



# Different ontologies, different constructs? Instruments for gaming-related health problems identify different groups of people and measure different problems

A recommendation by **Charlotte Pennington**  based on peer reviews by **Daniel Dunleavy**  and **David Ellis** of the STAGE 2 REPORT:

Veli-Matti Karhulahti, Jukka Vahlo, Marcel Martončík, Matti Munukka, Raine Koskimaa, Mikaela von Bonsdorff (2022) Ontological Diversity in Gaming Disorder Measurement: A Nationally Representative Registered Report. Missing preprint\_server, ver. 3, peer-reviewed and recommended by Peer Community in Registered Reports.

<https://doi.org/10.31234/osf.io/qytrs>

Submitted: 23 May 2022, Recommended: 06 July 2022

#### Cite this recommendation as:

Pennington, C. (2022) Different ontologies, different constructs? Instruments for gaming-related health problems identify different groups of people and measure different problems. *Peer Community in Registered Reports*, 100209.

<https://doi.org/10.24072/pci.rr.100209>

Published: 06 July 2022

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/>

---

Screening instruments that aim to provide diagnostic classifications of gaming-related health problems derive from different ontologies and it is not known whether they identify equivalent prevalence rates of 'gaming disorder' or even the same individuals. Underpinned by this, Karhulahti et al. (2022) assessed how screening instruments that derive from different ontologies differ in identifying associated problem groups. A nationally representative sample of 8217 Finnish participants completed four screening measures to assess the degree of overlap between identified prevalence (how many?), who they identify (what characteristics?) and the health of their identified groups (how healthy?). The results indicate that measures based on the ICD-11, DSM-5, DSM-IV, and self-assessment appear to be associated with lower mental health. However, these measures of gaming-related health problems differed significantly in terms of prevalence and/or overlap, suggesting that they identify different groups of people and that different problems or constructs are being measured by different instruments. These findings are important because they contribute to the rapidly growing literature on the 'fuzziness' of constructs and measures relating to technology use. The authors recommend that researchers working with these measures should: (a) define their construct of interest; and (b) evaluate the construct validity of their instruments. Being able to answer these questions will enhance

research quality and contribute to strengthened meta-analyses. Importantly, this will prevent hype around gaming-related disorders, allowing researchers to communicate clearly and appropriately without risk of confusing related yet different constructs. The Stage 2 manuscript was evaluated by two of the reviewers who assessed it at Stage 1. Following revision, the recommender judged that the manuscript met the Stage 2 criteria and awarded a positive recommendation. To ensure that the manuscript met the requirements of the [PCI RR TOP guidelines](#), prior to this acceptance an email communication was sent to the authors by the recommender to ensure that study data were openly available on a temporary OSF link before the final data archive is fully validated by the Finnish Social Sciences Data Archive (FSD). This is noted in the recommended preprint. **URL to the preregistered Stage 1 protocol:** <https://osf.io/usj5b> **Level of bias control achieved:** Level 6. *No part of the data or evidence that was used to answer the research question existed prior to Stage 1 in-principle acceptance.* **List of eligible PCI RR-friendly journals:**

- [Addiction Research & Theory](#)
- [F1000Research](#)
- [Peer Community Journal](#)
- [PeerJ](#)
- [Royal Society Open Science](#)
- [Swiss Psychology Open](#)
- [WiderScreen](#)

#### **References:**

1. Karhulahti V.-M., Vahlo J., Martončík M., Munukka M., Koskimaa R. and Bonsdorff M. (2022). Ontological Diversity in Gaming Disorder Measurement: A Nationally Representative Registered Report. *Peer-reviewed and recommended at Stage 2 by Peer Community in Registered Reports*  
<https://psyarxiv.com/qytrs>

## **Reviews**

### **Evaluation round #2**

DOI or URL of the preprint: <https://psyarxiv.com/qytrs>

Version of the preprint: v2

#### **Authors' reply, 01 July 2022**

- Please note that a DOCX version with line numbers available in the alternative URL provided and line numbers have been removed from the Preprint PDF.

