A bifactor analysis of the Weight Bias Internalization Scale

Meadows, Angela; Higgs, Suzanne

DOI:
10.1016/j.bodyim.2020.02.013

This is an Accepted Manuscript of an article published by Elsevier in Body Image.

Version of record: Vol 33, pp 137–151
https://doi.org/10.1016/j.bodyim.2020.02.013

Title: A bifactor analysis of the Weight Bias Internalization Scale: What are we really measuring?

Authors: Angela Meadows,[a1] Suzanne Higgs[b]

[a]School of Psychology, Western University, London, Ontario N6A 5C2, Canada

[b]School of Psychology, University of Birmingham, Birmingham, B15 2TT, UK

AM: drameadows@gmail.com

SH: s.higgs.1@bham.ac.uk

Corresponding author: Angela Meadows

[1] Work conducted while A Meadows was at University of Birmingham, UK

Abstract

Internalized weight stigma (IWS) has been linked with disordered eating behavior, both directly, and as a mediator of the relationship between experienced weight stigma and maladaptive coping. However, the construct of IWS is highly correlated with the related constructs of body image and global self-esteem, and the three constructs may better be represented by underlying trait self-judgment. This overlap is not generally accounted for in existing studies. The present study investigated the shared variance between self-esteem, body image, and IWS in an international sample of higher-weight individuals. Bifactor analysis confirmed that the intermediary role of IWS in the relationship between experienced stigma and self-reported eating behavior was largely accounted for by aspects of body image and global self-esteem. Greater conceptual clarity in the study of IWS is needed to understand the mechanisms via which societal weight stigma impacts on individuals' self-directed judgments and downstream health-related behaviors.

Keywords:

Internalized weight stigma; Body image; Self-esteem; Self-judgment; Disordered eating; Bifactor analysis

1. Introduction

Higher-weight individuals face prejudice and discrimination in employment, education, healthcare settings, and a wide range of everyday interpersonal situations (Puhl & Heuer, 2009). In addition, some individuals internalize societal anti-fat attitudes and stereotypes – that is, they devalue *themselves* because of their weight, a phenomenon known as internalized weight stigma (IWS; Durso & Latner, 2008). High levels of IWS have been linked with poorer health and wellbeing (Hilbert, Braehler, Haeuser, & Zenger, 2014; Latner, Barile, Durso, & O'Brien, 2014), and with a range of maladaptive coping strategies, including disordered eating behavior (Durso, Latner, & Hayashi, 2012; Schvey, Roberto, & White, 2013), avoidance of exercise (Mensinger & Meadows, 2017; Pearl, Puhl, & Dovidio, 2015) and experiential avoidance (Lillis, Luoma, Levin, & Hayes, 2010; Palmeira, Pinto-Gouveia, & Cunha, 2017). Internalized weight stigma also appears to be an important mediator of the effects of experienced weight stigma on downstream health and behavioral outcomes (Durso, Latner, & Hayashi, 2012; O'Hara, Tahboub-Schulte, & Thomas, 2016; Pearl et al., 2015).

In their development of the Weight Bias Internalization Scale (WBIS), the first validated measure of IWS, Durso and Latner (2008) operationalized IWS as requiring both endorsement of negative social stereotypes about higher-weight individuals and ascribing those negative stereotypes to the self, with resultant deleterious impact on self-worth. They distinguished the construct from both global self-esteem, due to the specificity of IWS to the domain of weight and shape, and also from body image, a construct that is limited to feelings about one's body, but that is not directly a measure of one's perceived social value. Nevertheless, while an individual with low self-esteem or poor body image may not exhibit high levels of IWS, the reverse is unlikely to be true: self-devaluation due to weight is likely to be reflected in scores on measures of self-esteem and body image. Indeed, initial construct validity of the WBIS was partly demonstrated by correlation with measures of global self-esteem (r = -.68) and weight-related body image concerns and emotions (r = .74), and these relationships were not weakened after controlling for BMI (Durso & Latner, 2008). Other studies in both clinical and community samples have confirmed the strong association between WBIS scores, global self-esteem, and a range of body image measures, including body satisfaction (Burmeister, Hinman, Koball, Hoffmann, & Carels, 2013), body appreciation (Carels et al., 2019), body image flexibility (Webb & Hardin, 2016), appearance orientation (Hübner et al., 2016), body dissatisfaction (Pearl & Puhl, 2016), body-related shame (Burmeister et al., 2013; Mehak, Friedman, & Cassin, 2018; Webb & Hardin, 2016), body surveillance (Mehak et al., 2018), and appearance anxiety (Mehak et al., 2018). As IWS, body image, and global self-esteem are all self-directed judgments, it is possible that part of this commonality is due to an underlying self-judgment factor. The impact of these constructs on health behavior may well be due to the shared variance between them; that is, an individual who tends to judge themselves negatively across numerous domains may engage in fewer health behaviors or more unhealthy behaviors than one who tends to judge themselves more positively.

