The effects of isolated game elements on adherence rates in food response inhibition training

Alexander MacLellan[1], Charlotte R. Pennington[2], Natalia Lawrence[3], Samuel J. Westwood[4], Andrew Jones[5], Anna Slegrova[1], Beatrice Sung[1], Louise Parker[1], Luke Relph[1], Jessica O. Miranda[1], Maryam Shakeel[2], Elizabeth Mouka[2], Charlotte Lovejoy[3], Chaebin Chung[3], Sabela Lash[3], Yusra Suhail[4], Mehr Nag[4] and Katherine S. Button[1]

1. Department of Psychology, University of Bath
2. School of Psychology, Aston University
3. Department of Psychology, University of Exeter
4. Department of Psychology, Kings College London
5. School of Psychology, Liverpool John Moores University

Author Note

Alexander MacLellan https://orcid.org/0000-0002-3927-3407

Charlotte R. Pennington https://orcid.org/0000-0002-5259-642X

Natalia Lawrence https://orcid.org/0000-0003-1969-6637

Samuel J. Westwood https://orcid.org/0000-0002-0107-6651

Andrew Jones https://orcid.org/0000-0001-5951-889X

Katherine S. Button https://orcid.org/0000-0003-4332-8789

Correspondence concerning this article should be addressed to Alexander MacLellan, Department of Psychology, University of Bath, Claverton Down, BA2 7AY, United Kingdom. Email: akem20@bath.ac.uk

Abstract

Introduction: Poor diet and the consumption of foods high in fat, sugar and salt are common causes of numerous health conditions and premature mortality. Computerised food response inhibition training (food-RIT) is a type of intervention found to aid weight loss and reduce snacking of these foods with the inhibition of motor responses proposed to operate through the process of stimulus devaluation. However, these interventions are repetitive by nature and suffer from a lack of adherence, leading some to propose gamification as a solution to increase engagement. The effect of gamification is unclear, however, with a lack of research investigating the effects of single game elements in improving adherence to interventions. The current study aims to investigate whether isolated common game elements (social or feedback) improve adherence, engagement and effectiveness of computerised food response inhibition training compared to standard non-gamified intervention. **Methods:** A sample of 252 participants (169 female) were randomly assigned to either a standard non-gamified food response inhibition training, a training gamified with feedback elements, or a training gamified with social elements. Participants completed measures of snacking frequency and food evaluation before and after a 14-day training period, during which they were instructed to complete their assigned training. Training adherence and daily training motivation was recorded during this training period. **Results:** There were no significant effects of adding either feedback or social gamification elements on training adherence, motivation, or effectiveness. **Conclusions:** This study finds no meaningful support for adding isolated game elements to food-RIT with a view to improving intervention adherence. These results raise questions about the magnitude of simple gamification effects, and future research may benefit from assessing the combined effects of multiple gamification elements in such cognitive interventions.

CRediT statement:

Alexander MacLellan: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Original Draft, Writing – Review and Editing, Visualization, Project Administration; **Charlotte Pennington:** Conceptualization, Methodology, Writing – Review and Editing, Supervision, Funding Acquisition; **Natalia Lawrence:** Conceptualization, Methodology, Resources, Writing – Review and Editing, Supervision, Funding Acquisition; **Samuel Westwood:** Conceptualization, Methodology, Writing – Review and Editing, Supervision, Funding Acquisition; **Andrew Jones:** Conceptualization, Methodology, Writing – Review and Editing, Supervision, Funding Acquisition; **Katherine Button:** Conceptualization, Writing – Review and Editing, Supervision, Funding Acquisition; **Gorilla.sc:** Software; **Anna Slegrova:** Investigation; **Beatrice Sung:** Investigation; **Louise Parker:** Investigation; **Luke Relph:** Investigation; **Jessica Oliveira Miranda:** Investigation; **Maryam Shakeel:** Investigation; **Elizabeth Mouka:** Investigation; **Charlotte Lovejoy:** Investigation; **Chaebin Chung:** Investigation; **Sabela Lash:** Investigation; **Yusra Suhail**: Investigation; **Mehr Nag**: Investigation.

Introduction

Poor diet is recognized as one of the leading causes of premature mortality (Roth et al., 2018), and though attempting to reduce calorific intake by dieting is common, it is often unsuccessful (Dohle et al., 2018). This may be due to prolonged calorie deficit increasing the reward value of high calorie food (Stice et al., 2019). Foods high in fat, sugar and salt (HFSS) generate strong reward responses that people susceptible to overeating can struggle to inhibit (Lawrence et al., 2022). Food-specific response inhibition training (F-RIT) targets these automatic responses by training participants to inhibit motor responses to HFSS foods in computerised tasks (e.g., Go/No-Go tasks). Such training has been shown to reduce participants' consumption of the targeted unhealthy food in lab-based studies (Jones et al., 2016), facilitate short-term weight loss in field studies (Veling et al., 2014; Lawrence et al., 2015), and reduce the palatability of HFSS food items (Chen et al., 2016; Veling et al., 2017).

F-RIT involves training participants to inhibit motor responses to unhealthy foods, pairing cues to withhold a motor response with HFSS stimuli, often in comparison with healthy food stimuli and non-food related controls (Jones et al., 2016). The mechanisms by which these interventions operate are still uncertain, though recent reviews propose that the learned motor response inhibition in response to HFSS foods conflicts with the reward responses elicited by the appetitive unhealthy food, which is resolved by a devaluation of the food stimulus (Chen et al., 2018; Lawrence et al., 2022; Quandt et al., 2019). In recent reviews, the effect of F-RIT on food devaluation has been demonstrated (Yang et al., 2022) and is supported by neuroimaging studies which find reductions in activity in reward- and attention-related regions of the brain (Chen et al., 2018; Stice et al., 2017). Though there is promising evidence of the potential benefits of F-RIT, there are questions about adherence rates to computerised cognitive training delivered outside of the lab (Kelders et al., 2012; Lumsden et al., 2017), as inconsistent usage of this intervention, or stopping altogether, can reduce the efficacy of training (Aulbach et al., 2021). Supporting this, Chen et al. (2019) found the effect of F-RIT on food choices to be reduced within weeks, suggesting the need for regular training.

RIT relies on the repetition of motor responses to similar stimuli hundreds of times over the course of an intervention, and engagement levels can waver over time. Engagement is not a well-defined term but can be interpreted as both the experience of completing a task, and the participant's behaviour when interacting with a task or intervention, such as how many sessions an individual completes or the timepoint at which they leave the study (Lumsden et al., 2017; Perski et al., 2017). Gamification – the process of adding game elements to tasks and systems without actually creating a game (Deterding et al., 2011) – presents a potential avenue to increase engagement with cognitive tasks and training. The rise in studies of gamified and game-like programs and tasks has resulted in several reviews to determine whether gamification can enhance intervention effectiveness, though the findings are mixed. Lumsden et al. (2017) and Vermeir et al. (2020) found tentative evidence that gamification can increase motivation and engagement with cognitive tasks, however whether this translates to increased adherence to an intervention schedule is unclear (Lumsden et al., 2017). Najberg et al. (2021) achieved protocol adherence rates of 95% with their food-based Go/NoGo and cued approach training, though there was no non-gamified control group, and the incentive for taking part may have been valuable enough to motivate participants to adhere to the protocol, with incentives known to improve recruitment and adherence rates (Abdelazeem et al., 2022). Aulbach et al. (2021) found a sharp decline in the number of participants using a gamified F-RIT app (FoodT) over the first 5 days of use in an opportunistic study of real-world usage with no incentive on offer. Furthermore, some studies have found that adding gamification elements may actually weaken training effects, which may be explained by game elements creating a distraction from the core stimuli (Forman et al., 2019). Careful consideration of how and when to add game-like elements is therefore important, but these elements are rarely examined in isolation (Seaborn & Fels, 2015; Vermeir et al., 2020), rendering their effects unclear. Game elements should be chosen from a theoretical perspective (Sailer et al., 2017; Vermeir et al., 2022), with self-determination theory (Deci & Ryan, 1985) commonly used to understand the potential role gamification elements play in increasing motivation and engagement in tasks.

