Evaluative Conditioning as a Body Image Intervention for Adolescents With Eating Disorders

Glashouwer, Klaske A.; Neimeijer, Renate A. M.; de Koning, Marlies L.; Vestjens, Michiel; Martijn, Carolien

Published in:
Journal of Consulting and Clinical Psychology

DOI:
10.1037/ccp0000311

Publication date:
2018

Evaluative Conditioning as a Body Image Intervention for Adolescents With Eating Disorders

Abstract

Objective: The aim was to investigate whether a computer-based evaluative conditioning intervention improves body image in adolescents with an eating disorder. Positive effects were found in earlier studies in healthy female students in a laboratory and a field setting. This study is the first to test evaluative conditioning in a clinical sample under less controlled circumstances. Method: Fifty-one adolescent girls with an eating disorder and a healthy weight were randomly assigned to an experimental condition or a placebo-control condition. The computerized intervention consisted of six online training sessions of 5 minutes, in which participants had to click on pictures of their own and other people's bodies. Their own pictures were systematically followed by portraits of friendly smiling faces. In the control condition, participants were shown the same stimuli, but here, a stimulus was always followed by another stimulus from the same category, so that own body was not paired with smiling faces. Before, directly after, three weeks after and 11 weeks after the intervention, self-report measures of body image and general self-esteem were administered. Automatic self-associations were also measured with an Implicit Association Test (IAT). Results: In contrast to our hypotheses, we did not find an effect of the intervention on self-report questionnaires measuring body satisfaction, weight and shape concern, and general self-esteem. In addition, the intervention did not show positive effects on implicit associations regarding self-attractiveness. Conclusions: These findings do not support the use of evaluative conditioning in its present form as an intervention for adolescents in clinical practice.

Keywords: Body image, Intervention, Evaluative conditioning, Eating disorders, Randomized clinical trial

Public Health Significance Statement: This study investigated a new intervention to improve body image in adolescents with eating disorders. Outcomes do not support the use of evaluative conditioning in its present form as a body image treatment for adolescents.

Evaluative Conditioning as a Body Image Intervention for Adolescents With Eating Disorders

Negative body image is a core characteristic of eating disorders (DSM-5), and is considered to be a key risk factor for the onset, maintenance and relapse of eating disorders (Fairburn, Peveler, Jones, Hope, & Doll, 1993; Stice & Shaw, 2002; Carter, Blackmore, Sutandard-Pinnock, & Woodside, 2004; Johnson & Wardle, 2005; Neumark-Sztainer, Paxton, Hannan, Haines, & Story, 2006). Body image is a complex construct encompassing thoughts, behaviors, feelings and evaluations related to one's body (Cash, 2011). A negative body image may express itself as a preoccupation and dissatisfaction with one's shape and weight. For those with a negative body image, weight and shape influence to a large extent how they judge themselves as a person. Some studies have shown substantial reductions in negative body image following interventions based on cognitive-behavioral therapy (e.g., Butters & Cash, 1987; Rosen, Reiter, & Orosan, 1995; McLean, Paxton & Wertheim, 2011), counter attitudinal therapy (e.g., Stice, Rohde, Butryn, Menke & Marti, 2015), and mirror exposure (e.g., Hildebrandt, Loeb, Troupe & Delinsky, 2012; Glashouwer, Jonker, Thomassen & de Jong, 2016). However, a recent meta-analysis of *stand-alone* interventions for body image showed that once corrections for several sources of bias were applied, existing interventions only led to small overall improvements in body image (Alleva, Sheeran, Webb, Martijn, & Miles, 2015). This points to the need for further improvement of current treatment approaches. Recent research has shown promising results for a body image intervention based on principles of evaluative conditioning in which participants learned to associate their body with positive social feedback (Martijn, Vanderlinden, Roefs, Huijding, & Jansen, 2010; Aspen et al., 2015). The aim of the present study was to investigate whether this evaluative conditioning could also help to improve negative body image in a clinical sample of adolescent girls with an eating disorder.

Evaluative conditioning refers to changes in the valence of an object (i.e., conditioned stimulus; CS) as a result of pairing the object with a positive or negative stimulus (i.e., unconditioned stimulus; US) (for a comprehensive review see De Houwer, Thomas, & Baeyens, 2001). Evaluative conditioning has already been extensively studied by researchers from diverse backgrounds using various stimuli and paradigms. Most relevant for the present study is the "picture – picture paradigm", originally developed by Levey and Martin (1975). These authors were the first to demonstrate that pairing a neutral picture (CS) with a previously liked picture (US) changes the evaluation of the neutral picture in a positive direction. Evaluative conditioning has also been applied to non-neutral objects such as the self (Baccus, Baldwin, & Packer, 2004; see also: Dijksterhuis, 2004 for related research). This research took place in a laboratory setting. Students had to click on self-relevant stimuli appearing on a computer screen (e.g., place of birth or first name; CS). After each self-relevant stimulus, a picture of a positive social stimulus (i.e., smiling face; US) was presented. Non-self-relevant stimuli were paired with non-smiling faces. Compared to those in a control condition, participants in the training condition showed an increase in positive automatic associations with the self (Baccus, Baldwin, & Packer, 2004). In a subsequent study, a similar intervention lead to a reduction in adolescents' aggressive feelings and intentions in response to social rejection (Baldwin, Baccus & Milyavskaya, 2010).

Martijn et al. (2010) investigated whether body satisfaction could be increased using an adapted evaluative conditioning procedure. They developed a computerized conditioning training task in which images of the participants' own body were used as CS and pictures of smiling faces were used as US. The purpose of the evaluative training was to teach individuals to associate their body with "new", more positive, evaluations which can counter or inhibit the "old" negative evaluations of their body, therefore increasing body satisfaction. This procedure can be considered a form of evaluative counter conditioning, since body

dissatisfied individuals have a negative evaluation of their own body (CS) The evaluative conditioning procedure was first tested in a controlled laboratory setting among healthy female students. In this study, 54 women with low and high body concern were randomly assigned to either an experimental or a placebo-control condition. Participants completed one session of the conditioning task in which they had to click (as fast as possible) on photographs of their own and other people's bodies. After clicking, the body picture disappeared and was replaced by a short presentation of a face with an emotional expression. In the experimental condition, pictures of their own body (CS) were consistently followed by pictures of smiling faces (US), whereas photographs of control bodies were followed by pictures of neutral or frowning faces. In the control condition, all body pictures were randomly followed by the same pictures of smiling, neutral, and frowning faces. Results showed that body satisfaction and general self-esteem increased directly after the training procedure for women in the experimental condition but not for those in the control condition. This evaluative training procedure was subsequently tested in a field experiment among 39 female students at risk for developing an eating disorder (Aspen et al., 2015). This study was a randomized waitlist-controlled trial in which the experimental group received four sessions of the conditioning training within a 4-week period. The training sessions were administered in a controlled setting under supervision. Again, women in the experimental group showed a decrease in shape and weight concern as well as an increase in self-esteem following the training procedure, as compared to those in the waitlist-control group. Importantly, despite the brevity of the training (4 x 5 minutes), improvements with respect to body image were maintained even at 4-week and 12-week follow-ups.