While the authors of the WBIS noted that, despite the overlap, these constructs may have different real-world implications, these very high correlations are a cause for concern, and raise the question of how much additional variation in individual behaviors or health outcomes are explained by IWS, above and beyond that attributable to the lower levels of global self-esteem and body image manifested by individuals who are high in IWS. Indeed, both self-esteem (Mann, Hosman, Schaalma, & de Vries, 2004) and body image (Mond et al., 2013; Stice & Shaw, 2002) are important predictors of mental and physical health and wellbeing, and a range of health behaviors, including disordered eating, exercise, and substance abuse. To date, only one study of IWS and eating behavior has attempted to account for these potential overlapping constructs. Controlling for depressive symptoms, endorsement of anti-fat stereotypes, and global self-esteem, the WBIS explained an additional 9% of variance in total scores on the Eating Disorder Examination Questionnaire in 100 treatment-seeking ‘obese’ patients with binge eating disorder (Durso, Latner, White, et al., 2012); however, the majority of the difference was accounted for by the weight- and shape-concern subscales, that is, appearance-related constructs. Thus, it is unclear whether or to what extent IWS, at least as measured by the WBIS, is a useful construct in explaining maladaptive eating behaviors, beyond its underlying associations with body image and self-esteem.

As the literature in this area proliferates, there is a danger of succumbing to a “jangle fallacy” – a situation where different labels are applied to essentially the same construct, which is subsequently treated as two distinct phenomena (Kelley, 1927). Rather than facilitating research, a jangle fallacy may result in over-complication and hinder progress. Further, use of measurement-level sum-scores in multiple regression analyses to demonstrate incremental validity, is not alone sufficient to confirm distinctiveness, as such analytical techniques fail to account for measurement error and are prone to unacceptably high Type 1 error rates (Westfall & Yarkoni, 2016).

The purpose of the present study was to investigate the shared variance between self-esteem, body image, and IWS, as measured by the WBIS, to determine whether the three constructs are better represented by a common negative self-judgment factor than as three unique dimensions, and to determine the extent of any additional unique variance contributed by the individual constructs. To this end, bifactor analysis was used to model both the specific variance contributed by self-esteem, body image, and IWS, and that contributed by a more global construct of negative self-judgment in an international sample of higher-weight individuals. The resulting construct-specific and general factors were then tested as mediators in the relationship between experienced weight stigma and disordered eating behavior. Bifactor analysis is a form of confirmatory factory analysis, described in more detail below, and, as such, accounts for measurement error in the model. It will thus be possible to ascertain whether IWS, as measured by the WBIS, continues to predict disordered eating behavior while controlling for shared variance with self-esteem and body image. Given the very high correlations between these measures, we hypothesised that a common underlying self-judgment factor would emerge that would explain a significant proportion of variance in disordered eating behavior. We made no *a priori* hypotheses regarding the extent of residual variance remaining for the domain-specific factors nor the strength of their relationship with eating behavior. However, if indeed the unique contribution of the WBIS no longer significantly predicts disordered eating behavior in a bifactor model, this would suggest either that IWS is not a useful phenomenon to study in its own right, at least within the realm of problematic eating behavior, or, more likely, that its operationalisation in the WBIS does not adequately capture the components that distinguish IWS from already established predictors of disordered eating. Either of these outcomes would have implications for the study of IWS and the design of interventions targeting it as a means of reducing harmful downstream outcomes.

