

Dear Dr Karhulahti and the Managing Board,

Thank you for the opportunity to submit a revised version of this manuscript to *PCI Registered Reports*, which has now been uploaded to the Open Science Framework and is available at this link:

https://osf.io/f3yab/?view_only=33e5875516d144ed98509c7871242b31. I include the complete text of Dr Karhulahti's comments and the three reviews by Drs Moshirnia, Chin, and Macey in *black italics*; my point-by-point responses in *purple*; and amended or newly added text in *blue* below.

Leon Y. Xiao

--

Dr Karhulahti's Recommender Comments

Dear Leon Xiao,

Thank you for submitting the revised Stage 1 of the manuscript. You've done comprehensive work and carefully responded to all requests. All the reviewers are satisfied with the revisions, and I agree that the manuscript is now closer to the IPA. However, I must request a few more revisions before we meet the PCI RR standards.

1. *Regarding the philosophy of hypothesis testing in general, there should always be a good reason for testing a hypothesis (PCI RR criterion 1B). Although a lot of relevant information is included in the manuscript, we are still lacking explicit justifications, i.e., "why" does the study expect that each hypothesis will be true. With H1 and H2, for instance, I would agree that because loot boxes are 'banned' in the country, there is a good reason to expect virtually no loot boxes at all. But you should explicitly provide such rationale, e.g. "Based on loot boxes being banned in Belgium, there will be no loot boxes in Belgium". The justification can be inside or before/after the hypothesis, but there must always be a clear justification.*

Response 1: I am grateful to Dr Karhulahti for the detailed comments. I really appreciate the time that has been taken in preparing this, and I apologise for issues caused by my unfamiliarity with the PCI RR process.

The following justification has been added immediately before H1 and H2:

"Because loot boxes have been effectively banned by the Belgian Gaming Commission's public pronouncement of its interpretation of Belgian gambling law^[44], no loot boxes should be found amongst video games available in Belgium:..."

I also added these two justifications for H3 and H4, respectively:

“Assuming that the Belgian ban on loot boxes has been effective to some perceivable degree, then the availability of loot boxes in Belgium should be lower than previously observed in other countries that have not actively regulated this mechanic (e.g., the UK):

[...]

“For the Belgian ban on loot boxes to be deemed fully effective, it must not only reduce loot box availability through the usual domestic channel of downloading iPhone games from the Belgian Apple App Store but also prevent potential technical circumventions (e.g., downloading the games from the UK Apple App Store from within Belgium and purchasing loot boxes in that version of the game). Preventing such circumventions appears technically difficult and therefore unlikely to have been accomplished: ...”

2. As an example of the above issue: on page 8, after stating the hypotheses, a conflicting interpretation of H1 and H2 seems to occur: “Hypotheses 1 and 2 mean that a Belgian loot box prevalence rate of *less*than or equal to 2% will be found”, while the actual H1 and H2 state a *higher* prevalence. Lower prevalence, again, is suggested on page 13, but higher is suggested in the table at the end of the MS. Please clarify and justify which result is expected. Also, please note that, as currently H6 expects that at least 3 games will include loot boxes, this would conflict with expecting less than 3% loot box prevalence (H1 and H2). Clearly stating the rationale in each case will solve this.

Response 2: H1 and H2 now refer to 0% as before given Dr Karhulahti's Comment 10. There should no longer be any conflicting references on that point.

Regarding that H6 (now H4) appears to predict a 3% loot box prevalence rate, I do not actually know that those three games (which were published some time ago) will appear in the 100 highest-grossing list in Belgium by the time of data collection, so although it appears to be a contradictory on the face of it, I do not think it necessarily is.

3. Justification is also needed in Type 1 error control. It currently reads: “The rate of 2% was chosen instead of 0% to provide type 1 error control.” But one needs to justify why 0.02 was chosen as a control. I provided one such example in my previous letter (a previous study found 1 false positive, which was doubled to be safe). As I have now read your new commentary paper, which demonstrates no less than 22.9% (!) disagreement rate between two studies, it might be justifiable to use even more control than 2% (one of the examples in the commentary reminds me of the problematics in assessing the level of skill, when evaluating a mechanic to be considered gambling/loot box, see:

Lipton, M. D., Lazarus, M. C., & Weber, K. J. (2005). Games of skill and chance in Canada. *Gaming Law Review*, 9(1), 10-18.)

(I mention this paper just it in case it might be useful later)

To be clear, it is up to you to decide what error control to use, and you can definitely keep 2%, but whatever number you use, please justify it (at least briefly explain why it was chosen). For these hypotheses, perhaps also address the possibility that some loot boxes operate with a license and, if such evidence is found, they will not count in H1/H2. In general, it could be useful to refer to “unlicensed loot boxes” maybe?