Considering these promising pre-clinical findings, we decided to translate this computer-based evaluative conditioning training into an intervention for clinical practice. In the current study, we investigated its effectiveness as an intervention for improving body

image in a clinical sample of adolescents with eating disorders. Eating disorders typically begin during adolescence. The development of effective treatments for this age group may help to interrupt the chronic course of eating disorders (Schmidt et al., 2016). The present study used a crossover design in which participants (N = 51) were randomly divided across an experimental condition and a placebo-control condition. Since we expected a clinical population to have a more negative body image than populations with subthreshold/subclinical symptoms, the amount of experimental training was increased to six evaluative conditioning sessions to be given over a 3-week period. To enhance the acceptability and feasibility of intervention implementation, the training sessions were not administered in a controlled setting, but online via personal computers at home, in order to minimize patient burden. Primary outcome measures included self-report questionnaires of body satisfaction, weight and shape concern, and general self-esteem. These were assessed at baseline, post intervention and again after three and 11 weeks. In addition, we included an Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) at pre- and post-intervention to investigate the effect of the training on automatic associations related to self-attractiveness (cf. Baccus et al, 2004; Dijksterhuis, 2004). We hypothesized that the experimental group would show a greater improvement on the primary outcome measures at post intervention than the control group; and we explored whether these changes would be maintained at three- and 11-week follow-up.

Method

Participants

Fifty-one adolescent girls with eating disorders (M_{age} = 16.73, SD = 2.45) were recruited through the Department of Eating Disorders of Accare, a facility for child and adolescent psychiatry in the Netherlands. All participants included in the study were at least 12 years old, had a good comprehension of the Dutch language, and were suffering from an

eating disorder as diagnosed by health care professionals of Accare using the (Dutch) child version of the Eating Disorder Examination (ChEDE; Bryant-Waugh, Cooper, Taylor, & Lask, 1996; Decaluwé & Braet, 1999). Participants were undergoing treatment for anorexia nervosa of the restrictive type (n = 15), anorexia nervosa of the purging type (n = 5), atypical anorexia nervosa (n = 7), bulimia nervosa (n = 9), or another specified eating disorder (n = 15; i.e., 8 with features of AN-R, 4 with features of AN-P, 2 with features of BN, 1 with features of BED). Participants could only participate if they had a healthy weight, as we wanted to exert caution with regard to recruiting those in the unhealthy weight range. Since Body Mass Index (BMI; weight/height^2) in children changes substantially with age, an age-related cut-off score is necessary to be able to compare the BMIs of adolescents. Adjusted BMI scores were therefore calculated ((actual BMI/Percentile 50 of BMI for age and gender) x 100; cf. Le Grange et al., 2012). The 50th percentile of BMI for age and gender was obtained from the Netherlands Organization for Applied Scientific Research (TNO, 2010). Participants with adjusted BMI scores between 85% and 140% were included in the study (cf. Van Winckel & Van Mil, 2001; M_{BMI_adj} = 98.05, SD = 7.64, range = 87.78 – 120.88). Participants who were diagnosed with anorexia nervosa were first required to gain enough weight to obtain a minimal adjusted BMI of 85% before they could participate in this study. Participants were randomly divided between the experimental condition (n = 25) and the control condition (n = 26). Groups did not differ significantly from each other in terms of age or adjusted BMI. The study protocol was approved by the Medical Ethical Committee of the University Medical Center Groningen (UMCG; NL51113.042.15) and the trial was pre-registered in the Dutch Trial Register (NTR5451). Participants (and, if younger than 18 years, their parents or a guardian with parental authority) actively gave informed consent before the start of the study.

Measurements

Negative body image. Body dissatisfaction was indexed with the 6-item Body Image States Scale (BISS; Cash, Fleming, Alindogan, Steadman, & Whitehead, 2002). BISS items were scored on a visual analogue scale (ranging from 0-100). In our sample, Cronbach's α (internal consistency) of the BISS at pre-intervention, post-intervention, 4-week follow-up and 11-week follow-up varied between .89 and .95. Higher scores indicate higher body satisfaction.

Shape and weight concern were measured with the 5-item weight concern and 8-item shape concern subscales of the Eating Disorder Examination Questionnaire (EDE-Q; Fairburn & Beglin, 2008). These subscales include items assessing the affective-evaluative dimension (e.g., body dissatisfaction, fear of gaining weight) and the cognitive-behavioral dimension (e.g., importance of and preoccupation with shape/weight) of body image, as defined by Cash (2011). We adjusted the original time-window of 28 days to 21 days to match our study design. Items measured negative body image during the last 21 days and were answered on a 7-point scale ranging from 0 (no days) to 6 (every day). We adapted the wording of some items slightly to make them appropriate and understandable for the adolescent age group. The weight and shape concern subscales showed good internal consistency within this study with α's at all assessment points varying between .86 and .97. Means score per subscale were calculated in such a way that higher scores indicate higher shape and weight concern.

Self-esteem. General self-esteem was measured with a Dutch adaptation (for adolescents) of the Rosenberg Self-Esteem Scale (RSES, cf. Mayer, Muris, Meesters, & Zimmermann-van Beuningen, 2009). Fifteen items based on the original RSES (Rosenberg, 1989) were rated on a five-point scale ranging from 0 (completely untrue) to 4 (completely true). After recoding the reverse-scored items, a total score was calculated and used as an index of self-esteem (range 0-60). The RSES showed good internal consistency in our sample,

with α's at all assessment points varying between .93 and .96. Increases in RSES scores are indicative of higher self-esteem.

