**This extremely complex and in-depth qualitative study aims to contribute to a taxonomy of gaming disorder. The study uses longitudinal qualitative methods to answer two questions that will help distinguish engaged video game play from gaming that is disordered or pathological. The report is extensively detailed but could benefit from some additional explanation to assist in understanding transferability and allow other researchers to extend the work in their own cultural contexts. I will structure the review by first providing some overarching suggestions, then answering some of the reviewer questions, and finally providing detailed feedback.**

**Thank you very much for the opportunity to review this report. I really look forward to the results.**

**Michelle Colder Carras**

**Department of International Health**

**Johns Hopkins Bloomberg School of Public Health**

**mcarras@jhu.edu**

**1. General feedback**

**I would especially recommend some clarifying figures to describe the temporality of data collection from various sources, the analysis timelines, and how different analyses feed into subsequent steps. The authors could consider reporting according to JARS criteria for qualitative or** [mixed methods studies](https://apastyle.apa.org/jars/mixed-methods)**. Although their approach is strongly qualitative, the temporal aspects seem closer to mixed methods.**

One major question: Why open interviews vs. semi-structured? This is a concern; a major data stream relies on experts who are not part of authorship team. The clinicians could, frankly, do a terrible job or conduct interviews very differently from one another. Especially since in-depth knowledge from qualitative studies in this area is scarce, greater structure in this study would be very useful. Even if the clinician interviews are later reviewed in panel interviews by separate experts, it’s too late if important questions are left out. Perhaps just an interim review of data to ensure that interviews are being conducted in a useful way? As I look again at the Instructions for Clinician Interview, it occurs to me that another area of difference might be the cultural awareness of the interviewer and their ability to speak the native language. For example, in the US many clinicians are not native speakers of English and come from cultures that are very different (e.g., collectivist). Without an understanding of the interviewers or more structure around assessing culture, collected data might look very different.

Another important question: The clinician interviews don’t include a specific question about comorbidities, and screening scales as used in the diaries don’t capture psychiatric diagnoses (for the most part) the way clinician interviews do. Adding this to the interview or asking the clinician to include information from billing/medical record entry would provide one piece of concrete information that could easily be incorporated into the taxonomy.

The temporality also has implications for use of screening scales in diaries completed four times a year. Depending on the scale, these are not likely to pick up conditions that have occurred greater than 30 days ago but after the previous diary entry, as most scales have a 2 week to 30-day time period. The temporality of states of gaming and mental health is also not very clearly described outside of the Programmatic Components. It might be good to weave in how the non-absorbency of the state of “problematic gaming”/gaming disorder, etc. might affect understanding of samples, data collection, and analysis.

The creation of the taxonomy itself could use some clarity as well. How will clustering be initiated? How will cultural differences (eg samples) play in? I would recommend cutting the template in half and displaying just a few codes in a hierarchy to illustrate what a cluster might look like. As an alternative, perhaps an example taxonomy that combines data from various sources might be useful.

It would be useful if the cultural context sections used a common template – they should all report on basically the same things. Some themes to include across all countries could be statistics re: “problematic” gaming/gaming disorder, prevalence of game play, ethnic diversity of country population, size of gaming industry (eg # employed), strength of gaming lobby, gaming-related regulation, integration of esports into college and pro athletics.

I also have two thoughts that are NOT related to the science but are more general suggestions for mental health researchers:

(1) Consider reaching out to potential partners in developing countries, as these are typically underrepresented in mental health research

(2) Consider partnering with or having more leadership by people with the experience of problematic gaming. The positionality statements are great, but there are researchers who have publicly described their struggles, e.g. [Halley Pontes](https://www.halleypontes.com/). It looks like his current website doesn’t describe this, but it was previously on his blog.

**2. Reviewer questions**

**1A. The scientific validity of the research question(s).**

*…defined with sufficient precision as to be answerable through quantitative or qualitative research. They should also fall within established ethical norms.*

The questions are clearly defined and fall within ethical norms.

**1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.**

*…This criterion addresses the coherence and credibility of any a priori hypotheses. The inclusion of hypotheses is not required– a Stage 1 RR can instead propose estimation or measurement of phenomena without expecting a specific observation or relationship between variables. However, where hypotheses are stated, they should be stated as precisely as possible and follow directly from the research question or theory.*

Hypotheses will be developed through qualitative work, so this is n/a.

**1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable).**

*This criterion assesses the validity of the study procedures and analyses, including the presence of critical design features (e.g. internal and external validity, blinding, randomisation, rules for data inclusion and exclusion, suitability of any included pilot data) and the appropriateness of the analysis plan.*

The soundness is there, but some points of clarity would help (defined in 2nd section). The feasibility is also there and has been well thought out, including alternative sampling plans and dropout*.* The sample size is more than adequate, but one aspect that didn’t seem quite clear was the potential for samples to change, e.g. the esports sample developing pathology, or the pathological sample to start playing esports. A sentence or two about how this might be handled would provide clarity.

