



Expanding the Intervention Potential of Tabletop Role-Playing Games

A recommendation by [Veli-Matti Karhulahti](#)  based on peer reviews by [Charlotte Pennington](#) , [Matti Vuorre](#)  and [Matúš Adamkovič](#)  of the STAGE 2 REPORT:

Joël Billieux, Loïs Fournier, Lucien Rochat, Iliyana Georgieva, Charlotte Eben, Marc Malmdorf Andersen, Daniel L. King, Olivier Simon, Yasser Khazaal, Andreas Lieberoth, Jonathan Bloch (2024) Can playing Dungeons and Dragons be good for you? A registered exploratory pilot program using offline Tabletop Role-Playing Games (TTRPGs) to mitigate social anxiety and reduce problematic involvement in multiplayer online video games. OSF, ver. 2, peer-reviewed and recommended by Peer Community in Registered Reports.

<https://osf.io/htbwa>

Submitted: 05 September 2024, Recommended: 21 January 2025

Cite this recommendation as:

Karhulahti, V.-M. (2025) Expanding the Intervention Potential of Tabletop Role-Playing Games. *Peer Community in Registered Reports*, 100889. [10.24072/pci.rr.100889](https://doi.org/10.24072/pci.rr.100889)

Published: 21 January 2025

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

The human capacity and need for play has been recognized as a central psychotherapeutic component for a long time (e.g. Winnicott 1971). More recently, experts have started developing specialized digital gameplay to be used as therapeutic tools and even utilizing existing videogames for similar purposes (see Ceranoglu 2010). On the other hand, the concerns about some players becoming overinvolved in videogames also led the World Health Organization to include “gaming disorder” in the 11th edition of the International Classification of Diseases, which echoes the nuance required to address human-technology relationships in general. In the present registered report, Billieux et al. (2024) make use of analog structured role-play in a new intervention aiming to mitigate social anxiety and problematic gaming patterns in online gamers. The authors carried out an exploratory pilot to test a 10-week protocol over three modules inspired by the well-known Dungeons & Dragons franchise. Through a multiple single-case design with a 3-month follow-up, the authors carried out the pilot with 20 participants (two dropped out), which enabled collecting valuable early data about the feasibility and potential of the intervention. Weekly psychometric assessments indicated that the intervention may support participants and it should continue to be studied in order to comprehensively evaluate its effectiveness. Feedback obtained from participants and other parts of the pilot project enabled identifying elements (e.g., difficulty adjustments), which can be given attention usefully in intervention development. Without question, the study by Billieux et al. (2024) is to be commended for highly transparent documentation

of the project, which allows future researchers and intervention developers to learn from the pilot beyond initial results. Among other relevant materials, the supplements provide a detailed case study and item-level psychometric reporting, which can help the development of technical solutions and new hypotheses. Taken together, the study by Billieux et al. (2024) paves the way for open science invention piloting in the field. The Stage 2 manuscript was evaluated by three experts (areas: experimental methods, psychopathology, gaming) via in-depth peer review across two rounds. Based on the authors' careful responses and revisions, the revised manuscript was judged to meet the Stage 2 criteria and was awarded a positive recommendation. **URL to the preregistered Stage 1 protocol:** <https://osf.io/h7qat>

Level of bias control achieved: Level 6. *No part of the data or evidence that was used to answer the research question was generated until after IPA.* **List of eligible PCI RR-friendly journals:**

- [Collabra: Psychology](#)
- [F1000Research](#)
- [Peer Community Journal](#)
- [PeerJ](#)
- [Royal Society Open Science](#)
- [Swiss Psychology Open](#)

References:

1. Billieux, J., Fournier, L., Rochat, L., Georgieva, I., Eben, C., Malmdorf Andersen, M., King, D., Simon, O., Khazaal, Y., Lieberoth, A. & Blocha, J. (2024) Can playing Dungeons and Dragons be good for you? A registered exploratory pilot program using offline Tabletop Role-Playing Games (TTRPGs) to mitigate social anxiety and reduce problematic involvement in multiplayer online videogames [Stage 2 RR]. Acceptance of Version 2 by Peer Community in Registered Reports. <https://osf.io/z32d5>
2. Ceranoglu, T. (2010). Video Games in Psychotherapy. *Review of General Psychology*, 14 (2). <https://doi.org/10.1037/a0019439>
3. Winnicott, D. (1971/2009). *Playing and Reality*. Routledge.