Response 3: The justification that Dr Karhulahti has suggested is now fully set out in the manuscript. In addition, the possibility that some loot boxes might be licensed is now mentioned and will be accounted for. I will keep the very apt ‘unlicensed loot boxes’ terminology in mind for future use:

“A Belgian loot box prevalence rate of 0% should be found amongst all games studied and amongst those games studied that were deemed suitable for underage players. However, considering that one previous loot box prevalence study identified an 1% false positive rate through its data collection process [4], to provide for type 1 error control, Hypotheses 1 and 2 will be accepted even if up to two Belgian games studied are identified as containing paid loot boxes (i.e., a prevalence rate of up to 2% will be deemed as effective elimination of loot boxes from the Belgian market). Considering that some video games might contain loot boxes which are duly licensed by the Belgian Gaming Commission (whose confirmation will be sought by the present study), any games that are so licensed will be excluded from the sample for the purposes disconfirming Hypotheses 1 and 2.”

4. Regarding H4 and H5, there seems to be no justification and, as you say, the numbers are based on mere intuition (theoretically, you could craft 101 unique hypotheses for the prevalence and one of them would be true!). Therefore, I must ask you remove these hypotheses (unless a good justification is found). When a research field is not yet at a stage where hypotheses can be crafted based on good existing knowledge, more work is needed before hypothesis testing can be started. See e.g.,

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. Perspectives on Psychological Science, 16(4), 744-755.

That said, you can surely provide an estimation of how many loot boxes there might be, but that wouldn’t be hypothesis testing (there would be no hypothesis to test, but RRs can also be used for transparent estimation in order to provide unbiased estimates).

Although this goes beyond the present study, I could imagine that reviewing the global literature and data on gambling regulation in general could yield a reasonable prior concerning the effect of successful regulation strategies. For instance, we do know that illegal gambling exists around the world, but perhaps less so in successfully regulated countries. Whether loot box regulation is useful or successful could be, in the future, based on this

*previous knowledge of regional (online) gambling regulation. But again, I highlight that using this kind of (or similar) approach in the present study could lead to new issues and the need for new reviews, for which removing H4 and H5 is likely the better option (again: you may still discuss the level of prevalence when the results are known, but you cannot make related *confirmatory* claims).*

Response 4: H4 and H5 have been removed. I intended effectively to preregister how I will interpret the Belgian loot box prevalence rate that will be found as I am conscious of the significant amount of discretion that I would have after data collection (law being a discipline where the same percentage can be interpreted in very different ways by different researchers). I now do this as part of how I will interpret the results of H3; I hope this would be acceptable:

“... how strongly this recommendation will be put by the present study in the Discussion section depends on the identified Belgian loot box prevalence rate that will be identified (a reduction to below 50% will be deemed as effective, whilst a reduction to below 25% will be deemed as very effective).”

5. Binomial testing in H3 should be ok now and the justification for 0.65 is reasonable. However, I am still concerned about the power analysis. 0.15 has been chosen as the effect size but there is still no justification for why is that effect suitable? So again, we need an explicit justification (PCI RR criterion 1C). What would be the smallest effect that is meaningful for this study and why? The paper I referred to previously also provides a good overview regarding different ways for defining a meaningful effect.

Dienes, Z. (2021). Obtaining evidence for no effect. *Collabra: Psychology*, 7(1), 28202. <https://doi.org/10.1525/collabra.28202>

In other words, how few (= how much less than 0.65) loot boxes should there be in Belgium for that number to be relevant at all? This question needs to be answered before statistical power can be calculated.

Response 5: Thank you for pointing this out. I must admit to not having much guidance to go on for this point, due to a lack of relevant literature in law and regulation as to the smallest effect size of interest. I therefore attempted to justify the smallest effect size of interest that I have chosen based on, in my opinion, how satisfied the end user (in this case, ‘policymakers’) would consider the effect size as below. Ideally, I suppose I could have surveyed some policymakers to determine the smallest effect size of interest to them, but I do not think doing so would be practicable.

