

Social Desirability Study 2 Methodology/Results

X

2023-10-07

Procedure

All ratings were conducted online using Qualtrics (2014), utilizing the Kuncel & Tellegen (2009) rating scale. Fifteen of the most decisive items from Study 1 were chosen, focusing on the three functional forms - linear, quasi, and asymmetric.

Due to the observed difficulty and length of the task in Study 1, we implemented carelessness items and additional questions to gauge carelessness and task difficulty. Because of the tedium and cognitive complexity of the requested task, we also computed consecutive non-differentiating responses as well as intra-individual response variability estimates (see, for example, Dunn et al., 2018; Marjanovic et al., 2015) via the careless R package (Yentes & Wilhelm, 2023) in R version 4.3.1 (R Core Team, 2023).

Inter-rater agreement analysis using Krippendorff's alpha was conducted to understand rater agreement between items for the same trait, as well as the three functional forms found in study 1, for the Kuncel and Tellegen (2009). Interrater agreement was calculated for these 15 items and 3 functional forms for study 1 and study 2.

An ordinal method was used when calculating Krippendorff's alpha using an ordinal method. Krippendorff's alpha can range from -1 to 1, where -1 indicates complete disagreement and 1 indicates complete agreement (Krippendorff, 2004).

Individual trait ratings ranged from 0.016 to 0.437 for study 2, and ranged from -0.022 to 0.607 for study 1. The functional form krippendorff alpha was calculated as the average of the 5 trait alphas for each functional form category. *should we do an aggregation or not?* Linear functional forms had the most interrater agreement (study 1: 0.448; study 2: 0.227), with asymmetric (study 1: 0.374; study 2: 0.263) and quasi (study 1: 0.069; study 2: 0.133) following.

should we calculate some sort of consistency/agreement for edwards and see if those are better

Krippendorff suggested that 0.667 was the lowest alpha to consider interrater agreement, while it is preferred to have interrater agreement that is greater than or equal to 0.800. Using these standards, no item had an acceptable interrater agreement alpha coefficient. This indicates that there was little interrater agreement among all ratings.

this is just a dump of thoughts - definitely needs to be reworded/made more coherent Although no ratings were beyond the threshold for interrater agreement. The linear ratings were more consistent than the asymmetric and quasi ratings. Also note that in study 2, the 3 negatively worded items were less consistent than positive items, bringing the linear average down. Even with 3 negatively worded items (Study 1 having a more positive range of 0.289 to 0.441, while study 2 had a range of 0.084 to 0.113), the mean ratings were still above the other functional form categories, suggesting that linear traits are cognitively easier to understand as well as agree upon.

The asymmetric ratings had one negatively worded item, affecting the overall average. The item "Worry about things" had very low agreement for both study 1 (0.176) as well as study 2 (0.016), which was the lowest rating. Based on interrater agreement, both linear and asymmetric items were easier to rate on 5 different levels compared to quasi items.

Quasi items were the lowest overall average and had some of the lowest agreement despite having no negatively worded items. In particular the item ‘Am Always Busy’ had the lowest ratings for any item out of all 15 for both study 1 (0.033) and study 2 (0.057). Based off the functional forms from both studies on this item, it was one of the most archetypal quasi functional form - peaking in the middle and having fairly even tail ends. The quasi ratings suggest that more neutral items are harder to agree upon and be inconsistent among raters.

Between both studies, linear functional forms stayed consistent (see graphs), while asymmetric functions being fairly consistent - some differences, with study 2 creating more linear like graphs in comparison to study 1. However quasi functional forms from study 1, changed quite a bit. The two items “Am always busy” and “Am a creature of habit” stayed in a quasi form, but flattened out. While the item “Always on the go” changed from a quasi form to a more asymmetric form, and the items ‘Enjoy wild flights of fantasy’ and “Relaxed most of the time” flipped into a more linear functional form. These changes between functional forms could be due to the sample, such as demographic differences (Midwest versus Northeast). *I think there is more to say on this, but need to think about it more My point in comparing these is that if the whole purpose of Kuncel/Tellegen is to gather more information/monotonicity, then is it really valid if its not really reproducible? I think the quasi functional forms really bring out validity and consistency issues for monotonicity. The asymmetric forms were honestly mostly linear anyway... definitely had some less linear like 2nd study worry about things, but that one also had really poor interrater reliability anyway... it seems like more neutral/least linear are posing issues which is basically the bread and butter of their paper in my opinion*

Inconsistently rating these items could pose issues if using for assessment. Unsure how to move on from here, but a thought.

##Dooley, K. (2017). Questionnaire Programming Language. Interrater Reliability Report. Retrieved July 5, 2017 from: <http://qpl.gao.gov/ca050404.htm>. ##Hayes, A.F. My Macros and Code for SPSS and SAS. Retrieved July 6, 2017 from: <http://afhayes.com/spss-sas-and-mplus-macros-and-code.html> ##Hayes, A. F. & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. Communication Methods and Measures 1,1:77-89. ##Jason Osborne. Best Practices in Quantitative Methods. Krippendorff, K. (2004). Content analysis: An introduction to its methodology. Thousand Oaks, California: Sage. ##Krippendorff, K. (2011). “Computing Krippendorff’s alpha-reliability.” Philadelphia: Annenberg School for Communication Departmental Papers. Retrieved July 6, 2011 from: http://repository.upenn.edu/cgi/viewcontent.cgi?article=1043&context=asc_papers