- We added new text, based on the reviews, only very briefly, as we were afraid of the overall word count. If desired, we could still add a recommendation for more careful and comprehensive measurement development. However, we also want to avoid negatively pointing at (or implying toward) the authors of the scales that we used. The problems that we discuss are general problems, and hopefully we can make progress collaboratively as a field – everyone makes mistakes. (We also tried to avoid using scale names in the MS as much as possible.)

- Regarding the open data, we understand that this may require further discussion due to the repository being able to do the final processing of files only after summer holidays. The corresponding author can be contacted to discuss this and seek solutions if needed.

Please see the attached files for more details.

[Download author's reply](#)

[Download tracked changes file](#)

## Decision by [Charlotte Pennington](#) , posted 28 June 2022

### Minor Revision

Dear Veli-Matti Karhulahti and co-authors,

I have now received two peer-reviews of your Stage 2 Registered Report submission: "Ontological Diversity in Gaming Disorder Measurement: A Nationally Representative Registered Report". As you will see, both reviewers are very positive about your submission and request some minor revisions and some discussion/thought before your manuscript is accepted.

Reviewer 1 makes a good suggestion about the conclusion of your paper with regards to measurement problems: "How do we prevent it happening again with other phenomena or technologies?". This is worth thinking about in your response.

Reviewer 2 asks about the open data, but this was solved through a temporary, accessible link. The data will be made permanently available via FSD once verified/approved. An update on this would be helpful and an updated statement (and link to the data when possible) should be added to the manuscript.

Minor points from myself:

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

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

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?

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

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.

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").

Yours sincerely,

Dr Charlotte R. Pennington

[Recommender]

## Reviewed by David Ellis, 19 June 2022

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?

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.

'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.

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.

## Reviewed by Daniel Dunleavy , 17 June 2022

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.

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.

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.

## Evaluation round #1

DOI or URL of the preprint: <https://psyarxiv.com/qytrs>

### Authors' reply, 31 May 2022

Dear Recommender and PCI Board,

Thank you for the insightful feedback before external review. We have finished the pre-review revisions and a new PDF has been uploaded in the same DOI location.

1. Open data: the data review time in the FSD varies. I do believe we can make the anonymous data temporarily available in an open location if the FSD review is not completed before decision. However, we cannot keep the data available in the other location permanently, as in the privacy statement we promised to archive the data via the FSD in particular. If more details regarding this issue are needed, I may have to consult our ethics committee for a statement.

2. The footnote has been revised and it is now more explicit that a) permission to proceed concerned H3 and b) new instruments have no effect on any of the present analyses and none of them are reported in this study.

3. We have exploratorily reproduced all registered hypotheses without those who failed the first control. The R file has been updated with these analyses. Table 4 is now preceded by an explanation of the analyses.

4. As we did not specify at Stage 1 how many or which (different) endorsement criteria would be tested, technically speaking, we feel that this should not be considered a deviation (as we did test different endorsements with THL1). The note in the previous cover letter was primarily to highlight that we are aware that numerous different endorsement combinations could be tested and compared. Only THL1 felt justified (i.e., it added coherence to test two THL1 endorsements in related to all hypotheses). If either reviewers or recommender/PCI feel that there is a need to test other endorsement criteria for a specific scale (regarding prevalence, prevalence difference, overlap comparison, mental/physical health, health comparison, etc.), we are naturally happy to do.

We have also revised based on the minor comments in the file. The sampling section has been moved to the beginning of Methods, as requested, and Table 1 now reports exact n's. As the reviewers might wonder why the Stage 1 section structure has been changed, it would be good to inform them that this change was due to a pre-review request by the recommender. A new tracked changes document is attached.

On behalf of the team,

Veli-Matti Karhulahti

[Download tracked changes file](#)

### Decision by [Charlotte Pennington](#) , posted 24 May 2022

#### Revision prior to Stage 2 review

Dear Veli-Matti Karhulahti and co-authors,

Thank you for submitting your Stage 2 Registered Report "Ontological Diversity in Gaming Disorder Measurement: A Nationally Representative Registered Report" for consideration by PCI Registered Reports.

