Journal of Behavior Therapy and Experimental Psychiatry 76 (2022) 101727

Contents lists available at ScienceDirect

Journal of Behavior Therapy and Experimental Psychiatry

journal homepage: www.elsevier.com/locate/jbtep

Gamification of cognitive bias modification for interpretations in anxiety increases training engagement and enjoyment

Elske Salemink [a,b,*], Suzanne R.C. de Jong [c], Lies Notebaert [d], Colin MacLeod [d], Bram Van Bockstaele [d,e]

[a] Department of Psychology, University of Amsterdam, P.O. Box 80140, 3508, TC, Utrecht, the Netherlands
[b] Department of Clinical Psychology, Utrecht University, the Netherlands
[c] Department of Department of Clinical, Neuro- and Developmental Psychology, Vrije Universiteit Amsterdam, Van der Boechorststraat 7, 1018, BT, Amsterdam, the Netherlands
[d] School of Psychological Science, University of Western Australia, 35 Stirling Highway, WA 6009, Crawley, Australia
[e] Research Institute of Child Development and Education, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018, WT, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Gamification
Interpretation bias
Cognitive bias modification
Anxiety

ABSTRACT

Background and objectives: Interpretation bias plays a crucial role in anxiety. To test the causal role and potential clinical benefits, training procedures were developed to experimentally change interpretation bias. However, these procedures are monotonous and plain, which could negatively affect motivation and adherence. The aim of this study was to make the interpretation training more engaging and enjoyable, without compromising its effectiveness, through gamification.

Methods: The training was gamified by including extrinsically and intrinsically motivating elements such as points, scores, time-pressure, fun and adaptive elements (training at an individually challenging level). A 2 (Type: Gamified vs. Standard) x 2 (Training Valence: Positive vs. Placebo) between-subjects design was used with random allocation of 79 above-average anxious individuals. Post-training, we assessed the liking and recommendation of the training task, interpretation bias (Recognition task and the Scrambled Sentence Task) and anxiety.

Results: Participants experienced the gamified training tasks as more engaging and enjoyable than the standard tasks, although it was not recommend more to fellow-students. Both positive training conditions (gamified and standard) were successful in eliciting a positive interpretation bias when assessed with the Recognition task, while only the standard positive training impacted on interpretations when assessed with the Scrambled Sentence Task. No differential effects were observed on anxiety.

Limitations: The study involved only a single-session training and participants were selected for high trait (and not social) anxiety.

Conclusions: The gamified training was evaluated more positively by the participants, while maintaining the effectiveness of eliciting positive interpretations when assessed with the Recognition task. This suggests that gamification might be a promising new approach.

1. Introduction

According to cognitive models of anxiety, anxious individuals are disproportionately likely to interpret ambiguous information as negative or threatening, rather than as benign (e.g., Mathews & MacLeod, 2005). For example, when giving a presentation, (socially) anxious individuals will more readily interpret smiling among people in the audience as a sign that they said something silly, while non-anxious individuals are more likely to interpret this as a sign that the audience is feeling entertained. Crucially, this negative interpretation bias has been argued to play a causal role in anxiety, with negatively biased interpretations maintaining or exacerbating anxiety (for a review, see Hirsch et al., 2016). Empirical evidence for this causal relation comes from Cognitive Bias Modification for Interpretation (CBM-I) studies, in

* Corresponding author. Department of Clinical Psychology, Utrecht University, P.O. Box 80140, 3508, TC, Utrecht, the Netherlands.
E-mail addresses: E.Salemink@uu.nl (E. Salemink), s.r.c.dejong@vu.nl (S.R.C. de Jong), Lies.Notebaert@uwa.edu.au (L. Notebaert), Colin.MacLeod@uwa.edu.au (C. MacLeod), b.d.vanbockstaele@uva.nl (B. Van Bockstaele).

https://doi.org/10.1016/j.jbtep.2022.101727
Received 23 September 2020; Received in revised form 29 October 2021; Accepted 6 February 2022
Available online 7 February 2022

which experimentally induced changes in interpretation bias resulted in changes in anxiety. The ambiguous scenario training approach (Mathews & Mackintosh, 2000) is the most frequently used CBM-I paradigm. In this paradigm, short ambiguous textual scenarios are presented, and a crucial word in the last sentence is presented as a word fragment. Participants' task is to complete this word fragment. In a positive training, the correct solution of the fragment is consistent only with the positive interpretation of the scenario. Initially, CBM-I training was developed to examine the causal role of interpretations in anxiety. Subsequently, clinical application of CBM-I have been evaluated as a potential therapeutic strategy, by examining whether multiple sessions of CBM-I training can reduce dysfunctional anxiety.

A meta-analysis (Menne-Lothmann et al., 2014) found that after positive CBM-I training, recipients are more likely to endorse positive interpretations than negative interpretations of ambiguity, and that positive CBM-I increased positive interpretations from before to after training. More recently, a meta-analysis specifically in youth examined CBM-I's impact on anxiety and found that positive CBM-I had a moderate effect on bias and a small effect on anxiety (Krebs et al., 2018). This is consistent with the results from a review of multiple CBM meta-analyses, showing that CBM-I has significant effects on interpretation bias, and that there is good evidence that CBM (attention and interpretation bias modification combined) has effects on anxiety symptoms in adults (Jones & Sharpe, 2017).

One of the factors that may contribute to the success or failure of CBM-I is participants' motivation to train. While CBM-I has low costs and is easily accessible, participants often find it boring (Beard et al., 2012) and would not recommend it to friends nor train again in case of emotional problems (de Voogd et al., 2017). Consequently, people may be less engaged in the training and may lose their motivation to complete sessions, which could be especially disadvantageous given that completing more training sessions increases the beneficial effects (Menne-Lothmann et al., 2014). Keeping people motivated to complete CBM-I may thus be crucial to maximize its effects. The goal of the present study was to make CBM-I more enjoyable, which could in turn help increasing motivation to train.

