

## Individual differences in online research: Comparing lab-based and online administration of a psycholinguistic battery of linguistic and domain-general skills

Psycholinguists increasingly conduct studies online, outside of controlled laboratory environments. Data collected online have shown promising validity and consistency, and many group-level effects replicate online (e.g., Germine et al., 2012). Yet it is not necessarily the case that paradigms that produce sensible data at the group level will also be reliable in an individual difference paradigm, which requires precision at the participant level (Hedge et al., 2018).

We report performance on fifteen tasks from a psycholinguistic individual difference battery, including assessments of linguistic experience (antonym production, author recognition, idiom recognition, vocabulary, prescriptive grammar and spelling), speech production (maximal speech rate, Rapid Automatized Naming, verbal fluency), and domain-general skills (Corsi span, digit span, nonverbal IQ). 149 native Dutch speakers aged 18-30 took part in the study in the lab and 515 demographically similar participants completed the battery online.

The descriptive data from both settings (online and in-lab) is highly comparable, with most tasks showing <5% difference between settings (see Figure 1). We fit a Bayesian mixed model predicting a participant's performance by the interaction between task and setting and a random intercept by participant. Most tasks show a very small effect of setting and almost all credible intervals cross 0. Despite this, there are a few tasks worth noticing (Figure 2). In the Spelling task, participants online had better scores ( $\beta = 0.139 [-0.0605, 0.3347]$ ) compared to in the lab ( $\beta = -0.1426 [-0.3778, 0.0871]$ ). On the other hand, participants in-lab had better scores for nonverbal IQ ( $\beta = 0.0848 [-0.1507, 0.3183]$ ) compared to online ( $\beta = -0.0915 [-0.2925, 0.105]$ ).

We can thus conclude that the underlying distributions are similar across both settings, but do the tests measure the underlying cognitive constructs in the same way? To assess this, we fit a confirmatory factor analysis (Figure 3) and followed the procedure for establishing measurement invariance (Meredith, 1993). Similar to the results from the mixed models, the Spelling test and the Antonym Production test had different results by setting. In all other tests, we found no systematic differences in how the tests measure the underlying constructs based on whether they are conducted in the lab or online.

In both analyses, we found higher scores in the spelling task from online participants, suggesting that they might have sought external help (e.g. by looking up answers) when unsupervised. At the same time, we see a role of motivation (Christianson et al., 2022). Online participants may be less motivated for difficult tasks like nonverbal reasoning (Raven's Advanced Matrices) and under-stimulating tasks like Antonym Production, leading to lower scores. However, since nearly all credible intervals from the model cross 0 and measurement variance is established, our results indicate that there is no reason to assume that participants tested online will underperform compared to lab-based testing. Overall, we conclude that there is reason for optimism in the future of online research into individual differences.

Christianson, K., Dempsey, J., Tsiola, A., & Goldshtein, M. (2022). What if they're just not that into you (or your experiment)? On motivation and psycholinguistics. In *Psychology of Learning and Motivation* (Vol. 76, pp. 51–88). Elsevier. <https://doi.org/10.1016/bs.plm.2022.03.002>

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, 19(5), 847–857. <https://doi.org/10.3758/s13423-012-0296-9>

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3), 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>

