Trends in Psychology
https://doi.org/10.1007/s43076-025-00487-3

ORIGINAL ARTICLE

Reply to "Fundamental Flaws in the Design and Reporting of Chew and Neo (2024)"

Peter K. H. Chew[1]

Accepted: 15 July 2025

Abstract

Although loot boxes are structurally and psychologically similar to gambling, they are available to minors and relatively free from government regulations. The extant research has generally found weak positive relationships between the purchase of loot boxes, and internet gaming disorder and problem gambling severity. However, most of the research do not include a comparison group to quantify the unique harm of loot boxes. Indeed, our study found that internet gaming disorder and problem gambling severity were not significant predictors of resources (i.e., money and time) spent on games with loot boxes after controlling for resources spent on games without loot boxes (Chew & Neo, 2024). In response, an authors' commentary provided feedback on the following issues: (a) fundamental error in the survey materials, (b) unclear and unjustified inclusion and exclusion criteria, (c) misinterpreting null results as evidence of absence, (d) statistical approach and overcontrolling, (e) lack of engagement with open science. In this paper, I elaborated on some of the decisions made in our study and strengthened the conclusions via Bayesian statistics. Overall, the authors' concerns are unwarranted and I reiterate our conclusion that (a) there is a lack of evidence for the unique harm of loot boxes and (b) it might be worth examining the effects of microtransactions in general (Chew & Neo, 2024).

Keywords Internet addiction disorder · Pathological gambling · Loot boxes

Loot boxes are a form of microtransaction that allow gamers to exchange money for a randomized reward (Montiel et al., 2022). Although they are structurally and psychologically similar to gambling (Drummond & Sauer, 2018), they are available to minors and relatively free from government regulations (McCaffrey, 2019). In general, research has found that the purchase of loot boxes is positively correlated with internet gaming disorder ($r=0.25$) and problem gambling severity ($r=0.26$) (see Garea et al., 2021 for a meta-analysis). However, in our study, we argued for the importance of a comparison

✉ Peter K. H. Chew
peter.chew@jcu.edu.au

[1] James Cook University, 149 Sims Drive, Singapore 387380, Singapore

Published online: 29 July 2025

group to quantify the unique harm of loot boxes (Chew & Neo, 2024). Indeed, we found that internet gaming disorder and problem gambling severity were not significant predictors of resources (i.e., money and time) spent on games with loot boxes after controlling for resources spent on games without loot boxes. Instead, the best predictor for resources spent on games with loot boxes is resources spent on games without loot boxes. In other words, individuals who invest resources in a game are more likely to do so for another game, albeit more so if loot boxes are present. Given that the negative results were inconsistent with the loot box literature, we anticipated difficulties with publishing the results and expected increased scrutiny and critique by other researchers. My responses to the commentary are documented in the following sections.

Fundamental Error in the Survey Materials

First, the authors argued that the survey materials were flawed because an incorrect example of a game was provided. Specifically, *League of Legends* was used an example of a game without loot boxes when it contains loot boxes since 2016. This is indeed an error and we are grateful to the first author for bringing this to our attention. Subsequently, we added this as a limitation but highlighted that its effect (if any) is mitigated by the provision of a definition of loot boxes and clear instructions (e.g., Please answer the following questions according to your engagement ONLY with games without loot boxes…). As mentioned in our study, the best way forward is to correct this error in future research (Chew & Neo, 2024).

Second, the authors argued that the variables used in our study (e.g., money spent on games with loot boxes) (Chew & Neo, 2024) was inferior to that used by majority of the researchers (e.g., money spent on loot boxes). In contrast, I would argue that the latter method is inferior because some games contain both loot boxes and non-loot box items. To estimate the amount of money spent on loot boxes, participants must first distinguish between the amount of money spent on loot boxes versus non-loot box items. This process increases cognitive load and the risk of errors in recall. In fact, it would be easier for participants to recall the amount of money spent on different types of games (i.e., games with loot boxes vs. games without loot boxes). This issue is shown in Fig. 1. More important, the accuracy of the variables in our study can be verified. For example, while Google Play Store provides a breakdown of the amount of money spent on games, it does not provide a breakdown of the amount of money spent on loot boxes versus non-loot box items. However, it should be noted that both methods are subjected to the same limitations such as the social desirability bias (Crowne & Marlowe, 1960) and errors in recall. In the future, it would be optimal to obtain and analyze the relevant data from game companies directly.

