Evaluative Conditioning Makes Slim Models Less Desirable as Standards for Comparison and Increases Body Satisfaction

Carolien Martijn  
Maastricht University

Paschal Sheeran  
University of Sheffield

Laura W. Wesseldijk  
Maastricht University

Hannah Merrick and Thomas L. Webb  
University of Sheffield

Anne Roefs and Anita Jansen  
Maastricht University

Objective: The present research tested whether an evaluative conditioning intervention makes thin-ideal models less enviable as standards for appearance-based social comparisons (Study 1), and increases body satisfaction (Study 2).

Design: Female participants were randomly assigned to intervention versus control conditions in both studies \((n = 66 \text{ and } 39)\). Intervention participants learned to associate thin-ideal models with synonyms of fake whereas control participants completed an equivalent task that did not involve learning this association.

Main outcome measures: The dependent variable in Study 1 was an implicit measure of idealization of slim models assessed via a modified Implicit Association Test (IAT). Study 2 used a validated, self-report measure of body satisfaction as the outcome variable.

Results: Intervention participants showed significantly less implicit idealization of slim models on the IAT compared to controls (Study 1). In Study 2, participants who undertook the intervention exhibited an increase in body satisfaction scores whereas no such increase was observed for control participants.

Conclusion: The present research indicates that it is possible to overcome the characteristic impact of thin-ideal models on women’s judgments of their bodies. An evaluative conditioning intervention made it less likely that slim models were perceived as targets to be emulated, and enhanced body satisfaction.

Keywords: body satisfaction, implicit associations, evaluative conditioning, social comparison

Analyses of the popular media have revealed that the ideal body shape for women has become increasingly thinner—a trend that is mirrored by fashion and celebrity models (Seifert, 2005). Indeed, the ideal body shape has become so extreme that photographs of female models typically are digitally altered in order to make them look even younger and thinner (Reaves, Hitchon, Park, & Yun, 2004). For most women, therefore, the ideal body shape as promoted in the media is not achievable and far removed from biological reality. Still, a vast amount of research has demonstrated that for certain women exposure to thin-ideal models is related to higher levels of body dissatisfaction and more disordered eating behavior (for meta-analyses, see Grabe, Ward, & Hyde, 2008; Groesz, Levine, & Murnen, 2002).

Social comparison is the key mechanism underlying the impact of thin-ideal models on body satisfaction is (e.g., Dijkstra, Gibbons, Buunk, 2010; Tiggeman & Slater, 2004). When most women compare themselves with a beauty ideal, they perceive a discrepancy between the self and socially prescribed standards of attractiveness, which in turn makes them feel bad about themselves (Strahan, Wilson, Cressman, & Buote, 2006). Evidence suggests that greater frequency of appearance-related social comparisons is associated with increased body dissatisfaction (e.g., Cattarain, Thompson, Thomas, & Williams, 2000; Want, 2009). Moreover, women who are dissatisfied with their bodies are especially likely to engage in comparisons with unrealistic beauty ideals (Heinberg & Thompson, 1995).

Given the pervasive influence of exposure to thin models on women’s body dissatisfaction, there is a clear need for interventions that prevent such influence. Holt and Ricciardelli (2008) classified interventions to prevent body dissatisfaction according to six categories developed by Stice and Shaw (2004): psychoeducational (e.g., increasing knowledge about nutrition and exercise), sociocultural resistance skills (e.g., critical thinking about
mass media), healthy weight-control skills (e.g., teaching healthy weight-control practices), stress and coping skills (e.g., stress management skills), self-esteem enhancement (e.g., strategies to promote self-esteem), and body satisfaction enhancement (e.g., promoting body acceptance). Although the interventions were generally effective at increasing knowledge, Holt and Ricciardelli concluded that there was little evidence that the programs reduced or prevented body image concerns. Similarly, a meta-analysis by Littleton and Ollendick (2003) found that even relatively intensive programs (involving an average of eight sessions) had a negligible impact on body satisfaction. 