Self-determination theory proposes that three psychological needs drive intrinsic motivation; that is, motivation without a need for external reward. The need for competence refers to the desire to feel success when interacting with an environment; relatedness refers to a desire to belong to a group and feel coherent within a social structure; and autonomy refers to the desire to both be free to choose to perform an action and feel performing that action is consistent with one's personal values (Sailer et al., 2017). Ryan et al. (2006) found these constructs predicted future video game-playing behaviour, though whether gamifying otherwise serious tasks affects these motivational constructs is an area needing more research (Seaborn & Fels, 2015). A survey of gamified work-related apps, such as those used for productivity and task management, found no effect of gamification on measures of autonomy, relatedness, or competence (Mitchell et al., 2021), though the apps were varied in purpose and nature of the gamified elements. In a review of the computerised cognitive training literature, Vermeir et al. (2020) found the most common game elements to be those related to achievement and progression, such as point-based systems and feedback loops, and immersion elements such as a story or theme. These elements can be mapped to fulfilling psychological needs as set out by self-determination theory, with Sailer et al. (2017) finding specific elements (e.g. points and leader boards) were rated higher on corresponding psychological needs (e.g. competence). However, there remains a paucity of research investigating the effectiveness of single gamified elements (Mazarakis, 2021). Though recent research has investigated the effects of elements in isolation, as well as when different elements are combined in simple and single session tasks, further work is needed to determine the effects of isolated game elements and whether they lead to changes in intrinsic motivation across more complex and longer interventions (Mazarakis & Brauer, 2022).

The current study therefore aims to examine the effects of isolated gamification elements on F-RIT engagement, adherence, and effectiveness in comparison to a well-established non-gamified intervention control. Our first gamified group adds social elements, by allowing participants to pick and join a team to contribute to, which is thought to increase motivation by fulfilling a need for social relatedness and providing a sense of relevance to their completion of the gamified task (Sailer et al., 2017). Our second gamified group adds feedback elements, thought to increase motivation by

addressing a psychological need for competence according to self-determination theory (Sailer et al., 2017). There are four specific research questions each with associated hypotheses:

RQ1 – Does gamification lead to improved training adherence and training motivation compared to the standard, non-gamified version of food response inhibition training?

H1a – Gamified training groups will have a significantly greater number of completed sessions compared to the non-gamified training control group.

H1b – Gamified training groups will report higher levels of training motivation compared to the non-gamified training control group.

RQ2 – Does gamification improve training effects on food evaluations and snacking?

H2a – There will be a larger decrease in the liking ratings for unhealthy items in gamified groups compared to the non-gamified training control group.

H2b – There will be a larger increase in liking ratings for healthy foods in gamified groups compared to the non-gamified training control group.

H2c – The gamified training groups will display a greater reduction in unhealthy food item snacking in the week following completion of the training compared to the control group.

RQ3 – Does training motivation and adherence mediate training response?

H3a – Pre- to Post-intervention differences in both unhealthy and healthy food item evaluations will be mediated by training adherence.

H3b - Pre- to Post-intervention differences in both unhealthy and healthy food item evaluations will be mediated by training motivation.

H3c – Pre- to Post-intervention differences in snacking frequency will be mediated by training

adherence.

H3d - Pre- to Post-intervention differences in snacking frequency will be mediated by training

motivation.

RQ4 – Is there equivalence between the gamification types for training adherence and motivation?

H4a – Training adherence rates will be equivalent between Feedback and Social gamified training

groups.

H4b – Training motivation will be equivalent between Feedback and Social gamified training groups.

Given the lack of previous work on the effect of gamification on specific components of motivation,

and potential equivalence of training effectiveness between single task gamification groups, we do not

propose to test any hypotheses, however we do state our intention to explore the effects of

gamification here to inform future research.

This project also included measures of stress, personality, dietary behaviours, video game experience

and inhibitory control for the purpose of student dissertation projects, but these do not form part of

any hypothesis tests or exploratory analysis herein.

Methods

Transparency and Openness Statement

All data are publicly available online via the University of Bath data repository archive:

https://researchdata.bath.ac.uk/id/eprint/1415, and materials and code are available on the OSF project

page for this study: https://osf.io/jdk5f/. This study was given In Principle Acceptance on 17th

November 2023 and the preregistered Stage 1 protocol is available at: https://osf.io/jspf3. In the

sections below, we report all manipulations, measures, and exclusions. This study meets Level 6 of

the PCI RR bias control (https://rr.peercommunityin.org/help/guide_for_authors).

Ethical Statement

This study was given a favourable opinion by the Biomedical Sciences Research Ethics Committee at

the University of Bath, with approval number 0260-2006. All participants provided informed consent.

Design

The study utilised a three-arm randomised controlled design, with intervention type as the 3-level

grouping variable (non-gamified food inhibition training [control], achievement-related gamified

inhibition training, social-oriented gamified inhibition training) and pre- and post-intervention as our

repeated measures variable. Participants were randomly allocated to groups using block

randomisation (Moher et al., 2010), with a block size of 3, and blinded to the other training conditions

in the study. Participants were recruited via research participation schemes (e.g. SONA Systems Ltd),

Prolific Academic (https://prolific.co/), and social media platforms (e.g. Twitter, Instagram).

Participants

The inclusion criteria for this study were as follows:

- Aged 18-65 with a Body Mass Index of 18.5 or above (suggesting a 'healthy' weight or

  above), consistent with previous research investigating food RIT training (e.g Lawrence et al.,

  2015).

- Participants reported snacking on either crisps, chocolate, biscuits and/or cake (foods high in

  sugar, salt and fat) at least three times per week, as measured on an unhealthy snacking Food

  Frequency Questionnaire, consistent with previous work investigating the effect of

  computerised response inhibition training (e.g. Lawrence et al., 2015).

- Had access to a stable internet connection and a personal computer or laptop.

The exclusion criteria for this study were:

- A current or previous clinical diagnosis of an eating disorder or diabetes, or self-identifying as having either an eating disorder or diabetes.

- Currently attending a formal weight loss programme or using weight loss medication.

- Currently attempting to quit smoking, due to the changes in appetite and food craving during nicotine withdrawal (Kragel et al., 2019).

Sample Size Estimation

Based on our resources, we estimated that it was possible to recruit 80 participants per group, for a sample size of 240 in total. This allowed us to detect an effect size of $f = 0.23$ with 90% power. Given previous literature finding a large effect of gamification on task engagement, $g = 0.72$ (which we approximate to a Cohens f value of 0.36), with no evidence of publication bias (Vermeir et al., 2020), we believe this to be an appropriate target sample size that yields informative results.

To measure our secondary hypotheses, an effect size of $f = 0.24$ was estimated for devaluation scores based on the previous work of the authors ($d = 0.48$, Lawrence et al., 2015). An a-priori power analysis (G* Power; Faul et al., 2007) indicated that to detect an interaction effect between 3 groups with 2 measurements, a total sample size of 60 is required to achieve 90% statistical power.

Finally, from a power analysis using the TOSTER R package (Lakens & Caldwell, 2021), we would be able to detect equivalence within the parameters $d = -0.46$ and $d = 0.46$ at 80% power with a sample size of 80 per group. We have been more lenient with our target power in this analysis to target relevant effect sizes which correspond to our previously stated effect size of interest (converting from f values of 0.23). Our total target sample size was therefore set at 240 participants.

We achieved this sample size with the final sample comprising 252 participants ($M_{AGE} = 34.94$, $SD = 14.10$, 67% female), which had 90% power to detect effects of $f = 0.22$.

Materials and Measures

GAMIFICATION IN FOOD-RIT

Training Conditions

A non-gamified standard food response inhibition training was used as our active control group and was taken from Lawrence et al. (2015). Pictures of 18 food items, 9 being healthy (e.g. fruit or vegetables) and 9 being high-energy density foods (defined as being greater than 4 kcal/g such as cakes and chocolate), and 18 non-food items (clothes) were presented on either the left or right of a screen. Stimuli were presented for 1250ms, followed by a 1250ms inter-stimulus interval. Participants pressed a key corresponding to the position of the stimuli on the screen ('c' for left and 'm' for right). Stimuli were presented on a white screen with a frame at the border, which turned bold on trials where the participant was instructed to inhibit their response ('no-go' trials, see Figure 1a). Healthy food items were always paired with the 'go' instruction and unhealthy items were always paired with the 'no-go' instruction, whilst non-food items were associated with 'no-go' instructions in 50% of trials. Each of the 36 images were presented once per block, with 6 blocks per training session.

The Feedback gamified food response inhibition training was identical in nature to the training protocol described previously, with the addition of points awarded or deducted based on participant's performance in each block. 10 points were awarded for correct 'go' responses and commission errors resulted in a 5 point deduction, with visual confirmation of each provided after each trial (see Figure 1c). At the end of each training session, participants were awarded a badge, decorated with a number of stars corresponding to the number of training sessions in a row they had completed. This manipulation was chosen as it is a frequently included gamification element (e.g., Vermeir et al., 2020, Vermeir et al., 2022).