Third, the authors also argued that money spent on non-loot box items "is irrelevant to loot box-specific research questions" (p. 7). This is a mistake that perpetuates the limitation in the literature. Loot boxes are similar to non-loot box items except that they contain mechanics similar to gambling (Drummond & Sauer, 2018). Consequently, if researchers are advocating for legislation against loot boxes, then they must show that loot boxes present harm *over and above* that of other in-game microtransactions (e.g., non-loot box items). In other words, it is the gambling-like

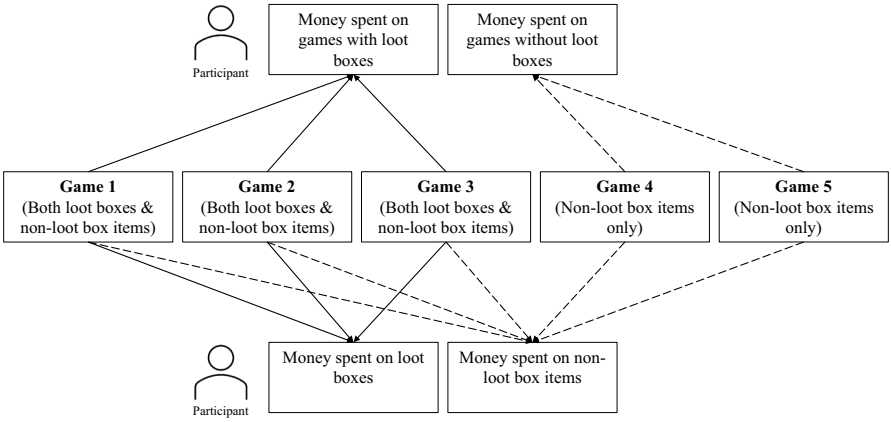

Fig. 1 A comparison of ease of recall between the variables used in our study (Top) and the variables suggested by the authors of the commentary (Bottom)

mechanics of loot boxes that are causing harm and hence, deserving of regulation. Otherwise, researchers should examine the harm of microtransactions in general and advocate for legislation against them.

Finally, the authors argued that money spent on games with loot boxes is confounded since some of that money could be spent on non-loot box items. They acknowledged that it is currently unclear the relative percentages of money spent across the two item types, which is a tacit agreement with my criticisms of the literature. However, they noted a strong correlation in our study between money spent on games with loot boxes and games without loot boxes and reasoned that there is no point examining the correlation between the two variables. Specifically, they argued that the "confound may explain why there is such a strong correlation between all spending on games with loot boxes and all spending on those without ($r(171) = 0.48$, $p < 0.001$ per Table 1 of Chew and Neo (2024))." (p. 7). There are numerous issues with this argument.

First, the authors did not make explicit the criteria used to determine the strength of a correlation. Cohen (1988) suggested that a Pearson r of 0.10 is weak, 0.30 is moderate, and 0.50 is strong. However, the strength of a correlation needs to be interpreted given the discipline and area of study (Akoglu, 2018). Specifically for psychology, a Pearson r of ≤ 0.30 is weak, 0.40 to 0.60 is moderate, and ≥ 0.70 is strong. In other words, a Pearson r of 0.48 is a moderate correlation and the two variables have a shared variance of 23.04% ($r^2 = 0.48^2 = 0.2304$). The overestimation of the strength of a correlation might cause the authors to overestimate the degree of the supposed confounding between the two variables. Furthermore, this might also cause the authors to overestimate the harm of loot boxes. For example, as mentioned earlier, the purchase of loot boxes is positively correlated with internet gaming disorder ($r = 0.25$) and problem gambling severity ($r = 0.26$) (Garea et al., 2021). If a correlation of 0.48 was interpreted as strong, would those two correlations be interpreted as moderate? In fact, an r of 0.26 is a weak correlation with a shared