One reason why interventions to date have not been very successful may be that the underlying process of social comparison was not tackled. Interventions that target the comparison process appear to have been more effective. For instance, Posavac, Posavac, and Weigel (2001) found that an intervention that stressed the inappropriateness of making comparisons with idealized images of beauty offset the impact of exposure to fashion models on body dissatisfaction. Strahan and colleagues (2008, Study 2) showed that when social cultural norms for appearance were challenged, women based their feelings of self-worth less strongly on appearance, resulting in less body dissatisfaction and less concern for others’ evaluations. Similarly, focusing on the motive to improve oneself rather than the motive to evaluate oneself overcame the impact of exposure to thin-ideal models (Halliwell & Dittmar, 2005). Thus, one approach to mitigating social comparison effects is to focus on perceivers (i.e., the women making the comparisons) and try to shape (a) the likelihood that they will make comparisons, or (b) the motive for those comparisons.

An alternative approach would be to question the standard of comparison itself. If thin-ideal models are not seen as highly desirable, then comparisons with these models should no longer undermine women’s body satisfaction. Changing the desirability of thin-ideal models is not likely to be easy, however. Part of the stereotype of thin-ideal models is that they are not only attractive but also happy and successful in life (Tiggeman, 2002). Merely challenging these positive stereotypical beliefs might not be very persuasive. This is because dual-process models of social comparison (e.g., Chaiken & Trope, 1999) point to a distinction between reflective (rational) and automatic (impulsive) components of social stereotypes. Targeting only the reflective component—for instance, by persuading people that thin-ideal models possess negative characteristics—may alter conscious evaluations of thin-ideal models. However, such interventions are unlikely to be effective in changing people’s underlying implicit associations between thin-ideal models and positive attributes. To tackle these associations requires an intervention that directly targets the automatic component, and alters the mental links that people have developed through experience.

The intervention developed and tested in the present research focused on modifying implicit associations concerning thin-ideal models. The intervention drew upon research on evaluative conditioning. In typical evaluative conditioning studies, affect is transferred to initially neutral stimuli (e.g., Levey & Martin, 1975; Staats & Staats, 1958; see De Houwer, Thomas, & Baeyens, 2001, for a review). For instance, geometric shapes may be repeatedly paired with positive stimuli such that evaluations of the shapes become more favorable (Bierley, McSweeney, & Vannieuwkerk, 1985). However, because thin-ideal models are associated with positive, rather than neutral, characteristics to begin with (e.g.,

**Study 1: Intervention Effects on Implicit Idealization of Thinness**

**Method**

A power analysis (conducted via G*Power) that assumed an effect size of $d = .73$ for the evaluative conditioning intervention (based on Martijn, Vanderlinden, Roefs, Huijding, & Jansen, 2010), indicated that a total of 62 participants (31 per condition) were required to have 80% power to detect a significant effect at $p < .05$ (2-tailed). In fact, 66 undergraduate women at the University of Sheffield took part in return for experimental credits ($M_{age} = 18.94, SD = 0.99$). Participants were randomly assigned to the intervention or control condition.

**Procedure**

**Evaluative conditioning task.** Participants took part in a study of “social attitudes” and were tested individually. First, participants undertook the intervention or control task, both of which involved completing a modified Implicit Association Test (IAT; Greenwald et al., 1998). In the intervention condition, participants learned to associate supermodels with “fake” and Dove models (i.e., normal-sized models that feature in advertisements for a popular cosmetics brand) with “real.” The categories Dove models or Real and Supermodel or Fake appeared at the top left and right hand of a computer screen, respectively. Participants had
to sort target words and images of female models that appeared in the center of the screen into the relevant categories. Targets were randomly selected from five pictures of Dove models, five pictures of supermodels, five synonyms of real (natural, true, genuine, authentic, sincere), and five synonyms of fake (artificial, phony, bogus, sham, false). Thus, on every trial, intervention participants responded by pairing supermodels with fake and Dove models with real. In the control condition, participants sorted the same target pictures into the categories Dove models or Verb and Supermodel or Noun. The five verbs used in this condition were go, make, perform, wash, and drive, and the five nouns were table, car, house, chair, and book. In this way, control participants did not learn the supermodel–fake and Dove model–real associations learned by participants in the intervention condition. Participants undertook 150 trials in both conditions.