The pseudo-socially (Social) gamified food response inhibition training was again identical in nature to the standard response inhibition training, with the addition of a social element at the beginning and end of each training session. Participants selected a team to join upon enrolment in the group: 'green', 'purple', 'yellow' or 'blue'. This then became their team for the duration of the training period, the border around the screen signalling a 'go' or 'no-go' trial changed to their selected team colour around the screen. At the beginning of each training session, participants were informed of their team's position in a league table (see Figure 1b). This position increased by one, remained constant,

or decreased by one at the end of each session. Socially oriented gamified elements are among the

least researched elements (Vermeir et al., 2020).

Figure 1a. Example Standard Non-Gamified Food RIT Control 'Go' Stimulus.

Figure 1b. Example Socially Gamified Food RIT League Table screen.

Figure 1c. Example Feedback Gamified Food RIT 'No-Go' correct response points feedback

Confirmatory Outcome Variables

Food liking and value judgement confidence were measured with a food evaluation task, in accordance with previous research (Lawrence et al., 2015). Twenty-seven pictures of food were presented – the same 9 healthy food and 9 high-energy density food images that were included in the inhibition training, with the other 9 being untrained food images. Participants were asked to rate how much they liked each food image on a 100mm visual analogue scale, anchored at the extremes of "not at all" (0) and "very much" (100), and then asked how confident they were in the evaluation they just gave. Participants selected their rating on the scale with their mouse cursor, which was reset to the midpoint (a rating of 50) at the start of each trial. Both liking and confidence ratings were collected pre- and post-intervention.

Snacking frequency was measured with a Snacking Food Frequency Questionnaire (FFQ; Churchill & Jessop, 2011). This was presented as an 8-item, 7-point Likert-type scale, asking about how frequently different foods (i.e., crisps, cakes, chocolate, and biscuits) were consumed over the previous 2 weeks. Answers ranged from 1 ("not at all") to 7 ("More than 4 times per week"). Responses were summed to form a total score, with higher scores indicating greater snacking.

Adherence was measured as the number of training sessions out of 10 during the 14-day period completed by a participant with an accuracy of >80% on both go and no-go trials.

Training motivation and enjoyment was measured by a questionnaire given at the end of each training session. Scores for each questionnaire were summed to give a daily score, with overall training motivation approximated as the mean of all daily scores. The questionnaire was adapted from Lumsden et al. (2017) and included the following items: "1. How enjoyable did you find that?; 2. How frustrating did you find that?; 3. How mentally stimulating did you find that?; 4. How repetitive did you find that?; and 5. How willing would you be to do that again tomorrow?" Participants recorded their answer to each item on a 100mm visual analogue scale with "not at all" at one end, to "very" at the other with no subdivisions. After reverse scoring questions 2 and 4, mean item scores were calculated for each participant, with higher values indicating greater subjective enjoyment.

Exploratory Outcome Variables

Intrinsic motivation was assessed by the Intrinsic Motivation Inventory (IMI; Deci & Ryan, 2005) to refer specifically to the training task (see Appendix A). This is a 19-item questionnaire with answers ranging from 1 ("Not all") to 7 ("Very true") with items measuring relatedness (e.g. "I felt connected with the others taking part in this study"), autonomy (e.g. "I believe I had some choice about completing each training session"), and competence ("I think I did pretty well at this training, compared to other participants"). After reverse scoring relevant items, scores were summed to create total scores for each subscale, with higher scores indicating greater sense of relatedness, autonomy, and competence.

Lifestyle factors was assessed with a 100mm visual analogue scale as before, set at the midpoint (a value of 50), anchored at either end with "No" (0) and "Yes" (100). Questions asked about a participant's sleep quality from the previous night ("Did you get enough sleep last night?"), whether they were currently hungry ("Are you currently hungry?") and whether they were hungover ("Are you currently hungover") were shown after each training session.

Attention checks

To protect against careless responding (Jones et al., 2022) participants were asked the multiple-choice question, "What planet do you live on?" with the available response options of: "Earth, Mars, Mercury, Saturn" after the food evaluation task (taken from Pennington et al., 2023). Analyses were conducted with and without participants who failed this attention check.

Procedure

We used the Gorilla Experiment Builder (www.gorilla.sc) to create and host our experiment (Anwyl-Irvine et al., 2018).

Potential participants first completed a screening stage consisting of completing the Food Frequency Questionnaire (Churchill & Jessop, 2011). They were also asked their height and weight, and this was then converted to yield their BMI (kg/m2). Data collection occurred over two testing sessions: baseline and post-intervention. Heights below 122cm and above 213cm (48 and 84 inches), or weights below 34kg and above 227kg (75 and 500 pounds) were deemed implausible, and therefore not included in any descriptive statistics (Das et al., 2005). At baseline, we recorded participants' self-reported weight, food evaluations and snacking frequency, alongside measures of stress, personality, dietary behaviours and inhibitory control for student dissertation projects. Between the two testing sessions, participants were asked to complete 10 sessions of the online training intervention over 14-days. Finally, participants competed all measures again post-intervention. Participants were paid £6 per hour, in line with the guidelines of Prolific Academic and the Lead institution.

Planned Statistical Analyses

Data Exclusion Criteria

Outliers on outcome measures were retained for analyses and participants who completed less than 10 training sessions were retained in order to achieve an unbiased effect estimate of the intervention

(McCoy, 2017). Participants who averaged below 80% accuracy on training sessions or whose accuracy score was greater than 2 standard deviations higher than the group mean were excluded from the analysis. Missing data was assessed for randomness using Little's test (Little, 1986), and a missing dummy variable was created and tested for associations with group type.

Planned Analyses

All analyses were carried out in R Studio, with details of all packages used included in the Supplementary Materials. The null hypothesis was rejected when $p < .05$, or when p > .05 for equivalence tests.

RQ1 – Does gamification lead to improved training adherence and motivation compared to non-gamified training?

All hypotheses relating to this question were tested with a one-way factorial Analysis of Variance (ANOVA) with training group as the factor, and mean motivation score and number of training sessions completed as the two dependent variables.

RQ2 – Does gamification improve training effects for food evaluations and snacking?

All hypotheses addressing this research question were tested with a series of 3 (Control, Social Gamification, Reward Gamification) x 2 (Time, Pre – Post) mixed design ANOVA tests. Significant main effects were followed up with an analysis of simple main effects to investigate the direction of the effect found.

RQ3 – Does motivation and adherence mediate training response?

Mediation analyses were conducted using a causal steps approach, as suggested by Hayes and Rockwood (2017), for each hypothesis separately. Significance of the mediation model was determined using the bootstrap method, based on 5000 bootstrap samples (consistent with the

recommended number from Hayes, 2018). Mediation coefficients were established with separate

regressions: first with intervention group entered as a predictor (dummy coded with the control set as

the reference category) and change in food item evaluations and in unhealthy food snacking frequency

as the two outcome variables. Secondly the direct effect of intervention group on change in training

engagement or motivation was established, followed by establishing the indirect effect with both

intervention group and change in the mediator as our predictor variables and change in food

evaluation score and snacking frequency as our two outcome variables. For all models, baseline

scores of our variables entered as covariates and to establish the significance and confidence intervals,

the R package '*mediation*' was used (Tingley et al., 2014).

RQ4 – Is there equivalence between the gamification types for training adherence and motivation?

All hypotheses for this research question were tested with two one-sided *t*-tests, using the R package

TOSTER (Caldwell, 2022). Equivalence was concluded when the 95% confidence intervals fell

between $d = -0.46$ and $d = 0.46$.

Planned Exploratory Analysis

Whilst not formally testing a specific hypothesis, we also investigated whether gamification resulted

in fewer outliers (scores greater than 2 standard deviations from the group mean) during training

performance.Second, we explored whether there were differences between the training groups on

measures of perceived competence, autonomy, and relatedness, in line with Sailer et al.'s (2017)

categorisation of gamification elements with self-determination theory. Third, we explored whether

gender moderates the relationship between training effects and adherence rates based on previous

findings that suggest gender moderates training efficacy in a largely theme-based gamification task

(Forman et al., 2021).Finally, we explored relationships between change in food item evaluation

confidence, food liking ratings and training adherence.