Automatic self-associations. Automatic associations related to self-attractiveness were assessed with an Implicit Association Test (IAT), a computerized reaction time task originally designed by Greenwald et al. (1998) to measure the relative strengths of automatic associations between two target categories and two attribute categories. In this study, target categories were "I" and "Other", and each category consisted of five stimulus words (I: I, mine, own, myself, self; Other: they, their, other, you, themselves). Attribute categories were "Beautiful" and "Ugly", and again, each category consisted of five stimulus words (Beautiful: beautiful, radiant, nice, pretty, attractive; Ugly: ugly, boring, stupid, dull, unattractive; stimuli are translated from Dutch). Stimuli across categories were matched on the number of syllables and characters. The IAT consisted of seven blocks (see Table 1).

Stimuli from all four categories appeared in randomized order in the middle of a computer screen and participants were instructed to sort them with a left or right response key. The category labels stayed visible in the upper left and right-hand corners of the screen for the duration of the whole task. The premise here is that the sorting becomes easier when a target and attribute that share the same response key are strongly associated than when they are weakly associated. Before the start of a new sorting task, written instructions were presented on the screen. Following a correct response, the next stimulus was presented with a 500 ms delay. Following an incorrect response, the word 'wrong' appeared shortly above the stimulus, and the stimulus remained on the screen until the correct response was given. The order of the blocks was fixed across participants to reduce method variance.

Raw response latencies of the IAT were transformed into D-scores using the D-algorithm (D1; Greenwald, Nosek, & Banaji, 2003). Error latencies were replaced by the response latencies of the correct responses that participants made after the error (and reaction

times above 10,000 ms) were discarded. D-scores were calculated by subtracting mean reaction times of Block 6 from Block 3 and Block 7 from Block 4. These two difference scores were divided by the pooled standard deviations based on all responses in the specific blocks and the mean was used as D-score (cf. Greenwald et al., 2003). Because there is still debate about the best way to calculate IAT scores, we repeated the analyses without dividing by the pooled SD (raw-score; Blanton, Jaccard, & Burrows, 2015). Outcomes did not differ markedly from analyses on the D-scores. The split-half reliability of the IAT was good in the present sample, with Spearman-Brown corrected correlations between test-halves of .86 and .89 at baseline and post intervention respectively (D-scores based on trials 1, 2, 5, 6, 9, 10 etc. vs. 3, 4, 7, 8, 11, 12 etc.). D-scores were computed such that higher scores reflect a stronger association between *I* and *beautiful* (and *other* and *ugly*).

Secondary outcome measures. We developed a questionnaire to measure Perceptions of Social Approval for Appearance (PSAA). Participants were asked to indicate (on a visual analogue scale where 0 = not at all and 100 = totally) to what extent they expected others to think that nine characteristics (e.g. attractive, beautiful) applied to their appearance and figure. After recoding the reverse-scored items, a mean score was calculated (range 0-100). The scale showed good internal consistency in our sample, with α's at pre- and post-intervention of .86 and .92 respectively. Higher scores indicate a more positive perception of social approval.

We also included the 5-item restraint and 5-item eating concern subscales of the EDE-Q as secondary outcome measures. The subscale items were adjusted in a similar way as the rest of the EDE-Q (see prior description). The restraint and eating concern subscales showed good internal consistency within this study with α's at pre- and post-intervention varying between .81 and .86. Higher scores indicate higher restraint and eating concern.

Finally, during all assessments and after each training session, participants were asked to indicate (on a visual analogue scale where 0 = not at all and 100 = totally) how satisfied

they were at that moment with their body and with themselves in general. These items were included to be able to explore the course of symptoms in more detail over time.

Evaluative Conditioning Intervention

Each training session consisted of 192 trials. Participants in the experimental condition were asked to click (as quickly as possible) on body pictures appearing on the computer screen at one of four places in a quadrant (see Martijn et al., 2010; for an illustration of the evaluative conditioning intervention). Body pictures comprised the two pictures taken of the participant at pretest and four standard pictures of two other girls (see *Stimuli* below). Each body picture was presented 16 times and presentation was counterbalanced across the four positions in the quadrant. After clicking on a body picture (either self or other), it disappeared, and a second picture of a face was presented for 400 ms in the same place. Pictures of the participants' bodies were always (100%) followed by a smiling face (64 trials). Pictures of the other girls' bodies were followed by pictures of neutral (50%, 64 trials) or frowning (50%, 64 trials) faces. Each session took about three to five minutes to complete. Participants in the control condition were presented with the same stimuli as in the experimental condition, but now a stimulus was always followed by another stimulus from the same category (e.g., own body picture 1> own body picture2; smiling face 1 > smiling face 2, etc.). This way, there was no link between body pictures and certain facial expressions.

An online log allowed us to determine whether participants carried out the training sessions as instructed. We also analyzed the reaction times from the six training sessions in the experimental and control conditions to check for compliance. Participants that completed the study always performed all of the training sessions. However, when taking into account the participants who dropped out, the average percentage of completed training sessions was 95.33 % for the experimental condition and 92.31 % for the control condition. In addition, results indicate that participants generally completed the training sessions in a conscientious

manner (RT: mean = 802 ms, SD = 189 ms, range = 514 – 1472 ms; mean % of trials > 3 s = 0.8 %).

Stimuli. Two full body pictures (front, profile) were taken of each participant against a white wall. Participants had been instructed to choose their favorite clothing prior to the session. Although participants were photographed fully clothed, they were instructed that their body shape should be clearly visible. In the front picture, participants looked into the lens. They could smile, but not to show their teeth. Participants selected the two pictures that they liked best. The four standard pictures of two other girls (acquaintances of the researcher, both with adjusted BMIs within the healthy range) were similar to the participants' body pictures, although they had been instructed to wear neutral clothing. The faces were selected from the NimStim Facial Stimuli Set2 (Tottenham et al., 2009) and consisted of 16 female and 16 male faces.

Procedure

This study had a crossover design in which participants were randomly allocated to an experimental group or a control group. Randomization occurred automatically when a new account was created via the online training platform. We did not use stratification strategies. The experimental training procedure consisted of six evaluative conditioning sessions spanning a 3-week period. Participants in the control condition received six sessions of the placebo training within an equivalent time-frame. After the placebo training was completed, participants in the control group received six additional sessions of the experimental training. Information about the design and drop-out rate is summarized in Figure 1.

