|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | Ian Hussey  E Ian.Hussey@UGent.be  T +32 470 396 842  Faculty of Psychology and Educational Sciences  Henri Dunantlaan 2  Gent 9000  Belgium |
| Christian S. Crandall, PhD, Editor in Chief  Personality and Social Psychology Bulletin | |  |
|  |  |  |
|  |  |  |  |
| date  2020-07-06 |  |  | |
|  |  |  |  |

Dear Prof Crandall,

Please find attached our manuscript “The Implicit Relational Assessment Procedure demonstrates poor internal consistency and test-retest reliability: A meta-analysis” for consideration.

Measures of implicit attitudes have become very common in contemporary psychology. The Implicit Association Test remains the most popular implicit measure, but many alternatives have been developed. Many of these, including the Implicit Relational Assessment Procedure (IRAP), promote their ability to capture not only the strength of association between concepts (e.g., “self” and “good”) but also relational information about how these concepts are linked (e.g., “I *am* good” vs “I *should be* good”; Remue et al. 2013). However, concerns over the psychometric reliability of implicit measures are long standing. Given recent revigorated awareness of the importance of measurement in psychology (e.g., Flake and Fried, 2018; Hussey & Hughes, 2020), we think it is important that this issue be properly addressed.

When we read Greenwald & Lai’s (2020) recent meta-analysis of the reliability of multiple implicit measures, we were surprised by their estimates of the IRAP’s internal consistency and test-retest reliability. While their estimates were rather low, our combined experience with the IRAP suggests that its true reliability is likely to be even lower than the published literature indicates. Indeed, our suspicious about its very poor psychometrics are a key reason why neither of us employ the measure in our own work any longer.

On this basis, we elected to combine all IRAP data we have collected over the past eight years in our two separate labs in order to conduct a file-drawer meta-analysis of all published and unpublished studies. Our collated dataset consists of mostly unpublished studies, and its sample size is roughly the same size as the total sample size included in Greenwald & Lai’s (2020) meta-analysis (i.e., our study doubles the *N* of the published IRAP literature). We have also made all this raw and processed data publicly available for reuse, along with all our R code for data processing and analysis. We applied recent guidelines for best practices in assessing both test-retest reliability and internal consistency for reaction-time based behavioural tasks (Parsons et al., 2019), and report the results of multiple approaches in order to consider the robustness of our estimates to experimenter degrees of freedom.

Our results demonstrate that the IRAP’s internal consistency is poor (α = .51, 95% CI [.46, .56]) and its test-retest reliability is unacceptably low (ICC = .20, 95% CI [.05, .34]). These estimates are significantly lower than those reported in Greenwald & Lai’s (2020) meta-analysis, perhaps due to publication bias in the published literature. Whereas Vahey et al.’s (2015) meta-analysis of the IRAP’s criterion validity concluded that it has “potential for clinical assessment”, we would strongly disagree. Our data suggests that the IRAP’s poor reliability makes it a poor candidate for both basic or applied use, and that researchers should be very cautious when choosing to use it in their work or interpret its results from published studies. We also include a number of moderator meta-analyses to assess task features that might be related to improved reliability. However, our results suggest that there are no simple fixes for the IRAP.

We hope that this manuscript will be of great interest to your readers given PSPB’s long history of publishing work on both measurement and the use of implicit measures.

Kind regards,

Ian Hussey