variance of 6.76%. More information about the Pearson correlation can be found in an elementary statistics textbook (e.g., Gravetter & Wallnau, 2007). Second, the correlation between the two variables was presented as part of the standard reporting procedure as the descriptives statistics for multiple regression and is not an aim of our study. Lastly, as shown in Fig. 1, I am aware that money spent on games with loot boxes might include spending on non-loot box items. Consequently, a hierarchical multiple regression was used, where money spent on games without loot boxes was controlled before the key predictors (i.e., internet gaming disorder and problem gambling severity) were added to the model.

Unclear and Unjustified Inclusion and Exclusion Criteria

The authors claimed that the inclusion and exclusion criteria are unclear and unjustified. The criteria are elaborated here. First, the inclusion criteria (at least 18 years and have experience playing games with and without loot boxes) was stated in the information sheet for potential participants. They also confirmed that they meet both inclusion criteria before proceeding with the study. Second, multivariate outliers were not defined and explained because it was assumed that readers would have a working knowledge of the assumptions of multiple regression. In our study, multivariate outliers refer to participants with a Mahalanobis distance value larger than the critical value from a Chi-sq table. These outliers were removed before data analysis (Chew & Neo, 2024). More details can be found in Tabachnick and Fidell (2012). Finally, participants with missing data were those who indicated their consent but did not complete any items of the study. It was impossible to conduct a missing completely at random test (Little, 1988) and imputing missing values using the expectation–maximization procedure (Dong & Peng, 2013). Consequently, these participants were removed before data analysis.

Misinterpreting Null Results as Evidence of Absence

The authors argued that the study incorrectly interpreted an absence of evidence as evidence of absence (Chew & Neo, 2024) and that a Bayesian approach should be used. I agree with their assessment. For example, I often supplement my analyses with Bayesian statistics when requested by editors and reviewers (e.g., Chew & Au, 2024; Chew & Yap, 2021; Chew et al., 2023, 2024). However, despite being aware of the limitations of the Frequentist approach (Wagenmakers, 2007), I do not include Bayesian statistics by default for two reasons. First, Bayesian statistics is seldom taught as part of the statistics and research methods curriculum in most universities (Faulkenberry et al., 2020). Consequently, readers might not be able to understand Bayesian statistics in journal articles. Second, the use of Bayesian statistics is currently not the norm in psychology. For example, a literature review found that although the use of Bayesian statistics has been increasing over the years, only 1579 articles used them between 1990 and 2015 (van de Schoot et al., 2017). The additional use of statistics from an approach outside of the norm requires more

explanations that might not fit within the space constraints of some journals. However, given the authors' comments, I have adopted the Bayesian approach to reanalyze the data.

The data from the study was reanalyzed using JASP (Chew & Neo, 2024; JASP Team, 2023). Only results for money spent on games with loot boxes were presented since that appears to be the focus in the commentary and that a similar pattern of results was found for weekday time and weekend time spent on games with loot boxes. A Bayesian linear regression was conducted with age, gender (1 = female and 2 = male), employment status (1 = employed and 2 = unemployed/student), income, money spent on games without loot boxes, internet gaming disorder, and problem gambling severity as the predictors and money spent on games with loot boxes as the outcome variable. Given the number of predictors, a large number of models were specified for comparison. Since the models were ordered by predictive adequacy, only the first five models and the models containing internet gaming disorder and problem gambling severity (i.e., the key indicators of harm) were presented (see Table 1).