Patients undergoing treatment at the Department of Eating Disorders of Accare who fulfilled the inclusion criteria were informed about the study by their therapist. Those who expressed an interest in participating were then contacted by the researcher to schedule an appointment for the pre-intervention assessment, photoshoot and first training session. All

participants were told that they would receive an intervention which had resulted in positive effects on body image in previous studies among individuals without eating disorders. They were told that they would be allocated to either a "short version" (i.e. the experimental group receiving six sessions) or a "long version" (i.e. the control group receiving 12 sessions; first six placebo sessions and subsequently six experimental sessions) of the intervention. Participants were informed that the training sessions could also contain elements that might not be effective, but we did not emphasize this information. The researcher only became aware of which condition the participant was allocated to after the first training session had been completed. The researcher then told the participant whether she was in the "short" or the "long" condition, so that the participant knew how many training sessions to expect. In general, participants had positive expectations of the training procedure, and were not aware of which condition they had been assigned to, only whether they received the long or the short version of the training. After the data collection was completed, participants were debriefed by email.

Baseline measures were completed by the participant in the following order: BISS, EDE-Q, RSES, short questions, PSAA, IAT. After this, the body pictures were taken. The researcher immediately edited and uploaded the photograph in an online program and the participant completed the first training session at the end of the appointment. The first assessment took approximately 45-60 minutes. Participants completed the remaining training sessions and assessments online via their personal computers at home in order to minimize participant burden. Participants received automatic invitations via e-mail when a training session or assessment was scheduled, and reminders were sent when someone did not participate. If a participant did not respond, the researcher tried to contact her via e-mail or phone. Three weeks and 11 weeks after their last training session, participants again completed the self-report measurements using an online survey. The IAT was only included

in the pre (T1) and post (T2) assessments so as to keep the assessments as short as possible and therefore increase the feasibility of the study. Participants received a small gift for their participation. The intervention was implemented in addition to the participants' regular treatment for their eating disorders.

Statistical Analyses

To test the short-term effects of the intervention on body satisfaction, weight and shape concern, general self-esteem, and automatic associations related to self-attractiveness, five separate ANCOVAs were performed with Condition (experimental, placebo) as a between-subject factor and T2-scores on the BISS, EDE-Q weight concern, EDE-Q shape concern, and the IAT as dependent variables. The T1 score of each dependent variable was included as a covariate. To correct for multiple testing, the alpha criterion was set at .01 (p = .05/5). We repeated these analyses for our secondary outcome measures: eating concern, dietary restraint and perceptions of social approval for appearance. We decided to repeat the ANCOVAs for the primary and secondary outcome measures using Bayesian hypothesis testing. This allowed us to quantify the evidence regarding the null hypothesis for each outcome measure. Statistical analyses were conducted using the free software JASP using default Cauchy priors (JASP Team, 2017). To facilitate the interpretation, we reported Bayes factors expressed as $BF_{01,}$ grading the intensity of the evidence that the data provide for H_0 (i.e. condition has no effect on the outcome measure over and above T1 scores of the dependent variable) versus H_1 (i.e. condition effects the outcome measure over and above T1 scores of the dependent variable).

In addition, to test whether the expected effect of the intervention was replicated in the control condition (in which the experimental training sessions were administered after the placebo training), we planned four additional ANCOVAs on body satisfaction, weight and shape concern, and general self-esteem, using the post-experimental training scores as

dependent variables, i.e. T2 for the experimental condition and T3 for the control condition. Again, Condition (experimental, control) was included as a between-subject factor and the pre-scores were included as covariates, i.e. T1 for the experimental condition and T2 for the control condition (see Figure 1 for an overview of the design).

Finally, to explore the longer-term effects of the intervention, four separate repeated measures ANOVAs were conducted in the total sample with Time (pre-training, post-training, 3-week follow-up, 11-week follow-up) as a within-subject factor and scores on the four primary outcome measures as dependent variables. For the control condition, we used scores at T2 as pre-training to keep the time of assessment before the experimental training consistent with that of the experimental condition. Polynomial trend analyses were used to examine the development of the scores on the dependent measures over time.

Missing Data and Drop-outs

During the course of the intervention, 10 participants dropped out before T2 (19.6%), and another five participants dropped out after T2 (total drop-out % = 29.4%). Drop-outs did not differ significantly from those who completed the intervention on any of the pre-intervention scores of the primary outcome measures. Missing data were estimated using multiple imputation (Schafer & Graham, 2002). Missing data were imputed 40 times using a linear regression model (IBM SPSS Statistics 24). Imputation was based on all predictors that were included in the model as well as other variables (e.g., age) in order to impute as accurately as possible. We report the pooled results.

The data of three participants were excluded from the IAT analyses because their mean reaction times exceeded the cutoff criterion of 2.5 *SDs* above the grand mean of the task (*M* = 829 ms, *SD* = 136 ms, threshold = 1171 ms) or because the error rates exceeded the cutoff criterion of 2.5 *SDs* above the grand mean of the task (*M* = 6.25 %, *SD* = 4.93 %, threshold = 18.6 %).

Results

Short-term Intervention Effects

Primary outcome measures. The experimental condition and the control condition did not differ significantly from each other on pre-intervention scores of the primary outcome measures (BISS: $t(49) = -.76$, $p = .45$; EDE weight concern: $t(49) = .95$, $p = .35$; EDE shape concern: $t(40.14) = 1.47$, $p = .15$; RSES: $F(1, 48) = t(49) = -.89$, $p = .38$; IAT: $t(46) = .26$, $p = .80$). In all five ANCOVA's, scores at pre-intervention were significantly and strongly related to scores at T2 (BISS: $F(1, 48) = 95.26$, $p < .001$, partial $\eta^2 = .66$; EDE weight concern: $F(1, 48) = 96.26$, $p < .001$, partial $\eta^2 = .66$; EDE shape concern: $F(1, 48) = 178.79$, $p < .001$, partial $\eta^2 = .78$; RSES: $F(1, 48) = 286.02$, $p < .001$, partial $\eta^2 = .85$; IAT: $F(1, 45) = 18.10$, $p = .015$, partial $\eta^2 = .27$). However, none of the analyses showed significant effects of condition on the primary outcome measures (BISS: $F(1, 48) = .42$, $p = .64$, partial $\eta^2 = .01$; EDE weight concern: $F(1, 48) = .78$, $p = .58$, partial $\eta^2 = .02$; EDE shape concern: $F(1, 48) = .26$, $p = .72$, partial $\eta^2 = .01$; RSES: $F(1, 48) = .24$, $p = .74$, partial $\eta^2 = .01$; IAT: $F(1, 45) = .61$, $p = .57$, partial $\eta^2 = .01$). To summarize, in contrast to our expectations, we found no evidence that the experimental training procedure leads to positive short-term effects on body satisfaction, weight and shape concern, general self-esteem, or automatic associations related to self-attractiveness. Since we did not find any effects of the training on primary outcome measures, we did not conduct the additional ANCOVAs once participants in the control condition had also received the experimental training sessions. Table 2 provides an overview of means and standard deviations for the primary outcome measures at all assessment points. In order to examine body satisfaction and self-esteem over the course of the six training sessions, we also report the means and standard deviations of the single items measuring state body satisfaction and self-esteem after each training session per group.