One way of increasing enjoyment is through gamification; that is "the use of game design elements in non-game contexts" (Deterding et al., 2011, p. 9). Gamification aims to increase task user engagement and positive outcomes (e.g., user activity) by adding motivating elements (Deterding et al., 2011; Hamari et al., 2014). While, to our knowledge, there have been no previous attempts to gamify CBM-I training, Wartena and Van Dijk (2013) described the development of Bias Blaster, where a standard computer game was used as a positive reinforcer to complete training trials. Participants were required to complete a number of standard training trials before being able to (continue to) play a computer game. A pilot study with seven patients with a psychotic disorder showed that they felt competent in playing but the effectiveness of modifying interpretation bias and relevant symptoms was not tested (Van der Krieke et al., 2014). Moreover, the training itself was not gamified (i.e., the setting was gamified but not the task), thus potentially limiting increases in task engagement. To our knowledge, there are no studies in which the CBM-I training itself was gamified.

Gamification research has identified several factors that can increase task motivation. These include enjoyment (Boendermaker et al., 2015), fantasy and sensory stimuli (Garris et al., 2002), positive feedback (Jin, 2012), possibility to improve ones' own performance (Fasola & Mataric, 2010), and flow (i.e., a state that is reached when engaging in optimally challenging activities that are enjoyable and absorbing) (Andersen, 2012; Csikszentmihalyi, 2000; Fasola & Mataric, 2010; Jin, 2012). Factors that enhance flow (and therefore motivation) are concrete goals, understandable and intuitive rules, feedback systems, and enough attractive features to block distractors (Garris et al., 2002; Jin, 2012).

In our study, these gamification factors, identified as improving motivation, were used as guidelines to develop a gamified paradigm. To increase enjoyableness and flow of the training, the training was presented as a shooting game with sound effects and visual feedback. Ambiguous textual scenarios were presented at the top of the screen and two word fragments descended down. Participants had to identify the fragment that provided a meaningful completion to the scenario, move a canon to aim for this word, and press a key to shoot it. The key that would shoot the word was the first letter missing in the word fragment. To further improve flow, original scenarios were shortened. To increase challenge, a block was declared "game over" when participants made too many errors or when time ran out. Furthermore, to increase competition, we added a scoring system per block, and providing the possibility to beat one's own highest score. Finally, the speed of the gamified training was adaptive, with the speed of the word fragments dropping down increasing as participants' performance improved.

This gamification of CBM-I was intended to increase engagement and motivation, without compromising the impact of the training on interpretation bias and anxiety. In the field of Cognitive Bias Modification for Attention (ABM), there have been studies that gamified the training but then failed to change the targeted bias and symptoms (Boendermaker et al., 2016; Pieters et al., 2017, but see Notebaert et al., 2015; 2018). Research focussing on a gamified cognitive training in a school setting suggested that some motivating elements (including real-time scoring during the play) may distract from the core training task resulting in reduced task performance (Katz et al., 2014). Thus, while a gamified CBM training has the potential to increase the training effects in the long-run by increasing adherence to the training and reducing drop-out, it is pivotal that the training remains effective in modifying the targeted bias. Therefore, to examine whether gamifying the training did not compromise its effectiveness, we compared the effects of the gamified CBM-I training with the standard training on interpretive bias and anxiety.

The main questions addressed in this proof-of-principle study were whether gamified CBM-I training (compared to standard CBM-I training) was more engaging and enjoyable, and whether the gamified, positive training was as effective as the standard, positive training in eliciting a more positive interpretation bias and reducing anxiety compared to the placebo training conditions. Above average anxious individuals were randomly allocated to one of four different training conditions varying on two dimensions: Training type (gamified or standard) and Training Valence (positive or placebo). We tested i. whether the gamified training conditions were more engaging and enjoyable than the standard training conditions; ii. whether the positive training conditions (gamified and standard) were equally effective in eliciting a stronger positive interpretation bias than the placebo conditions; and iii. whether the positive training conditions (gamified and standard) were equally effective in reducing anxiety following a stressor and in anticipation of an upcoming social event compared to the placebo conditions (Mackintosh et al., 2013; Murphy et al., Salemink et al., 2007b; 2009).

2. Methods

2.1. Participants

Students from the University of Amsterdam completed the trait version of the State and Trait Anxiety Inventory (STAI-T, van der Ploeg et al., 1980) as part of a mass screening and students who scored within the top 50% (scores >39) in that sample were invited to participate. The study was called "How quick are you to find the right words" and was advertised as a study that involved computer tasks and questionnaires.

A total of 82 participants completed the experiment.[1] Using

[1] A trait anxiety assessment during the actual lab session showed that 18 participants scored below the cut-off at that point in time. Given that their scores were above 39 at the screening and to preserve power, the participants were included in the analyses.

conventional values of 0.80 for power and .05 for alpha, power analyses using G*power showed that a minimal sample size of $N = 52$ was required to detect large effects ($f = 0.40$; Otkhmezuri et al., 2019) for gamified versus standard training on enjoyment, $N = 76$ to detect expected medium-sized effects ($f = 0.33$; Hallion & Ruscio, 2011) for positive versus placebo training on interpretive bias, and $N = 68$ to detect the expected small to medium-sized interaction ($f = 0.175$; Hallion & Ruscio, 2011) between training valence and pre-versus post--assessment (correlation between repeated measures of 0.50) on anxiety vulnerability. One participant was excluded because of difficulties understanding the instructions and resulting low accuracy during training, and two because they were inadvertently included (scoring below 39 on the STAI-T during screening). Our final sample thus consisted of 79 participants (17 men, *M*age = 21.9, *SD* = 3.3). A computer program randomized eligible participants into one of the four groups based on their participant number. Participants were blind to training type and valence, and the experimenter was blind to training type (and not present in the room when participants completed the study). Groups did not differ significantly in age and sex distribution, *p*-values >.32. The study was approved by the Ethics Committee of the University of Amsterdam (2015-DP-4395).

2.2. Materials

2.2.1. Interpretation training: Standard CBM-I

The standard CBM-I was based on the Dutch translation of the ambiguous scenario training (Mathews & Mackintosh, 2000; Salemink et al., 2009). All scenarios were related to ambiguous social situations and consisted of three lines, which were presented line by line on a white screen by pressing the space bar. The final line of each scenario contained a word fragment and this fragment could be completed in one way. The solution was consistent with only one resolution of the scenario's ambiguity. Participants were required to press the space bar when they knew the correct solution and to type in the first missing letter. In the positive training group, the correct solution was always consistent with the positive interpretation of the ambiguity, while in the placebo group, the correct solution was consistent with the positive interpretation in half of the trials and consistent with the negative interpretation of the ambiguity in the other half of the trials. After each trial, participants responded to a yes/no comprehension question and received feedback (Correct vs. Incorrect). Participants undertook this standard CBM-I training for 20 min (fixed time period) and completed on average 85 trials (*SD* = 3.0).