First, the models containing internet gaming disorder and problem gambling severity individually have low posterior probabilities. Second, after observing the data, the odds in favor of these two models decreased substantially. Finally, compared to the models containing internet gaming disorder and problem gambling severity, the observed data is about 1,250,000,000 times (1/ 0.0000000008) to 129,870,129.9 times (1/ 0.0000000077) more likely under the most probable model (i.e., Model 1 in Table 1), respectively. Model 3 deserves additional attention since it includes problem gambling severity. However, it has a low posterior probability of only 0.048. Although the odds in favor of this model increased by a factor of 8.489 after observing data, the observed data is about 3.086 times (1/0.324) more likely under Model 1 than Model 3.

Table 1 Model comparisons with the first five models and models containing internet gaming disorder and problem gambling severity with money spent on games with loot boxes as the outcome variable

Models	P(M)	P(M|data)	BF_M	BF_{10}	R^2
Money_GwoLB	0.018	0.448	44.676	1.000	0.230
Money_GwoLB + Income	0.006	0.159	31.63	1.066	0.248
Money_GwoLB + PGSI	0.006	0.048	8.489	0.324	0.237
Age + Money_GwoLB	0.006	0.030	5.112	0.199	0.232
Employment + Money_GwoLB	0.006	0.028	4.867	0.190	0.232
.	.	.	.	.	.
.	.	.	.	.	.
Problem Gambling Severity	0.018	0.0000000035	0.0000001910	0.0000000077	0.029
.	.	.	.	.	.
.	.	.	.	.	.
Internet Gaming Disorder	0.018	0.0000000003	0.0000000192	0.0000000008	0.001

Money_GwoLB = Money Spent on Games Without Loot Boxes; P(M) = Prior Probability of Model; P(M|data) = Posterior Probability of Model; BF_M = Bayes Factor of Model Odds; BF_{10} = Relative Predictive Adequacy

The posterior summaries of the variables were also inspected. After observing the data, the posterior probability of including internet gaming disorder is reduced from 0.500 to 0.107 and for problem gambling severity, the probability is reduced from 0.500 to 0.165. Furthermore, internet gaming disorder and problem gambling severity have a $BF_{inclusion}$ of 0.120 and 0.197, respectively. In other words, the data has reduced our odds of including internet gaming disorder by a factor of 8.333 (1/0.120) and for problem gambling severity, by a factor of 5.076 (1/0.197). Taken together, internet gaming disorder and problem gambling severity should be excluded as predictors once key variables (e.g., money spent on games without loot boxes, income) are included in the model. Given the results, I reiterate our conclusion that (a) there is a lack of evidence for the unique harm of loot boxes and (b) it might be worth examining the effects of microtransactions in general (Chew & Neo, 2024).

On a related note, the Bayesian approach is not just useful for examining the evidence in favor of the null hypothesis. Specifically, the Frequentist approach to null hypothesis statistical testing using the strictly defined cutoff of 0.05 as a test of significance has been criticized for decades (Cohen, 1994). Yet, this approach remains the norm in psychology and is extensively used by researchers, including the authors of the commentary. The Bayesian approach offers numerous benefits over the Frequentist approach and should be used as the dominant instead of a complementary approach to hypothesis testing (see Wagenmakers et al., 2018 for a summary). Given the authors' commitment to best practices, I encourage them to ignore the norms in psychology and use Bayesian statistics exclusively in their future papers.

Statistical Approach and Overcontrolling

The authors suggested that the study could have inadvertently controlled for unobserved mediators (e.g., emotional investment in games), resulting in the nonsignificant results (Chew & Neo, 2024). They also provided a hypothetical casual model that included an unobserved mediator. Although this is an interesting possibility, it should be noted that any regression models that have controlled for a few variables could have inadvertently controlled for unobserved mediators. More important, this possibility remains speculation on the part of the authors, and it is unclear how this constitutes a limitation of our study. In fact, this appears to be a direction for future research. Consequently, I urge the authors to collect their own data to test their model rigorously in a future study.