Reviews

Evaluation round #2

Reviewed by [Matti Vuorre](#) , 07 January 2025

The authors have adequately addressed my main concerns: The intent of the study and scope of results is now communicated with more clarity. The authors have also done a commendable job in clarifying what has changed and where—I was not previously aware of the supplementary letter. As such I would be happy to see this write-up recommended.

Respectfully signed
Matti Vuorre

Reviewed by [Matúš Adamkovič](#) , 20 January 2025

I've reviewed the revised manuscript and the accompanying letter to the reviewers. I'm satisfied with the changes the authors have made and have no further substantial input. Great job!

Matúš Adamkovič

Reviewed by [Charlotte Pennington](#) , 16 January 2025

I am satisfied that all of my comments have been addressed and the final Stage 2 report is more balanced and overall is very transparent. I approve of recommendation at this final stage.

Congratulations to the authors on a really strong and interesting research study!

Charlotte

Evaluation round #1

DOI or URL of the preprint: <https://osf.io/uwpy7>

Version of the preprint: 1

Authors' reply, 20 December 2024

[Download author's reply](#)

[Download tracked changes file](#)

Decision by [Veli-Matti Karhulahti](#) , posted 31 October 2024, validated 02 November 2024

Revision

Dear Joël Billieux and co-authors, Thank you for your detailed and informative Stage 2 manuscript. We were lucky to have all three reviewers return to assess the work. In general, the reviewers are impressed but also see a few issues that we must address. Below I spell out the key points that I'm expecting to be addressed, overlapping with some (but not all) reviewer feedback. 1. At Stage 1, we discussed three options: a) set criteria to confirm effectiveness, b) set criteria to confirm feasibility, c) not set any criteria and make the study fully exploratory. We agreed with the last option. As two reviewers remind us, this means that the study cannot, at any place, communicate the intervention to be effective and/or feasible; such claims always depend on selected criteria and should they be made, the criteria and their justifications need to be mutually agreed at Stage 1. Therefore, I kindly ask you to remove words and phrases that claim feasibility or effectiveness, including terms like "high" that are a matter of subjective evaluation. The study is very strong without such claims; a detailed exploration like this doesn't need to speculate about success. The readers can decide themselves how effective and/or feasible the reported results appear to them. I will carefully re-review all words/phrases on the next round and if debatable sections remain, we can discuss them separately one by one. [To further clarify: it's different to register analyses vs interpretations; this paper did the former but not the latter – and it's perfectly ok!]

2. As reviewers note, there's an unusually large number of changes to Stage 1 content. I tried to double check all of them, and my impression is that they don't meaningfully change planned content and are thus acceptable. However, in the future (especially if you have other recommenders), such changes might not be accepted (e.g., when the original comparison to "medical interventions" has been changed to "psychological treatments", I don't see this problematic but others might consider it reframing the work). Please consider this in your next RR.

3. A major related point concerns effect sizes. At Stage 2, the paper has introduced new benchmarks into a Stage 1 section; these new labels are linguistically very favourable (already 0.61 = large). Even though the footnote reports these benchmarks as unplanned, this has huge impact by framing results to a positive light post hoc, especially because readers of the final article cannot see the originals (without separately seeking Stage 1). Because PCI RR doesn't generally support any standardized benchmarks but rather encourages authors to think and justify how and why certain effects are meaningful (or meaningless), I propose a compromise to remove all benchmark labels and simply report the effects via some clear and neutral way. As the cited Vannest and Ninci say: "For those engaging in research, a benchmark is only useful when a body of literature has been meta-analyzed for the typical results... Benchmarks are not as useful as direct interpretations of the change in relationship to factors related to client needs, context, and prior intervention work."

4. Following the above quote, I'd personally like to see in more detail *what* the reported changes were, beyond averages. For example, Nuyens et al. (2023) have recently noted that IGDT-10 conflates core and peripheral criteria and is not well positioned to distinguish between problematic and non-problematic gaming. This isn't a problem for the pilot, but for a scientifically curious reader (like myself - I'm genuinely excited about the intervention mechanisms), what matters are the very symptoms that changed. As is well known in the field, e.g. #1 ("When you were not playing, how often have you fantasized about gaming, thought of previous gaming sessions, and/or anticipated the next game?") does not refer to any problems, and #8 ("Have you played to relieve a negative mood [for instance helplessness, guilt, or anxiety]?") can be fully met by just gaming to relax when tired. I find it problematic to speak of "recovery" in cases where such experiences are absent. On the other hand, reporting e.g. #9 (Have you risked or lost a significant relationship because of gaming?) would be highly valuable effectiveness information. A matrix with specific symptom distribution changes would be informative and help us understand where the effects come from and further specify the possible mechanisms (this is why we do pilots, after all).

