**Identifying Gaming Disorders by Ontology: A Nationally Representative Registered Report**

Thank you for allowing me the opportunity to review this RR. I am glad to see work in this area going through the RR process and appreciate the authors’ efforts in detailing the rationale, methodology and analysis plan with a good level of detail. I have included my comments in line with the reviewer criteria below, as well as some more general observations/comments at the end. I wish the authors all the best with their research and hope my comments are helpful in their research endeavours.

Signed: Dr Linda Kaye

**The scientific validity of the research question(s)**

These are important questions and I am glad to see the conceptual bases of “gaming addiction” being queried as this is an ongoing concern in this field. The RQs therefore seem relevant although I have some observations about some of the specific hypotheses which put these in operation (see below).

It would perhaps be helpful in the rationale to make explicit reference to the distinction between core addiction criteria/symptoms (e.g., tolerance, withdrawal, mood modification, etc) and gaming-related problems (e.g., interference and impairments in life events) which are noted to often be conflated within empirical work (Colder Carras & Kardefelt-Winther, 2018). E.g., Colder Carras and Kardefelt-Winther (2018) identify four classes of player; IGD class, normative class, engaged class, and concerned class. However, their latter two classes would be misclassified based on their reporting of gaming-related problems only. Also Myrseth and Notelaers (2018) identify five classes; never symptoms, rarely symptoms, occasionally symptoms, problem gamers, and disordered gamers. Again, this suggests that different classes of players may have distinct patterns of gaming-related behaviour/gaming-related problems. Including some explicit statements about how each of the instruments being used are able to measure gaming addiction symptoms vs gaming-related problems would be helpful here I think. This would also be important in relation to some of the hypotheses. For example, H3 makes suggestions about those with “gaming disorders” from DSM 4 (GAS7) & self assessment (THL1) but up to this point, we don’t have much insight to know whether these two measures make clear the addiction symptoms from the health-related issues and so isn’t clear why this specific hypothesis is directed in this way.

Following the previous point, I also think it is potentially problematic to be forming hypotheses which expect poorer health in those who self-assess compared to general population (H3b). That is, people with gaming-related problems (which may be picked up from the THL1 scale) could equally be representative of the general population and not specifically distinct in the way those who are “addicted” may be from a general population. Further H3c expects health related outcomes between those with DSM-4 and THL1 scale to be similar. I am not sure I follow this rationale, based on one perhaps being incorporating more core addiction-related symptoms (DSM-4) and the other not (THL1). The hypotheses here may therefore require additional rationale or be removed if this reasoning is difficult to articulate

**The logic, rationale, and plausibility of the proposed hypotheses (where a submission proposes hypotheses)**

Figure 1 is helpful to articulate the various scenarios. However, I feel some further rationale behind this is needed about why four groups of GD are expected to be derived from the measures. Why “severe”, “medium”, “minor” & “v minor” ; is this based on existing classification levels of these measures?

In the hypotheses (e.g., H1 including H1a-H1d), the term “remarkably” is used to explain expected differences in prevalence rates. Arguably “significantly” would be a more typical term here.

There could be more context given about the “health-related problems” before it gets to the methodology where it becomes clear it is referring both to physical and mental health

**The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable)**

Analytic decisions etc are well detailed and provides insight into how these correspond to RQ/hypotheses. The analysis appears feasible but I feel more information about the sampling strategy would help know whether the sample size is feasible and whether it is representative (see next section)

A mixture of monothetic and polythetic approaches to cut-offs is proposed. It is good that these different scoring approaches are noted. However given recent findings that different scoring approaches are better or worse than others at capturing differences between “problematic gaming” groups (Connolly et al., 2021), it may be worth considering whether a consistent approach should be used (or do analyses which test both monothetic and polythetic for all analyses)

Connolly, T., Atherton, G., Cross, L., & Kaye, L. K. (2021). The Wild West of measurement: Exploring problematic technology use cut off scores and their relation to psychosocial and behavioural outcomes in adolescence. *Computers in Human Behavior, 125*, e106965. https://doi.org/10.1016/j.chb.2021.106965

“we include two control questions (Oppenheimer et al. 2009) in the survey and remove those responses that fail both. Participants who report not having played videogames within the past six months will not fill out the gaming-related screening instruments, and they will be considered not meeting the problem criteria that concern the present study”- I find this a little problematic in the way this is written. The last part of this sentence by implication suggests that those people who have played games in the last 6 months do meet the “problem criteria” which I feel is not a justified assumption.

**Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses**

There is currently not sufficient information about the recruitment strategy in relation to where people will be sampled from. If a representative player sample is sought, then some details about the relevant contexts they can be sought from is required. This information would be important especially in ensuring that any replication efforts are matched in this regard

**Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).**

Description of the outcome measures provides information about descriptors of the scoring ranges as reference points. There is not a control group but that is not unusual for research on this topic which is conducted in this way. The authors however may consider whether it is worth recruiting a control group to take the Physical and mental health measures as a control however especially as some of their hypotheses are making comparisons to the general population. Without a control group, this comparison may not be achieved.

**Presentational comments**

In the Introduction, the term “abnormal technology use” is used. I would advise the authors to avoid this term and instead perhaps use “problematic technology use” or similar

The cognitive load in reading this is quite heavy as acronyms are used quite a lot to describe the different instruments (and sometimes interchangeably between the initials of the scale and the diagnostic criteria they refer to). It would be most helpful to reduce acronyms down to aid readability