Manuscript Number: JCBS-D-22-00252     
  
The Implicit Relational Assessment Procedure is not very sensitive to the attitudes and learning histories it is used to assess  
  
Dear Dr. Hussey,  
  
Thank you for submitting your manuscript to the Journal of Contextual Behavioral Science. The AE, Dr. Rogge, received 2 reviews of your manuscript and provided his own feedback. Based on these reviews (found below) we have decided to ask you to revise and resubmit. We are asking that the revised manuscript be submitted by Mar 06, 2023.  
  
When you resubmit your manuscript please take care to note all comments along with how they were addressed or why they were not addressed in a separate Response to Reviewers file.  Also, please ensure that no identifying information is included in the Response to Reviewers, as this would unmask the reviewers and delay processing of your manuscript considerably. For example, do not sign the Response to Reviewers or provide it on letterhead.  
  
To submit your revised manuscript, please log in as an author at <https://www.editorialmanager.com/jcbs/>, and navigate to the "Submissions Needing Revision" folder.    
  
Thank you for the opportunity to consider your work. Please contact me, Michael Levin, if you have any concerns or questions about this decision, the revision process, or about JCBS in general.  
  
Regards,       
Michael Levin     
Editor-in-Chief    
Journal of Contextual Behavioral Science  
  
Associate Editor and Reviewer comments:  
  
Dear authors,  
  
Thank you for submitting your manuscript on the IRAP to JCBS and for your patience with the peer review process. I have now obtained comments from two reviewers and have read your manuscript carefully myself, and we all see notable strengths in your study and clear potential for your manuscript to impact the field. Both reviewers also noted a number of ways the manuscript could be improved with sets of helpful comments and suggestions. I am therefore recommending a revise and resubmit with major revisions at this point. In addition to the reviewer comments, I have two comments that I would ask you to address.  
  
1) PROVIDE MORE DETAILS ON THE METHOD: As someone who has never used an IRAP, I actually went to the [osf.io](http://osf.io/) listing for supplemental materials for the main dataset to find the details on each IRAP’s stimuli, task parameters, and responding rules (using the link for that information provided in your manuscript). Unfortunately, I was unable to find a document in the [osf.io](http://osf.io/) project with all of those details. This is problematic as it will leave many readers unable to critically evaluate the results being presented. Thus, although I would ask you to make changes in response to all of the comments and suggestions raised by reviewers, I would ask you to pay particular attention to those comments requesting additional details. I would also ask that you provide more information on the IRAPs included in the current analyses – possibly creating your own online supplemental material to document their various methods.  
  
2) PROVIDE MORE DISCUSSION OF SAMPLE SIZES AND POWER: My second concern is the sample sizes of the individual studies. Although a sample of 709 respondents represents a large sample, when split between 16 distinct IRAP studies, that leaves roughly 44 respondents per study – a somewhat concerningly small sample size. As an outsider to the IRAP literature, that alone made me wonder if the data you used were optimized for finding a common pattern but possibly underpowered to find distinct patterns for the various targets being assessed in the individual IRAPs. I would ask that you address this concern within your discussion – possibly as a limitation and/or future direction. Of course, if you feel that should not be a concern for readers, then I would ask you to explain to readers why having a bunch of notably small samples is not a threat to the conclusions being drawn.  
  
Please be sure to clearly and carefully explain the changes you made in response to each of the comments and suggestions within your response letter. I look forward to seeing your revised manuscript. Thank you again for your patience.  
  
Sincerely,  
  
Ron Rogge  
Associate Editor, JCBS  
  
  
REVIEWER 1 COMMENTS  
Thank for you the opportunity to review this interesting manuscript on the generic pattern of IRAP results and the implications for conducting future IRAP studies. In my opinion, this paper will have a significant impact within the substantive area of research using the IRAP to assess implicit cognitions, attitudes, and learning histories. My comments focus on the how to improve the presentation of the concept and results and in support of the authors stated recommendations to aid future research with the IRAP.  
  
The authors of this study compiled a relatively large (N=753) dataset of published and unpublished IRAP studies across two independent laboratories and make these data, and corresponding code, publicly available. This is a strength of the study, not only for its transparency but also because it can serve as a model for replication or further examination in case other investigators which to evaluate the presence of a generic IRAP effects within studies that exist inside and outside these two laboratories.  
  
