Stage1 RR review response

Author's response. *A warm thank you for the opportunity to revise this manuscript. We are grateful to the reviewers for their precious time and willingness to read and comment on this long manuscript. Below, we address all the reviewers' suggestions one by one and lay out the revisions made. Please note that, in order to improve readability and reduce text load, we have included only those comments that requested a response.*

Additional note: we have added Footnote 1 regarding a new event that took place during revisions. We have carefully assessed this opportunity from all perspectives and see no ethical or practical obstacles for including such additional line of invitation distribution. However, if reviewers or the recommender perceive this problematic, we can discard Footnote 1 and pursue the study without it.

Reviews

By Christopher Ferguson Ferguson

I appreciated the opportunity to review this RR. I have just a few comments below that I wonder if these may improve some of the ideas in here. First, when mentioning the WHO "gaming disorder", it might be helpful to note the specific controversies around this. Many scholars directly opposed the WHO (Aarseth et al., 2017) as did the APA and Psychological Society of Ireland's media psychology divisions in an open statement (https://div46amplifier.com/2018/06/21/an-official-division-46-statement-on-the-who-proposal-to-include-gaming-related-disorders-in-icd-11/). The authors do a nice job discussing the wider controversies, but I think it's good to note specific criticisms of the WHO's decision. Related to this conceptually, is the lack of interest (by the WHO and, frankly, most researchers) over the larger issues of behavioral overuse disorders...why is there so little interest in overeating, work, exercise, shopping or religion "addiction" (Griffiths, to his credit, has papers on dance and fishing "addiction".) There's an important question about why we're always hyperfocused on technology (and I'm not convinced the answer is that technology is inherently more addictive than these other things.

Author's response: We agree it is important to further highlight how other behaviors, with which some people have problems in a similar manner, have not received diagnostic attention to the same degree that gaming does and this has

been pointed out by various authoritative groups. The point has now been added to the MS with references.

This I think is also a practical issue for the paper...particularly given lack of clear guidance on what constitutes "gaming disorder"...responses by both clinicians and gamers may simply invoke societal moral panics in open response formats. I'm thinking in particular of the clinician ratings here. Given how easy it is to portray anything addictive...see the offline friends addiction scale (Satchell et al), I'd like to see a bit more attention to this.

Author's response: We agree it is highly relevant to remind especially the clinicians to be reflective of their assessment and not simply count listed diagnostic symptoms or repeat confirmatory ideas of addiction. To stress this, we have revised the instructions for clinicians so that they are explicitly asked to report "how" and "why" they choose certain views. We also stress that all expert data will be explicitly discussed in a panel, with a goal to understand each case idiographically (rather than simply relying on numeric expert reports). As a final step, we further apply interpretive analysis to all the collected data—including raw data, secondary expert data, and panel-generated tertiary data—which will allow us to flesh out and triangulate the complete process of knowledge generation.

We also highlight that even though we apply some gaming disorder scales, the scale results will not be interpreted as reflective of gaming disorder; rather, we are exploratively curious how participants fill them and how they relate to all other data.

I think the authors include a lot of good and important variables about gaming contexts and health, but I think it will be possible to find spurious correlations here, without controlling for other variables. I assume that T1 health/mental health data will be controlled. However, particularly as we have some decent information to suggest that academic and family stress is causal for problematic gaming, I think some of these larger contexts would need to be measured and controlled.

Author's response: This is a highly important suggestion and we have accordingly added these measures.

The authors need to set a smallest effect size of interest (SESOI) for their quantitative analyses, otherwise they may inadvertently misinterpret "statistically significant" noise effects as hypothesis supportive when they should not be interpreted as such. See Ferguson & Heene (2021) for discussion and guidance. In

conclusion, I support the authors' efforts here and hope my suggestions are constructive.