In the absence of any prior guidance on what effect size would constitute a ‘legally meaningful’ and ‘socially beneficial’ regulatory measure, a smallest effect size of interest of Hedges’ $g = -.15$ is proposed based on the potential usefulness of the results to the end users^[76]. The intended end users would be

the policymakers in other countries who might be considering taking the same regulatory action that Belgium has already taken: importantly, besides the Belgian Gaming Commission having issued its interpretation of Belgian gambling law and thereby threatened criminal prosecution of non-compliant companies implementing paid loot boxes, little else appears to have been done by the Belgian Gaming Commission in terms of enforcement; therefore, thus far, the costs that have been incurred by Belgium in its attempt to regulate loot boxes have been relatively low. Accordingly, realistic policymakers seeking to expend a similarly low amount of resources to regulate loot boxes would likely not expect a particularly high reduction to loot box prevalence in Belgium (e.g., for the loot box prevalence rate to be reduced by at least 50 percentage points, i.e., to 15.0% or lower). Nonetheless, these policymakers would likely still expect some perceivable reduction (e.g., for the loot box prevalence rate to be reduced by at least 10 percentage points, i.e., to 55.0% or lower) before being persuaded to emulate the Belgian ban, considering that some regulatory costs have been incurred by Belgium and that Belgian consumers have been given the (potentially incorrect) impression that loot boxes have been effectively eliminated from the market. Recognising that some policymakers might be more hesitant to restrict players' freedom to purchase loot boxes and video game companies' commercial interests and therefore be more cautious when relying on the results (e.g., they might view a reduction of 10 percentage points or less as being insufficiently persuasive), it is proposed, conservatively, that the vast majority of policymakers would likely consider a reduction of at least 15 percentage points (i.e., for the loot box prevalence rate to be 50.0% or lower) as demonstrating the effectiveness of Belgium's loot box ban (as implemented in its relatively low-cost manner) and be persuaded to potentially emulate the Belgian regulatory actions in their own countries.

6. Still related to H3, there seem to be 2-sided and 1-sided tests, both, carried out for no reason. Please delete the 1-sided tests, which are duplicates.

7. Regarding RQ1 and RQ2, I would suggest combining them along the following lines: "Has the Belgian ban succeeded in eliminating paid loot boxes from mobile games?" You can answer this RQ via both H1 and H2.

8. Regarding RQ3, I would suggest simplifying it along the lines "Has the Belgian ban on paid loot boxes been effective?" This is what you test with H3.

9. If you remove H4 and H5, you can also remove RQ4. Of course, the results will still include the prevalence rate, so scholars can speculate about the exact effectiveness of the ban post hoc. RQ5 is good.

Response 6: These changes (regarding Comments 6–9) have been made as recommended. Thank you for pointing these out, and I apologise for the need for such basic issues to be identified.

10. ... going briefly back to the hypotheses, with error control now in H1 and H2, you have rephrased them as “More than two...” Please note that hypotheses are not statistical statements, but they apply to the world in general. Essentially, we’re expecting the absence of loot boxes, and error control only reflects our awareness that testing and methods aren’t perfect. So, the hypotheses can well be “The highest-grossing iPhone games in Belgium do not contain paid loot boxes” -- and this will be accepted even if 1 or 2 do contain loot boxes, but only because we acknowledge the possibility of error in analysis/methods (alternatively you can include the justification directly inside the hypothesis, as exemplified in #1)

Response 7: I have changed this back to ‘none’ of the games.

11. The above applies to H3 as well. I would suggest following Macey’s suggested wording with the necessary modification: “Of the highest-grossing iPhone games, fewer will contain paid loot boxes in Belgium than in countries that have not banned loot boxes.”

Response 8: This suggested change has been adopted.

12. If you wish, H6 could also be clarified more (but this is optional, as it’s not making explicit statistical statements): “Games known to contain paid loot boxes will continue to offer them for sale even when the phone is within geographical and jurisdictional Belgium.”

Response 9: H6 (now H4) has been changed to:

“Hypothesis 4: UK iPhone games known to contain paid loot boxes will continue to offer them for sale even when the phone is within geographical and jurisdictional Belgium.”

13. Following the above, it is also not clear what outcome will corroborate H6 (or null). E.g., if one game continues to offer loot boxes but 2 games do not, what would be the conclusion? The criteria for interpreting the results for a hypothesis must always be clear.

Response 10: I now set this out as follows:

“... Hypothesis 4 will be accepted, if loot box purchase is possible within one or more of the games using any of the abovementioned methods. The interpretation will be that the law can be easily circumvented by dedicated players; the Belgian Gaming Commission should therefore consider ways to force video game companies to better enforce compliance with the law.

However, if loot box purchase is not possible within one or more of the games using any of the abovementioned methods, the interpretation is that the law could not be circumvented in the simple ways that have been attempted, although other potential circumventions remain untested and possible. The present study will conclude that companies might have taken *some* technological measures to prevent circumventions of the Belgian ban, although further evidence would be required to confirm this (e.g., contacting the relevant company to request for confirmation of the compliance actions that have been taken)."

A few smaller notes/suggestions, which may be considered.

- On page 2 it reads: "there are two types of loot boxes" --> perhaps rephrase into "loot boxes can be divided in two types" (because there are dozens of different types of loot boxes)

Response 11: This has been changed as suggested.