Results

Participants

Participant characteristics and descriptive statistics are presented in Table 1. One participant failed the attention check and was excluded from all analyses. Pre-registered analyses with this participant included were consistent with the reported analyses below and are included in the Supplementary materials. Consistent with our registered plan, for RQ's 2 and 3, and all exploratory analyses investigating food item ratings and snacking frequency, we excluded 51 participants who scored less than 80% accuracy average across all training sessions, and 9 participants who did not complete a single training session. Analyses were conducted with these participants included and are reported in the Supplementary materials. Little's test suggested data was missing completely at random, $\chi^2 = 0$, $p = 1.0$.

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

Variable		Control (*n*=85)	Feedback (*n*=85)	Social (*n*=81)
			Training group	
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)

Note: BMI = Body Mass Index; FFQ = Food Frequency Questionnaire.

RQ1 – Does gamification lead to improved training adherence and training motivation compared to the standard, non-gamified version of food response inhibition training?

One-way factorial ANOVAs were conducted in accordance with our registered analysis plan. There were no significant differences between training groups on the number of training sessions completed, $F(2, 248) = 0.17$, $p = .848$, $\eta_p^2 = .001$, or on mean motivation scores $F(2, 214) = 1.40$, $p = .250$, $\eta_p^2 = .012$, failing to support either H1a and H1b.

Unplanned exploratory analysis - exploring the effect of recruitment method:
Given that the adherence data is a count variable, and with evidence of overdispersion, a negative binomial regression was also conducted, exploring the effect of including recruitment method (research participation schemes, Prolific Academic, or social media platforms) as a predictor (Table 2). There was a significant effect of being recruited through Prolific compared to the social media ($exp(b) = 0.64$, $p < .001$, 95% CI [0.47, 0.83]); and university participation panels ($exp(b) = 0.45$, $p = .002$, 95% CI [0.28, 0.74]) on adherence rates. This suggests that those recruited through Prolific, or university participation schemes were less likely to adhere to the training than those recruited through the social media. There was still no significant effect of training group, and no interactions between training group and recruitment method. Given this effect, we determined recruitment method should be included as a covariate in our analyses for RQ2 and RQ3, and so report outcomes for our registered hypothesis tests both with and without it.

In summary, we found no evidence that adding either Feedback or Social game elements improved adherence or motivation compared to a standard F-RIT training.

GAMIFICATION IN FOOD-RIT

Table 2. Results of negative binomial regression of training group and recruitment method on adherence.

	Coefficient	SE	p	LCI	UCI
Control vs Feedback Group	1.13	0.15	.413	0.84	1.52
Control vs Social Group	1.06	0.15	.705	0.76	1.43
General vs Prolific recruitment	**0.64**	**0.14**	**.001**	**0.49**	**0.83**
General vs University RPS recruitment	**0.45**	**0.25**	**.002**	**0.28**	**0.74**

RQ2 – Does gamification improve training effects on food evaluations and snacking?

A series of mixed 3 (Group: control, feedback, social) x 2 (Time: pre-post training) ANOVAs were conducted for Healthy and Unhealthy food liking ratings and food item snacking frequency, with results presented in Table 3. All distributions met parametric assumptions, as assessed by Shapiro-Wilk tests, (all p's > .05). There was a significant main effect of Time for each outcome, indicating that participants' ratings for both Healthy, $F(1, 125) = 41.67$, $p < .001$, $\eta_p^2 = .260$ and Unhealthy food items, $F(1, 125) = 4.76$, $p = .03$, $\eta_p^2 = .039$ increased from pre- to post-training, as well as a reduction in snacking frequency, $F(2, 125) = 20.52$, $p < .001$, $\eta_p^2 = .141$. There was no significant Group x Time interaction effect on either Healthy or Unhealthy food item liking ratings, and therefore we found no support for either H2a or H2b (see Fig 2). Healthy and Unhealthy food item confidence ratings did not change significantly over time, nor was an interaction effect found, with these results presented in the Supplementary materials.

There was a significant two-way interaction between Group and Time for snacking frequency $F(2, 125) = 3.84$, $p = .024$, $\eta_p^2 = .054$. Follow up simple main effects indicated a significant effect of Group at baseline after correcting for multiple comparisons, $F(2, 188) = 4.64$, $p = .022$, $\eta_p^2 = .047$, but no significant effect of Group at the post-intervention testing session, $F(2, 125) = 0.07$, $p > .999$, $\eta_p^2 = .001$. Pairwise comparisons indicated that pre-training FFQ scores for the Feedback group were significantly higher than the Control group ($p = .008$), but no other comparisons were significant, suggesting baseline differences drove this interaction effect, thus failing to support H2c.

GAMIFICATION IN FOOD-RIT

Unplanned exploratory analysis – including recruitment method as a covariate

Including recruitment method as a covariate changed several results. There was no longer a significant main effect of Time for Healthy food item ratings, $F(1,123) = 2.10$, $p = .149$, $\eta_p^2 = .002$; Unhealthy food item ratings, $F(1,123) = 0.02$, $p = .883$, $\eta_p^2 < .001$; or snacking frequency, $F(1,123) = 0.43$, $p = .513$, $\eta_p^2 = .001$. As before, there was no significant main effect of Group, and the Group*Time interaction remained significant for snacking frequency, $F(2, 123) = 4.07$, $p = .020$, $\eta_p^2 = .02$, again suggesting baseline group differences drove this interaction. No other Group x Time interactions were significant. There was also a significant main effect of recruitment method for Healthy food item ratings, $F(2, 123) = 3.97$, $p = .021$, $\eta_p^2 = .05$; and snacking frequency, $F(2, 123) = 13.60$, $p < .001$, $\eta_p^2 = .14$. Finally, there were no significant interactions between time and recruitment method for Unhealthy food item ratings, $F(2, 123) = 1.45$, $p = .237$, $\eta_p^2 = .002$; and snacking frequency, $F(2, 123) = 2.27$, $p = .108$, $\eta_p^2 = .01$, but there was a significant interaction for Healthy food item ratings, $F(2, 123) = 5.19$, $p = .007$, $\eta_p^2 = .01$. Paired *t*-tests for each recruitment method found that participants recruited through Prolific Academic showed a significant increase in Healthy item food ratings from pre ($M = 47.15$, $SD = 18.16$) to post-training ($M = 56.18$, $SD = 16.61$), $t(84) = -6.93$, $p < .001$, $d = -0.75$, but there were no significant differences for the other two recruitment methods.

These results suggest the Group x Time interaction for snacking frequency is robust, and Healthy item ratings and Snacking Frequency differed between the recruitment streams.
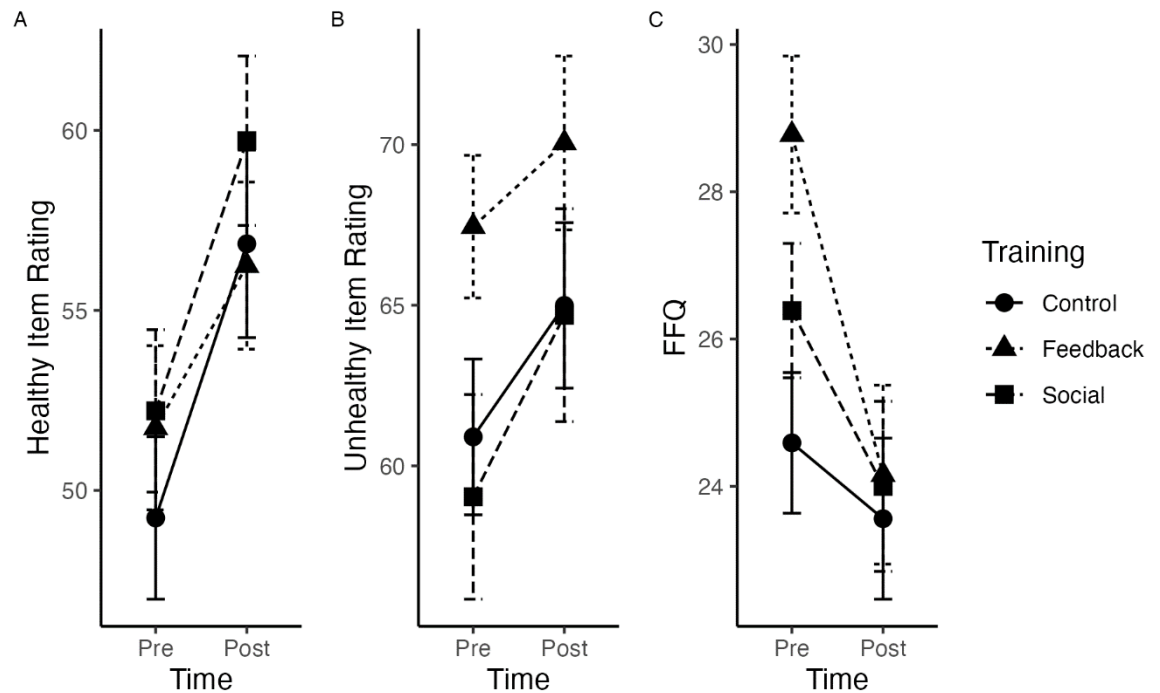
GAMIFICATION IN FOOD-RIT