Lack of Engagement with Open Science

The authors asserted that there is a lack of engagement with open science since I did not respond to the request for data and the study was not preregistered, precluding corrections to the supposed fundamental flaws of the study. Indeed, I received a request for data from the first author of the commentary. However, I was made aware of their preprint where they engaged in ad hominem attacks (Xiao et al., 2024). As such, I did not respond to the request for data. Given the updated commentary, I

have made the data available on the Open Science Framework (https://osf.io/vbfsm/?view_only=f49fd95a135f4a7b9905f8d2f4be413d).

Second, the authors appear to adopt an uncritical view towards preregistration. Specifically, the authors have not provided any evidence that preregistration corrects errors and improves science. In fact, preregistration is not a panacea to the problems plaguing psychology and science in general. For example, while a study found that preregistered studies tended to have power analyses and larger sample sizes, the study did not find that preregistered studies have lower number of statistical errors than non-preregistered studies (van den Akker et al., 2024). Despite some benefits, the study concluded that there is no "robust evidence that preregistration prevents *p*-hacking and HARKing (Hypothesizing After the Results are Known)" (van den Akker et al., 2024, p. 5424). Most important, preregistration comes with a wide range of risks and costs that must be considered before it is adopted (see Pham & Oh, 2021 for a summary). Specifically, it is essential for researchers to adopt a critical view towards preregistration instead of blindly using it as a form of virtue signaling (e.g., "my study is good because it is preregistered"). Finally, and perhaps due to these issues, preregistration is not a requirement for publication in *Trends in Psychology*. It is disingenuous to take an uncritical stance towards preregistration, expect it even though it is not a requirement, and upon finding it absent, frame it as a criticism of our study.

Lastly, the authors concluded that our study is fundamentally flawed and could mislead policymakers. Furthermore, they recommended that meta-analyses should exclude the study. The authors' concerns are unwarranted. First, all studies, including our study, have its share of limitations and no single study can provide a definitive answer to a research question. Instead, conceptual and direct replications are required to improve our understanding of a phenomenon. Second, given the limited evidence for the unique harm of loot boxes and the premature advocation for legislation by some researchers, I would argue that some policymakers are *already* misled. I suggest policymakers ensure that their consultants have a basic working knowledge of statistics and are relatively free from any conflicts of interest (declared or undeclared). This procedure would help policymakers obtain the most accurate information about the effects of loot boxes. Finally, there might be some misunderstanding of the process of conducting meta-analyses. More information can be found in a textbook on meta-analysis (Borenstein et al., 2009). The authors could also refer to my meta-analyses as examples (Chew, 2022; Chew et al., in press).

In conclusion, I am thankful to the authors for providing an opportunity to elaborate on some of the decisions made in our study and strengthen the conclusions via Bayesian statistics (Chew & Neo, 2024). It is commendable that the authors actively engage with open science practices and worry about the policymaking process. Nevertheless, it is worth ensuring that the fundamentals are correct. To that end, I recommend that the authors engage with textbooks on elementary statistics (e.g., Gravetter & Wallnau, 2007), multivariate statistics (e.g., Tabachnick & Fidell, 2012), and meta-analysis (e.g., Borenstein et al., 2009).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data Availability N.A.

Declarations

Informed Consent N.A.

Conflict of interest The authors declare that they have no conflict of interest.

References

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine, 18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis* (1st edition). Wiley.