Before sending this for in-depth peer-review, there are a few reassurances and edits required. I explain these in detail below and also provide in-text comments on your Stage 2 preprint to make it easier to see where I think changes/clarity is required.

Yours sincerely,

Dr Charlotte R. Pennington

**Recommender Comments:**

**Major**

1. Open data. You provide a valid justification as to why the data cannot be made openly available at this point in time. To adhere to the [PCI RR TOP guidelines](#), data should be made publicly available, or a legal/ethical justification should be included. You state that the verification of a dataset can take approximately 3-months. Can the data be made openly available by the time of (potential) Stage 2 acceptance, do you think? Could the data be uploaded to the OSF, or do Blendi not allow for this?

2. You contacted me to explain the error regarding one item in the PROMIS Global Physical Health 2 scale (GPH-2) meaning that they couldn't test half of H3. Along with the Managing Board, we agreed that this was OK and to continue with Stage 2 analyses. However, in the Stage 2 submission, a footnote explains further that this error happened because additional measures were added, which weren't explained or signed off by us: namely, anxiety, depression, and a question about the war in Ukraine (see below). Are any of these questions analysed in the Stage 2 manuscript? Are any of the questions combined with other questionnaire indices? For clarity, I would appreciate if you could update the footnote to explain that the Managing Board signed off on the questionnaire-item measure and then explain that extra control measures were included and their reasoning. These two things are mutually exclusive from my point of view of 'signing' one off.

*"as an extra control measure, our team agreed to enlarge the survey with two additional measures: anxiety (validated Finnish translation of GAD-2: Kujanpää et al. 2014) and depression (validated Finnish translation of BDI-6: Aalto et al. 2012). To add further means for assessing the effects of the drastic world events, we included a single item that asks the participants to self-report the negative mental health impact of the war in Ukraine. As a byproduct of these last-minute changes and several extra test iterations, a mistake occurred in our team and an erroneous GPH-2 item—PROMIS Global Health item #09 instead of #06, which is very similar in wording—ended up being included in the final survey. We noticed this soon after the data had been collected and immediately contacted the recommender who, after discussing with the managing board, advised us to proceed without confirmatory GPH-2 analysis in H3. We thus report physical health exploratively in this section with only one GPH-2 item ("GPH-1")."*

3. The confirmatory analyses look good and I have reproduced them. Thank you for providing the R-script, too. Further information is, however, required for the exploratory analyses to allow me to send this out for in-depth review. First, what are the Results when participants are removed for mischievous responding? It seems imperative to me that you need to check whether the Results hold when individuals are removed for failing the first control item. Second, what are the exploratory analyses outlined in Table 4? You need to explain to the reader what these are in the immediate text where they follow: I looked up and down the manuscript but couldn't quite work out what these analyses were referring to: a reminder for the reader would be very helpful.

4. You state the below in your cover letter. You need to make transparent ANY deviations from Stage 1 to Stage 2 in the main text, so this needs to be acknowledged within the manuscript itself:

*"One further point is worth noting: in the Stage 1 we promised to test different endorsement criteria for instruments to see if the results vary. We soon realized that even testing \*one\* set of different criteria for a single instrument results in a massive set of analyses, with lots of excessively lengthy reporting (and R code). The 2-page long Table 4 illustrates this, with only one alternative endorsement option tested. Therefore, we decided not to analyze and report further endorsement options to keep the article length reasonable and workload manageable. Regardless, we do fulfill our Stage 1 promise by reporting alternative endorsement criteria for THL1."*

I recap some of these comments via in-text comments your Stage 2 manuscript (attached). Additional, minor editing points are provided for you there also.

I understand you are concerned about word count, so I'd be happy for additional analyses etc. to be reported in supplementary files and uploaded to the project's OSF Page. Nevertheless, the manuscript itself needs to read clearly enough for the reader to understand exactly what you are referring to (i.e. clarifying the exploratory analyses in Table 4). [Download recommender's annotations](#)