2. Method

2.1. Participants and procedure

Participants were 384 adults who self-identified as 'overweight' or 'fat.' Purposive sampling was used to obtain a sample likely to have a range of views on the acceptability of societal weight stigma and both positive and negative emotions about their own body weight. Invitations to participate in the survey “Life experiences of overweight individuals” were posted on social media and Internet forums related to weight, weight-loss, health, nutrition, fitness, plus-size fashion, and the size acceptance movement. Participants completed the survey anonymously via a dedicated survey platform (Qualtrics.com). After providing consent, a screening question asked participants for their height and weight, and BMI was automatically calculated. Individuals with a self-reported BMI below 25 kg/m^2 (n = 5) were excluded from the study and thanked for their time. While self-identification of high-weight status is often either an equally or more consistent predictor of cognitive, affective, and behavioral correlates than is objective BMI (M. S. Lee & Dedrick, 2016; Lin, Latner, Fung, & Lin, 2018; Major, Hunger, Bunyan, & Miller, 2014), this two-step inclusion criteria, involving both self-classification as ‘overweight’ and a BMI in the ‘overweight’ category, has been used previously as a more conservative sample selection procedure (Durso, Latner, & Hayashi, 2012; Hunger, Blodorn, Miller, & Major, 2018; Pearl & Puhl, 2016). Thus, the final sample size was 379. All participants were entered into a prize draw to win a £50 Amazon voucher (or local equivalent). The study was approved by the University of Birmingham Ethical Review Committee.

2.2. Measures

2.2.1. Experienced weight stigma

Experiences of weight stigma was assessed using the 50-item Stigmatizing Situations Inventory (SSI; Myers & Rosen, 1999). The SSI has excellent internal reliability in clinical and non-clinical samples, and in US and international populations; it is positively associated with psychological distress, body dissatisfaction, and disordered eating, and has good discriminant validity across weight and eating pathology categories (Brauhardt, Rudolph, &

Hilbert, 2014; Myers & Rosen, 1999; Vartanian, 2015). Participants rate the frequency with which they have experienced stigmatising events across 11 domains, such as being the target of nasty comments from various sources, being stared at, avoided, or excluded, and outright discrimination. Frequency of experiences was rated on a four-point scale: 0 (*never*), 1 (*once in your life*), 2 (*more than once in your life*), and 3 (*multiple times*). This scoring method has previously been shown to be easier for participants to use than the original 10-point frequency scale and to have high internal validity (α = .96) .

2.2.2. Internalized weight stigma

Internalized weight stigma was assessed with the 11-item Weight Bias Internalization Scale . The WBIS has excellent internal reliability and good convergent and predictive validity in clinical and non-clinical samples of men and women (Durso & Latner, 2008; M. S. Lee & Dedrick, 2016), and has been validated in several European populations (Gomez & Baile, 2015; Hilbert, Baldofski, et al., 2014; Hübner et al., 2015; Innamorati et al., 2017).[1] Items are scored on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Higher scores indicate a greater degree of IWS.

2.2.3. Explicit anti-fat attitudes

Explicit anti-fat attitudes, that is, negative attitudes toward fat others, were measured using the 7-item Dislike subscale of the Anti-Fat Attitudes Questionnaire (AFAQ; Crandall, 1994). The Dislike subscale has good internal reliability in international samples (Crandall, D'Annello, Sakalli, Lazarus, Nejtardt, & Feather, 2001) and small to moderate associations with other measures of prejudicial attitudes and with conservative ideological beliefs (Crandall, 1994; Crandall et al., 2001; Magallares, 2014). Unlike other measures of negative

[1] Small differences in item-total correlations were observed in the Italian sample and a two-factor structure emerged in the Spanish version – however, psychometric properties of the composite scale were excellent.

attitudes among members of stigmatised groups, the AFAQ indicates little in-group bias among higher-weight individuals, who score similarly on the Dislike subscale as normative-weight individuals (Crandall, 1994). Anti-fat attitudes have been shown to correlate with disordered eating behavior in both clinical and non-clinical samples (Barnes, Ivezaj, & Grilo, 2014; Pepper & Ruiz, 2007). Items are scored on a 10-point Likert scale from 0 (*very strongly disagree*) to 9 (*very strongly agree*). Higher scores indicate stronger anti-fat attitudes.

