a few months. I struggled with the steps to correctly process and organise the data...I think I am also so used to having a step-by-step as a part of our coding tests etc...”

Implementing academic rigour (coded in 14/33% of participants)

- “I think I found this challenging because I so rarely come across non-significant results, which makes me feel like I did something wrong.”

The process of the task was rewarding (coded in 13/31% of participants)

- “I felt accomplished whenever I worked out a solution to a step that I had been stuck at, and I feel confident that I can get around problems by using the internet, as long as I know the basics and the sort of direction to head in...”

Discussion

- Like identifying a statistical test (Allen et al., 2016), participants reported that they struggled to identify the wrangling steps they needed to reach their analysis goal without step-by-step instructions (Sundin et al., 2021).
- Teaching data possibly mirrors journal publication bias (Franco et al., 2014); participants assumed the non-significant model answer was a mistake.
 - No concrete figures, but we know textbooks misinterpret NHST (Cassidy et al., 2019).
- Students do not automatically hate research methods and statistics (Counsell & Cribbie, 2020); participants found the process and outcome of the task rewarding.

Teaching implications

- Provide students with plenty of scaffolded data skills practice *before the dissertation*, working from the wrangling to analysis steps to generalise across research questions and datasets (McAleer et al., 2022).
- Include both significant and non-significant findings in teaching examples; reinforce how to appropriately interpret non-significant findings.
- Reflect on the software you teach students; question what skills you want them to develop, whether they can access it after graduation, and what employers want.

Limitations

- **Artificial scenario**
 - We suggested a two-hour limit for fair pay, but students reported feeling pressured like there were test conditions.
- **Selection bias**
 - Two layers of selection bias, these findings represent the best-case scenario.
- **Limited to cognitive task data**
 - Data skills are relatively universal but considering employability routes, students more likely to work with messy questionnaire data.

Thank you!

- A huge thank you to our student research assistants who transformed the data collection for the project: Heejung Choi, Sara Eftekhari, Rebecca Smith, Layla Clark, Ellie Fraser, Olivia Jackson, Pui Sang Leung, and Lois Weston-Camor.
- Pre-print coming soon!
- All questions welcome.
- Please come chat about data skills and research methods over the conference.

james.bartlett@glasgow.ac.uk
[bartlettje.bsky.social](https://bsky.app/profile/bartlettje.bsky.social)



References

- Allen, P. J., Dorozenko, K. P., & Roberts, L. D. (2016). Difficult Decisions: A Qualitative Exploration of the Statistical Decision Making Process from the Perspectives of Psychology Students and Academics. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00188>
- Allen, P. J., Fielding, J. L., East, E. C., Kay, R. H. S., Steele, C. S., & Breen, L. J. (2021). Using StatHand to train structural awareness and promote the development of statistic selection skills. *Scholarship of Teaching and Learning in Psychology*, 7(1), 53–67. (2019-75005-001). <https://doi.org/10.1037/stl0000177>
- Cassidy, S. A., Dimova, R., Giguère, B., Spence, J. R., & Stanley, D. J. (2019). Failing Grade: 89% of Introduction-to-Psychology Textbooks That Define or Explain Statistical Significance Do So Incorrectly. *Advances in Methods and Practices in Psychological Science*, 2(3), 233–239. <https://doi.org/10.1177/2515245919858072>
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*. <https://doi.org/10.1126/science.1255484>
- Gurung, R. A. R., & Stoa, R. (2020). A National Survey of Teaching and Learning Research Methods: Important Concepts and Faculty and Student Perspectives. *Teaching of Psychology*, 47(2), 111–120. <https://doi.org/10.1177/0098628320901374>
- McAleer, P., Stack, N., Woods, H., DeBruine, L., Paterson, H., Nordmann, E., Kuepper-Tetzel, C. E., & Barr, D. J. (2022). *Embedding data skills in research methods education: Preparing students for reproducible research*. PsyArXiv. <https://doi.org/10.31234/osf.io/hq68s>
- Sundin, L., Sakr, N., Leinonen, J., Aly, S., & Cutts, Q. (2021). Visual recipes for slicing and dicing data: Teaching data wrangling using subgoal graphics. *Proceedings of the 21st Koli Calling International Conference on Computing Education Research, Koli Calling '21*, 1–10. <https://doi.org/10.1145/3488042.3488063>
- Thibault, R. T., Bailey-Rodriguez, D., Bartlett, J. E., Blazey, P., Green, R. J., Pownall, M., & Munafò, M. R. (2024). A Delphi study to strengthen research-methods training in undergraduate psychology programs. *Advances in Methods and Practices in Psychological Science*, 7(1). <https://doi.org/10.1177/25152459231213808>

