

Dear Prof Dr Malte Elson,

Thank you very much for the opportunity to submit a second revised version of this stage 1 registered report proposal to *Peer Community In Registered Reports (PCI RR)*.

I include your recommender comments in *black italics* and my point-by-point responses in **purple** below.

In the interest of open science, I intend to publicly archive the review history for this manuscript regardless of publication outcome.

Yours sincerely,
Leon Y. Xiao

--

Recommender Comments from Prof Dr [Malte Elson](#)

28 Jan 2024 21:24

Revision invited

Dear Dr. Xiao,

Thank you for your submission to PCI-RR. I now had the opportunity to read your revised manuscript and the response to the reviewers. Let me first apologise for the considerable delay in my response. I understand that this may have caused changes to the timeline of your project, with the original date of the first sampling of games being January 18. There were unfortunate private and professional circumstances on my side that I failed to manage to provide a timely response. Again, my sincerest apologies.

Response 1: There is absolutely no need to apologise. I hope things are better now. I appreciate also that, importantly, you are volunteering your time and knowledge. The timeline was perhaps a bit ambitious even considering the use of scheduled review to expediate the process. It is unfortunate that registered reports as a format, because of how it frontloads the reviewing time (whose length cannot be accurately estimated), might not be suitable for extremely time-sensitive projects.

I have indeed preregistered the previously revised version of the manuscript and started data collection on 18 January 2024. I believe that, technically, this would mean that we would be at bias control level 2 now, rather than level 6, as I have observed some of the data (as of 29 January 2024, about 25 games have been fully coded). In addition, I have collected the complete data for all games on whether their Apple App Store product pages are disclosing the presence of loot boxes, although I have not calculated the exact percentage rate of compliance. I would say that I do have the feeling that it will not be very high as I have types many more 0s than 1s. I am unsure whether for that one variable, the bias control level would indeed be 1 instead because, despite not having performed the 'analysis' per se, which would be the percentage calculation, I have developed an instinctive feeling from my experience collecting the data.

I am doing this as a registered report because I want to do the most transparent and robust study possible (by getting peer reviews prior to data collection), so I am not

personally bothered by the reduction of the bias control level. I do believe I am not proposing any substantive changes to the method as registered prior to data collection on 18 January 2024 (although I am amending the hypothesis testing point below and adding a section on DCMS's and Ukie's responses to my request for them to preregister their potential interpretations). I believe this would constitute 'robust' bias control measures: in effect, the study is being performed in accordance with what will hopefully later be approved and registered.

This is a strange situation that should perhaps be referred to the Managing Board if the Recommender deems it appropriate.

I have decided not to send out again your revised manuscript for re-review. The points raised by the reviewers in round 1 were quite clear, and it felt like an undue use of their time to ask them again to check your response to them.

Overall, I am quite happy with your points, counterpoints, and changes to the manuscript. Referring to the major points in my previous letter:

1. I no longer see any issues with using Ukie and non-Ukie games, given their claims that you cite, but also their normative influence on the gaming industry in the UK. I also appreciate the changes you made to the title. Further, I also find it acceptable that you do not scale your sample against the number of players per games, and essentially give each game the same "weight" in your assessment. However, I expect that this importance point will receive extensive attention in the discussion of your findings, whatever they will be.

Response 2: Of course. I will discuss this in due course as it is indeed a very important point that Prof Przybylski initially raised: it is plausible (and likely true) that the 100th ranked game has a much smaller player base and earns a lot less, when compared to the 1st ranked game. We unfortunately do now have good data on this.

2. You offer a useful definition of what you consider a lootbox "event" in your study. This is now much clearer, and also how you will proceed playing and encountering lootboxes, if they exist.

3. To be frank, I still have some concerns with regards to your use of hypothesis tests and cutoffs. In your response letter, you argue that the cutoffs are arbitrary, but that you will use them regardless to prevent yourself from changing your interpretation the final %, as different people might interpret the same rate quite differently (industry person vs advocacy group). Further, you argue that the observed rate should not be generalised beyond the top 100 grossing games, and you argue that it is not a sample, but a population.

I can follow your reasoning for the first point, though I would have expected an elaboration of this point in the manuscript itself rather than merely the response letter. Certainly, one could argue that rules are rules, and that therefore any deviation from 100% is a failure of self-regulation. But then: I am sure that the same industry, or other industries, also do not perfectly self-regulate with regards to other issues, although one might generally say that, on average, the system works. It would probably be best to offer your line of reasoning to the reader.

Regarding the second point, I will admit I remain unconvinced. I can only imagine that it will be excruciatingly difficult for you to write the results section of your manuscript in a

way that does not generalise beyond the two populations examined (the January and the July top 100 grossing games). Simply considering the title of the current manuscript, "Assessing compliance with UK loot box industry 1 self-regulation on the Apple 2 App Store", readers might not get the impression this is merely a study on two specific lists of games at two arbitrary points in time. Similarly, the abstract says "[c]onclusions will be drawn as to whether the measures have been complied with by companies to an adequate degree", which certainly suggests you seem to think your findings can be used to infer something about companies in this industry generally (which I'd agree with). Finally, although this may be more a linguistic habit, you do refer to these lists of games as "samples" throughout the manuscript (the word is mentioned 19 times, whereas "population" is not mentioned once).

I do not think this point is super important for the manuscript or the empirical work. Looking at the top 100 grossing games makes sense, as these are certainly the games where compliance with the regulation would be more important. However, I do not see what is special about the top 100 grossing games that would justify excluding games not on this list from your interpretation. At the very least, you have not yet provided a compelling argument for this.

Response 3: I think I was preoccupied by the fact that I have personally changed the compliance status of two quite popular games (that I thought may appear in the sample again, and I think one of them, *F1 Clash*, is in fact included now that data collection has started). This has actually been addressed (or rather will be addressed) by the separate reporting of compliance rates amongst games that have previously been studied and those that have not, as suggested by Prof Chambers. I was additionally concerned that previous research publicly identified certain top games as being non-compliant, which may have caused them to do better now. The same did not occur to less popular games. I was also conscious that the most popular games were more likely to be complained about by players and other stakeholders who may have complained of non-compliance and caused changes, which is less likely to occur with the lower-ranked games. Those are the reasons why I thought, and indeed continue to think, that the results should not be interpreted beyond the top-100 without certain important caveats. Prof Dr Elson is right that preventing that over-interpretation entirely (but just by myself but also by others) is presumably impossible and overly cautious, so it would be better to acknowledge those limitations, rather than try to artificially limit the interpretation.

I also agree that the use of a formal hypothesis testing framework for research like this is probably inappropriate and was likely initially done by me to fit my studies into certain moulds (*e.g.*, the PCI RR study design table and what quantitative researchers expect to see). I have removed the hypotheses and simply ask the research questions. I do however continue to include how I will interpret the results depending on the percentages: I have added details as to why I am making these preregistrations. As there is now technically no hypotheses and quantitative hypothesis testing, the study design table (which I have tried to amend but is finding difficult to fit qualitative research questions and methods in) has been removed.

In addition, since the last revision, I received the responses from both DCMS and Ukie as to preregistering their interpretations of the results. Both have, unsurprisingly, refused. I have amended the section that previously discussed this in the future stance as this has since happened. I also provide details of their respective responses.

4. Finally, I appreciate that you have shifted from a programmatic RR to a standard one.

With kind regards

Malte Elson