2.2.2. Interpretation training: Gamified CBM-I

The gamified CBM-I training contained shortened versions of the scenarios from the standard version. These were presented on a screen with the background portraying an arcade game console (see Fig. 1 for a screen shot and the link https://doi.org/10.1016/j.jbtep.2022.101727 for a short video). At the top of this screen, the scenario with a missing word was presented. Two word fragments descended down the console screen in a zig-zag manner, bouncing off the sides of the screen and off each other. The fragments in each pair were matched on valence, but only one fragment could yield a word that meaningfully fitted the scenario. Participants were instructed to "shoot" the correct word fragment by using the keypad arrows to move a canon and pressing the first missing letter of that particular word fragment to shoot. Sounds and visual feedback demonstrated whether the correct fragment was shot. Points were earned for correct responses. The training was adaptive: When trials were completed correctly, the word fragments on subsequent trials descended faster.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.jbtep.2022.101727

"Game over" could occur in two ways. First, the maximum playing time of 5 min could run out, which was shown by a green horizontal bar filling the screen from left to right. Second, the game was also over when too many errors were made. Each time an error was made, a red bar started to fill from the right, which could be undone by three successive correct trials. When the red error bar met the green time bar, the game was over. After each game, participants' scores were shown and they were encouraged to improve their high score in the next game.

Similar to the standard CBM-I, the target word fragment was always consistent with the positive interpretation in the positive gamified training, and it was equally often consistent with the positive and negative interpretation in the placebo condition,. Participants played for 20 min (fixed time period) and completed on average 121 trials (*SD* = 1.6).

2.2.3. Trait anxiety

Trait anxiety was assessed with the Dutch translation of the trait

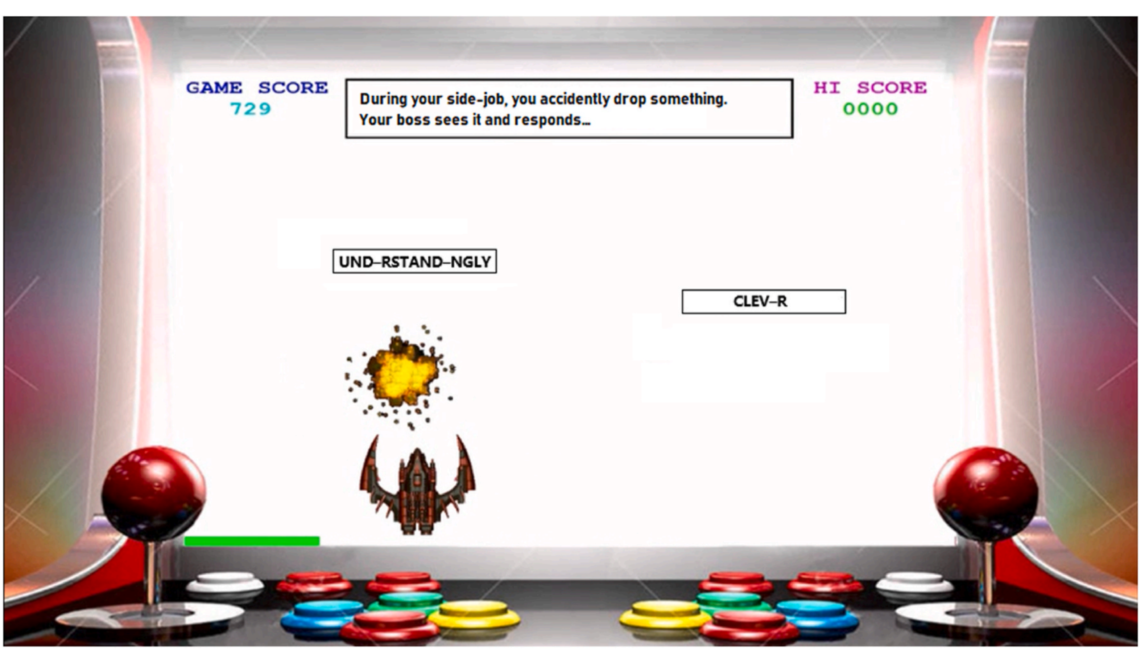

Fig. 1. Screenshot of the gamified training (scenario and word fragments translated from Dutch to English).

version of the STAI (STAI-T; van der Ploeg et al., 1980). The STAI-T consists of twenty statements, each scored on a 4-point Likert scale ranging from 1 = "almost never" to 4 = "almost always".

2.2.4. Training Task Evaluation Questionnaire

A custom-designed Training Task Evaluation Questionnaire was developed to assess participants' evaluation of the training in two ways: i. by reporting the degree to which they personally liked the training; and ii, by reporting whether they would recommend the experiment to a friend. The assessment of the personal liking of the training task consisted of eight statements. Participants indicated to what extent they experienced the training as exciting, unpleasant, uninteresting, challenging, boring, interesting, entertaining, and monotonous on 4-point Likert scales ranging from 1 (absolutely not) to 4 (very much). The four negative statements were reversed and a sum score was calculated. This Personal Enjoyment Score could range from 8 to 32 to indicate the personal liking of a training task. In addition, participants were asked if they would advise a fellow student to participate in this experiment, based on the training part of the experiment. Participants had to circle either 'Yes' or 'No'. The percentage of 'Yes' answers was used as the Recommendation Score reflecting the percentage of participants that would recommend the experiment to other students.

2.2.5. Interpretation bias assessment I: Recognition task

Interpretation bias was assessed with the Recognition task developed by Mathews and Mackintosh (2000; Dutch translation by Salemink et al., 2009). Participants received ten social stories, all of which were ambiguous in terms of valence. Each story had a title and consisted of three lines. A word fragment was presented in the final sentence. Participants were asked to complete the fragment as quickly as possible, yet the valence of the story remained ambiguous. Then participants responded to a yes/no comprehension question, followed by feedback (correct versus wrong answer).