Method: Instructions

...You can calculate the Simon effect by subtracting the mean RT to congruent trials from the mean RT to incongruent trials for each participant. Positive effects indicate participants respond faster on average to congruent trials, whereas negative effects indicate participants respond faster on average to incongruent trials.

We would like you to analyse the data to answer the following research question: Is there a difference in the Simon effect under a distraction compared to no distraction? We would like you to:

1. For each participant, calculate the mean response time (RT) for congruent trials and the mean RT for incongruent trials.
2. Create a new variable for the Simon effect by subtracting the mean RT to congruent trials from the mean RT to incongruent trials.
3. Calculate the mean and standard deviation (to two decimal places) for the new Simon effect variable for the distraction and no distraction groups, in addition to the sample size of each group.
4. Apply a statistical test you consider most appropriate to address the research question “Is there a difference in the Simon effect under a distraction compared to no distraction?”

Method: Procedure

1. Information sheet and consent form
2. Statistics confidence pre-task (0 - 100)
3. Complete the analysis task using the easy or hard Simon task data set
4. Report the mean, SD, and sample size for each group in their analysis
5. Complete the free-text reflection questions
6. Demographics
7. Statistics confidence post-task (0 - 100)

Demographics

Variable	N (%)
<i>Mean age</i>	24.57 (7.1)
<i>Gender</i>	
Female	29 (69.05%)
Male	10 (23.81%)
Non-binary	3 (7.14%)
<i>University</i>	
University of Glasgow	16 (38.1%)
University of Southampton	11 (26.19%)
Other	5 (11.9%)
King's College London	4 (9.52%)
University of Edinburgh	4 (9.52%)
Durham University	2 (4.76%)
<i>Degree</i>	
Undergraduate student.	22 (52.38%)
Postgraduate student.	17 (40.48%)
Graduated.	3 (7.14%)

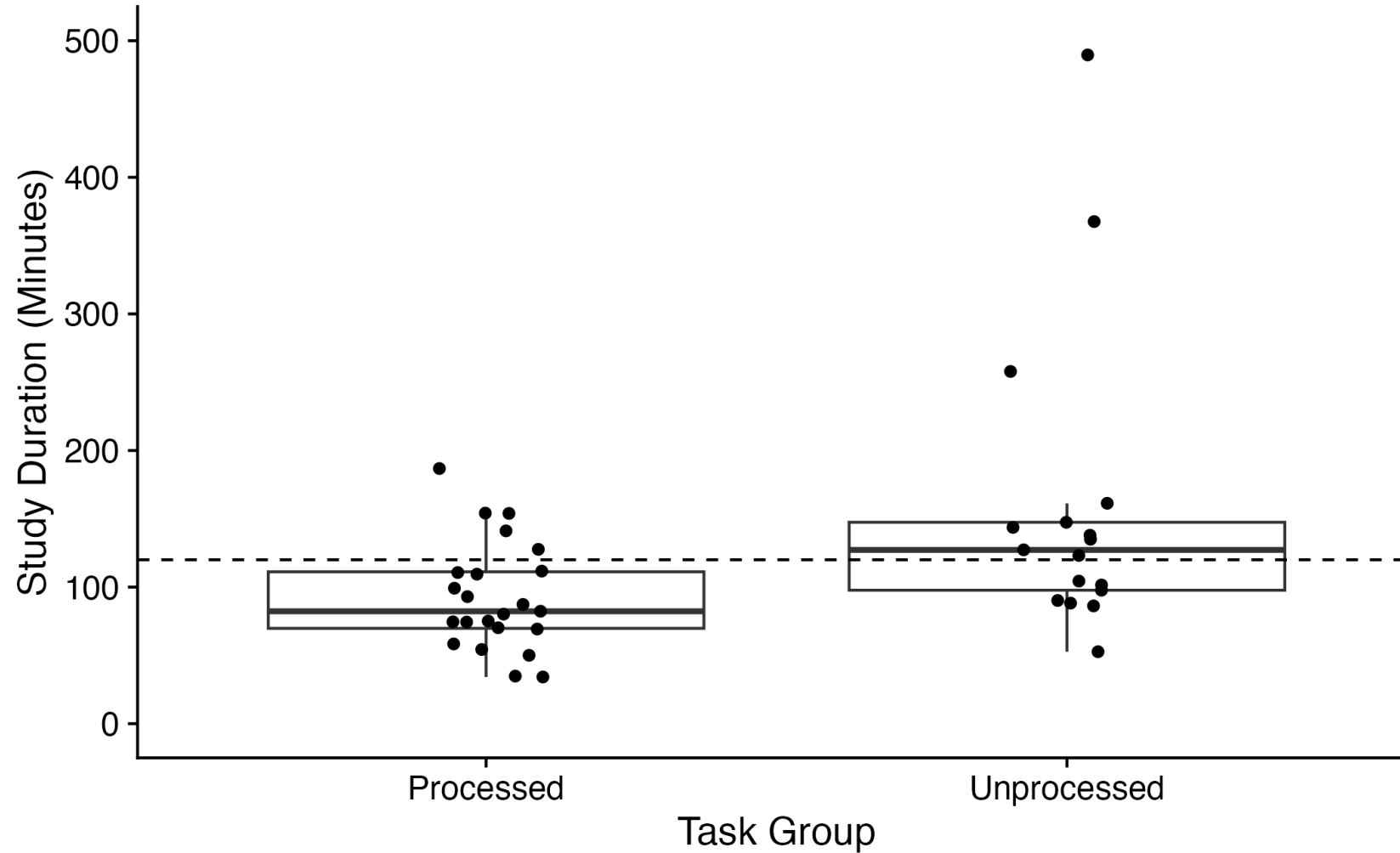
Note: Values do not sum to 42/100% as participants could select multiple.



