**Editor’s comment**

Dear Dr. Toffalini,

I have now received the reviews of your revised paper and I am happy to inform you they are, in general, favourable and suggest that, subject to minor revisions, your paper is suitable for publication.

Please address the minor comments raised by reviewer 2.

*RESPONSE: Thank you for your positive feedback and the opportunity to improve our manuscript. We are happy to hear that the reviewers have generally appreciated our work. We have now revised the manuscript according to the comments of Reviewer 2: their points are addressed at page 3, page 5, and especially in the Conclusions at pages 27-28.*

I would like to receive your revision within 30 days. If you feel you are unable to meet this deadline, please let me know.

When you revise your manuscript, please highlight the changes you make in the manuscript by using the track changes mode in MS Word or by using bold or coloured text.

Please ensure you supply a blinded copy of your manuscript, and only a blinded response letter with no identifying information.

IMPORTANT: Before you resubmit your revised manuscript, please ensure that it still conforms to our requirements indicated in the instructions for authors (http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1464-066X). In particular, please pay attention to our word limits for manuscripts. Short Report papers must not exceed 3000 words, 20 references and a 150 word abstract; Empirical Papers and Review articles must not exceed a maximum 6000 words, 30 references and a 200 word abstract.

*RESPONSE: We tried to follow all of these indications in our revision process. However, to adequately address the Reviewer 2 points and add relevant references, our final paper is composed of 5455 words in the main text plus 682 words in the references (27 entries), totalling 6137 words.*

**Reviewer(s)' Comments to Author:**

**Reviewer: 1**

**Comments to the Author**

This tutorial is very elaborate. The examples are well-designed, and the r code illustration is easy to follow for a beginner.

*RESPONSE: We thank the reviewer for the positive consideration of our work.*

**Reviewer: 2**

**Comments to the Author**

The paper presented is, in my opinion, of great utility both theoretically and practically. From the theoretical point of view, it has the merit of showing the fundamental importance of verifying the assumptions involved in using specific clustering algorithms before their application, advocating in the psychological research community a deeper reflection (and urging caution) on the possible erroneous inferences that can be made from data that break even moderately the necessary assumptions. On the other hand, from a practical point of view, I have had the opportunity to try the R code and the Shiny app several times, and I think it has a particular potential from an educational and informational point of view: explaining in words the risks of underestimating the violation of statistical assumptions is not as useful as visualizing directly the probability of incurring false positives. Some aspects can be improved, such as great sensitivity in the format required for uploaded data or some rigidity in the manipulation of actual datasets, but the importance of a hands-on approach to the risks of carefree cluster analysis cannot, in my opinion, be underestimated.

*RESPONSE: We thank the reviewer for appreciating our work and highlighting the positive aspects in it.*

However, I believe the tone of the paper could be softened somewhat. While it is true that cluster analyses are sometimes conducted almost mechanically (e.g., without taking into account multivariate outliers), there are also times when these analyses are carried out with reasonable caution and robust methodology.

Indeed, the simulation studies presented in the paper introduce two different cluster detection procedures and algorithms, with the analyses performed separately. Authors who advocate for person-centered approaches particularly emphasize the need to confirm the convergence of results from different analyses before asserting the actual presence of clusters in the sample. For example, Asendorpf and colleagues (2001, 2015) propose a two-step process of cluster analysis, where initially a hierarchical cluster analysis using the Ward method is performed, selecting the most plausible solution not only based on the inspection of the dendrogram but also through comparison of multiple indices, following a stability check via bootstrapping. After this, by seeding the k-means procedure with the centroids obtained from the hierarchical partition deemed most plausible, it is recommended to verify the agreement between the cluster categorizations of the different algorithms. While it is true that each of these methods has its own specific assumptions, and that these are rarely satisfied, using all of these methods together can give researchers greater confidence in the accuracy of detecting clusters that actually exist. This approach significantly reduces the probability of error, especially since it is unlikely that all these different indices and procedures would converge to the same solution if the clusters were spurious.

*RESPONSE: We thank the reviewer for these suggestions. We have now mentioned these important recommendations and cited the relevant articles. We have slightly amended our manuscript in multiple points (particularly pages 3, 5, and more extensively at pages 27-28), but these suggestions are especially reported in the revised concluding remarks, where we now highlight the importance of adopting multiple criteria, methods, and indices before making decisions in applied settings.*

Interestingly, I see that the type 1 errors detected by the simulation often indicate an incorrect presence of two clusters, wrongly rejecting the actual one-cluster solution. However, again the relevant literature on the person-centered approach often identifies non-trivial solutions in this domain as having three, four, or five clusters (Asendorpf, 2015; Herzberg & Roth, 2006), leading researchers to often disregard two-cluster solutions, thereby likely reducing a portion of the false positives detected.

*RESPONSE: In fact, the prevalence of the two-cluster solution when incurring type I error may depend on the methods and statistical criteria being used. This is the case, for instance, using k-means and the silhouette index on variables presenting a homogeneous positive manifold. However, Figures 3-4 in the submitted manuscript suggest that false positives may be associated with varying numbers of detected clusters both when using k-means and GMM.*

I believe that the paper has a truly important role in promoting attention when doing cluster analysis in psychology, but I equally think it could benefit from the addition of few references to the psychological literature that considers the convergence of different procedures and analysis results as a way of increasing confidence in identifying qualitatively different groups of people.

*RESPONSE: In the revised manuscript, we have added four new relevant references in line with the reviewer remarks, and including those suggested by the reviewer.*