Figure 2 - Line Graphs of Food Item Ratings and Snacking Behaviour Pre- and Post-Training

Table 3. Results of mixed 3 (Group) x 2 (Time) ANOVAs for food item ratings and snacking behaviour

	F	p	η_p^2
Healthy Food Item ratings			
Group	0.35	.707	.006
Time	**41.67**	**< .001**	**.260**
Group x Time	0.46	0.632	.009
Unhealthy Food Item ratings			
Group	1.11	0.334	.017
Time	**4.76**	**.031**	**.039**
Group x Time	0.40	.987	< .001
Snacking Behaviour			
Group	2.14	.122	.024
Time	**20.52**	**< .001**	**.041**
Group x Time	**3.84**	**.024**	**.016**
Healthy Rating Confidence			

23

GAMIFICATION IN FOOD-RIT

Group	0.26	.770	.004
Time	0.13	.711	.001
Group x Time	0.59	.555	.009
Unhealthy Rating Confidence			
Group	0.28	.756	.004
Time	1.86	.176	.017
Group x Time	0.75	.471	.015

Note: Bold denotes p < .05

Unplanned exploratory analysis – Training manipulation check and the effect on weight

Consistent with previous research with food response inhibition training (e.g. Lawrence et al., 2015), we looked for evidence of a stimulus-response relationship during the training sessions. Reaction times for 'go' trials were significantly quicker for Healthy food ($M = 858.43$ms, $SD = 69.89$ms) than filler images ($M = 872.94$ms, $SD = 77.97$ms), $t(184) = 5.46$, $p < .001$, $d = 0.40$, which is consistent with participants learning to respond ('go') to healthy foods. There were no significant differences in commission errors between Filler ($M = 0.97$, $SD = 0.04$) and Unhealthy ($M = 0.97$, $SD = 0.03$) food NoGo trials, $t(184) = -1.75$, $p = .082$, $d = -0.13$ which may suggest weak learning effects to inhibit responses to unhealthy foods.

A linear regression investigating the effect of both training group and recruitment method on overall training accuracy found no significant effects (see Supplementary materials for model parameters), suggesting there were no performance differences between the groups, and participants from all recruitment methods performed similarly during the training task.

Finally, we conducted a mixed 3 (Group) x 2 (Time) ANOVA to investigate the effect on participant's self-reported weight. There was a significant main effect of Time, $F(2, 121) = 4.03$, $p = .046$, $\eta_p^2 = .03$, indicating that participants' weight decreased across the training period, but there was no significant main effect of group or two way interaction suggesting gamification did not influence weight change.

GAMIFICATION IN FOOD-RIT

In summary, we found that adding Feedback or Social game elements did not accentuate changes in healthy or unhealthy food item evaluations, but those assigned to the Feedback group experienced a greater decrease in snacking behaviour.

RQ3 –Does training motivation and adherence mediate training response?

Coefficients for each step in the mediation models are presented in Table 4. Bootstrapped confidence intervals found no significant mediating effect of training adherence or motivation scores on baseline to post-intervention differences for Healthy or Unhealthy food liking judgements or change in Snacking frequency, thus contrary to H3a, H3b, H3c, and H3d. However, two direct effects in the regression models were of note: First, assignment to the Feedback group, as compared to the control, predicted a greater reduction in snacking frequency, b = -4.54, p = .007, 95% CI [-7.71, -1.26]. Second, adherence to training predicted a decrease in Unhealthy food liking judgements, b = -0.70, p = .015, 95% CI [-1.26, -0.14].

Unplanned exploratory analysis – including recruitment method as a covariate

When including recruitment method in the models, outcomes for the tests of mediation effects were unchanged, and support for the effect of motivation on the change in healthy item ratings (see Table 4) remained, as did the effect of assignment to the Feedback group on change in Snacking frequency, b = -4.51, p = .005, 95% CI [-7.64, -1.39]. There were some differences in the models, namely the effect of assignment to the Social gamification group on mean daily motivation scores met the significance threshold, b = 9.55, p = .016, 95% CI [1.78, 17.32], whilst the effect of training adherence on change in unhealthy item ratings instead suggested only weak evidence for this effect, b = -0.53, p = .063, 95% CI [-1.09, 0.02].

Table 4. Parameter estimates for mediation models

GAMIFICATION IN FOOD-RIT

	Model	Estimate (SE)	LCI	UCI	p
Feedback vs Control	Adherence	0.15 (0.87)	-1.58	1.88	.866
Social vs Control	Adherence	0.46 (0.88)	-1.28	2.19	.603
Feedback vs Control	Motivation	3.51 (4.03)	-4.47	11.49	.386
Social vs Control	Motivation	7.67 (4.10)	-0.45	15.80	.064
Adherence	Δ Healthy Ratings	0.04 (0.28)	-0.51	0.59	.898
Feedback vs Control	Δ Healthy Ratings	-2.00 (2.71)	-7.37	3.36	.462
Social vs Control	Δ Healthy Ratings	0.37 (2.73)	-5.03	5.78	.892
Indirect effect		0.01	-0.51	0.97	.97
Motivation	**Δ Healthy Ratings**	**0.17 (0.06)**	**0.05**	**0.29**	**.006**
Feedback vs Control	Δ Healthy Ratings	-2.51 (2.69)	-7.83	2.82	.353
Social vs Control	Δ Healthy Ratings	-1.23 (2.77)	-6.76	4.20	.645
Indirect effect		0.60	-0.86	2.48	.39
Adherence	**Δ Unhealthy Ratings**	**-0.70 (0.28)**	**-1.26**	**-0.14**	**.015**
Feedback vs Control	Δ Unhealthy Ratings	0.53 (2.76)	-4.94	6.00	.848
Social vs Control	Δ Unhealthy Ratings	0.67 (2.78)	-4.84	6.17	.810
Indirect Effect		-0.10	-1.45	1.14	.85
Motivation	Δ Unhealthy Ratings	-0.04 (0.06)	-0.17	0.09	.506
Feedback vs Control	Δ Unhealthy Ratings	0.25 (2.87)	-5.44	5.94	.890
Social vs Control	Δ Unhealthy Ratings	-0.41 (2.96)	-6.27	5.45	.890
Indirect Effect		-0.11	-0.87	0.99	.88
Adherence	Δ Snacking	-0.08 (0.17)	-0.42	0.25	.618
Feedback vs Control	**Δ Snacking**	**-4.54 (1.66)**	**-7.81**	**-1.26**	**.007**
Social vs Control	Δ Snacking	-2.72 (1.67)	-6.02	0.58	.105
Indirect Effect		-0.01	-0.35	0.36	.92
Motivation	Δ Snacking	-0.03 (0.04)	-0.11	0.04	.419
Feedback vs Control	**Δ Snacking**	**-4.69 (1.70)**	**-8.05**	**-1.33**	**.007**
Social vs Control	Δ Snacking	-2.81 (1.66)	-6.04	0.87	.141
Indirect Effect		-0.11	-0.65	0.36	.71

In summary, we found no evidence that adherence or motivation mediated the relationship between training group and change in Healthy 'liking' ratings, Unhealthy 'liking' ratings or snacking behaviour. We again found assignment to the Feedback group predicted a greater change in snacking behaviour, and additionally found that greater adherence to the training predicted a decrease in Unhealthy food item ratings.

RQ4 – Is there equivalence between the gamification types for training adherence and motivation?

Two one-sided *t*-tests found that training adherence rates, $t(164) = -.40, p = .345, g = 0.06$, and mean daily motivation scores, $t(164) = 1.41, p = 0.920, g = 0.24$, were statistically equivalent between the two groups, suggesting there were no differences equivalent to or larger than $d = 0.46$ in these two outcomes between the gamification types.

Pre-registered Exploratory Analyses

Full results of our pre-registered exploratory analyses are presented in the Supplementary materials, though we present the main findings here.

A binomial logistic regression investigating the effect of training group on whether a participant fell 2 standard deviations away from their group training accuracy mean was not significant $\chi^2(2) = 1.59, p = .451$. This suggested that gamification did not change training performance.