Outcomes of Bayesian hypothesis testing were in line with the outcomes of the frequency statistics showing that the observed data are 1.43 to 3.23 times more likely under H_0 than under H_1 (BISS: $BF_{01} = 3.22$; EDE weight concern: $BF_{01} = 1.43$; EDE shape concern: $BF_{01} = 1.48$; RSES: $BF_{01} = 3.23$; IAT: $BF_{01} = 2.76$). Results indicate that there is moderate evidence favoring H_0 over H_1 for BISS and RSES (Lee & Wagenmakers 2013; adjusted from Jeffreys 1961). The strength of the evidence for the other outcome measures is "anecdotal" (i.e. inconclusive).

Secondary outcome measures. In all three ANCOVA's, scores at pre-intervention were significantly and strongly related to scores at T2 (EDE restraint: $F(1, 48) = 45.61$, $p < .001$, partial $\eta^2 = .48$; EDE eating concern: $F(1, 48) = 116.28$, $p < .001$, partial $\eta^2 = .70$; PSAA: $F(1, 48) = 55.35$, $p < .001$, partial $\eta^2 = .53$). However, again, none of the analyses showed significant effects of Condition (EDE restraint: $F(1, 48) = .29$, $p = .71$, partial $\eta^2 = .01$; EDE eating concern: $F(1, 48) = 1.09$, $p = .42$, partial $\eta^2 = .02$; PSAA: $F(1, 48) = 2.58$, $p = .20$, partial $\eta^2 = .05$). We therefore found no evidence that the intervention leads to positive short-term effects on restraint eating, eating concern and perceived social approval for appearance.

Outcomes of Bayesian hypothesis testing were in line with the outcomes of the frequency statistics showing that the observed data are 0.84 to 3.10 times more likely under H_0 than under H_1 (EDE restraint: $BF_{01} = 3.09$; EDE eating concern: $BF_{01} = 3.10$; PSAA: $BF_{01} = 0.84$). There is moderate evidence favoring H_0 over H_1 for EDE restraint and EDE eating concern. The strength of the evidence for the PSAA is inconclusive.

Longer-term Intervention Effects

RM-ANOVAs showed main effects of Time for all primary outcome variables (BISS: $F(2.69, 134.68) = 7.00$, $p = .002$, partial $\eta^2 = .12$; EDE weight concern: $F(2.41, 120.29) = 13.05$, $p < .001$, partial $\eta^2 = .21$; EDE shape concern: $F(2.19, 109.66) = 14.02$, $p < .001$, partial $\eta^2 = .22$; RSES: $F(2.12, 103.11) = 5.95$, $p = .033$, partial $\eta^2 = .10$). For all primary

outcome variables Mauchly's test of sphericity was significant. Consequently, Huynh-Feldt corrected tests are reported for these variables. Polynomial contrasts showed significant linear trends for all variables (Fs > 9.18, ps < .022, partial η^2s >.15), but not quadratic or cubic trends. These outcomes indicate a general improvement over time on the outcome measures across groups.

Discussion

The present study was the first to investigate the effectiveness of evaluative conditioning as a body image intervention for adolescents with eating disorders. In contrast to our hypotheses, we did not find an effect of our intervention on self-report questionnaires of body satisfaction, weight and shape concern, and general self-esteem. Moreover, the intervention did not result in more positive implicit associations related to self-attractiveness, as measured by an IAT. State items measuring body satisfaction and general self-esteem during the intervention indicate that both groups remained stable over the course of the training sessions. Additional Bayesian hypothesis testing confirmed the outcomes of the frequency statistics showing no effects of the intervention on any of the outcome variables. Results indicate that the evidence was moderate for body satisfaction and general self-esteem, favoring the null hypothesis over the alternative hypothesis. The strength of the evidence concerning the other primary outcome measures should be interpreted as inconclusive.

The present findings do not support our hypotheses and are not consistent with pre-clinical studies showing a positive effect of evaluative conditioning on body image and self-esteem (Martijn et al., 2010; Aspen et al., 2015). This could indicate that we failed to create positive enough evaluations related to body image to counter participants' initially (highly) negative evaluations. As a result, body satisfaction may not have increased in the experimental group as compared to the control group. This explanation is consistent with the literature showing that evaluative conditioning is more successful for CSs that are

evaluatively neutral than for CSs that have a marked valence (Hofmann, De Houwer, Perugini, Baeyens & Crombez, 2010). This is especially the case for negative evaluations, which are usually easier to learn and harder to unlearn than positive evaluations (De Houwer et al., 2001). Self-report measures indicate that our clinical sample of eating disorder patients was characterized by more severe body dissatisfaction than prior pre-clinical samples (Martijn et al., 2010; Aspen et al., 2015). This might explain why we failed to "counteract" these negative evaluations in the present sample. Although we already increased the dose of the intervention from four to six sessions, it is possible that more sessions are needed in order to achieve an effect. Future research should investigate whether this is the case.

However, important methodological differences between the present study and prior pre-clinical studies might also explain why the outcomes of our study differed from the two pre-clinical studies. In the process of translating laboratory experiments into a clinical intervention, changes are made to make the intervention suitable, feasible and, acceptable for use in clinical practice. In the present study, we allowed participants to wear their own clothes instead of standardized clothes during the photoshoot. Moreover, training sessions and measurements were not administered in a controlled setting, but (for the most part) online via personal computers at home. It should also be noted that the intervention was tested in an adolescent sample rather than an adult sample. The relatively simple and repetitive training procedure might have been too "boring" for the adolescent age group that is used to advanced computer games. Furthermore, the intervention was administered next to treatment as usual, while this was not the case in pre-clinical studies. Finally, although the sample was rather homogeneous - all participants were adolescent girls with an eating disorder and with a healthy weight - we observed substantial variance in body image indices within groups. Consequently, it could be that the experimental training procedure did work to some degree, but that the effect of evaluative conditioning was too small to show an effect over and above

the inevitable noise that comes with implementing an intervention in clinical practice. It may be more fruitful to "turn back the clock" in future clinical studies by administering the training sessions in a controlled setting rather than online at home. It would also be interesting to test the intervention in an adult clinical sample.