In the second part of this recognition task, participants saw the title of the ambiguous story, together with two versions of the final sentence presented in a random order. These sentences represented a possible positive and possible negative interpretation. Participants rated each interpretation for its similarity in meaning to the original story using a 4-point scale ranging from 1 (very different in meaning) to 4 (very similar in meaning). An interpretive bias index was calculated by subtracting the average similarity ratings of negative interpretations from the average similarity ratings of positive interpretations (Salemink et al., 2010b), with positive scores reflecting a more positive interpretation bias.

2.2.6. Interpretation bias assessment II: Scrambled Sentence Task

A shortened version of the Dutch Scrambled Sentence Task with cognitive load (SST; Everaert et al., 2014; de Voogd et al., 2017) was used to test transfer of training effects to a different measure of interpretation bias (see also Bowler et al., 2012). Participants unscrambled twenty socially evaluative scrambled sentences, using five out of the six presented words. For example, the sentence "*am winner born loser a I*" could be solved as "I am a born winner" or "I am a born loser". Participants were required to keep a five-digit number in mind to impose a cognitive load to prevent deliberate report strategies (Everaert et al., 2014). At the beginning of the task, the number was presented for as long as participants needed. Then, they solved the first 10 scrambled sentences using pen and paper, after which they were asked to reproduce the five-digit number. This procedure was repeated for the next ten items, with a different number to remember. Interpretation bias was indexed by the ratio of positively unscrambled sentences over the total number of sentences completed (Everaert et al., 2014), with higher values representing more positive interpretations.

2.2.7. Anagram stressor task

In the anagram stressor task (e.g. MacLeod et al., 2002), participants were presented with fifteen difficult anagrams (number of letters ranged between 4 and 8) and were asked to solve them. Participants were told that the task is a good measure of their language proficiency and a reliable predictor of their future success in various domains. It was emphasized that students generally perform well on this task. As the majority of anagrams was in fact quite difficult to solve, students were led to believe that they did not do well on the task. To increase stress, participants were told they were being videotaped and there was a time limit of 20 s per anagram with a countdown clock presented during the last 10 s.

Anxious mood state was measured before and after the stressor with two Visual Analogue mood Scales (VAS) consisting of 10 cm horizontal lines. The first was anchored "insecure" to "self-confident", and the second "stressed" and "relaxed". VAS scales can measure feelings in a relatively sensitive and reliable way (Cella & Perry, 1986). The scores were recoded such that higher scores represented more insecurity and more stress, and the scores were averaged into a single anxious mood score.

2.2.8. Anticipated stressful situation task

Participants were told that they would be meeting two unknown individuals for a 5 min conversation. Anticipated anxiety for this social event (Murphy et al., 2007) was measured using 10-cm VAS scales assessing their predicted feelings of stress during the upcoming social interaction (anchored from "very stressed" to "very relaxed") and their predicted performance in that situation (anchored from "very poorly" to "very well"). Ratings were recoded such that higher scores represented more stress and worse anticipated performance, and averaged into a single anticipated anxiety score.

2.3. Procedure

Participants were tested individually in lab cubicles. After participants gave informed consent, they completed the STAI-T. Participants then performed their assigned CBM-I task. This was followed by the Recognition task, the SST, and the anagram stressor task. VAS scales measuring anxiety were completed directly before and after the anagram stressor task. Then, the Training Task Evaluation Questionnaire was completed, followed by the anticipated social situation test.[2] After the experiment, participants were debriefed and rewarded with credits or money.

2.4. Data analyses

The data were analyzed within a frequentist, Null-Hypothesis Significance Testing and a Bayesian framework (Krypotos et al., 2017). For the Null-Hypothesis Significance Testing framework, IBM SPSS Statistics for Windows Version 24 (IBM Corp, 2016) was used.

Effect sizes were reported as partial eta squares ($\eta_p{}^2$) for the Analyses of Variance (ANOVAs) with 0.01, 0.06, and 0.14 indicating small, medium, and large effects, respectively. For t-tests, Hedges' *g* were reported with 0.20, 0.50, and 0.80 considered as small, medium, and large effect sizes, respectively.

As the second and third hypothesis concern the prediction that the gamified and standard CBM-I training are as effective in affecting interpretive bias and anxiety, a Bayesian framework (e.g., Krypotos et al., 2017) was added to the frequentist analyses. Bayes factors quantify the amount of evidence that the data provide for one of two hypotheses, which are the null and alternative hypotheses. Given that our hypotheses were that the two positive training conditions (gamified and standard) were equally effective in eliciting a stronger positive interpretation bias (Hypothesis 2) and reducing anxiety (Hypothesis 3)

[2] There were some additional measures, which are not part of the current manuscript.

compared to the placebo conditions, our null hypothesis is the model without the higher order interaction effect with Training Type and the alternative hypothesis is the model with the main effects and interaction effect with Training Type. To test for the effect of the interaction term with Training Type in the Bayesian repeated measures ANOVAs, we used the Effects function across matched models. All models were tested using JASP, version 0.14.1.0 (JASPTeam, 2020), using default settings. We report Bayes factors quantifying evidence under the null hypothesis H_0, relative to the alternative hypothesis H_A, denoted as BF_{01}. The larger this Bayes factor, the more relative evidence there is for the null hypothesis compared to the alternative. A $BF_{01} = 3$ indicates that the data are three times more likely under the null hypothesis than the alternative hypothesis, while the opposite is true for $BF_{01} = 0.333$. The following evidence categories (Wetzels et al., 2011) were used for BF_{01}: 3–10 = substantial evidence for H_0, 1–3 = anecdotal evidence for H_0, 1 = no evidence, 1/3–1 = anecdotal evidence for H_A, and 1/3–1/10 = substantial evidence for H_A.

3. Results

3.1. Training Task Evaluation Questionnaire

To examine whether the gamified versions of CBM-I were evaluated more positively than the standard versions, the two outcome measures from the Training Task Evaluation Questionnaire were examined (Table 1). A 2 (Training Type: gamified versus standard) x 2 (Training Valence: positive versus placebo) ANOVA on the Personal Enjoyment Scores yielded a significant main effect of Training Type, $F(1, 73) = 5.3$, $p = .03$, $\eta_p^2 = 0.07$. No other effects were significant. Participants experienced the gamified CBM-I tasks as more enjoyable and engaging than the standard CBM-I tasks.