Linear regression models were conducted to explore the effect of gamification on intrinsic motivation, as defined by Sailer et al.'s (2017) categorisation of gamification elements according to self-determination theory. There was evidence of an effect of adding Feedback elements on a measure of perceived autonomy at completing the training tasks, $b = 1.90, p = .047$, 95% CI [0.03, 3.78], though there was no effect of gamification on measures of perceived competence or relatedness, $p > .05$.

GAMIFICATION IN FOOD-RIT

Given there was only an effect of adherence on change in Unhealthy food item liking ratings, as determined in our analyses for RQ3, a moderated regression investigating the interaction between gender and adherence was only conducted with change in Unhealthy food item liking ratings as an outcome. As in Forman et al. (2021) we restricted our gender variable to men and women for this analysis. Adherence data was centred, and there was no significant evidence of an interaction between gender and adherence to the training, $b = 0.20$, $p = .737$, 95% CI [-0.99, 1.40].

We explored the relationship between changes in food item liking ratings, rating confidence, and training adherence with Spearman's rank correlations given the non-normal distribution of our adherence data (see Table 5). Change in Unhealthy food confidence ratings were positively associated with a change in the liking rating of that item ($rho = .21$, $p = .015$), suggesting as individuals were less confident in their ratings, their liking of those items reduced. The change in Unhealthy item ratings was also positively associated with change in healthy item ratings ($rho = .21$, $p = .010$) suggesting that changes in endorsement of these items occurred in tandem. There were no other significant relationships between these variables, all $p > .05$.

Table 5. Correlation analyses between food liking ratings, rating confidence and training adherence.

	Coefficient	p	LCI	UCI
Δ Healthy Ratings - Δ Healthy Ratings Confidence	.05	.561	-.09	.19
Δ Healthy Ratings - Adherence	.01	.873	-.13	.16
Δ Unhealthy Ratings - Δ Unhealthy Ratings Confidence	**.21**	**.015**	**.08**	**.35**
Δ Unhealthy Ratings - Adherence	-.15	.090	-.29	.01
Δ Healthy Ratings - Δ Unhealthy Ratings	**.21**	**.010**	**.07**	**.33**
Δ Snacking - Adherence	.07	.413	-.07	.21
Δ Snacking - Δ Healthy Ratings	-.07	.436	-.21	.07
Δ Snacking - Δ Unhealthy Ratings	-.03	.754	-.17	.11

Unplanned Exploratory Analysis – The influence of lifestyle factors on training performance and engagement.

Finally, we explored the relationship between lifestyle factors, daily training accuracy and daily motivation with Pearsons's correlations (see Table 6). Daily motivation was positively associated with training accuracy ($r=.19$, $p < .001$), but no other significant relationships were present, all $p > .05$.

Table 6. Results of correlation analyses between lifestyle factors, daily training accuracy and daily motivation

	Coefficient	p	LCI	UCI
Hunger - Accuracy	-.05	.209	-.12	.03
Tiredness - Accuracy	.07	.068	-.00	.14
Hungover - Accuracy	.01	.683	-.06	.09
Daily Motivation - Accuracy	**.19**	**< .001**	**.13**	**.26**
Hunger – Daily Motivation	.02	.659	-.06	.09
Tiredness – Daily Motivation	.05	.262	-.03	.11
Hungover – Daily Motivation	.03	.430	-.04	.10

Discussion

With little research investigating the effect of single gamified elements on engagement motivation for computerised cognitive interventions (Mazarakis & Brauer, 2022), and given the calls for freely available software to gamify experiments (Long et al., 2023), we used Gorilla, a widely used experiment builder with an integrated game builder, to investigate the effect of isolated game elements (feedback or social) on adherence, motivation, and the effectiveness of computerised food response inhibition training. Contrary to our hypotheses, we found no statistically significant evidence suggesting that the addition of isolated game elements (feedback or social) influenced either training adherence or daily motivation compared to a non-gamified control intervention. We found a reduction in snacking frequency across the training period for all groups. We also found liking ratings increased for both healthy and unhealthy food items from pre- to post-training, the latter of which was against our predictions.

Against our main hypothesis, adding isolated game elements did not significantly improve training adherence and motivation. Whilst the two gamified groups exhibited slightly higher adherence and motivation scores on average, this was not significantly different from the control group. There was weak evidence for an effect of assignment to the Social gamification group on daily training motivation, though this met significance criteria only after adjusting for recruitment method. These results are somewhat consistent with previous reviews suggesting that gamified tasks may be more engaging or motivating than non-gamified tasks (Lumsden et al., 2016; Vermeir et al., 2020) but this does not appear to translate into improved adherence rates (e.g., Lumsden et al., 2017). Given the evidence for equivalence between the two gamification groups on measures of adherence and motivation and lack of effect in our registered analyses, it is probable that isolated game elements may produce effects too small ($f^2 = 0.04$ for adherence and $f^2 = 0.11$ for motivation) to be detected with our sample size. Our sample size estimation was informed by previous work finding large effects of gamification (Vermeir et al., 2020), with previous research mainly implementing multiple gamified elements. 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.

We selected feedback and social elements based on Sailer et al.'s (2017) classification of gamified elements into measures of intrinsic motivation: competence, relatedness and autonomy. In exploratory analyses, we found that participants assigned to the Feedback group displayed significantly higher levels of perceived autonomy; that is, they felt it was their choice to complete training sessions rather than something they felt obliged to do as part of the study. However, there were no significant differences between the groups on the expected concepts of competence or relatedness. The social gamified elements aimed to recreate social elements described by Sailer and colleagues to foster a sense of relatedness and obligation to others through picking a team, the colour scheme of the programme adapting to that theme, and a competitive element through a leaderboard. However, given that there was no real social interaction with other members of the team, it may be

that these social elements were not effective. As social elements are amongst those rarely implemented and measured (Vermeir et al., 2020), further research is required on how to best implement them.

There was also a lack of evidence that adding feedback elements increased feelings of competence at completing the training, against expectations. Unlike adding social elements, there is precedence for each of the feedback elements in isolation, with Lumsden et al. (2017) adding a points variation to a similar task relying on motor inhibition They found participants in the points variant group had the highest average enjoyment rating of all conditions, however we found no such effects, either in daily motivation or in the intrinsic motivation measure completed in the post-training testing session. Our findings are, however, consistent with that of Mekler et al. (2017) who found no effect of gamification on measures of intrinsic motivation. Nevertheless, whilst self-report measures of motivation and enjoyment showed no differences, those assigned to the feedback group had a higher average accuracy on the training task in our exploratory analysis, suggesting improved performance whilst completing the task.

Recreational computer games are likely intrinsically motivating as players find themselves in a cycle of being presented with a challenge and expending effort to surmount that challenge (Deterding, 2015). The F-RIT task in this study may not have been challenging enough to induce a feeling of competence or achievement in having completed the task with a high accuracy rate, given the consistent ratio of 'Go' to 'NoGo' stimuli. Previous research suggests that the addition of gamified elements is unlikely to produce any additional benefit if the task is not sufficiently challenging (Mekler et al., 2017).

Taken together, our findings may suggest an improved engagement with the training task whilst it was being performed, though this does not relate to motivation to complete the task. This has implications for gamification of non-adaptive cognitive tasks, such as F-RIT, which often requires a consistent pairing of healthy images with 'Go' responses and unhealthy images with 'NoGo' responses. Such training may benefit more from other gamified elements that either improve the visual appeal of the training or increase task complexity.

Our hypotheses that gamification would accentuate changes in healthy (H2a) and unhealthy (H2b) food liking ratings were also not supported. Instead, ratings for healthy and unhealthy items increased across the training period for all three groups. We had expected ratings for unhealthy foods to decrease over time, consistent with many studies demonstrating devaluation effects for trained no-go foods (Lawrence et al., 2022; Yang et al., 2022). Indeed, we observed an association between greater training adherence and the expected devaluation of unhealthy foods here, aligning with previous reports that more training is associated with greater reductions in liking and intake of unhealthy no-go foods (Aulbach et al., 2021), including a study very similar to this one (Adams et al., 2021). It may be that the lack of overall unhealthy food devaluation resulted from a lack of robust learning of unhealthy food-no-go associations during training (our manipulation check failed to demonstrate such learning effects) because previous research suggests that attention to and memory (awareness) of food-no-go contingencies is an important determinant of reductions in liking and preference (Chen and Veling, 2022; Camp and Lawrence, 2019).