Implicit idealization measure. Next participants undertook a task that measured their implicit idealization of thin models. Participants completed a standard 7-block IAT (see Greenwald, Nosek, & Banaji, 2003, Table 1). Category labels appeared at the top left side and top right-hand side of the computer screen. Participants’ task was to classify target words and pictures that appeared in the center of the screen into the relevant categories using the e and i keys, respectively.

The categories were Dove models and Supermodels and My ambition and My lowest standard. Two sets of five pictures (different to those used in the previous task) were used to represent the Dove model and supermodel categories. Five phrases represented My ambition (my ideal, my goal, my target, my aspiration, my dream) and My lowest standard (my nightmare, the pits, rock bottom, my fear, my hell). Blocks 1, 2, and 5 were practice blocks (20 trials) and involved classifying relevant targets into My ambition-My lowest standard, Supermodel-Dove model, and Dove model-Supermodel categories, respectively. The critical blocks were (a) blocks 3 and 4 where the categories were Supermodel + My ambition-Dove model + My lowest standard (60 trials in total) and (a) blocks 6 and 7 where the categories were Dove model + My ambition-Supermodel + My lowest standard (60 trials in total). Words and pictures from all four categories appeared in the center of the screen during these blocks. The difference in response times to blocks 3 and 4 versus blocks 6 and 7 was used to compute the implicit idealization measure. In this way, we attempted to index how enviable were supermodels as standards for appearance-related social comparisons.

Results and Discussion

A funnel debriefing procedure designed to assess participants’ awareness of the experimental hypotheses (Chartrand & Bargh, 1999) indicated that participants believed both tasks measured their social attitudes. No participant believed that the initial (intervention or control) task could have influenced his or her performance on the second (implicit idealization) task. Thus, any effects of the intervention cannot be attributed to participants’ expectations or their desire to please the experimenter.

The RTs from blocks 3, 4, 6, and 7 were treated in accordance with the improved scoring algorithm (Greenwald et al., 2003) to compute scores on the second IAT implicit idealization task. Higher scores indicate greater idealization of thin models. As predicted, participants in the intervention condition—who learned to associate supermodels with fake and Dove models with real—showed less idealization of thin models compared to control participants (Ms = −.03 and .30, SDs = .43 and .38, respectively), F(1, 64) = 10.84, p = .002, ηp² = .15. Thus, the intervention was successful in rendering thin-ideal models less desirable as comparison standards for appearance.

Study 2: Intervention Effects on Body Satisfaction

Study 2 tested whether the intervention effects observed in Study 1 extend to influencing women’s body satisfaction, as measured by a validated, self-report scale (the Body Image States Scale; Cash, Fleming, Alindogan, Steadman, & Whitehead, 2002). Pilot research indicated that the intervention used in Study 1 should be modified to better conceal the purpose of the study as the dependent variable in Study 2 was a standard, self-report measure (and so responses could more easily be faked if participants guess the true purpose of the study). This was not an issue in Study 1 because the dependent variable was an implicit measure that did not rely on introspection and has been found low in susceptibility to self-presentation distortion (Kim, 2003). The new intervention and control tasks were conceptually equivalent to those used in Study 1 and used similar images and words. However, the tasks were disguised as a study of consumer evaluations of “summer fashion.”

Method

The research question was tested in a 2-within (Time: pretest vs. posttest) by 2-between (Conditioning: experimental vs. control) mixed factorial design with body dissatisfaction as the dependent variable. Participants were 39 female students at Maastricht University (Mage = 21.51; SD = 1.83). Participants were randomly assigned to the intervention or control condition.

