Dear Mateo,

Thank you and the reviewers for the timely consideration of our Stage 2 Registered Report and the valuable comments made by both you and the two original reviewers of this work. We have revised the manuscript accordingly and outline our response to each comment below. We believe this has improved the quality of the paper substantially and hope this satisfies the requirements for acceptance as a Stage 2 Registered Report.

Kind Regards,

Alexander MacLellan and co-authors.

**Reviewer 1**

**Comment:** In the Stage 2 ms the authors present the results of the study, following the preregistered protocol in every detail and clarifying the few occasions in which they have departed from the initial plan and why. The interpretation of the results does not go beyond what the preregistered analyses permit. Therefore, in my humble opinion, the paper could be accepted essentially as is. I only have a small number of recommendations that the authors might want to consider, although they shouldn't feel obliged to include them in the final version of the manuscript.

*Response:* Thank you for your kind and constructive comments. We respond to each of your minor concerns below and hope these are satisfactory.

**Comment:** I have the feeling that the authors fail to acknowledge the added value of pre-registration in their study. In the general discussion, they point out that based on previous research they expected gamification to have a greater impact. But how many of those previous studies were pre-registered and followed a registered report format? Plenty of research shows that RRs are far less biased and non-registered studies. Is it possible that previous research on gamification suffers from bias and therefore the reported effect sizes are inflated to an unknown extent? Even in the literature on response inhibition training, there is some evidence that, once corrected for publication bias, the average power could be substantially lower than .50, suggesting that the actual effect sizes are lower than anticipated by researchers in their power analyses (see https://onlinelibrary.wiley.com/doi/10.1111/obr.13338).

*Response:* Thank you for this valuable comment; we agree that the Registered Report format is a strength of this study and have now included the following discussion of this on Page 31:

*"Finally, our study addresses previous issues identified in this field, such as low statistical power in response inhibition training studies (Navas et al., 2021) and low study quality in the gamification literature (Vermeir et al., 2020), by conducting this study through the Registered Report publication model. Such model has been found to reduce publication bias (Scheel et al., 2019) and improve research quality compared to non-registered reports (Soderberg et al., 2021). As such, this study may provide more reliable effect size estimates for this research field and contribute to meta-analytic tests of the utility of gamification in response inhibition training."*

**Comment:** The effect of the manipulation on weight loss is analyzed, but no descriptive information is provided, possibly because this analysis was not pre-registered. But this information can be useful in many different ways (e.g., for future meta-analyses). I'd encourage the authors to report the descriptives of weight loss either in a table or a figure.

*Response*: We report participants' mean weight (and standard deviations) in Table 1 ("Weight (kg)"), along with the other measures we analysed, as follows:

Table 1. Participant characteristics and descriptive statistics (Means, *SD*) for all main outcome variables as a function of training group.

| | | Training group | | |
|---|---|---|---|---|
| Variable | | Control (*n*=85) | Feedback (*n*=85) | Social (*n*=81) |
| Age | | 34.57 (13.78) | 34.86 (14.18) | 35.74 (14.51) |
| N Female (%) | | 56 (66%) | 59 (69%) | 54 (67%) |
| Baseline BMI | | 26.50 (5.60) | 27.39 (6.11) | 26.64 (5.57) |
| Weight (kg) | Pre-Training | 75.47 (16.47) | 77.56 (19.19) | 76.18 (17.24) |
| | Post-Training | 73.17 (20.15) | 73.34 (22.22) | 77.15 (18.01) |
| FFQ | Pre-Training | 25.27 (8.45) | 28.64 (7.93) | 25.62 (7.50) |
| | Post-Training | 23.57 (7.31) | 24.52 (8.28) | 23.76 (7.37) |
| Healthy Food Liking | Pre-Training | 49.28 (18.00) | 50.87 (17.27) | 52.61 (17.33) |
| | Post-Training | 55.45 (17.45) | 56.38 (14.51) | 60.12 (15.30) |
| Unhealthy Food Liking | Pre-Training | 61.61 (20.02) | 68.01 (17.53) | 58.20 (23.82) |
| | Post-Training | 65.21(18.93) | 69.03 (19.07) | 65.19 (20.96) |
| Mean Training Sessions | | 3.09 (3.32) | 3.40 (3.65) | 3.16 (3.76) |
| N >1 training session (%) | | 42 (49.4%) | 45 (52%) | 30 (37.0%) |
| Average Daily Motivation | | 49.31 (19.78) | 50.01 (19.36) | 54.47 (16.94) |

**Comment:** In the analysis of RQ4 it is perhaps worth noting that the SESOI entered in the power analysis was relatively large. The conclusion that the effects of both manipulation are identical relies on the assumption that effect sizes below d = 0.46 are too small to matter.