Rather, our findings suggest that participants showed stronger learning of the healthy food-go association, which can increase food liking (Chen and Veling, 2022), and this may have generalised to all foods. Another potential reason for the lack of devaluation effects was the long time lag between the final session they completed (e.g., if they stopped training after the second day) and the next evaluation of food items (e.g., day 15). Evidence suggests that training-related increases in preference for 'go' foods is more robust and lasts longer than decreased preference for 'no-go' foods, although the latter has been shown to last for up to one week (Chen & Veling, 2022). Nevertheless, caution is required in interpreting the present increased ratings of foods over time, given this disappeared after including recruitment method as a covariate. Our exploratory analysis found that participants recruited via Prolific Academic increased their liking ratings for all food items, even untrained foods, with no differences found in those recruited by other methods. It may be that Prolific participants are more experienced in completing cognitive tasks and adopted a different strategy during the training, e.g., only attending to the 'Go' signals or trials. 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.,

GAMIFICATION IN FOOD-RIT

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.

Consistent with previous research, snacking frequency decreased for all groups. This may suggest that the training worked as expected in reducing consumption of no-go foods (albeit without their devaluation) or it may reflect general effects of participation, such as expectancy effects. For example, participant expectations about the intervention may affect their behaviour or responses, or self-monitoring may increase as a result of answering questions about their diet (Lawrence et al., 2022). This may also explain the reduction in weight across all groups. Compared to the control group, participants assigned to the Feedback group showed significantly lower snacking frequency (consistent with H2c), but we are cautious about over interpreting this result given it was not mediated by adherence or motivation (H3c) and may be due to baseline differences.

In summary, we did not find support for our hypotheses that gamification would influence training effects, though we cannot make further inferences as to whether gamification may improve or hinder training effects (as in Forman et al., 2019) given the generally poor adherence to the intervention.

Strengths, limitations, and future directions

One strength is that our study did not incentivise completion of the training sessions. Incentives have been shown to improve adherence rates (Abdelazeem et al., 2022), potentially conflating the effect of training with the monetary reward. Our findings, therefore, offer a more accurate estimate of the effect of implementing single gamified elements in F-RIT on training adherence and motivation. Another strength was the inclusion of a non-gamified control group, which allowed us to isolate any additive effect of gaming elements as well as address the dearth of non-gamified control groups in previous gamification research (Johnson et al., 2016, Vermeir et al., 2020). 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.

There are, however, several limitations which can inform future research in this area. First, we recruited participants from a variety of sources, namely Prolific Academic, university research participation panels, and generally through researcher adverts, contact lists and social media. These groups differed in their demographics (e.g., BMI), and adherence, which required us to conduct exploratory analyses to assess the influence of such recruitment procedures. Future research should either account for this in any pre-specified analysis plans or seek to recruit with only one method; yet with this latter recommendation, researchers should acknowledge that their participant sample is likely to differ based on their chosen recruitment method, which may affect generalisability.

Although we screened our sample for relevant characteristics such as snacking frequency and implemented strict exclusion criteria to ensure those taking part in weight loss interventions did not take part, we did not specify any dieting intentions as part of our inclusion criteria. The majority of previous research in this field has been conducted in a motivated sample (e.g. those wishing to reduce unhealthy eating behaviours or with a desire to lose weight, with some studies not offering any incentives), and it may be that, although our results are appropriate for a casual user, they are not representative of those who are most likely to utilise F-RIT. Capturing participant willingness or desire to either change their diet or lose weight would be a useful addition to future research. App-based F-RIT shows promise as an accessible method of delivery (Aulbach et al., 2021) and would represent a logical next step when exploring the effect of gamification in this field. Apps provide other means of increasing adherence, such as reminders and notifications and are instantly accessible through mobile devices, yet research finds relatively low adherence rates (e.g., Aulbach et al., 2021), and so may benefit from the addition of multiple game elements. 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.

Conclusions

This study set out to investigate whether single element gamification improved adherence and motivation to F-RIT and whether these may mediate the training's effectiveness. We found no meaningful evidence for the effect of adding single game elements on food-RIT adherence or motivation. There was also a lack of evidence that gamification alters training effects (e.g., food liking ratings, snacking behaviour), which may be explained by the generally poor adherence rates across the sample. With a view to increasing adherence and motivation for cognitive interventions, such as F-RIT, we recommend that future research increases the challenge or difficulty of the task used, tailor the recruitment method to motivated samples who are likely to benefit from such interventions and consider the impact of different recruitment methods on the measured outcomes.

References

Abdelazeem, B., Abbas, K. S., Amin, M. A., El-Shahat, N. A., Malik, B., Kalantary, A., & Eltobgy, M. (2022). The effectiveness of incentives for research participation: A systematic review and meta-analysis of randomized controlled trials. *PLOS ONE*, *17*(4), e0267534. https://doi.org/10.1371/journal.pone.0267534

Adams, R. C., Button, K. S., Hickey, L., Morrison, S., Smith, A., Bolus, W., Coombs, E., Randolph, S., Hunt, R., Kim, D., Chambers, C. D., & Lawrence, N. S. (2021). Food-related inhibitory control training reduces food liking but not snacking frequency or weight in a large healthy adult sample. *Appetite*, *167*, 105601. https://doi.org/10.1016/j.appet.2021.105601

Anwyl-Irvine, A.L., Massonié, J., Flitton, A., Kirkham, N.Z., Evershed, J.K. (2019). Gorilla in our midst: an online behavioural experiment builder. Behavior Research Methods. https://doi.org/10.3758/s13428-019-01237-x

Aulbach, M. B., Knittle, K., van Beurden, S. B., Haukkala, A., & Lawrence, N. S. (2021). App-based food Go/No-Go training: User engagement and dietary intake in an opportunistic observational study. *Appetite*, *165*, 105315. https://doi.org/10.1016/j.appet.2021.105315

Caldwell, A. R., aut (2022). "Exploring Equivalence Testing with the Updated TOSTER R Package." *PsyArXiv*. doi:10.31234/osf.io/ty8de, https://doi.org/10.31234/osf.io/ty8de.

Carbine, K. A., & Larson, M. J. (2019). Quantifying the presence of evidential value and selective reporting in food-related inhibitory control training: a p-curve analysis. Health Psychology Review, 13(3), 318–343. https://doi.org/10.1080/17437199.2019.1622144

Chen, Z., Veling, H., Dijksterhuis, A., & Holland, R. W. (2016). How does not responding to appetitive stimuli cause devaluation: Evaluative conditioning or response inhibition? *Journal of Experimental Psychology: General*, *145*, 1687–1701. https://doi.org/10.1037/xge0000236

GAMIFICATION IN FOOD-RIT

GAMIFICATION IN FOOD-RIT

Chen, Z., Veling, H., Dijksterhuis, A., & Holland, R. W. (2018). Do impulsive individuals benefit more from food go/no-go training? Testing the role of inhibition capacity in the no-go devaluation effect. *Appetite*, *124*, 99–110. https://doi.org/10.1016/j.appet.2017.04.024

Chen, Z., Holland, R. W., Quandt, J., Dijksterhuis, A., & Veling, H. (2019). When mere action versus inaction leads to robust preference change. *Journal of Personality and Social Psychology*, *117*, 721–740. https://doi.org/10.1037/pspa0000158

Chen, Z., & Veling, H. (2022). Toward a better understanding of durable behavior change by food Go/NoGo training. Current Opinion in Behavioral Sciences, 48, 101212. https://doi.org/10.1016/j.cobeha.2022.101212

Das, S. R., Kinsinger, L. S., Yancy, W. S., Wang, A., Ciesco, E., Burdick, M., & Yevich, S. J. (2005). Obesity prevalence among veterans at Veterans Affairs medical facilities. *American Journal of Preventive Medicine*, *28*(3), 291–294. https://doi.org/10.1016/j.amepre.2004.12.007

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer US. https://doi.org/10.1007/978-1-4899-2271-7

Deterding, S. (2015). The Lens of Intrinsic Skill Atoms: A Method for Gameful Design. *Human–Computer Interaction*, *30*(3–4), 294–335. https://doi.org/10.1080/07370024.2014.993471

Dohle, S., Diel, K., & Hofmann, W. (2018). Executive functions and the self-regulation of eating behavior: A review. *Appetite*, *124*, 4–9. https://doi.org/10.1016/j.appet.2017.05.041