> **Author's response:** *We discussed this with the team at length and decided not to preregister any hypotheses, despite the project producing large amounts of quantitative scale-based data as well. We considered adding one more programmatic component with SESOIs clearly stated and hypothesized, but ultimately decided against it because the sample size with the present approach is not fixed and different time points would add too much uncertainty to hypothesis testing (control of freedom and power analyses would be virtually impossible).*
>
> *That said, although we do not pursue confirmatory findings, it remains a possibility for any other researchers to do so later: because the data are archived in FSD, any research team will have the chance to preregister hypotheses before accessing the data (and this will have very a high confidence level because every data access will be marked in the FSD archive).*

By Michelle Colder Carras

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. Although their approach is strongly qualitative, the temporal aspects seem closer to mixed methods.

> **Author's response:** *We have now carefully followed JARS guidelines for qualitative studies and added a visual summary of the protocol with a timeline, as well as two additional clarifying figures.*

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.

> ***Author's response:*** *These are all important points. In fact, our team members being present in the interview was one of the means for us to assess the quality of data, but we have now added an explicit quality check after each clinical interview, which allows us to provide systematic feedback to the experts or recruit new ones in case the quality is low. Regarding cultural awareness, our three language groups (Finnish, Korea, Slovak) are so small that it would be very unlikely for a clinical expert to be able to carry out a professional interview in any of these languages without cultural awareness (which is different from English). However, we have added a note for exceptions, e.g. if some of our participants happen to be immigrants, we will seek extra cultural expertise for these cases.*

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.

> ***Author's response:*** *This is a good point; indeed, the idea for the clinical assessment was to evaluate mental health but we have now made it explicit (also in the clinical instructions) that we wish potential mental disorders to be evaluated and previous diagnoses to be disclosed.*

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.

> ***Author's response:*** *We agree temporality is highly important, but also very difficult, if not impossible, to tackle comprehensively. First, we highlight that our primary data are qualitative and in-depth, i.e. the annual long interviews and periodic diaries will go through temporal changes in life explicitly. Recall bias remains a challenge and does not allow day-to-day reporting, but this remains a limitation for any 3-year study.*

That said, we fully agree that because we are cycling multiple scales (= not possible to administer the full set of all scales every period), it would be useful to control scale outcomes more frequently. As a solution, we have decided to create single-item recall measures, which will be used to control changes in areas that are not measured at the time (as part of diaries). For example, when social health is not measured, we will use a single item to inquire whether there has been a change in this regard since the previous diary.

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.

Author's response: *These are all valuable recommendations. Accordingly, we have created an example taxonomy based on our pilot data and systematized the cultural context sections.*

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

Author's response: *We fully agree and this is our plan for Duplication; we have further highlighted this.*

(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. It looks like his current website doesn't describe this, but it was previously on his blog.

Author's response: *We have carefully considered this important comment and explored it in several meetings. In brief, we believe that seeking and adding (more)*

co-authors with gaming problems after peer review could be considered problematic in many ways, e.g. possibly perceived tokenistic. Positionality and reflexivity are really important for us, and while we are not doing participatory research, we want to have participant involvement done in the right way. We believe there are several helpful ways to ensure that external voices of lived experiences will be even more part of the process; we address this below, but first a note about our own team composition.

One of our regional representatives, Marcel (who is in a leading role throughout the entire study in Slovakia) has openly shared problematic gaming patterns. Matti (the first author) and Yaewon (leading role in Korea) have both described negative experiences related to gaming, but not considering them clinically problematic. Some team members, like Tiina, have no gaming background at all. These kinds of different renderings or understandings of our diverse lived experiences with intensive gaming, we believe, represent the spectrum that our ontological project is trying to map out. We have carefully pursued diversity in our team when it comes to gaming habits and backgrounds (in addition to other domains), which we hope to enable open, multiperspective dialogue when it comes to generating and interpreting the data.

That said, we acknowledge that our team size is limited and cannot fully represent all perspectives, for which we have also added the following plan: "We are prepared to apply member checking by recruiting researchers and non-researchers with treatment-seeking gaming backgrounds to comment on our interpretations." Because this would be done as audits (due to data sensitivity) and it is not clear which individuals would be available or qualified for the task over the next 4 years—and it is currently unsure at what step this procedure would be most efficient—we have framed it as "prepared to" instead of providing a fixed plan. In sum, we agree that receiving more external viewpoints would contribute to the project and look forward to utilizing this help over the years.