Despite the strengths of the present design (we were the first to study a clinical group using a randomized placebo-controlled design and including a behavioral outcome measure), there are some limitations which should also be taken into consideration. Most notable is the lack of a manipulation check. It is reassuring that reaction time data indicate that participants generally carried out the training tasks in a conscientious manner. Nevertheless, future studies should test whether the evaluative conditioning training successfully changes the valence of the CS. This could be examined, for example, by using an evaluative priming task in which the body stimuli are included as primes. This would make it possible to determine whether the training procedure was effective but did not influence the outcome measures, or whether the training task itself did not work. A second limitation is the small sample size of this study, increasing the chance of type-II errors. To be able to quantify the evidence regarding the null hypothesis for each outcome measure, we repeated the analyses with Bayesian hypothesis testing. These analyses indicate that we can be quite confident that the training procedure did not influence body satisfaction and general self-esteem. However, the strength of the evidence concerning the other primary outcome measures is inconclusive. A third limitation of this study is the diagnostic heterogeneity of the sample which might have hampered the detection of intervention effects. However, it should be noted that prior studies with similar diagnostic heterogeneity have found significant reductions in negative body image (e.g., Stice, Rohde, Butryn, Menke & Marti, 2015; Hildebrandt, Loeb, Troupe & Delinsky, 2012). Finally, although the standard pictures of the control bodies were adapted to the age category of the participants, it was not feasible to adapt the face stimuli. Consequently, the face stimuli that

were used as feedback in the training were of an older age (approximately 20-30 years) than the participants (M_{age} = 16.73, *SD* = 2.45). This age difference could have made the intervention less effective, especially since it has been shown that the nature of the relationship between the CS and US is important (*belongingness*; De Houwer et al., 2001). Evaluative conditioning works best when the relationship between the CS and US is believable and relevant. Smiling faces of "older" people may be less believable or relevant to adolescents than smiling faces of people their own age.

Conclusions

Our study did not provide evidence for the effectiveness of evaluative conditioning as an intervention for body image in adolescents with eating disorders. Despite positive findings in pre-clinical samples, we did not find any positive effects of evaluative conditioning on body image, either in terms of self-report indices or a more implicit (automatic) measure of self-associations. Although participants generally improved over the 14-week course of the study, these changes cannot be attributed to the intervention. Present findings do not, therefore, support the use of evaluative conditioning (in its present form) as an intervention in clinical practice, at least not in its present form for the adolescent age-group. Moreover, these outcomes highlight the need to stringently test promising pre-clinical interventions in patient samples before implementing them in clinical practice.

References

Alleva, J. M., Sheeran, P., Webb, T. L., Martijn, C., & Miles, E. (2015). A meta-analytic review of interventions designed to improve body image. *PLoS ONE, 10,* e0139177. doi:org/10.1371/journal.pone.0139177

Aspen, V., Martijn, C., Alleva, J., Nagel, J., Perret, C., Purvis, C. . . . Taylor, C. B. (2015). Decreasing body dissatisfaction using a brief conditioning intervention. *Behaviour Research and Therapy, 69,* 93-99. doi:10.1016/j.brat.2015.04.003

Baccus, J. R., Baldwin, M. W., & Packer, D. J. (2004). Increasing implicit self-esteem through classical conditioning. Psychological Science, 15, 498–502. doi:10.1111/j.0956 –7976.2004.00708.x

Baldwin, M. W., Baccus, J. R., & Milyavskaya, M. (2010). Computer game associating self-concept to images of acceptance can reduce adolescents' aggressiveness in response to social rejection. *Cognition and Emotion, 24,* 855-862. doi:10.1080/02699930902884386

Bryant-Waugh, R., Cooper, P. J., Taylor, C. L., & Lask, B. D. (1996). The use of the eating disorder examination with children. A pilot study. *International Journal of Eating Disorders, 19,* 391–397. doi:10.1002/(SICI)1098- 108X(199605)19:4<391::AID-EAT6>3.0.CO;2-G.

Butters, J. W., & Cash, T. F. (1987). Cognitive-behavioral treatment of women's body-image dissatisfaction. *Journal of Consulting and Clinical Psychology, 55,* 889-897. doi:10.1037/0022-006X.55.6.889

Carter, J. C., Blackmore, E., Sutandard-Pinnock, K., & Woodside, D. B. (2004). Relapse in anorexia nervosa: A survival analysis. *Psychological Medicine 34,* 671–679. doi:10.1017/S0033291703001168

Cash, T. F. (2011). Cognitive-behavioral perspectives on body image. In T. F. Cash & L. Smolak (Eds.), *Body image: A handbook of science, practice, and prevention* (pp. 39-47). New York, NY: Guilford Press.

Cash, T. F., Fleming, E. C., Alindogan, J., Steadman, L., & Whitehead, A. (2002). Beyond body image as a trait: The development and validation of The Body Image States Scale. *Eating Disorders: The Journal of Treatment & Prevention, 10,* 103–113. doi:10.1080/10640260290081678

Decaluwé, V., & Braet, C. (1999). Dutch translation of the child eating disorder examination authored by C. G. Fairburn, Z. Cooper & R. Bryant-Waugh. Unpublished manuscript.

De Houwer, J., Thomas, S., & Bayens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127,* 853–869. doi:10.1037/0033–2909.127.6.853

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology,* 86, 345-355. doi:10.1037/0022-3514.86.2.345

Fairburn, C. G., & Beglin, S. J. (2008). Eating disorder examination questionnaire (EDE-Q 6.0). In C. G. Fairburn (Ed.), *Cognitive behavior therapy and eating disorders* (pp. 309-314). New York, NY: Guilford Press.

Fairburn, C. G., Peveler, R. C., Jones, R., Hope, R. A. & Doll, H. A. (1993). Predictors of twelve-month outcome in bulimia nervosa and the influence of attitudes to shape and weight. *Journal of Consulting & Clinical Psychology, 61,* 696-698. doi:10.1037/0022-006X.61.4.696

Glashouwer, K. A., Jonker, N. C., Thomassen, K., & de Jong, P. J. (2016). Take a look at the bright side: Effects of positive body exposure on selective visual attention in women

with high body dissatisfaction. *Behaviour Research and Therapy, 83,* 19-25. http://dx.doi.org/10.1016/j.brat.2016.05.006

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197-216. doi:10.1037/0022-3514.85.2.197

Hildebrandt, T., Loeb, K., Troupe, S., & Delinsky, S. (2012). Adjunctive mirror exposure for eating disorders: a randomized controlled pilot study. *Behaviour Research and Therapy, 50,* 797-804. http://dx.doi.org/10.1016/j.brat.2012.09.004

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010) Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136(3),* 390–421. DOI: 10.1037/a0018916

JASP Team (2017). *JASP (Version 0.8.4) [Computer software]*.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.) Oxford, UK: Oxford University Press.