- On page 3 it reads: "and therefore does not possess real-world monetary value" --> consider specifying e.g., "direct real-world monetary value" (because accounts can still be sold onward, right?)

Response 12: It is factually correct that user accounts can always be sold (although doing so would almost always contrary to the End User Licensing Agreement). I have added 'direct' to before 'real-world monetary value.'

- It reads that "The following hypotheses will be preregistered at <[OSF registry link]>" but since this is an RR, you don't need to separately register hypotheses.

Response 13: Of course. This has been removed.

- On page 11 it reads: "A 'paid loot box' will be defined as being either an Embedded-Isolated random reward mechanism or an Embedded-Embedded random reward mechanism" --> please explain to the reader what these concepts mean

Response 14: These two concepts are now defined as follows:

"A 'paid loot box' will be defined as being either an Embedded-Isolated random reward mechanism (which are video game mechanics that players must pay real-world money to activate and which provide randomised rewards that do not possess direct real-world monetary value) or an Embedded-Embedded random reward mechanism (whose activation also must be paid for by players with real-world money but which do provide

randomised rewards that possess direct real-world monetary value), as defined by Nielsen & Grabarczyk (2019) [7]."

- On page 11 it reads: "95.4% of games were coded through gameplay and only 4.6% of games had to be coded through internet browsing." Does this refer to games or games with loot boxes? As the % of how many games in general must be coded via internet depends on the prevalence rate (e.g., with 0.5 prevalence one would need to code 50% of data with internet), the % of loot boxes found would be more informative.

Response 15: This did refer to all video games studied. I have changed this to the alternative framing that has been suggested: X% of loot boxes were found through gameplay, whilst X% of loot boxes were identified through internet browsing.

"Secondly, in the most recent loot box prevalence research of Xiao et al. (2021), of the 77 games amongst the 100 highest-grossing UK games that were found to contain loot boxes, 73 games' loot boxes were identified through gameplay (94.8%), whilst only 4 games' were determined through internet browsing (5.2%), so the potential bias caused by coding games that must be coded through internet browsing as not containing loot boxes would be very minor [5]."

- On page 15, I would still suggest removing the following sentence: "... and conclude that the Belgian measure was likely ineffective." Already the anecdotal evidence cited in the manuscript shows that the ban has had some effect (= some companies adjusted their design), so it feels wrong to conclude that the measure was likely ineffective, unless direct evidence is found (and justified what effectiveness would be). The section reads very well otherwise, so I suggest just dropping this sentence.

Response 16: I take Dr Karhulahti's point. I have deleted '... was likely ineffective and...' I hope this resolves this issue.

I hope this feedback is helpful and I believe we are very close to IPA after the above revisions have been implemented. Please let me know if any of the comments are unclear -- I will be happy to clarify.

Response 17: These feedbacks have been incredibly helpful. Thank you very much.

- Veli-Matti Karhulahti

--

Review 1 by Dr Andrew Moshirnia

I would like to thank Mr. Xiao for his revisions. My concerns have been addressed and I reiterate my recommendation of the piece.

The Dutch reversal is an interesting addition and might we worth investigation if firms perceive uncertainty in the law and the likelihood of leniency based on that uncertainty.

The Minecraft and Roblox discussion is necessary and well addressed. Of course it opens a whole new line of inquiry (merging modding/playbour with loot-box/consumer protection), but one outside the scope of the current experiment.

--

Review 2 by Dr Jason Chin

I enjoyed reading the response of the author and the other reviewers' reviews. The author responded satisfactorily to all my comments and questions.

There were methodological issues I missed in my initial review due to my lack of knowledge about the subject matter (e.g., the third party involvement issue). I am happy to leave it to the editor, reviewers, and author to work out those details.

I always sig my reviews,

Jason Chin (ORCID: 0000-0002-6573-2670)

Reviewed by Joseph Macey, 24 Mar 2022 12:29

--

Review 3 by Dr Joseph Macey

I thank the author for their comprehensive response to the comments of myself and the other reviewers, I believe all points have been addressed and am happy to accept the revised submission.

A small note, in their response to my 2nd point, the author queried as to whether further action is required:

"For transparency, I would be happy to add that "I have argued elsewhere..." before the section on overregulation (i.e., immediately following the section quoted above), if Dr Macey may think that would improve the fair presentation of the arguments."

I do not believe any further changes are necessary, as I feel my original concern was appropriately addressed and is no longer valid, as such the author is free to make any further changes based on their own judgement.

Response 18: Thanks again to Drs Moshirnia, Chin, and Macey for their time and helpful feedback.