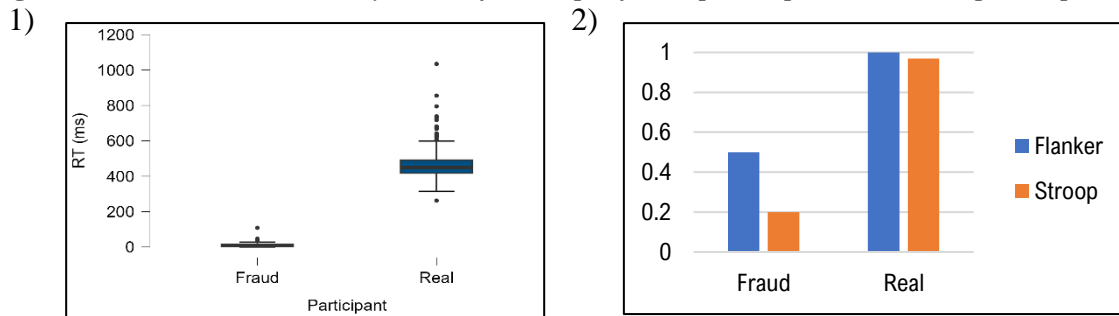


Beware the bots and button-mashers: small-scale linguistics studies are not immune to fraudulent and low-quality participations

Ariane Sénécal¹ and Laura Sabourin¹; ¹ERPLing Lab, University of Ottawa Department of Linguistics/Laboratoire PÉLing, Département de linguistique de l'Université d'Ottawa

Participant recruitment is an ongoing challenge for researchers. Solutions include sourcing from university student pools, offering compensation or thank you gifts when funds and ethics allow, using social media to spread word-of-mouth, or moving suitable studies online. This last option could not only grant access to a greater number, but possibly also a better diversity—age, level of education, economic status, etc.—of participants, resulting in a more representative and ecologically valid sample (Huber & Gajos, 2020). These past years, COVID-related public health measures contributed to a boom in online research (De Man et al., 2021), and may have led to an even greater diversity in potential participants' profiles (Arachar & Rand, 2021). However, a growing number of less attentive (*ibid.*) or outright dishonest respondents (Griffin et al., 2022; Lawrence et al., 2023; Pozzar et al., 2020) could threaten the quality and validity of online-gathered data. How can researchers reap the benefits of online research while protecting the integrity of their studies? Though this question has garnered considerable attention in disciplines with a longer history of recruitment via the Internet and where surveys are a common tool, this is less the case in domains where in-person lab studies are the norm, such as in our own field of psycholinguistics. We therefore examine this issue through the lens of a small-scale behavioural study aimed at multilinguals, which we designed and hosted through the Gorilla experiment builder platform (www.gorilla.sc; Anwyl-Irvine et al., 2020). Recruitment efforts for this endeavour were split between admissible students sourced from a pool and the general public; the former earned class credit, while the latter were offered a \$10 CAD gift card for their time. As part of outreach efforts, the principal investigator (PI) shared the recruitment poster via social media. Of the 15 participations collected within the 48 hours that followed, all were fraudulent. This was evident from the minimal, irrelevant and/or inconsistent information entered in their language background questionnaires, as well as their 'button-mashing'—hitting keys blindly (or always the same key) and rapidly throughout the tasks, yielding chance-level accuracy rates and far below cutoff reaction times (Welford, 1980). To prevent further harm, the study link was subsequently shared only through email or direct messaging with the PI. Bot-generated participation requests and some brazen fraudsters persisted; fortunately, these could be fielded before they compromised the data. The student pool yielded slightly better outcomes: still, of the 21 participants recruited through this stream, 7 were excluded due to the low quality of their contribution. Through this experience and a survey of the literature, we identify the red flags of deceitful participations, consider tactics to weed out impostors, and brainstorm ways to attract and retain quality participants for future online psycholinguistics studies.

Figure 1. Average flanker task reaction times of a sample fraud participant vs. a real participant
Figure 2. Global task accuracy rates of a sample fraud participant vs. a real participant



References

- Anwyl-Irvine, A.L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J.K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52, 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Arechar, A. A., & Rand, D. G. (2021). Turking in the time of COVID. *Behavior Research Methods*, 53(6), 2591–2595. <https://doi.org/10.3758/s13428-021-01588-4>
- De Man, J., Campbell, L., Tabana, H., & Wouters, E. (2021). The pandemic of online research in times of COVID-19. *BMJ Open*, 11(2), e043866–e043866. <https://doi.org/10.1136/bmjopen-2020-043866>
- Griffin, M., Martino, R. J., LoSchiavo, C., Comer-Carruthers, C., Krause, K. D., Stults, C. B., & Halkitis, P. N. (2022). Ensuring survey research data integrity in the era of internet bots. *Quality & Quantity*, 56(4), 2841–2852. <https://doi.org/10.1007/s11135-021-01252-1>
- Huber, B., & Gajos, K. Z. (2020). Conducting online virtual environment experiments with uncompensated, unsupervised samples. *PloS One*, 15(1), e0227629–e0227629. <https://doi.org/10.1371/journal.pone.0227629>
- Lawrence, P. R., Osborne, M. C., Sharma, D., Spratling, R., & Calamaro, C. J. (2023). Methodological Challenge: Addressing Bots in Online Research. *Journal of Pediatric Health Care*, 37(3), 328–332. <https://doi.org/10.1016/j.pedhc.2022.12.006>
- Pozzar, R., Hammer, M. J., Underhill-Blazey, M., Wright, A. A., Tulskey, J. A., Hong, F., Gundersen, D. A., & Berry, D. L. (2020). Threats of Bots and Other Bad Actors to Data Quality Following Research Participant Recruitment Through Social Media: Cross-Sectional Questionnaire. *Journal of Medical Internet Research*, 22(10), e23021–e23021. <https://doi.org/10.2196/23021>
- Welford, A. T. (Alan T., & Brebner, J. M. T. (John M. T.). (1980). *Reaction time*. Academic Press.