**1D. 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**

*…contains sufficient detail to be reproducible and ensure protection against research bias…*

This is not applicable as the study is a qualitative phenomenological approach. That said, the authors address epistemologies and potential bias through positionality statements, which is a big credit. A few additional points could be clarified for transferability/extensibility.

**1E. 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).**

*This criterion addresses, where applicable, the extent to which the proposal pre-specifies data quality checks that will reveal whether the results are able to answer the research question(s).*

Because of the flexibility built into the study through open interviews by clinicians, I feel the study would benefit from some additional details about data quality as outlined below.

**Key issues to consider at Stage 1**

* Have the authors clearly distinguished work that has already been done (e.g. preliminary studies and data analyses) from work yet to be done?
	+ **this could be improved and preliminary/prior work used to explain details in recruitment and analytic approach**
* Regardless of whether the study has received ethical approval, have the authors adequately considered any ethical risks of the research?
	+ **Yes, however, it would improve their ability to minimize risks if they provided clear resources (e.g., crisis hoteline) for participants who experience crisis-level discomfort through answering research questions.**
	+ **They could also refer people to treatment sources if it is determined that the non-treatment-seeking groups are in need.**

**3. Specific feedback**

Intro: Please clarify where your previous work comes in and how far along the work is

Lines 104-125: Would be good to be specific about where the health fits in (e.g., seems to be in RQb)

Line 129 Please define kinesthetic a bit more

Line 142 Please explain what a phenomenological forest is

Line 149-50: It’s unclear why the samples may come from Sweden and Czech.

Line 154: Consider calling Duplication Extension or something similar to better explain.

Table 1: Please clarify ages

Line 160: “sought treatment” – themselves or family members (eg the parents of 16 year olds)

Line 163 typo-Motor should be motive, I think

Line 165 It would be good to somewhere indicate the prevalence of “disordered” gaming by country and any statistics on treatment seeking that can be found

Line 168 Please clarify the idea of self-identifying as a player of esports games

Line 178 “to playing alone” might be more clear as “only to playing” so as not to be confused with “playing only by oneself”

Line 180 What are the implications of comparing adolescents to adults?

Line 192 and elsewhere: Native language – how will data collection in the native language include or exclude native speakers, both participants and clinicians/experts?

Line 200: Why would team members be present in clinical interviews? What kind of problems might this introduce or avoid?

Line 207 Please describe the translation process.

Line 210 Does carrying out the interview in the native language exclude speakers of any particular languages?

Line 215 Will filling out the diary every 4 months capture any episodes of e.g., depression between the 4-month periods?

Line 268 It would be good if all authors had a positionality statement, especially since some of the author team may come from countries where the anti-gaming disorder lobby is very strong.

Line 272-284 Could you please describe the exact purpose of the expert panels? What is the expectation, given that theoretically clinicians are able to make ICD diagnoses themselves? Where will the panelists be recruited from? Perhaps this should be separated out. This is an area where a figure detailing the flow of data collection and analysis would be especially useful.

Line 288 As the pilot is the source of the phenomenological interview approach, please provide one or two more sentences describing the pilot’s approach.

Line 304-305 Please give a little more detail at this point about why the health scales are not included in the analysis. They are the only strong source of quantitative data and part of the original research question, so the approach to including data about health in analysis could be better justified and described throughout.

Table 2: This might benefit from a diagram that shows the temporality and repetition of data collection that would allow the dependencies to be described (e.g., panel discussion dependent on clinical interviews, different participants for those, phenom interviews collected from a different sample by researchers, etc). It’s a very complex design and this clarity would help.

Line 314 Please provide a brief description of the definition and goals of multiverse ethnography; the report outlines only the methods. Please also provide a bit more explanation of what a structural synthesis might contain.

Line 339: The taxonomization is by far the most complex part of the study and seems a bit like a black box to me. It may be that this is a bit beyond my skills in reviewing qualitative studies, but perhaps the process of initiating clustering, for example, could be described more.

Line 343 How might a participant be clustered more than once, and how might this be marked in the final model?

Line 345 I realize that the figure is meant only to illustrate what the taxonomy would look like, but it would greatly aid understanding to provide some made-up illustrative examples.

Line 350 The clusters are described as numbering 5-10 and are described as health clusters, but it is not clear how health is highlighted when so much of the analysis relates to meaning and design. Examples of the taxa might help with this. I’m trying to work it out in my head, and this is what I’m imagining:

Even providing an example of each level would be useful. I’m especially wondering where culture (and gender) might fit in.

Line 386 Please describe the abstraction techniques that might be used to protect participants’ privacy. I think it might be useful to add these to the consent form, even if it’s just 1-2 examples. E.g., if a participant uses their gamer handle or another’s gamer handle, these are easily searchable, so if these will be anonymized, this could be communicated to the participant.

Line 686 Typo; myst should be “midst”

Line 816 Where translation is used, please describe the process briefly.