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Forman, E. M., Manasse, S. M., Dallal, D. H., Crochiere, Rebecca. J., Loyka, C. M., Butryn, M. L., Juarascio, A. S., & Houben, K. (2019). Computerized neurocognitive training for improving dietary health and facilitating weight loss. *Journal of Behavioral Medicine*, *42*(6), 1029–1040. https://doi.org/10.1007/s10865-019-00024-5

Forman, E. M., Manasse, S. M., Dallal, D. H., Crochiere, R. J., Berry, M. P., Butryn, M. L., & Juarascio, A. S. (2021). Gender differences in the effect of gamification on weight loss during a daily, neurocognitive training program. *Translational Behavioral Medicine*, *11*(4), 1015–1022. https://doi.org/10.1093/tbm/ibaa091

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Hayes, A. F., & Rockwood, N. J. (2018). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour Research and Therapy*, *98*, 39–57. https://doi.org/10.1016/j.brat.2016.11.001

Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions*, *6*, 89–106. https://doi.org/10.1016/j.invent.2016.10.002

Jones, A., Di Lemma, L. C. G., Robinson, E., Christiansen, P., Nolan, S., Tudur-Smith, C., & Field, M. (2016). Inhibitory control training for appetitive behaviour change: A meta-analytic investigation of mechanisms of action and moderators of effectiveness. *Appetite*, *97*, 16–28. https://doi.org/10.1016/j.appet.2015.11.013

Jones, A., Hardman, C. A., Devlin, N., Pennington, C. R., & Robinson, E. (2022). Weight-based discrimination in financial reward and punishment decision making: Causal evidence using a novel experimental paradigm. *International Journal of Obesity*, *46*(7), Article 7. https://doi.org/10.1038/s41366-022-01109-z

Kelders, S. M., Kok, R. N., Ossebaard, H. C., & Gemert-Pijnen, J. E. V. (2012). Persuasive System Design Does Matter: A Systematic Review of Adherence to Web-Based Interventions. *Journal of Medical Internet Research*, *14*(6), e2104. https://doi.org/10.2196/jmir.2104

GAMIFICATION IN FOOD-RIT

Kragel, E. A., Sweitzer, M. M., & Davis, J. M. (2019). The Effect of Brief Mindfulness Training on Brain
   Reactivity to Food Cues During Nicotine Withdrawal: A Pilot Functional Imaging Study.
   Mindfulness, *10*(11), 2272–2276. https://doi.org/10.1007/s12671-019-01201-y

Lawrence, N. S., O'Sullivan, J., Parslow, D., Javaid, M., Adams, R. C., Chambers, C. D., Kos, K., &
   Verbruggen, F. (2015). Training response inhibition to food is associated with weight loss and
   reduced energy intake. *Appetite*, *95*, 17–28. https://doi.org/10.1016/j.appet.2015.06.009

Lawrence, N. S., Porter, L., & Staiger, P. K. (2022). The 'Go's and the 'No-Go's of response-inhibition
   training to food: Lessons learned from trials. *Current Opinion in Behavioral Sciences*, *48*, 101229.
   https://doi.org/10.1016/j.cobeha.2022.101229

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/10.1080/01621459.1988.10478722

Lumsden, J., Skinner, A., Coyle, D., Lawrence, N., & Munafo, M. (2017). Attrition from Web-Based
   Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques. *Journal of
   Medical Internet Research*, *19*(11), e8473. https://doi.org/10.2196/jmir.8473

Mazarakis, A. (2021). Gamification Reloaded. *Current and Future Trends in Gamification Science*, *20*(3),
   279–294. https://doi.org/10.1515/icom-2021-0025

Mazarakis, A., & Bräuer, P. (2023). Gamification is Working, but Which One Exactly? Results from an
   Experiment with Four Game Design Elements. *International Journal of Human–Computer
   Interaction*, *39*(3), 612–627. https://doi.org/10.1080/10447318.2022.2041909

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of
   individual gamification elements on intrinsic motivation and performance. *Computers in Human
   Behavior*, *71*, 525–534. https://doi.org/10.1016/j.chb.2015.08.048

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D.,
   Egger, M., & Altman, D. G. (2010). CONSORT 2010 Explanation and Elaboration: Updated

guidelines for reporting parallel group randomised trials. *BMJ*, *340*, c869.

https://doi.org/10.1136/bmj.c869

Najberg, H., Rigamonti, M., Mouthon, M., & Spierer, L. (2021). Modifying food items valuation and

weight with gamified executive control training. *Royal Society Open Science*, *8*(5), 191288.

https://doi.org/10.1098/rsos.191288

Pennington, C. R., Monk, R. L., Heim, D., Rose, A. K., Gough, T., Clarke, R., Knibb, G., Patel, R., Rai, P.,

Ravat, H., Ali, R., Anastasiou, G., Asgari, F., Bate, E., Bourke, T., Boyles, J., Campbell, A., Fowler,

N., Hester, S., … Jones, A. (2023). The labels and models used to describe problematic substance use

impact discrete elements of stigma: A registered report. *Psychology of Addictive Behaviors*, No

Pagination Specified-No Pagination Specified. https://doi.org/10.1037/adb0000919

Perski, O., Blandford, A., West, R., & Michie, S. (2017). Conceptualising engagement with digital

behaviour change interventions: A systematic review using principles from critical interpretive

synthesis. *Translational Behavioral Medicine*, *7*(2), 254–267. https://doi.org/10.1007/s13142-016-
0453-1

Quandt, J., Holland, R. W., Chen, Z., & Veling, H. (2019). The role of attention in explaining the no-go

devaluation effect: Effects on appetitive food items. *Journal of Experimental Psychology: Human

Perception and Performance*, *45*, 1119–1133. https://doi.org/10.1037/xhp0000659

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C.,

Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J.,

Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., … Fuster, V. (2020). Global

Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study.

Journal of the American College of Cardiology, *76*(25), 2982–3021.

https://doi.org/10.1016/j.jacc.2020.11.010

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental

study of the effects of specific game design elements on psychological need satisfaction. *Computers

in Human Behavior*, *69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, *74*, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

Soderberg, C.K., Errington, T.M., Schiavone, S.R. et al. Initial evidence of research quality of registered reports compared with the standard publishing model. Nat Hum Behav 5, 990–997 (2021). https://doi.org/10.1038/s41562-021-01142-4

Stice, E., Yokum, S., & Waters, A. (2015). Dissonance-Based Eating Disorder Prevention Program Reduces Reward Region Response to Thin Models; How Actions Shape Valuation. *PLOS ONE*, *10*(12), e0144530. https://doi.org/10.1371/journal.pone.0144530

Stice, E., Yokum, S., Nelson, T. D., Berkman, E., Veling, H., & Lawrence, N. (2022). Efficacy of a combined food-response inhibition and attention training for weight loss. *Current Opinion in Behavioral Sciences*, *46*, 101168. https://doi.org/10.1016/j.cobeha.2022.101168

Tingley D, Yamamoto T, Hirose K, Keele L, Imai K (2014). "mediation: R Package for Causal Mediation Analysis." Journal of Statistical Software, 59(5), 1–38. http://www.jstatsoft.org/v59/i05/.

Veling, H., van Koningsbruggen, G. M., Aarts, H., & Stroebe, W. (2014). Targeting impulsive processes of eating behavior via the internet. Effects on body weight. *Appetite*, *78*, 102–109. https://doi.org/10.1016/j.appet.2014.03.014

Veling, H., Lawrence, N. S., Chen, Z., van Koningsbruggen, G. M., & Holland, R. W. (2017). What Is Trained During Food Go/No-Go Training? A Review Focusing on Mechanisms and a Research Agenda. *Current Addiction Reports*, *4*(1), 35–41. https://doi.org/10.1007/s40429-017-0131-5

Vermeir, J. F., White, M. J., Johnson, D., Crombez, G., & Van Ryckeghem, D. M. L. (2020). The Effects of Gamification on Computerized Cognitive Training: Systematic Review and Meta-Analysis. *JMIR Serious Games*, *8*(3), e18644. https://doi.org/10.2196/18644

Vermeir, J. F., White, M. J., Johnson, D., Crombez, G., & Van Ryckeghem, D. M. L. (2022). Gamified Web-Delivered Attentional Bias Modification Training for Adults With Chronic Pain: Protocol for a

Randomized, Double-blind, Placebo-Controlled Trial. *JMIR Res Protoc*, *11*(1), e32359.

https://doi.org/10.2196/32359

Yang, Y., Qi, L., Morys, F., Wu, Q., & Chen, H. (2022). Food-Specific Inhibition Training for Food

Devaluation: A Meta-Analysis. *Nutrients*, *14*(7).