1) The biggest concern I have is whether the generic IRAP pattern, in and of itself, is the result of the method/task or an indication of an actual bias in participant attitudes or learning histories. For example, if there was a true cultural bias for faster response times on "thin-positive-true" relative to any and all other combinations, shouldn't this be reflected as such in the IRAP? And wouldn't a test comparing IRAP scores from this trial type to a self-reported or other external measure of this potential cultural bias also be a valid test or check of this cultural bias (as the authors note in their Discussion)? The same goes for other category stimuli. If so, then it would be difficult to attribute variance to the method or task vs. a true bias in that the stimuli are already and inherently confounded before taking the IRAP in question. The authors reference studies by Finn et al (2016, 2018) that used non-evaluative stimuli, which is a critical test of this issue.  
They also present results from a small, nonsense word IRAP in Figure 2. I think more needs to be said about these studies to firm up the argument in the Introduction and Discussion that there is a generic IRAP pattern that exists in all IRAP studies regardless of the valenced attribute stimuli and response options used. This will be key for future research to tease apart. I also know the authors are attempting to disaggregate the within (trial type) and between (category) variance components in their ANOVA but they are doing so only on valenced attributes and response options.  
  
Another major implication of this work is that, in the context of a generic IRAP pattern existing in all IRAP studies, researchers should not be using inferential tests that rely on a null hypothesis of zero given that the generic IRAP pattern will inflate scores away from zero regardless of category domain. They also caution against within-person tests of differences between IRAP trial types given the generic IRAP pattern of results. The authors do a nice job of explaining these points in the manuscript and offer a template in Table 1 for what would be an appropriate analytic approach in IRAP research. They also provide helpful recommendations for how to model future IRAP results.  
  
2) I have some additional questions about scenarios a-c presented in Figure 4. For scenarios a & b, would the authors recommend an approach that residualizes the generic IRAP effect, like models detrending a time-series, prior to conducting planned inferential tests on domain-level variables? Seemingly, any remaining variance could be explained by parameters of interest to IRAP researchers without concerns raised by a generic pattern or trend in such models. For scenario c, the authors state that tests of group differences (e.g. treatment vs. control) for individual trial types do not rely on a zero point and therefore would not be invalidated by the generic pattern of the IRAP. Do the authors claim or assume that each individual trial type is independent from all others? Or, that say, trial type four, as used in the scenario, is dependent based on the ordering in which it occurs in the generic IRAP pattern (e.g. after trial types one through three)? If the independence  
assumption is made, would the authors claim as valid/recommend a model where the four trial types are tested as factors simultaneously or serially with an alpha correction in conjunction with group status as a moderator? I understand there could be many ways to appropriately model IRAP data given a generic IRAP pattern but specifying more basic, key assumptions in the stated recommendations (e.g. independence of trial types, removing artifacts/trends in the data before analysis) would allow others to plan valid models that are devoid of capitalizing on the generic IRAP pattern.  
  
3) The authors limit their available data to IRAP studies that used valenced attribute stimuli (e.g. positive or negative) with "true" and "false" response options. While this structure constitutes the majority of IRAP studies I am aware of, it should be mentioned in the Discussion that results from this study should be limited to studies with this same structure.  
  
4) In a table, please report the bivariate correlations for each of the D-scores obtained for the four IRAP trial types. Similarly, please report the means and standard deviations of the D-scores obtained from each of the four IRAP trial types.  
  
5) I know I'm fighting a losing battle here but "data" is a plural term that requires the use of a plural verb (e.g. "were"). There are several instances of this throughout the paper.  
  
6) The first reference to a figure is Figure 2. I recommend re-labeling the figures to match their ordered presentation in the manuscript or delete the sentence, "See Figure 2 for a list of all domains.", on page 5.  
  
  
REVIEWER 2 COMMENTS  
Thank you for inviting me to review the manuscript titled "The Implicit Relational Assessment Procedure is not very sensitive to the attitudes and learning histories it is used to assess". I believe the manuscript contributes to the literature and could have the effect of improving the statistical inferences made by researchers employing the implicit relational assessment procedure. The recommendations made in this respect are particularly valuable. Below are some comments that I think need to be addressed.  
  
1) Stimulus functions for words and pictures are acquired based on a history of learning. The implicit relational assessment procedure is sensitive to the functions of these stimuli as well as the relationships between them. Thus, the title is not entirely accurate. A more appropriate title is "The implicit relational assessment procedure is not sensitive to stimulus relations only".  
  
2) p 4: The inclusion criterion that "the IRAP must employ single-word valenced attribute category stimuli" restricts the analyzed data sets to those including a variable hypothesized as being involved in producing the stated effect (i.e., valenced stimuli that may have valence overlap with response options). This places limits the generalizability of any claim to IRAP employing similar stimulus sets that are presented in a similar manner. Are claims about generic patterns in IRAP data still warranted?  
  
3) p 6: I assume that the data analyzed in the manuscript are from IRAPs that matched the generic description given in the measures section. This section should be concluded with a statement to the procedural similarity of the IRAPs from which data are analyzed.  
  
4) pp. 8-10: This is very tidy.  
  
5) p 13: Change "as-yet-known" to "as-yet-unknown".  
  
6) p 13: Change "driven by category stimuli" to "relationships between category stimuli and evaluative stimuli".  
  
7) p 16-18: I believe many researchers employing the IRAP will find both this section and Table 1 to be informative and helpful.