5. The above could apply to other measures as well. For example: if the intervention positively contributes to reduced social gaming, it would make perfect sense that some participants increase their loneliness scores (losing contact to close online friends). Such hypothesis too might benefit from item-specific analysis of loneliness scores. As another example, I checked LSAS and found items like these to be rated for avoidance: "Participating in a small group activity", "Talking face to face with someone you don't know very well", "Meeting strangers", "Entering a room when others are already seated", "Speaking up at a meeting", etc. Factually, anyone who participated in the intervention didn't avoid such situations; the scores should drop in these items with any (control) activity that involves such engagements. For scientists interested in the mechanism, imo the most valuable information would be the scoring distribution, allowing us to learn about the possible transfer network beyond the obvious. I won't address other measures but would encourage taking a look into their operationalisations too.

6. I really like the supplement case report. It's great to see, on an individual level, what changed and didn't change, in what context. It's unfortunate that it wasn't possible to triangulate it with any qualitative data that could've helped better understand reporting. We've found it highly informative to add open questions after symptom items to understand participants' reasoning, this can be easy to implement for interventions too. For outliers, I'd encourage some conventional negative case analysis to better understand why they occurred. I see some of this has already been done informally (in discussion); of course it's also possible that there's nothing more to find, in which case you can report that explicitly.

7. Minor things:

a. One reviewer comments about the changes not being visible. You can ignore that comment; for me the tracked changes were clear and the managing board (Chris) specifically asked not to add extra marks to the main document.

b. At the end of Stage 1, we discussed separately about final technical edits after agreeing that no hypotheses or feasibility are tested. I notice that in a few places "hypothesis" remains, especially p. 9 and it has also been added to conclusions. Please remove these to protect readers from misinterpreting the work as hypothesis testing. I

hope you find my above notes and the authors' feedback sufficiently clear and useful; the study operates with numerous critical small details so I wouldn't be surprised if there are some misunderstandings/errors involved in our readings. As earlier, you can contact me directly to consult with specific questions, to streamline the next review round. However, I will invite all reviewers back for a second review and am naturally going to consider their views carefully as well. Again: this is a highly interesting pilot that, over some years of further development, could become useful in practice. The present work is just the very first step – let's ensure that we don't hype the results but focus on what was learned, what can be improved, and how to next proceed optimally toward formal effectiveness testing. Veli-Matti Karhulahti

Reviewed by [Matti Vuorre](#) , 28 October 2024

I have read the Stage 2 RR "Can playing Dungeons and Dragons be good for you?" and provide my observations and evaluation below.

Overall it seems the authors have conducted the study as described in the Stage 1, but I am less satisfied by some aspects described below.

One of my main comments regarding the Stage 1 manuscript was that the study was framed to provide both a feasibility test and a test of an "initial" effect. This framing was not improved upon in later Stage 1 submissions, and I find this Stage 2 report to suffer from the same problem. Calling this a "registered exploratory pilot" that tests an "initial effect" is in my view not justified because the concrete differences between a "pilot" and an "initial effect" from a non-pilot and non-initial effect are not sufficiently clear. The motivation to provide comprehensive hypothesis tests of effects, instead, is clear throughout, e.g. in the conclusion that this is "the first quantitative study of the therapeutical use of TTRPG that endorses all the canons of open science, from pre-registration of design and hypotheses to open data and material."

Moreover, none of the tests of feasibility were specified in advance: In a version of the Stage 1 manuscript ("Current study"), I read that "Against this background, the current study proposes an exploratory pilot experiment that aims to test the feasibility (e.g., number of dropouts, ability of the participants to understand and engage in a tabletop role-playing game, ability of the participants to complete regularly the online assessment)". This sentence does not appear in the Stage 2 report, which instead states that "In terms of feasibility, we were interested in how many sessions participants would miss and how many participants would drop out from the program entirely, the ability of the participants to complete the weekly online psychometric assessment, and the ability of the participants to understand and engage in a tabletop role-playing game as well as to succeed in the various objectives of increasing difficulty implemented in the TTRPG program." It is not clear to me if those are exactly the same aims? And, it is not specified what would e.g. count as a low dropout rate; yet 10% of participants dropped out with 10 out of 18 remaining participants completing all sessions, and this was decided to count as "low dropout". I don't necessarily disagree but what would have counted as high dropout? These issues seem to me to go against the idea of (pre-)registering a feasibility study.