Software and Resources

Variable	N (%)
<i>Software</i>	
R/RStudio	24 (57.14%)
SPSS	10 (23.81%)
Excel	9 (21.43%)
jamovi	2 (4.76%)
JASP	1 (2.38%)
Python	1 (2.38%)
<i>Resource</i>	
Google	17 (40.48%)
PsyTeachR books	15 (35.71%)
Specific software or libraries/packages within software	12 (28.57%)
Generative AI	10 (23.81%)
Using/adapting previous code from courses	8 (19.05%)
Previous knowledge	7 (16.67%)
Stack overflow/other coding websites	4 (9.52%)
Lecture slides/recordings	3 (7.14%)
Other resource	3 (7.14%)
Previous notes	3 (7.14%)
Did not use resources	2 (4.76%)
Textbooks	2 (4.76%)
Resources available within analysis system (e.g. Help functions, cheat sheets)	1 (2.38%)
Unspecified workbook	1 (2.38%)

Quantitative Results: Completion Time



Qualitative: Bonus quotes

- *“It was a bit finnickty at first since I am often so reliant on SPSS, and so not having access to it was challenging at first. Fortunately, the questions themselves seemed quite straightforward, and so I was able to calculate the data without much issue. However, I am not accustomed to arranging columns and rows in Excel, so I decided to highlight the experimental group in red and use that as guidance instead. It was frustrating that I didn't have access to any tools that could calculate the statistical test for me, though.”*
- *“I felt frustrated that I couldn't get past the first step of data wrangling. i thought I had learned these basic steps throughout the year and performed very well in all the data skills assignments I have completed, but I realised that is because the instructions made it clear what approach I should take at each step of the process. When left to make my own decisions on what steps to take in the data wrangling process I felt overwhelmed and confused. It's quite disappointing.”*
- *“The 'using RStudio' part was not challenging. Making sure what I was doing was right was more challenging. First, I tried computing descriptive statistics with both distraction and no-distraction values in the same table and it was not working. I spent more time here because I HAD to find a way. DEFEAT IS ONLY FOR THE WEAK. Finally, I accepted I was weak, separated the distraction and no-distraction values to two different tables and 'very easily' computed the descriptive statistics. The statistical test part was easier (but I don't know, maybe I 'easily' did the wrong test!).”*