A chi-squared test on the Recommendation Scores was conducted to test whether the gamified CBM-I tasks (positive and placebo), compared to the standard CBM-I tasks (positive and placebo), would more often be recommended to fellow students. Results indicated that there was no significant difference in recommendation between the gamified and the standard CBM-I tasks (respectively 80.0% vs. 85.3%), $\chi^2(1) = 0.37$, $p = .54$.

3.2. Effects on interpretation bias

3.2.1. Recognition task

To examine the training effects on interpretation bias as assessed with the Recognition task, the bias scores were entered in a 2 (Training Type) x 2 (Training Valence) ANOVA (Table 1). There was only a significant, medium-sized main effect of Training Valence, $F(1, 73) = 6.1$, $p = .02$, $\eta_p^2 = 0.08$, indicating that individuals who completed a positive training interpreted ambiguous information more positively than individuals who completed a placebo training. We did not find any evidence that this effect was modified by Training Type. The 2-way interaction was not significant, $F(1, 73) = 0.1$, $p = .74$, $\eta_p^2 = 0.002$, $BF_{01} = 3.04$. The Bayes factor indicated that there is (anecdotal to) substantial evidence for the null hypothesis (model with main effects) compared to the alternative hypothesis (model with interaction term). Thus, the results from the Recognition task indicate that the positive training conditions were successful in modifying interpretations relative to the placebo training conditions and that the gamified and standard training tasks had no differential impact on interpretations.

3.2.2. Scrambled Sentence Test

To examine the training effects on interpretation bias as assessed with the SST, the SST scores were entered in a 2 (Training Type) x 2 (Training Valence) ANOVA. We only found a significant, medium-sized Training Type x Training Valence interaction, $F(1, 75) = 7.09$, $p = .01$, $\eta_p^2 = 0.09$, $BF_{01} = 0.21$. Independent samples t-tests were conducted for the gamified and standard training groups separately. Within the gamified training conditions, there was no significant difference between the positive and placebo condition in interpretation bias, $t(43) = -0.64$, $p = .53$, Hedges' $g = 0.18$. Within the standard training conditions, there was a significant difference between the positive and placebo condition, $t(32) = 3.4$, $p = .002$, Hedges' $g = 1.09$. Participants in the standard positive training group constructed more positive sentences than participants in the standard placebo training (Table 1). Contrary to the Recognition task, the SST results thus indicated that only the standard positive training (and not the gamified positive training) was successful in modifying interpretations.

3.3. Effects on anxiety

3.3.1. Anagram stressor task

A 2 (Training Type) x 2 (Training Valence) x 2 (Assessment Point: Pre-versus Post-Stressor) mixed ANOVA on the anxious mood scores yielded a significant, large, main effect of Assessment Point, $F(1, 73) = 74.2$, $p < .001$, $\eta_p^2 = 0.50$, indicating that the stress task successfully increased anxious mood from pre-to post-stressor (Table 1). However, no other effects were significant, $Fs<1.2$, $ps>.27$, including the Training Valence x Assessment Point interaction, $F(1, 73) = 0.32$, $p = .58$, $\eta_p^2 = 0.004$, and the three-way interaction, $F(1, 73) = 0.16$, $p = .69$, $\eta_p^2 = 0.002$, $BF_{01} = 3.08$, suggesting that the training manipulations had no impact on stress reactivity. The Bayes Factor indicated that there is (anecdotal to) substantial evidence for the null hypothesis (the gamified and standard training conditions do not differ) compared to the alternative hypothesis.

3.3.2. Anticipated stressful situation task

A 2 (Training Type) x 2 (Training Valence) ANOVA on the averaged anticipated anxiety scores[3] revealed no significant effects; Training Type, $F(1, 65) = 1.0$, $p = .32$, $\eta_p^2 = 0.02$; Training Valence, $F(1, 65) = 0.4$, $p = .54$, $\eta_p^2 = 0.006$, and the Training Type x Training Valence interaction effect, $F(1, 65) = 1.9$, $p = .18$, $\eta_p^2 = 0.03$, $BF_{01} = 1.27$ (Table 1). Participants' anticipated anxiety was thus not differentially affected by the different training regimes.

4. Discussion

Our study aimed to develop a gamified version of a CBM-I training that was 1) more engaging and enjoyable than the original training, and 2) as effective in eliciting a positive interpretation bias and reducing anxiety compared to the original training. As intended, participants experienced the gamified CBM-I tasks as more engaging and enjoyable than the standard tasks, indicating the potential of gamification of CBM-I training. However, there were no differences between the gamified and standard CBM-I tasks in the likelihood to recommend the training to fellow students. With respect to training effectiveness in affecting interpretations, the results were mixed. The gamified and standard positive training regimes did not differ in their capacity to elicit positive interpretations as assessed with the Recognition task, while the SST assessment indicated that the standard positive training outperformed the gamified positive training. Finally, none of the training conditions differentially affected stress-induced anxious mood or anticipated anxiety.

The first aim of our study was to develop a gamified version of the CBM-I training that was more engaging and enjoyable than the standard version. Based on the gaming literature, we included a points system (Mekler et al., 2013), enjoyable elements (Boendermaker et al., 2015), adaptiveness (Fasola & Mataric, 2010), indirect competition (Jin, 2012), sound effects (Garris et al., 2002), and understandable rules and goals (Garris et al., 2002) in the gamified CBM-I training. As participants

[3] Due to an experimenter error, data of 10 participants were missing for this task.

Table 1
Means and standard deviations of the Training Task Evaluation Questionnaire (Personal Enjoyment Score, Recommendation Score), Recognition Task, Scrambled Sentence Test, anxious mood in response to a stressor, and anticipated anxiety for a social situation as a function of Training Type (Gamified vs. Standard) and Training Valence (Positive vs. Placebo).