Consequently, I am not able to fully evaluate to what extent the feasibility results support the feasibility of the study protocol. It seems fine, but since no cutoffs for "fine" were reported in advance I don't think it is fair to describe those analyses as (pre-)registered.

Similar issues apply to results regarding "initial" effects: For example, what proportions of very large, large, etc. effects would support each of the hypotheses? Authors state in the design table that "Our study is not testing a specific theory or model" and "We will not provide "general" interpretation (unless in the unlikely case where all participants present with the same pattern of results)." This is fine, but none of the results should then be described as tests of hypotheses regarding some (general) effect. It seems unfair that the interpretation of the presented results will inevitably be colored by the advertised registration of this study, even though it doesn't extend to these "hypothesis tests".

- Stage 1 read "Participants with missing data will not be omitted from the analyses unless the number of measurement points per phase is < 3, as three measurement points per phase is considered the minimal standard to reach in a single-case methodology (Tate et al., 2015)." but Stage 2 does not have this sentence

and instead refers to a three-week baseline measurement. Those do not seem the same but it is possible that I do not understand this correctly; a clarification for the deviation would be appreciated.

- The statement "In this sense, our study confirms the soundness of single-case analysis to test treatment efficacy, as a more traditional group approach would only have emphasized the global (i.e., group-based) small or small to medium effect on symptoms, masking the important heterogeneity in the response to intervention." is incorrect and should be removed. There is no statistical quantification of heterogeneity, but simply a casual comparison of significant to not significant. A "more traditional group approach" using e.g. multilevel models would allow such quantification.

- In the design table authors say "We will not provide "general" interpretation (unless in the unlikely case where all participants present with the same pattern of results)." Yet they draw conclusions such as "The current study demonstrated that a 10-week structured TTRPG-based intervention is feasible and effective in reducing symptoms in a sample of sub-clinical socially anxious gamers." Any general conclusions such as that should be removed as per authors pre-specified plan.

- I found the Stage 2 report evaluation cumbersome because authors did not submit the required (or I did not find) manuscript with highlighted changes from Stage 1. I can tell that some language has changed substantially. I have closely read the submission and it seems that these changes are not to e.g. theoretical rationale, but it is harder to tell lacking the document with highlighted changes.

- For example, the write-up of the data analytic strategy section has changed a lot from Stage 1 to Stage 2 rendering its evaluation unnecessarily difficult.

- Please include figures and tables where they belong in the manuscript. I will not review future submissions with floats at the end of the document.

Respectfully signed

Matti Vuorre

Reviewed by [Matúš Adamkovič](#) , 14 October 2024

Dear Authors,

I was very pleased to read this Stage 2 report. Honestly, this is perhaps the most detailed and in-depth pilot/feasibility study I've ever read. The level of transparency and reporting practices is outstanding. I compared the Stage 1 protocol with the Stage 2 report, and all the deviations I could detect (and even a few more) were disclosed. The reasons for these deviations were understandable, and none of them altered the study design in a way that would make me question adherence to the preregistered protocol or the validity of the piloting procedure.

I have some minor suggestions to improve the paper (especially the discussion section), but frankly, I wouldn't mind if the paper was published in its current version. My suggestions are as follows:

- Please consider moving the footnotes regarding the deviations into one paragraph in the main text (perhaps merging them with the current paragraph on study limitations, as it already covers some of the deviations), so readers can easily see what was changed and why.

- The program had a very low attrition rate, which is very promising. The authors attribute the high engagement to (the appeal of) TTRPGs. However, I would appreciate a short discussion on what other factors (e.g., rapport with the game master, group dynamics, conscientiousness, etc.) might have contributed to the high engagement, and how these could be investigated and potentially accounted for in future studies.

- I would also suggest addressing an implementation issue—in my reading, the success of the program might have largely depended on having a highly skilled game master. Could the authors reflect on this and suggest potential solutions that could be applicable to a broad range of clinicians and other health care professionals?