2. Reviewer questions

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.

Author's response: *We have added a note about this relevant issue.*

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.

> *Author's response: We have added this to the final section.*

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.

> *Author's response: This has been addressed (see earlier).*

- 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**

  > *Author's response: We have now added details about this.*

- 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 hotline) 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.**

  > *Author's response: During the three ethics reviews, we developed step-by-step team guidelines for unexpected situations and how to correctly react in crisis-like events. This is not a public document but we will share it for review via a private link.*

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

> *Author's response: We have now clarified this.*

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

Author's response: We have now clarified this.

Line 129 Please define kinesthetic a bit more

Author's response: We have now clarified this.

Line 142 Please explain what a phenomenological forest is

Author's response: We have now removed this part.

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

Author's response: We have now explained this part.

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

Author's response: We have now explained this.

Table 1: Please clarify ages

Author's response: Clarified.

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

Author's response: Corrected.

Line 163 typo-Motor should be motive, I think

Author's response: Corrected.

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

Author's response: Good point, we have added this to cultural context.

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

> *Author's response: Corrected.*

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

> *Author's response: Corrected.*

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

> *Author's response: Corrected.*

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

> *Author's response: Corrected.*

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

> *Author's response: Corrected.*

Line 207 Please describe the translation process.

> *Author's response: Corrected.*

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

> *Author's response: Corrected.*

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

> *Author's response: We should highlight first (it is now presented more explicitly in our list of measures) that we cannot administer all scales every 4 months. Because we have numerous recurring scales, having every scale repeat each time would take all hours that the participants have reserved for the recurring diary -- and our primary interests are, indeed, in the open diary entries. Especially with children, data quality would suffer if they would have to repeatedly answer the same long questions.*

Our scale use plan for all groups includes the same scales once per year, cycling throughout the year between different scales (with some exceptions). This allows us to provide participants a mix of different content, which we believe contributes to data quality. This will naturally result in temporal gaps, i.e. periods not covered by scales, but we stress that the scales are only tertiary data sources in this study (after interviews and open diary entries) and we believe that 3-year annual responses will be sufficient for generating appropriate health information to support health taxonomization (with primary clinical interview data for adults and selected children).

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.

Author's response: *Corrected.*

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.

Author's response: *Details added.*

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.

Author's response: *Added.*

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.

Author's response: *We have now included the scales to analysis. Of note, considering that we have clinical experts carry out individual mental health assessment, we will prioritize such gold standard data over the scale outcomes (but scales will be consulted in detail).*

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.

Author's response: Added.

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.

Author's response: Added.

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.

Author's response: We fully sympathize with this comment because clustering such a vast amount of diverse qualitative types of data will necessarily involve elements that are difficult to clearly preregister (and it makes reviewing the process understandably difficult). We have tried to clarify the essential parts and created a model based on our pilot data. We privately share a non-public excel file to illustrate how, as an example, we clustered the health level based on 16 cases by their overlap (the data/codes are unfortunately in Finnish but color codes demonstrate entities like "depression" that formed our taxa).

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

Author's response: One of our example taxon involves "stress"; we hope this explains how a participant might belong to this group and after a year when the stressful situation is gone (e.g. at work) the participant could "move" to another group. We will visualize these temporal changes by arrows or other suitable means, which will be carefully considered when we craft the higher-detail taxonomy (also visually, i.e. graphic experts will be hired for that). Note that the current example illustration is very much abstracted and simplified because carrying out the full, comprehensive analytic work and results reporting for the pilot model would take the time of another full study.