Chew, P. K. H. (2022). A meta-analytic review of Internet gaming disorder and the big five personality factors. *Addictive Behaviors, 126*, 107193. https://doi.org/10.1016/j.addbeh.2021.107193

Chew, P. K. H., & Au, J. Q. M. (2024). Internet gaming disorder in the DSM-5: Engagement vs. addiction. *Journal of Technology in Behavioral Science, 9*(4), 652–660. https://doi.org/10.1007/s41347-023-00364-9

Chew, P. K. H., & Yap, B. L. M. (2021). Mortality salience in an offline and online setting. *Death Studies, 46*(9), 2085–2092. https://doi.org/10.1080/07481187.2021.1888826

Chew, P. K. H., & Neo, G. (2024). Internet gaming disorder and problem gambling severity do not predict money and time spent on games with loot boxes. *Trends in Psychology*. Advance online publication. https://doi.org/10.1007/s43076-024-00425-9

Chew, P. K. H., Yow, Y. J., & Tan, C. S. Y. (in press). Prevalence of problematic pornography use: A meta-analysis. *Sexual Health*.

Chew, P. K. H., Danni, B., Koh, S. Y., & Tay, G. M. S. (2023). Terror management: The effects of mortality salience on desire for money among Singaporeans. *Journal of Humanistic Psychology*. Advance online publication. https://doi.org/10.1177/00221678231203513

Chew, P. K. H., Tan, T. M. X., & Yeo, H. W. (2024). Effects of mortality salience on racism in Singapore. *Journal of Humanistic Psychology*. Advance online publication. https://doi.org/10.1177/00221678241231880

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist, 49*(12), 997–1003.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349–354.

Dong, Y., & Peng, C.-Y.J. (2013). Principled missing data methods for researchers. *Springerplus, 2*(1), 222–239. https://doi.org/10.1186/2193-1801-2-222

Drummond, A., & Sauer, J. D. (2018). Video game loot boxes are psychologically akin to gambling. *Nature Human Behaviour, 2*(8), 8. https://doi.org/10.1038/s41562-018-0360-1

Faulkenberry, T. J., Ly, A., & Wagenmakers, E.-J. (2020). Bayesian inference in numerical cognition: A tutorial using JASP. *Journal of Numerical Cognition, 6*(2), 231–259. https://doi.org/10.5964/jnc.v6i2.288

Garea, S. S., Drummond, A., Sauer, J. D., Hall, L. C., & Williams, M. N. (2021). Meta-analysis of the relationship between problem gambling, excessive gaming and loot box spending. *International Gambling Studies, 21*(3), 460–479. https://doi.org/10.1080/14459795.2021.1914705

Gravetter, F. J., & Wallnau, L. B. (2007). *Statistics for the behavioral sciences*. Wadsworth.

JASP Team. (2023). *JASP* (Version Version 0.18.1) [Computer software]. https://jasp-stats.org/

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. https://doi.org/10.2307/2290157

McCaffrey, M. (2019). The macro problem of microtransactions: The self-regulatory challenges of video game loot boxes. *Business Horizons, 62*(4), 483–495. https://doi.org/10.1016/j.bushor.2019.03.001

Montiel, I., Basterra-González, A., Machimbarrena, J. M., Ortega-Barón, J., & González-Cabrera, J. (2022). Loot box engagement: A scoping review of primary studies on prevalence and association with problematic gaming and gambling. *PLoS ONE, 17*(1), e0263177. https://doi.org/10.1371/journal.pone.0263177

Pham, M. T., & Oh, T. T. (2021). Preregistration is neither sufficient nor necessary for good science. *Journal of Consumer Psychology, 31*(1), 163–176. https://doi.org/10.1002/jcpy.1209

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics*. Pearson Education.

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods, 22*(2), 217–239. https://doi.org/10.1037/met0000100

van den Akker, O. R., van Assen, M. A. L. M., Bakker, M., Elsherif, M., Wong, T. K., & Wicherts, J. M. (2024). Preregistration in practice: A comparison of preregistered and non-preregistered studies in psychology. *Behavior Research Methods, 56*(6), 5424–5433. https://doi.org/10.3758/s13428-023-02277-0

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Xiao, L. Y., Ballou, N., & Eben, C. (2024). *Gambling with peer review: Fundamental flaws in the design and reporting of Chew and Neo (2024)*. OSF. https://doi.org/10.31219/osf.io/yud2x