- Finally, I would like to see more discussion on the generalizability of the program to clinical populations or

a mention of what aspects/features of the intervention might need to be adjusted for such populations. While the authors note that the pilot was conducted on a subclinical sample, a more detailed discussion would be helpful.

To conclude, the authors did a great job piloting the program, especially considering how well they managed the challenges of conducting longitudinal interventions in real-life settings. I'm looking forward to seeing other teams implement the procedure as a way to alleviate problematic behaviors and add evidence on its efficiency and usability.

Matúš Adamkovič

Reviewed by **Charlotte Pennington** , 10 September 2024

Overall, this is an excellent Stage 2 report that clearly and transparently reports the findings of this study and interprets the data in a balanced manner. I am particularly impressed by its clarity and the Discussion section which considers the impact of all necessary deviations to the protocol and interprets the data in line with these. I am also impressed by the Conclusion that outlines how this is the first pilot evaluation that follows open science principles throughout and responds to a call made in the field to conduct more robust and well-designed empirical studies. The clarity of this manuscript has made it one of the easiest Stage 2 reports that I have reviewed for PCI Registered Reports, and the authors should be very proud of this. My overall decision is for minor revisions to be made and some clarifications to be outlined in the response to reviewer's document for this manuscript to fully meet the Stage 2 requirements of PCI RR, as follows:

2A. Whether the data are able to test the authors' proposed hypotheses (or answer the proposed research question) by passing the approved outcome-neutral criteria, such as absence of floor and ceiling effects or success of positive controls or other quality checks.

I agree that this criterion is met. This is a pilot evaluation to test the feasibility and efficacy of a 10-week TTRPG programme and to assess whether it can reduce gaming involvement, problematic gaming, social anxiety, and perceived loneliness. The data collected are able to answer the primary and secondary research questions and all deviations are transparently outlined throughout – these deviations appear necessary to make the project feasible and the study contributes to the literature to suggest that the TTRPG intervention is feasible and may be used to reduce social anxiety and gaming disorder symptoms. I am also impressed that the authors have outlined each of these deviations in the limitations section (e.g., how they were unable to test a reduction in time spent gaming at the three-month follow up) and have interpreted their findings in line with these. I have one minor comments in this regard:

The Results state that "Only two participants dropped out of the TTRPG program...", but these participants were not replaced with other participants to adhere to the participant sample size outlined in the methods (n = 20, four groups of five participants). As such, it would be good to clarify the final sample size (n = 18) within the Method section itself (and elsewhere where relevant) to ensure this is clear. I also thought that this would be a useful addition to the Abstract.

2B. Whether the introduction, rationale and stated hypotheses (where applicable) are the same as the approved Stage 1 submission. This can be readily assessed by referring to the tracked-changes manuscript supplied by the authors.

I partially agree that this criterion is met. Looking at the tracked changes document, you can see that several changes are made to the Stage 1 Introduction and Methods sections that are not usually permissible at Stage 2. However, a closely look at these changes does show that many of these are simply rephrasings or expansions of previous sentences, so nothing 'untoward' or different is added. The authors should note that this isn't usually appropriate for RRs, however. **The Editor may also wish to specifically look closely at the "Current study" subsection to ensure they are happy with the changes made there.** I have the following comments

regarding some specific changes:

The Stage 1 Abstract used to state “Outcomes assessed include social skills, self-esteem, loneliness, assertive, and gaming disorder symptoms.” However, these are now split into primary and secondary outcomes: “Primary outcomes assessed include gaming disorder symptoms, time spent gaming, and social anxiety symptoms. Secondary outcomes assessed include assertiveness/social skills, self-concepts, and perceived loneliness”. I have looked at this closely and this change reflects the primary and secondary outcomes stated in the Stage 1 report, so the addition in the Abstract is to simply clarify the variables. One minor point here is that it would be good if the Abstract followed the same order as the Data Analytic Strategy with regards to these variables: specifically, the Abstract outlines gaming disorder prior to time spent gaming, but the Data Analytic Strategy presents these the other way around.

Most of the changes to the Methods text are outlined in footnotes which is excellent to see. However, were any of these changes discussed and approved with the Editor, e.g., the change not to have a clinical psychologist verify a IGD-10 diagnosis? Personally, I can see why this change was required and deemed not necessary but just a note (as an Editor of PCI-RR myself) that any changes should be discussed with the Editorial Board before implementation to ensure the Stage 1 IPA stands.