	Enjoyment Score	Recom.Score	Recognition Task	SST	Stress-induced anxious mood		Anticip. anxiety
	M (*SD*)	% yes	*M* (*SD*)	*M* (*SD*)	*M* (*SD*) Pre	*M* (*SD*) Post	*M* (*SD*)
Gamified							
Positive	22.90 (3.77)	89.5	0.73 (0.73)	0.59 (0.08)	4.85 (1.97)	6.22 (1.67)	5.84 (1.24)
Placebo	20.25 (4.73)	73.1	0.36 (0.79)	0.62 (0.21)	4.31 (2.49)	5.63 (2.38)	5.08 (1.74)
Standard							
Positive	19.82 (4.56)	76.5	0.99 (0.79)	0.75 (0.15)	3.80 (1.98)	5.61 (2.07)	4.93 (1.49)
Placebo	18.77 (4.02)	94.1	0.51 (0.65)	0.60 (0.12)	4.03 (2.47)	5.51 (1.98)	5.22 (1.75)

Note. Recom score = Recommendation score and SST = Scrambled Sentence Test.

enjoyed the gamified versions more than the traditional versions, our first aim was fulfilled. The finding of increased enjoyment in the gamified version is an encouraging basis for future studies comparing multiple sessions of standard and gamified training, as increased enjoyment could potentially increase compliance and mitigate against drop-out when performing extended, multi-session training, thus potentially increasing the effectiveness of such interventions (see Otkhmezuri et al., 2019 for a Virtual Reality CBM-I).

With respect to training effectiveness in eliciting a positive interpretation bias, results depended on the type of assessment task used. Following training, results from the Recognition task revealed that both the gamified and the standard positive training were equally successful, as participants who were exposed to the positive training conditions (gamified and standard) interpreted ambiguity in the Recognition task more positively than participants who were exposed to the placebo training conditions. In addition, Bayesian analyses suggested that the gamified and standard training tasks did not differ in their effectiveness of eliciting a positive interpretation bias after training. Given that the gamified and standard positive training versions did not differ, our gamification of CBM-I did not negatively affect its ability to elicit a positive interpretation bias assessed with the Recognition task. This is an encouraging finding, given that attempts to gamify other CBM protocols (attentional bias modification) have failed to affect the targeted bias (Boendermaker et al., 2016; Pieters et al., 2017). While changing a training paradigm to make it more engaging runs the risk of changing or removing elements that are crucial for its effectiveness and thus may render the training ineffective, our Recognition task findings indicate this was not the case for our gamified training.

However, results from the SST indicated that only the standard positive training was successful as only those participants (and not those in the gamified positive training condition) had more positive interpretations after training than participants in the placebo condition. This was supported by the results from the Bayesian analyses as they provided support for the alternative hypothesis (compared to the null hypothesis) that the gamified and standard training tasks differed in their capacity to affect interpretations assessed with the SST. On the one hand, given the inconsistent effects of standard scenario-based interpretation bias training on interpretations assessed with tasks that differ from the training (Salemink et al., 2007a; 2010a), it is promising that the current study did replicate the effects of the standard CBM-I scenario training on interpretation bias as assessed with the SST (Bowler et al., 2012). On the other hand, it raises the question why the effects of the gamified positive CBM-I training on interpretations depended on the type of assessment task, with only effects on interpretations observed when assessed with the Recognition task and not with the SST. Given the structure and procedure of the two assessment tasks, the scenario-based CBM-I training (i.e., short scenarios) overlaps more with the Recognition task than with the SST. Performance on the Recognition task thus reflects relatively near transfer, while the SST reflects relatively far transfer (Hertel & Mathews, 2011). A single gamified training session may be potent enough for the acquired more positive interpretation to influence the processing of situations requiring near transfer (i.e., Recognition task), but not far transfer (i.e., SST). In addition, the tasks differ with respect to the role of ambiguity. Interpretation bias typically involves the resolution of ambiguity, and while such ambiguity is part of the CBM-I training and the Recognition task, it is absent in the SST. Participants see an unambiguously positive and an unambiguously negative word in the SST, and they choose which word they want to use to create a sentence.

The positive CBM-I training conditions reduced neither anxious mood induced by a stressor nor anticipated anxiety for an upcoming social event in comparison to the placebo training conditions. As the standard and gamified positive CBM-I tasks did not differ from each other in this regard (Bayesian analyses provided some evidence in favour of the null hypothesis compared to the alternative hypothesis), it is a general issue and not specifically tied to gamification. This absence of effects on anxiety is consistent with the variety in effect sizes observed in the literature, where mixed effects of CBM-I on anxiety have been reported (meta-analyses: Hallion & Ruscio, 2011; Menne-Lothmann et al., 2014). As the positive training conditions were capable of eliciting a positive interpretation bias after training compared to the placebo conditions when assessed with the Recognition task, this begs the question which factors play a role in determining whether training-induced positive interpretations influence anxiety. Transfer appropriate processing models emphasize the importance of matching the training situation with the situation in which emotion is assessed (Hertel & Mathews, 2011), with a greater match resulting in better transfer of the learned contingencies to the test situation. It has been shown that optimizing the match between training and stress task content improves CBM-I's capacity to affect anxious mood in response to a stress task (Mackintosh et al., 2013). In the current study, the training scenarios involved social situations while the anagram stress task mainly involved a performance/test-anxiety component, which might account for the lack of transfer. Another potentially relevant factor is one's internal emotional state. Dispositionally anxious individuals will likely feel elevated anxiety and arousal when exposed to stressors, whereas this may not be the case during training. Increasing the contextual match by having participants perform the training in an anxious and aroused mood state might increase the chance of the trained interpretation bias transferring to a subsequent stressful situation, thus affecting emotional responding (e.g., Kuckertz et al., 2014; Nuijs et al., 2020).

Some limitations should be discussed. First, our study involved a single training session. More training sessions might be needed for the training to achieve far transfer. Also, future studies using multi-session gamified CBM-I training would allow to examine the effect of gamification on drop-out. Second, we selected participants who were in the upper half of the trait anxiety distribution. As the CBM-I training focused on social situations, it might have been more appropriate to select participants on *social* anxiety. In addition, the current lack of baseline measures (including a measure of motivation) is a limitation of the study and the college student sample limits generalizability of the findings. Finally, while participants liked the gamified versions better than the standard versions, the scores were not very high (22.9 and 20.3 on a scale from 8 to 32, see Table 1). This could in part be explained by the fact that evaluations depend on the comparison category. When comparing the gamified version with mainstream videogames, the experience could be disappointing, leading to low liking ratings (see Boendermaker et al., 2015; 2016). Conversely, when evaluating the standard training, participants may compare this task to more boring lab experiments, which would lead to a relatively inflated evaluation of the standard training. Future research could directly compare gamified with standard training in within-subject designs, so that both versions are evaluated against each other.