Johnson, F., & Wardle, J. (2005). Dietary restraint, body dissatisfaction, and psychological distress: A prospective analysis. *Journal of Abnormal Psychology, 114(1),* 119-125. doi:10.1037/0021-843X.114.1.119

Lee, M. D., & Wagenmakers, E. -J. (2013). *Bayesian cognitive modeling: A practical course.* Cambridge University Press

Le Grange, D., Doyle, P. M., Swanson, S. A., Ludwig, K., Glunz, C., & Kreipe, R. E. (2012). Calculation of expected body weight in adolescents with eating disorders. *Pediatrics,* 129(2), 438–46.

Levey, A. B., & Martin, I. (1975). Classical conditioning of human 'evaluative' responses. *Behaviour Research and Therapy, 4,* 205-207.

Martijn, C., Vanderlinden, M., Roefs, A., Huijding, J., & Jansen, A. (2010). Increasing body satisfaction of body concerned women through evaluative conditioning using social stimuli. *Health Psychology, 29,* 514-520. doi:10.1037/a0020770

Mayer, B., Muris, P., Meesters, C., & Zimmermann-van Beuningen, R. (2009). Brief report: direct and indirect relations of risk factors with eating behavior problems in late adolescent females. *Journal of Adolescence, 32,* 741-745. doi:10.1016/j.adolescence.2008.

McLean, S. A., Paxton, S. J., & Wertheim, E. H. (2011). A body image and disordered eating intervention for women in midlife: A randomized controlled trial. Journal of *Consulting and Clinical Psychology, 79,* 751-758. doi:10.1037/a0026094

Neumark-Sztainer, D., Paxton, S. J., Hannan, P. J., Haines, J., & Story, M. (2006). Does body satisfaction matter: Five-year longitudinal association between body satisfaction and health behaviours in adolescent females and males. *Journal of Adolescent Health, 29,* 244-251. doi:10.1016/j.jadohealth.2005.12.001

Rosen, J. C., Reiter, J., & Orosan, P. (1995). Cognitive-behavioral body image therapy for body dysmorphic disorder. *Journal of Consulting and Clinical Psychology, 63,* 263-269. doi:10.1037/0022-006X.63.2.263

Rosenberg, M. (1989). *Society and the adolescent self-image.* Middletown, CT England: Wesleyan University Press.

Schmidt, U., Adan, R., Böhm, I., Campbell, I. C., Dingemans, A., Ehrlich, S., . . . Zipfel, S. (2016). Eating disorders: the big issue. *Lancet Psychiatry, 3(4),* 314-315. doi:10.1016/S2215-0366(16)00081-X

Schafer, J. L., & Graham, J. W. (2002). Missing data. Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Stice, E., Rohde, P., Butryn, M., Menke, K. S., & Marti, C. N. (2015). Randomized controlled pilot trial of a novel dissonance-based group treatment for eating disorders. *Behaviour Research and Therapy, 65,* 67-75. http://dx.doi.org/10.1016/j.brat.2014.12.012

Stice, E., & Shaw, H. E. (2002). Role of body dissatisfaction in the onset and maintenance of eating pathology: A synthesis of research findings. *Journal of Psychosomatic Research, 53(5),* 985-993. doi:10.1016/S0022-3999(02)00488-9

TNO. BMI-for-age charts. In: TNO Growth Charts [Internet]. 2010 [cited 12 Sep 2017]. Available: https://www.tno.nl/nl/aandachtsgebieden/gezond-leven/prevention-work-health/gezond-en-veilig-opgroeien/groeidiagrammen-in-pdf-formaat/.

Tottenham, N., Tanaka, J., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. A. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168,* 242–249. doi:10.1016/j.psychres.2008.05.006

Van Winckel, M., & Van Mil, E. (2001). Wanneer is dik té dik? [When is fat too fat?]. In C. Braet & M. Van Winckel (Eds.), Behandelstrategieën bij kinderen met overgewicht [Treatment strategies in overweight children] (pp. 11–26). Houten/ Diegem: Bohn Stafleu Van Loghum.

Table 1

Description of the Implicit Association Test

Block	Left Label(s)	Right Label(s)	No. of Trials
1. Practice	I	OTHER	10
2. Practice	BEAUTIFUL	UGLY	10
3. Practice	I + BEAUTIFUL	OTHER + UGLY	20
4. Test	I + BEAUTIFUL	OTHER + UGLY	40
5. Practice	OTHER	I	10
6. Practice	OTHER + BEAUTIFUL	I + UGLY	20
7. Test	OTHER + BEAUTIFUL	I + UGLY	40

Table 2

Means and Standard Deviations at All Assessments Points per Group

	Experimental group		Control group	
	Original data	Imputed data	Original data	Imputed data
BISS				
Pre-intervention 1	26.33 (17.14)	-	30.53 (21.82)	-
Pre-intervention 2[a]	-	-	34.24 (18.41)	30.69 (18.18)
Post-intervention	26.76 (17.91)	27.41 (17.34)	40.15 (19.33)	39.01 (19.90)
3-week follow-up	29.77 (20.71)	30.45 (20.01)	42.01 (19.54)	39.19 (21.15)
11-week follow-up	29.70 (21.45)	31.95 (21.77)	40.85 (18.84)	38.55 (19.76)
EDE weight concern				
Pre-intervention 1	3.90 (1.67)	-	3.43 (1.82)	-
Pre-intervention 2[a]	-	-	2.98 (1.79)	3.20 (1.72)
Post-intervention	3.35 (1.75)	3.40 (1.70)	2.35 (1.50)	2.55 (1.53)
3-week follow-up	3.43 (1.77)	3.38 (1.71)	2.23 (1.80)	2.67 (1.85)
11-week follow-up	3.19 (1.99)	3.03 (1.93)	2.06 (1.76)	2.38 (1.84)
EDE shape concern				
Pre-intervention 1	4.86 (1.03)	-	4.26 (1.80)	-
Pre-intervention 2[a]	-	-	3.63 (1.89)	3.97 (1.78)
Post-intervention	4.53 (1.19)	4.54 (1.16)	3.16 (1.73)	3.47 (1.67)
3-week follow-up	4.41 (1.44)	4.33 (1.44)	2.94 (2.04)	3.37 (1.95)
11-week follow-up	4.03 (1.72)	3.92 (1.69)	3.05 (1.93)	3.26 (1.79)
RSES				
Pre-intervention 1	16.80 (10.47)	-	19.58 (11.68)	-
Pre-intervention 2[a]	-	-	22.21 (12.35)	19.31 (12.50)