Procedure

Participants were tested individually and told that they would take part in two unrelated studies. The first study supposedly concerned the development of a new questionnaire measure of “Students’ Life Satisfaction.” Participants were told that the reliability of the measure needed to be established and so they would have to complete the questionnaire twice—before and after an unrelated consumer behavior study on “summer fashion”. Inter-spersed among the questionnaire items was a standard measure of body satisfaction—the 6-item Body Image States Scale (BISS; Cash et al., 2002). The original 9-point response scales were replaced with 100 mm visual analogue scales so that responses to the BISS were measured in same way as other items in the Students’ Life Satisfaction questionnaire. Examples of BISS items are “Right now I feel extremely dissatisfied (0)—extremely satisfied (100) with my physical appearance” and “Right now I feel that I look a great deal better than the average person looks (100)—a great deal worse than the average person looks (0). One item correlated weakly with the other items and was discarded (item-total rs at pretest and posttest < .15). Reliability of the 5-item BISS was high (α = .87) and the mean score of the five items was computed to create a pretest body satisfaction measure (score vary from 0 to 100; higher scores indicate greater body satisfaction).
Next, participants took part in the “summer fashion” study. In actual fact, this was the intervention or control task. Participants were seated at a computer and informed that a picture of swimwear (a bikini or bathing suit) would appear in one of four quadrants of the computer screen; some pictures would show a woman wearing the swimwear whereas other pictures showed only the swimwear. Participants were told that the experimenter was interested in whether evaluations of swimwear differed depending on whether the swimwear was worn by a person or not. Participants’ task was to click on the picture as fast as possible (participants were led to believe that clicking speed indicated the amount of liking). The pictures of curvy and thin models were pretested and differed on the dimensions curvy-thin and natural-fake but did not differ on attractiveness.) After participants clicked on the picture, a word was presented briefly (400 ms) in that same quadrant before the next trial began. The type of word presented varied by condition.

In the intervention condition, the six pictures of curvy models were always followed by one of six “natural” words (pure, real, honest, fresh, sincere, natural), the six pictures of the thin models were always followed by one of six “fake” words (fake, phony, artificial, impure, faulty, unnatural), and the six pictures of swimwear were followed by one of six words referring to types of fabric (cotton, linen, silk, polyester, nylon, viscose). Each of the 18 pictures was followed by three presentations of the six words associated with the relevant picture category for a total of 324 trials; the order of presentation varied randomly per participant. In the control condition, the same pictures of curvy models, thin models, and swimwear were randomly followed by the natural, fake, or fabric words; again, there were 324 trials.

After finishing the intervention or control task, participants again completed the Students’ Life Satisfaction questionnaire.

Table 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Preintervention M (SD)</th>
<th>Postintervention M (SD)</th>
<th>Difference M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental (n = 17)</td>
<td>55.98 (14.70)</td>
<td>62.53 (13.13)</td>
<td>6.55 (7.74)</td>
</tr>
<tr>
<td>Control (n = 19)</td>
<td>56.16 (14.16)</td>
<td>56.75 (18.15)</td>
<td>.59 (8.99)</td>
</tr>
<tr>
<td>Total (N = 36)</td>
<td>56.07 (14.21)</td>
<td>59.48 (16.02)</td>
<td>3.41 (8.84)</td>
</tr>
</tbody>
</table>

Note. Body Satisfaction Scale ranges from 0–100 (BISS, adapted from Cash et al., 2002); higher scores indicate higher body satisfaction. Difference = scores at posttest minus scores at pretest; a positive score indicates a positive change in body satisfaction.

Awareness Check

As a check for awareness of the experimental hypotheses, participants were asked to describe the purpose of the study in their own words. Three participants in the intervention condition realized that the study concerned body satisfaction and were excluded from subsequent analyses. None of the remaining 36 participants guessed the true purpose of the study. We also checked for contingency awareness, that is, whether participants realized that specific words followed specific pictures in the intervention conditions. None of the 36 participants correctly identified that natural words always followed pictures of curvy models or that fake words always followed pictures of thin models. Thus, participants showed no awareness of the associations that we intended them to learn (see meta-analysis by De Houwer et al., 2001, for discussion of contingency awareness).