*Response:* For clarity we have specified in our Results that no differences larger than *d* = 0.46 were found between the groups. We have also addressed the need for future research to investigate what constitutes a meaningful effect size in the context of adherence and motivation.

*"Single element gamification may therefore produce effects smaller than we could detect, however, as even small effect sizes in computerised interventions may be meaningful (Carbine & Larson, 2019), future research investigating what constitutes a meaningful effect size for adherence and motivation would be a valuable addition to the literature."*

**Comment:** Figured 4 is not referenced in the main text.

*Response*: Thank you for alerting us to this. This oversight has now been addressed and is referenced on Page 19.

**Reviewer 2**

From the previous round of review

**Comment:** The authors have done a good job of replying to my comments. My only remaining thought is that, while I agree with the authors that measures of implicit associations are often difficult to change and can be frustrating, I would still recommend noting that they were not directly assessed in the discussion section of the manuscript when this comes around. It would be an interesting direction for future research. This might be particularly relevant where discussing that participants might not be aware of the n-go associations they learned and thus have not reported differences on self-reported measures.

*Response:* We have added a comment on page 30 about the lack of measure of implicit evaluations in this study, and that this might be interesting for future research. However, we wish to clarify that we did not show any evidence that participants learned the no-go associations we were trying to train at either the behavioural ('implicit') or explicit level (i.e. there was no improved inhibition to 100% no-go unhealthy foods vs. 50% no-go filler items, and we did not attempt to measure awareness of these associations). The section of the discussion referred to therefore discusses that participants might not have learned the no-go associations we were trying to train, rather than having learned them but not being aware of them.

*"We did not use any implicit measures of food evaluations in this study (as these are less sensitive to no-go training effects than explicit evaluations; Yang et al., 2022), though it may be interesting to include such measures in future research to help interpret any unexpected changes in explicit evaluations, such as the generalised increase in food liking seen in Prolific Academic participants here."*

**Comment:** The authors have now dropped the bayesian statistics which is fine. But I beleive there was also mention of lmer in the original version? Was this also dropped?

*Response*: The reviewer is correct we listed *lmer* as an R package we would use, though this package was not required when carrying out our pre-registered analysis or exploratory analyses and so has been removed from the manuscript. Additionally, we removed our Bayesian analysis plan from the Stage 1 report before gaining In Principle Acceptance.

**Comment:** Minor comment but the tables are a little messy and not formatted per APA or similar.

*Response: We thank the reviewer for highlighting this, and we have now formatted our tables in APA style.*

**Comment:** Violin plots add very little for me, as they seem very consistant with the numerical resutls. Consider moving these to a supplement to keep the manuscript streamlined.

*Response*: We agree with this comment and have moved these to the supplementary materials.

**Comment:** How meeting parametric assumptions, was this tested formally using shapro wilk or similar? or just by inspection of figures? Either is fine with me but I think it warrents mentioning.

*Response*: We have now clarified this was assessed via Shapiro-Wilk tests in the manuscript (page 19)

*"All distributions met parametric assumptions, as assessed by Shapiro-Wilk tests, (all p's > .05)."*

**Comment:** I would like to see more on the argument for single vs. multiple gamified elements to highlight why this research is needed in the discussion (e.g., it is briefly touched on page 29). For me, if multiple gamified elements typically produce big effects, but single ones here only produce small effects, it leaves a big open question about how the multiple gamified elements stack. Is it a cumultative or interactive effect etc. A larger scale study with more combinations of gamified effects would be something very interesting to suggest for future research.

*Response*: We agree this would be an interesting and valuable direction for future research, and have incorporated this suggestion on page 33.

*"Finally, single gamified elements may produce effects too small to be detected in our sample size. However, future research should systematically investigate combinations of game elements to identify the most optimal gamified intervention."*