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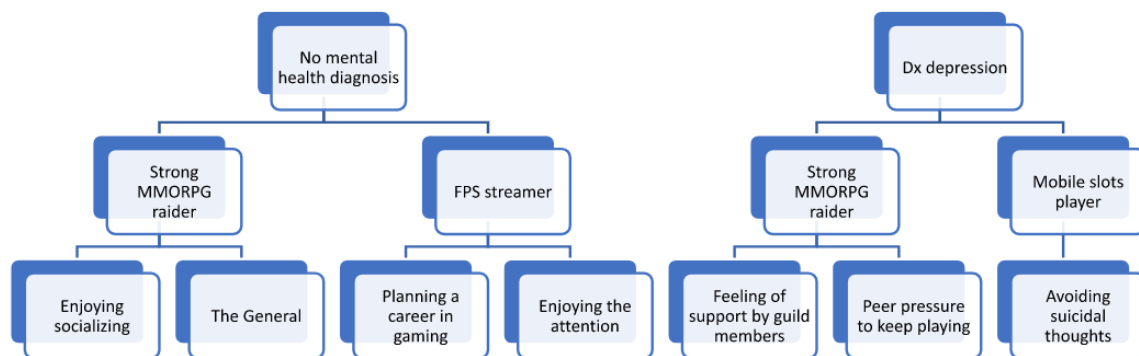
Even providing an example of each level would be useful. I'm especially wondering where culture (and gender) might fit in.

Author's response: *We really appreciate the above example, which inspired us to craft the pilot model. Because our example is based on actual (limited) Finnish data, we didn't model culture in it -- in the final version we will use color and shape codes to signal specifically cultured taxa. Again, we stress that our example falls far from the detailed expression that we pursue.*

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.

Author's response: *Added.*

Line 686 Typo; myst should be "midst"

Author's response: *Corrected.*

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

> **Author's response:** *Added.*

By Anonymous

Major

1) The manuscript emphasises the clinical relevance of quantitative research on gaming addiction to inform contemporary thinking and classification instruments. However, this is no further mentioned throughout the paper and it remains unclear whether the present analysis approach can meet these ambitions. In particular, the authors indicate that that the health scale data will not be used in later analyses, without providing reason for doing so. I believe a concrete plan of how the present research may inform gaming disorder classification (with or without analysing health outcomes) would positively contribute to the quality of this work. Besides, if health outcomes are not to be analysed, please provide reasons for doing so. I was surprised to see health removed from the analysis given it is a primary outcome of interest, with direct relevance to research question A: can passionate and pathological gaming be distinguished? A data analysis table may be utilized to demonstrate how the present research aims to fulfil the ambitious objectives.

> **Author's response:** *We agree and believe this largely a misunderstanding; namely, we do plan to use health data for analysis, but primarily that generated by clinical experts, which we consider a gold standard and more reliable than scale data. We have now clarified this in the MS and included scales in the health analysis in order to be able to triangulate with multiple data types. We apologize for our initial unclear phrasing.*

2) The present proposal sets out to measure participants 'health' via different measurement scales that remain to be named. Please define the to-be assessed constructs and what measures will be used to capture the said constructs. In the current state, the reader is left guessing which data may later be used for clustering analyses.

> **Author's response:** *We agree and have added a complete scale list of recurring scales.*

3) The sampling approach involves asking local esports organisations to recruit participants for the project. This introduces systematic bias and players who are doing well may be more inclined to participate in this research. While it may be hard

to circumvent this bias in a hard-to-reach population, this should be acknowledged in the manuscript and later analyses.

> *Author's response: We were not clear in the previous version what we meant by "utilizing the organizations," and we have clarified how we use an open call that is distributed in esports networks (i.e. they don't recruit participants). That is, the esports orgs' role is similar to those of schools and clinical orgs: they only **distribute** the call. We choose who to include/exclude. We have now also highlighted that we are explicitly interested in "healthy" players for this specific group, as it will widen the spectrum of intensive players (4h+/day) in the longitudinal design. They must, as an inclusion criterion, consider gaming not being harmful for them (i.e. purposive sampling), which makes them different, by purpose, from treatment-seekers. (But it remains possible that gaming becomes problematic for them too.)*

4) The authors anticipate 33% dropout in the first year and will sample another batch of participants after 12 months. However, no such procedure is described for the 2nd and 3rd year. How will later dropout be dealt with? Perhaps, oversampling is an appropriate way of dealing with anticipated dropout? Moreover, the fact that participants start at different times means later analyses will collate data from different time periods. How will time be accounted for?