The following sentence has been omitted from the Methods section: “Participants with missing data will not be omitted from the analyses unless the number of measurement points per phase is < 3, as three measurement points per phase is considered the minimal standard to reach in a single-case methodology (Tate et al., 2015”. However, it is now clarified within the Data Analytic Strategy section, so such change is fine (and I just wanted to note I had looked at this closely as usually changes to Stage 1 text are not permitted).

In the “Data Analytic Strategy”, the authors have removed the sentence “Any deviation from this pre registered data analytic plan will be discussed with the recommender and described and justified in the final version of the registered exploratory pilot”. I would clarify here that all deviations are now reported in the footnotes and were approved by the recommender(?).

2C. Whether the authors adhered precisely to the registered study procedures.

I agree that this criterion is met. In the main, the registered protocol is followed and, where necessary, any deviations appear to have been necessary and are outlined transparently in footnotes and in the Discussion section. The OSF page is extremely well organised and clear, with all materials, code, and data publicly available and presented in a FAIR manner.

2D. Where applicable, whether any unregistered exploratory analyses are justified, methodologically sound, and informative.

I agree that this criterion is met. The authors outline the findings for each of their primary and secondary outcomes and transparently delineate between these throughout.

2E. Whether the authors’ conclusions are justified given the evidence.

I partially agree that this criterion is met. I recommend the following revisions and/or clarifications:

Abstract: I recommend revising the final sentence of the Abstract to be a little more tentative given that this is (a) a pilot study and (b) there were $n = 18$ participants; specifically I recommend changing the term ‘can’ to ‘may’ in the following: “and can be used to reduce social anxiety and gaming disorder symptoms”.

Results: In the Results section, the authors report that 10/18 completers participated in all 10 sessions, and then go on to interpret this as “high”. Do you really think that a 55.56% completion rate of all 10 sessions is “high”? You may want to be a bit more tentative here to suggest that is it relatively high? I also note that this is ‘brushed over’ in the Discussion section by stating “and the majority participated in all 10 TTRPG sessions”). I do understand that 55% is still a ‘majority’ but it’s not as impressive, I’m afraid, as the authors currently make out. I think further discussion of this, perhaps in the Limitations section, is warranted – what steps could be taken by future research to ensure that this is either met or exceeded? It’s also noteworthy that you paid participants and this may have influenced their engagement and the success of the project; something I feel needs to be, at least, noted within the Discussion.

The following points relate to the Discussion:

The Discussion starts by outlining the positive findings regarding feasibility and participant's engagement with this programme. It then explains that one explanation is that TTRPGs share many features with videos games and similarly allow for fulfilling basic individual needs. I would clarify here that the inclusion criteria was being being a gamer with past or current experience of playing MMORPGs or online RPGs and, as such, this intervention is likely to work for those with experience with, or interest in, playing video games. That is to say, the generalizability of this intervention for those who are not gamers is unknown.

The second paragraph of the Discussion states that "Regarding reliable clinical change (post-intervention and follow-up), the effect was less pronounced", and then goes onto discuss this generally. It would be good to be more specific here – what specific findings demonstrate that this effect was less pronounced? (remind the reader). Similarly, in the Results section you clearly outline that many of these effects were short-lived, but this doesn't seem to come across within this Discussion section, which is key to evaluating the pilot.

What do you mean by the term 'diminished' in the following sentence: "Self-reported time spent gaming diminished post-intervention for a majority of the sample"? This could be clearer.

The following points relate to the figures:

Figure 2 states 5 participants in each of the 4 groups, but it would be good to update this to the final sample size with 5 participants in 3 groups and 4 participants in 1 group. Alternatively, you could put 'target' and then 'achieved' to clarify this.

Figure 3 presents a study flowchart but I am a little confused by the 2nd box which outlines those excluded from the study. Specifically, you state that $n = 123$ did not meet the inclusion criteria but then at the bottom you state that $n = 40$ did not meet 'other inclusion criteria'. What is this other inclusion criteria and why can't you combine this in the first 'did not meet inclusion criteria' statement? Otherwise this figure is fantastic and very clear.

I hope this comments prove useful to the authors in revising their manuscript and I want to reiterate that this is a really strong manuscript that I have enjoyed reading and learning from.

Signed,

Dr Charlotte R. Pennington,

School of Psychology, Aston University, Birmingham.