In conclusion, we were successful in developing a gamified CBM-I training that was more enjoyable than the standard training, while maintaining the effects on interpretations when assessed with the often-used Recognition task, but not when assessed with the SST. None of the training versions had an impact on anxiety. The promise of gamified CBM-I is the potential of increasing the degree to which recipients experience the training task as engaging and enjoyable, though the effects on interpretative bias and anxiety warrant further research.

Funding

This study was supported by a UWA Research Collaboration Award awarded to Lies Notebaert, Patrick Clarke, Bram Van Bockstaele, Reinout Wiers, and Elske Salemink. Colin MacLeod is supported by an ARC Laureate Fellowship (FL170100167) and Elske Salemink by a NWO VIDI grant (VI.Vidi.195.041) from the Netherlands Organization for Scientific Research.

CRediT authorship contribution statement

Elske Salemink: Conceptualization, Methodology, Software, Data curation, Formal analysis, Supervision, Writing – original draft, Writing – review & editing. **Suzanne R.C. de Jong:** Conceptualization, Methodology, Data curation, Supervision, Writing – original draft, Writing – review & editing. **Lies Notebaert:** Conceptualization, Methodology, Writing – review & editing. **Colin MacLeod:** Conceptualization, Methodology, Software, Writing – review & editing. **Bram Van Bockstaele:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

All authors acknowledge that they have exercised due care in ensuring the integrity of the work. Further, none of the original material contained in the manuscript has been submitted for consideration nor will any of it be published elsewhere except in abstract form in connection with scientific meetings. We have no conflicts of interest to disclose.

References

Andersen, E. (2012). Optimizing adaptivity in educational games. In *Proceedings of the international conference on the foundations of digital games* (pp. 279–281). https://doi.org/10.1145/2282338.2282398

Beard, C., Weisberg, R. B., & Primack, J. (2012). Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): A qualitative study. *Behavioural and Cognitive Psychotherapy, 40*, 618–633. https://doi.org/10.1017/S1352465811000671

Boendermaker, W. J., Maceiras, S. S., Boffo, M., & Wiers, R. W. (2016). Attentional bias modification with serious game elements: Evaluating the shots game. *JMIR Serious Games, 4*, Article e20, 10.2196/games.6464.

Boendermaker, W. J., Prins, P. J., & Wiers, R. W. (2015). Cognitive Bias Modification for adolescents with substance use problems – can serious games help? *Journal of Behavior Therapy and Experimental Psychiatry, 49*, 13–20, 10.1016/j.jbtep.2015.03.008.

Bowler, J. O., Mackintosh, B., Dunn, B. D., Mathews, A., Dalgleish, T., & Hoppitt, L. (2012). A comparison of cognitive bias modification for interpretation and computerized cognitive behavior therapy: Effects on anxiety, depression, attentional control, and interpretive bias. *Journal of Consulting and Clinical Psychology, 80*, 1021–1033. https://doi.org/10.1037/a0029932

Cella, D. F., & Perry, S. W. (1986). Reliability and concurrent validity of three visual-analogue mood scales. *Psychological Reports, 59*, 827–833. https://doi.org/10.2466/pr0.1986.59.2.827

Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-Bass.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9–15). https://doi.org/10.1145/2181037.2181040

Everaert, J., Duyck, W., & Koster, E. H. (2014). Attention, interpretation, and memory biases in subclinical depression: A proof-of-principle test of the combined cognitive biases hypothesis. *Emotion, 14*, 331–340. https://doi.org/10.1037/a0035250

Fasola, J., & Mataric, M. J. (2010). Robot motivator: Increasing user enjoyment and performance on a physical/cognitive task. In *2010 IEEE 9th international conference on development and learning* (pp. 274–279). doi:10.1109/DEVLRN.2010.5578830.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*, 441–467. https://doi.org/10.1177/1046878102238607

Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological Bulletin, 137*, 940–958. https://doi.org/10.1037/a0024355

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In *System sciences (HICSS), 2014 47th Hawaii international conference* (pp. 3025–3034), 10.1109/HICSS.2014.377.

Hertel, P. T., & Mathews, A. (2011). Cognitive bias modification: Past perspectives, current findings, and future applications. *Perspectives on Psychological Science, 6*(6), 521–536. https://doi.org/10.1177/1745691611421205

Hirsch, C. R., Meeten, F., Krahé, C., & Reeder, C. (2016). Resolving ambiguity in emotional disorders: The nature and role of interpretation biases. *Annual Review of Clinical Psychology, 12*, 281–305. https://doi.org/10.1146/annurev-clinpsy-021815-093436

IBM Corp. (2016). *IBM SPSS Statistics for Windows, version 24.0*. IBM Corp.

JASP Team. (2020). *JASP (version 0.14.1.0) [Computer software]* https://jasp-stats.org/.

Jin, S.-A. A. (2012). Toward integrative models of flow": Effects of performance, skill, challenge, playfulness, and presence on flow in video games. *Journal of Broadcasting & Electronic Media, 56*, 169–186. https://doi.org/10.1080/08838151.2012.678516

Jones, E. B., & Sharpe, L. (2017). Cognitive bias modification: A review of meta-analyses. *Journal of Affective Disorders, 223*, 175–183. https://doi.org/10.1016/j.jad.2017.07.034

Katz, B., Jaeggi, S., Buschkuehl, M., Stegman, A., & Shah, P. (2014). Differential effect of motivational features on training improvements in school-based cognitive training. *Frontiers in Human Neuroscience, 8*, Article 242. https://doi.org/10.3389/fnhum.2014.00242

Krebs, G., Pile, V., Grant, S., Degli Esposti, M., Montgomery, P., & Lau, J. Y. F. (2018). Research review: Cognitive bias modification of interpretations in youth and its effect on anxiety: A meta-analysis. *Journal of Child Psychology and Psychiatry, 59*, 831–844. https://doi.org/10.1111/jcpp.12809