Post-intervention	16.59 (11.51)	16.69 (11.01)	23.00 (11.31)	21.76 (11.61)
3-week follow-up	18.00 (13.07)	18.38 (12.71)	24.63 (12.53)	22.65 (12.79)
11-week follow-up	19.85 (14.23)	21.23 (14.34)	24.75 (11.91)	24.04 (12.22)
IAT				
Pre-intervention (T1)	.26 (.51)	-	.22 (.43)	-
Post-intervention (T2)	.29 (.41)	.30 (.44)	.30 (.34)	.31 (.62)
VAS body satisfaction				
Pre-intervention (T1)	18.18 (16.65)		31.58 (25.71)	
Session 1	17.45 (17.05)		27.37 (24.38)	
Session 2	23.32 (19.79)		31.53 (24.61)	
Session 3	18.91 (16.51)		32.95 (24.43)	
Session 4	21.41 (20.24)		35.21 (26.12)	
Session 5	21.50 (20.97)		34.26 (25.07)	
Session 6	19.27 (18.55)		37.11 (26.71)	
Post-intervention (T2)	19.32 (14.64)		36.11 (25.22)	
VAS self-esteem				
Pre-intervention (T1)	29.36 (23.76)		33.00 (28.25)	
Session 1	25.77 (20.73)		29.63 (22.66)	
Session 2	29.32 (22.81)		32.89 (23.89)	
Session 3	27.05 (23.57)		35.42 (25.17)	
Session 4	29.23 (26.99)		38.53 (26.91)	
Session 5	27.05 (23.96)		34.16 (22.22)	
Session 6	26.32 (23.69)		37.58 (21.86)	
Post-intervention (T2)	26.91 (23.40)		33.63 (22.49)	

Note. BISS = Body Image States Scale (range 0-100, higher scores indicate higher body satisfaction), EDE = Eating Disorder Inventory (range 0-6, higher scores indicate higher weight and shape concern), RSES = Rosenberg Self-Esteem Scale (range 0-60, higher scores indicate higher self-esteem), IAT = Implicit Association Test (higher scores indicate a stronger automatic association between *I* and *beautiful* (and *other* and *ugly*), VAS = Visual Analogue Scale (range 0-100, higher scores indicate higher body satisfaction / self-esteem).
[a]The second measurement before the start of the experimental intervention training (pre-intervention 2) was assessed only in the control condition.

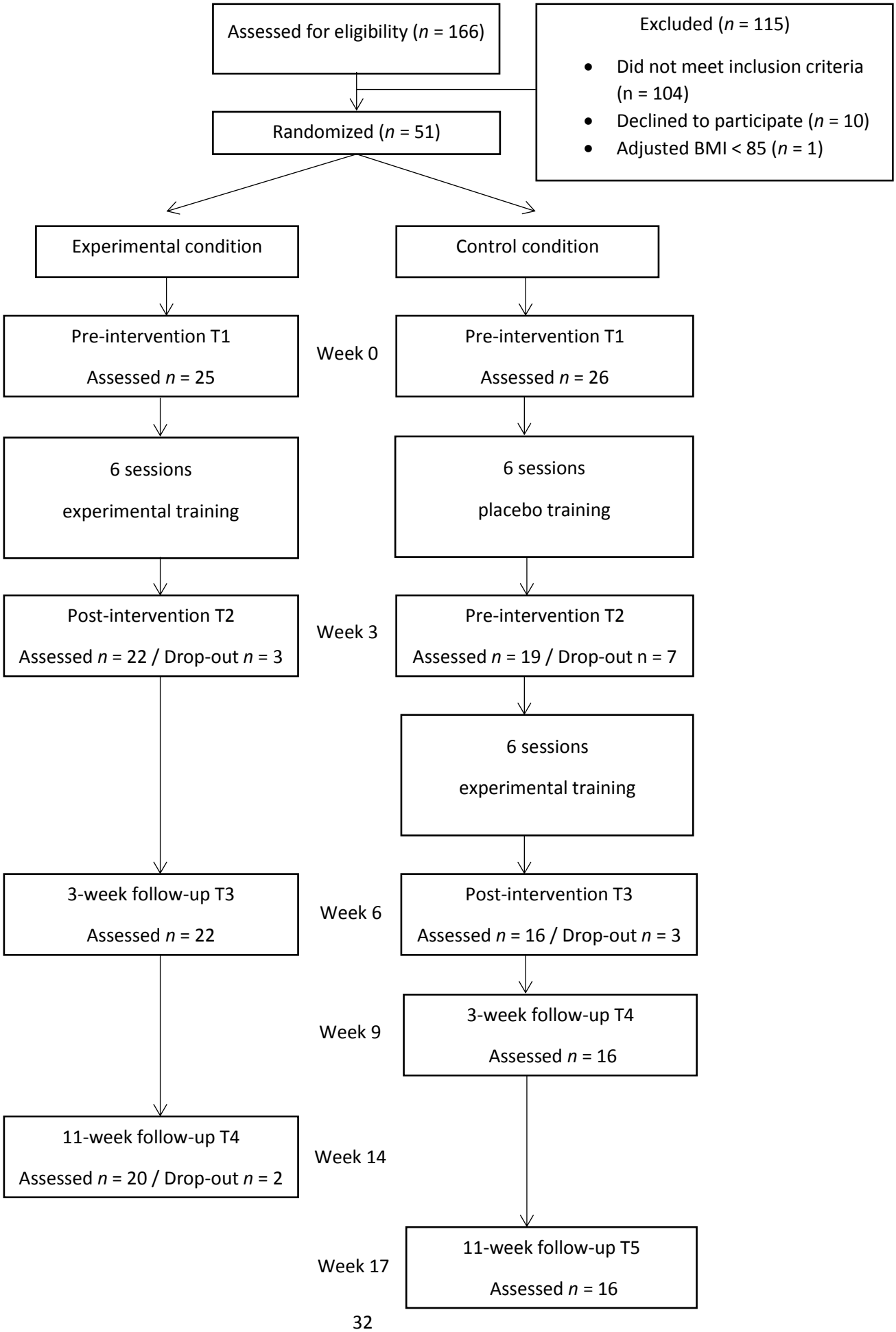

Figure 1. Study Design