Results

We expected that body satisfaction would increase in the intervention condition (where curvy models were associated with natural words and thin models were associated with fake words) but would remain unchanged in the control condition (see Table 1 for mean body satisfaction scores at pretest and posttest for the intervention and control conditions). We first established that body satisfaction did not differ between conditions before the intervention or control task; this proved to be the case, $F(1, 34) < 1$, ns. Next, body satisfaction scores were submitted to a 2-within (Time: preintervention vs. postintervention) by 2-between (Condition: intervention vs. control) repeated measures ANOVA (see Table 1). There was no main effect of Condition, $F(1, 34) < 1$, ns. There was, however, a main effect of Time, $F(1, 34) = 6.44, p = .02, \eta^2_p = .15$ (equivalent to Cohen’s $d = .82$). Using this effect size as the basis for computing the required sample size in Study 2, analyses conducted via G*Power indicated that 18 participants ($n = 9$ per condition) were needed to observe a significant effect at 90% power ($p < .05, 2$-tailed). Moreover, a power analysis using the effect size that we observed for the significant interaction term in Study 2 ($\eta^2_p = .12$ or $d = .73$) indicates that we had a 99% chance of detecting a significant effect at $p < .05$ (2-tailed). In sum, concerns about statistical power in Study 2 would seem to be unfounded.

General Discussion

The present research started from the position that targeting social comparison processes might be an effective strategy for enhancing body satisfaction. However, whereas previous research focused on the agents of social comparison (i.e., the motives and characteristics of the women making the comparisons, and the inappropriate comparisons they make; Halliwell & Dittmar, 2005; Strahan et al., 2008), the present focused on the targets of social comparison (i.e., thin-ideal models). We developed an intervention...
that integrated ideas from evaluative conditioning and stereotype change in an effort to render thin-ideal models less enviable as standards for appearance-based social comparisons. Two studies tested whether the intervention could (a) reduce implicit idealization of thin models, and (b) enhance women’s body satisfaction. Study 1 showed that intervention participants who learned to associate supermodels with synonyms of fake were less likely to view supermodels as targets to be emulated. That is, intervention participants exhibited weaker implicit associations between supermodels and concepts such as my goal or my ambition compared to controls. Study 2 deployed an equivalent intervention and observed a significant increase in body satisfaction using a validated, self-report scale. Intervention participants showed improved body satisfaction after having learned to associate thin-ideal models with fake; no such improvement was observed for control participants.

The present study is not the first to use evaluative conditioning to try to alter women’s body satisfaction. Martijn and colleagues (2010) repeatedly paired pictures of participants’ own bodies with positive social stimuli (i.e., smiling faces) and found significant increases in body satisfaction. This finding echoes previous research demonstrating that evaluative conditioning can improve global self-esteem (Baccus, Baldwin, & Packer, 2004; Dijkstra-huis, 2004). The present study is the first (to our knowledge), however, that intervened on the targets of comparison rather than the comparisons that women make.

As with any new program of research, there are limitations in the present studies that should be acknowledged. First, relatively small samples of undergraduates took part in the research. Power analyses indicated that the studies had adequate sample sizes for testing our hypotheses, however. Moreover, the use of college student participants seemed appropriate as we wished to provide “proof-of-concept” for the present intervention. Future studies should use larger and more representative samples and also test effects among subclinical or clinical populations (i.e., women at a high risk for developing an eating disorder or women diagnosed with a eating disorder). Second, the evaluative conditioning intervention used in the present research involved both learning to associate thin-ideal models with “fake” and normal-sized models with “real”. Thus, the present research does not indicate whether the present intervention indicates that an intervention that targets the implicit idealization of thin models as a component of the automatic system is effective in increasing body satisfaction. However, there may be elements of traditional interventions that target components of the reflective system (e.g., Halliwell & Dittmar, 2005; Posavac et al., 2001) that could profitably be combined with the intervention used here. It may well be the case that interventions are most effective in promoting body satisfaction if they target both the reflective and automatic systems. Research is needed to test this idea.

**References**


Received August 23, 2011
Revision received February 16, 2012
Accepted February 21, 2012