Krypotos, A. M., Blanken, T. F., Arnaudova, I., Matzke, D., & Beckers, T. (2017). A primer on bayesian analysis for experimental psychopathologists. *Journal of Experimental Psychopathology, 8*, 140–157. https://doi.org/10.5127/jep.057316

Kuckertz, J. M., Gildebrant, E., Liliequist, B., Karström, P., Väppling, C., Bodlund, O., Stenlund, T., Hoffman, S. G., Andersson, G., Amir, N., & Carlbring, P. (2014). Moderation and mediation of the effect of attention training in social anxiety disorder. *Behaviour Research and Therapy, 53*, 30–40. https://doi.org/10.1016/j.brat.2013.12.003

Mackintosh, B., Mathews, A., Eckstein, D., & Hoppitt, L. (2013). Specificity effects in the modification of interpretation bias and stress reactivity. *Journal of Experimental Psychopathology, 4*, 133–147. https://doi.org/10.5127/jep.025711

MacLeod, C., Rutherford, E., Campbell, L., Ebsworthy, G., & Holker, L. (2002). Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. *Journal of Abnormal Psychology, 111*, 107–123. https://doi.org/10.1037/0021-843X.111.1.107

Mathews, A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology, 109*, 602–615. doi:10.1037/0021-843X.109.4.602.

Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology, 1*, 167–195. https://doi.org/10.1146/annurev.clinpsy.1.102803.143916

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). Do points, levels and leaderboards harm intrinsic motivation? An empirical analysis of common gamification elements. In *Proceedings of the first international conference on gameful design, research, and applications* (pp. 66–73). doi:10.1145/2583008.2583017.

Menne-Lothmann, C., Viechtbauer, W., Höhn, P., Kasanova, Z., Haller, S. P., Drukker, M., van Os, J., Wichers, M., & Lau, J. Y. (2014). How to boost positive interpretations? A meta-analysis of the effectiveness of cognitive bias modification for interpretation. *PLoS One, 9*, Article e100925, 10.1371/journal.pone.0100925.

Murphy, R., Hirsch, C. R., Mathews, A., Smith, K., & Clark, D. M. (2007). Facilitating a benign interpretation bias in a high socially anxious population. *Behaviour Research and Therapy, 45*, 1517–1529. doi:10.1016/j.brat.2007.01.007.

Notebaert, L., Clarke, P. J. F., Grafton, B., & MacLeod, C. (2015). Validation of a novel Attentional Bias Modification Task: The future may be in the cards. *Behaviour Research and Therapy, 65*, 93–100. https://doi.org/10.1016/j.brat.2014.12.007

Notebaert, L., Grafton, B., Clarke, P. J. F., Rudaizky, D., Chen, N. T. M., & MacLeod, C. (2018). Emotion in motion: A gamified approach for the modification of attentional bias. *JMIR Serious Games, 6*(4), Article e10993. https://doi.org/10.2196/10993

Nuijs, M., Larsen, H., Bögels, S., Wiers, R. W., & Salemink, E. (2020). Context matters: The role of subjective arousal during attentional bias modification targeting socially anxious students. *Journal of Behavior Therapy and Experimental Psychiatry, 68*. https://doi.org/10.1016/j.jbtep.2019.101545

Otkhmezuri, B., Boffo, M., Siriaraya, P., Matsangidou, M., Wiers, R. W., Mackintosh, B., Ang, C. S., & Salemink, E. (2019). Believing is seeing: A proof-of-concept, semi-experimental study on using mobile virtual reality to boost the effects of interpretation bias modification for anxiety. *JMIR Mental Health, 6*, Article e11517. https://doi.org/10.2196/11517

Pieters, E. K., De Raedt, R., Enock, P. M., De Putter, L. M., Braham, H., McNally, R. J., & Koster, E. H. (2017). Examining a novel gamified approach to attentional retraining: Effects of single and multiple session training. *Cognitive Therapy and Research, 41*(1), 89–105, 0.1007/s10608-016-9803-z.

van der Ploeg, H. M., Defares, P. B., & Spielberger, C. D. (1980). *Handleiding bij de Zelf-Beoordelings Vragenlijst. [Manual for the Dutch adaptation of the STAI-Y]*. Swets & Zeitlinger.

Salemink, E., Van den Hout, M., & Kindt, M. (2007a). Trained interpretive bias: Validity and effects on anxiety. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 212–224. doi:10.1016/j.jbtep.2006.10.010.

Salemink, E., van den Hout, M., & Kindt, M. (2007b). Trained interpretive bias and anxiety. *Behaviour Research and Therapy, 45*, 329–340. https://doi.org/10.1016/j.brat.2006.03.011

Salemink, E., Van den Hout, M., & Kindt, M. (2009). Effects of positive interpretive bias modification in highly anxious individuals. *Journal of Anxiety Disorders, 23*, 676–683. doi:10.1016/j.janxdis.2009.02.006.

Salemink, E., van den Hout, M. A., & Kindt, M. (2010a). Generalisation of modified interpretive bias across tasks and domains. *Cognition & Emotion, 24*, 453–464. https://doi.org/10.1080/02699930802692053

Salemink, E., van den Hout, M. A., & Kindt, M. (2010b). How does cognitive bias modification affect anxiety? Mediation analyses and experimental data. *Behavioural and Cognitive Psychotherapy, 38*, 59–66. https://doi.org/10.1017/S1352465809990543

Van der Krieke, L., Boonstra, N., & Malda, A. (2014). Bias blaster: A game to beat interpretation bias in psychosis. *Psychiatric Services, 65*. https://doi.org/10.1176/appi.ps.650701, 961-961.

de Voogd, E. L., De Hullu, E., Burnett Heyes, S., Blackwell, S. E., Wiers, R. W., & Salemink, E. (2017). Imagine the bright side of life: A randomized controlled trial of two types of interpretation bias modification procedure targeting adolescent anxiety and depression. *PLoS One, 12*, Article e0181147, 10.1371/journal.pone.0181147.

Wartena, B. O., & Van Dijk, H. W. (2013). Bias blaster – aiding cognitive bias modification-interpretation through a bubble shooter induced gameflow. *Games for Health*, 47–60. https://doi.org/10.1007/978-3-658-02897-8_4

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6*(3), 291–298. https://doi.org/10.1177/1745691611406923