

Dear Recommender and Reviewers,

Thank you for the valuable feedback again. Please find our point-by-point responses below.

Recommender Dr Charlotte R. Pennington

Table 1: the formatting for the 95% CI on the top row should be centered also.

Fixed.

Tables should follow APA style, particularly if you want to go to a journal that has this requirement.

We've now changed formatting more toward the APA, but it's possible that we've still missed something. We'll be happy to further edit formatting in case the MS ends up being published in an APA-following journal.

In Table 2 there is a right-square bracket but no left-square bracket around the exploratory probabilities – please see R2's comment also about making this clearer in the table that these are exploratory probabilities; at first glance they can be mistaken for upper and lower confidence intervals. Looking at Table 2 alone, it is not clear what the 'overlap' values actually represent – could this be made any clearer?

We've clarified the exploratory part, following the APA guide, and elaborated on overlap in the title of the table.

For result H3c (Page 9), there is a plus and minus sign for the t-test result; please revise: $t(323.22)=+-2.72, p< .01$.

In this case, t was identical for both + and – and thus both are reported. We have added a note in the analysis plan section that we generally report the smaller test statistic (as advised by Lakens, 2022). We also clarified that in this anomaly both + and - were identical.

In case one wonders why we also changed one decimal from 6.22 to 6.19, this is because we reran all analyses with the recently updated R package, which gives this outcome.

Exploratory analyses, Page 9: “Although this might reflect the poor attention skills of respondents who have GRHPs...”. Is there evidence to suggest that this is the case? If so, please provide a reference to support this assertion; if not, I would remove this sentence from this section.

We have now revised this part into a more general form, i.e. we only wanted to highlight that reverse causality could also be possible (even if unlikely).

On Page 10, you mention that 'post hoc' power analyses are reported in the supplementary materials – do you mean 'sensitivity' power analyses here? Post-hoc power is essentially meaningless but sensitivity power analyses would provide the estimated effect size that could be found with N , power and alpha. See Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267. Relatedly, for some of these analyses you are comparing very large samples with very small ones (e.g., 8186 vs. 31) – is this appropriate? What was the ES that could be found for these analyses; I know this is reported in the supplementary materials but you may want to include this within the table too, given your conclusions (“The exploratory analyses regarding the mental health of gaming and non-gaming populations did not yield any meaningful differences. [...] This implies

a construct difference in terms of mental health, but confirmatory research is needed to corroborate it”).

We have removed the word “post doc”, which was indeed erroneous. We have also moved the power curves as the 4th appendix. We hope this will help readers to assess the reliability of the exploratory results. In some cases, small groups certainly make detecting small effects problematic.

R1 David Ellis

This is an excellent paper and it’s nice to see things in this area coming together coherently. As before, my comments are minor. Do you want to put the word validated in quotation marks (at least in the abstract) given that the scales now appear to be less valid?

We agree and have removed the word “valid” from all places in the MS. Because the scales do have psychometric validation studies for them, we fear that quotation marks might cause unnecessary debates.

On the other hand, these are very central scales for the field (GDT = 111 citations after 2021, IGDT-10 = 277 citations after 2017, GAS = 1446 citations after 2009). Perhaps popularity is even more critical when it comes assessing the relevance of such scales? We leave this undiscussed in the MS, as trends tend to change, but maybe historians will entertain themselves with this note in these review documents one day.

The sample size here remains a key strength as does the analysis, which is comprehensive and clear. I wonder if the first sentence of the discussion could be improved for clarity. In fact, starting from the second sentence might make more sense before returning to the point made in sentence one.

We agree and have revised this section.

‘Evidently, many people have some 466 problems with gaming sometimes, but this should not be confused with the prevalence of 467 related mental disorders.’ This reminds me of how researchers often conceptualise other technologies whereby normal use that can include some minor issues is conflated with problematic use (e.g. smartphones): ‘While it is easy to conflate heavy use with problem use, research into smartphone use should identify heavy use and problem use independently of one another’ (Andrews et al., 2015; p7) Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond self-report: Tools to compare estimated and real-world smartphone use. *PloS one*, 10(10), e0139004.

We have added a reference to this persistent challenge in the field. Because we have already exceeded the word limit of many journals, this was kept at one-sentence length.

Returning to this paper: ‘In sum, while the current technology use scales of different constructs seem unable to distinguish themselves from others, the scales of addictive gaming behaviors—standardly studied as a single construct—seem unable to identify mutual groups with shared problems. Presently, the field appears incapable of managing both, construct differences and similarities.’ This is an extremely powerful conclusion and likely has implications for measurement across psychology. The authors might want to touch briefly on how this has been allowed to happen in the first place. How do we prevent it happening again with other phenomena or technologies? This is hinted at in the conclusion but could be more explicit. For example, measurement development

appears to be rushed, and measures quickly become established with little fanfare. This is why the research reported here is so important.

We agree and have added a (long) sentence about this.

R2 Daniel Dunleavy

I thank the authors for the stage 2 submission of their manuscript. I hope the following comments, suggestions, and questions help strengthen and clarify components of this submission. Reporting: 1. In Table 1 and Table 2, the authors state: "Exploratory probabilities in square brackets" / "Exploratory differences in square brackets.". If this is common practice, please ignore my comment. However, I'd recommend using some other notation to enhance visibility. Asterisks might be misleading, given their common usage designating statistical significance. A dagger or other typographical mark (or perhaps just a superscript E, with a footnote explaining its meaning) might enhance visibility, without being misleading.

We agree and have revised the tables with suggested clarifications.

2. The authors appeared to have adhered to their proposed Stage 1 procedures/analyses. The exception (hypothesis 3) was reasonably explained and addressed (as much as they were able to) by the authors. I believe they have reasonably interpreted their results and drawn appropriate/justifiable conclusions. Code, Data, and other Materials: 1. Is there a link or persistent identifier to be able to access the relevant FSD data? I've tried the links provided, but don't quite seem to arrive at the relevant pages to (try to) access the data. Of course, this might be my mistake, since I'm relying on google chrome translation to help navigate the page. Any insight/help is welcome.

Unfortunately, the FSD staff are already on summer holidays, and we won't have the official link until later. We did ask and receive a confirmation letter about this, however:

To Whom It May Concern:

Veli-Matti Karhulahti et al. have submitted their data to be archived and shared for open scientific use via the FSD (www.fsd.tuni.fi/en/). These data have passed initial validity checks, and will be made available after further internal FSD validation unless issues occur that do not allow it.

Sincerely,

Jarkko Päivärinta
Senior Specialist
[Finnish Social Science Data Archive \(FSD\)](http://www.fsd.tuni.fi/en/)
FI-33014 Tampere University

2. I've been able to access the relevant r code and other materials on the OSF and it appears to be appropriate. Other Comments: I don't have any other concerns at this time. I thank the authors for their clearly written Stage 2 submission and the recommender for their consideration of the above review.

We thank the recommender and all reviewers (Stage 1 and Stage 2) for their accurate and rapid feedback during these busy times at the edge of summer. This has been a highly constructive and valuable review process in general.

- Authors