> *Author's response: This is a good point and we have added oversampling in the plan. Because our funding runs out in 2027, we cannot afford to start third and fourth rounds in 2025 and 2026 (we need 3 years after the first contact). Thus, participants must join in 2024 at the latest, which is the backline of additional recruitment.*
>
> *Time will necessarily different between participants, but we will document the times of interviews/diaries, which will be assessed by readers later. This note has been added.*

5) The analyses procedures are far too vague and "culture specific contextual factors" should be known prior to data collection. As such, the researchers need to specify which factors they aim to include in the analyses and for what reason.

> *Author's response: This is a valuable comment and we have added these elements in more detail in the first programmatic component.*

6) "Depending on the time of writing, we may utilize cross-sectional or longitudinal data" (l. 407) is too vague to be included in a registered report.

> *Author's response: Corrected.*

Minor

1) While I do appreciate the sentiment, the first paragraph of the analysis section does not provide information on to-be performed analyses (pp. 6-7, ll. 261-269). I believe this statement is of little interest to readers and should be devoted to the present analyses instead. More generally, I am not familiar with the term 'bayesiastic' and think you may be referring to 'bayesian'?

> *Author's response: We were afraid that "bayesian" would give the false expression that we will report actual statistical probabilities, thus "bayesiastic" -- but this has now been changed to the former.*
>
> *We agree the first paragraph might not suit all readers, but especially because another reviewer addressed issues related to this very paragraph, we would prefer keeping it.*

2) The authors mention that 45 participants from group C are selected for an additional phenomenological annual interview. Currently, there is no information on how these individuals will be selected.

> *Author's response: Added.*

3) Raters are repeatedly used throughout the procedure. To quantify their agreement, I would suggest documenting their inter-rater reliability coefficient.

> *Author's response: We have kept the clinical ratings with added quantification, but removed other rating elements. Because the quantified outcomes will be heavily dependent on discrepancies between individual expert viewpoints in each country (noting R1 advice as well), we do not aim to make direct comparisons but use these outcomes to inform (health) taxonomization.*
>
> *The data will be made open for reuse and other researchers can use the ratings in future studies and meta analyses, perhaps combining them with other expert reports.*

4) Please use consistent formatting throughout the manuscript.

> *Author's response: Corrected.*

5) Many, if not most gamers will be playing several games at a time. Likewise, the games played may change over the three years of participation. How will this be reflected in the research procedure and analyses?

> ***Author's response:*** *This is a good point; in addition to the original plan to code all games from the data, we will create participant-specific lists that show evolution in this regard.*

6) For the sake of readability, avoid abbreviations (e.g., IPA p. 9, l. 358) and/or make sure to define any acronyms.

> ***Author's response:*** *Corrected.*

7) The data statement and ethics (p. 10, ll. 381-389) is too vague. "Big qual" datasets and "non-open forms of participation" should be further defined. I would suggest focusing on the to-be performed analyses, rather than the underlying motivations (e.g., "our goal is to share as much as possible of the generated data").

> ***Author's response:*** *We have improved this section. Some vagueness remains because sharing qualitative health data is complicated (especially with minors), but the supplements (e.g., participants' privacy notification) now involve these details at length.*

8) Please combine supplementary files for readability.

> ***Author's response:*** *We tried to comply with this request, but unfortunately we encountered pragmatic challenges. Because we have 7 supplement types, each of which includes numerous sub-supplements using various fonts and file formats, it turned out to be impossible to create a highly readable combined supplement. That said, we have reformatted the manuscript and our supplements should be now easily accessible and clear.*