2.2.4. Self-esteem

Self-esteem was measured using the 10-item Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965). The RSE is the most widely used measure of global self-esteem and has demonstrated good internal and test-retest reliability and convergent, discriminant, and predictive validity in multiple international samples (Donnellan, Trzesniewski, & Robins, 2015). Items are scored on a 4-point Likert scale ranging from 0 (*strongly disagree*) to 3 (*strongly agree*). The maximum possible score is 30, and higher scores are indicative of higher self-esteem. The RSE correlates negatively with measures of experienced and internalized weight stigma and disordered eating cognitions and behaviors (Durso & Latner, 2008; Friedman et al., 2005; Griffiths et al., 1999).

2.2.5. Body image

Body image is a multifactorial construct encompassing distinct positive and negative perceptions and attitudes towards one's body (Thompson, 2004). In the present study, we assessed body image with an evaluative measure of global appearance satisfaction, namely the 7-item Appearance Evaluation subscale of the Multidimensional Body Self-Relations Questionnaire (MBSRQ; Brown, Cash, & Mikulka, 1990; Cash, 2000). This subscale measures feelings of body satisfaction and physical attractiveness and was selected as it does not presuppose pathology or introduce weight-specific terminology, which might produce confounding effects. It is scored on a 5-point Likert scale ranging from 1 (*definitely disagree*) to 5 (*definitely agree*), with higher scores indicating greater body satisfaction. The scale has good internal reliability in both male and female sample, good (males) to excellent (females) test-retest reliability, and has demonstrated convergent, discriminant, and construct validities in demographically and culturally diverse clinical and non-clinical samples (Brown et al., 1990; Cash, 2000; Mautner, Owen, & Furnham, 2000).

2.2.6. Eating behavior

Two measures were used to assess eating habits. The Dutch Eating Behavior Questionnaire (DEBQ; van Strien, Frijters, Bergers, & Defares, 1986) was used to evaluate habitual disordered eating patterns. The DEBQ comprises three subscales, which measure dietary restraint, emotional eating, and external eating – eating in response to external cues rather than bodily hunger signals. The dimensional structure of the DEBQ is measurement invariant by age, gender, and BMI status and has been confirmed in numerous international and diverse cultural samples, although women and higher-weight individuals tend to score higher on the subscales (Wardle, 1987). Psychometric properties are also excellent across cultural samples (Brunault et al., 2015; Dakanalis et al., 2013; van Strien et al., 1986; Wang, Ha, Zauszniewski, & Ross, 2018; Wardle, 1987). Items are scored on a 5-point Likert scale measuring frequency of the different styles of eating behaviors, ranging from 0 (never) to 5 (very often). The individual subscales are scored separately. Higher scores indicate more frequent disordered eating.

Additionally, current dieting behavior was assessed with a single item. Participants indicated whether they were currently dieting for weight loss, watching their food intake so as to maintain their current weight and prevent weight gain, or not dieting.

Cognitions and behaviors consistent with more severe eating pathology were assessed using the Eating Disorder Diagnostic Scale (EDDS; Stice, Telch, & Rizvi, 2000). Items are summed to produce a composite symptom count that can be used as a measure of overall eating pathology, with higher scores indicating more problematic cognitions and behaviors (Stice, Fisher, & Martinez, 2004). In addition to providing a total symptom score, the EDDS can be used to provide diagnostic indications of the presence of eating disorders. Presence of Binge Eating Disorder (BED) or Bulimia Nervosa (BN) was evaluated according to the criteria stipulated in the *Diagnostic and Statistical Manual of Mental Disorders-5* (DSM-5; American Psychiatric Association, 2013). A subset of the items also captures the frequency of binge eating episodes in the previous 3- (BE3) and 6-month (BE6) periods. The EDDS has good internal consistency in both clinical and non-clinical female samples, high test-retest reliability, excellent concordance with interview diagnoses of disordered eating, and good convergent validity with self-report measures of disordered eating behavior and general psychopathology (Stice et al., 2004, 2000). While not formally validated in adult males, the EDDS also had strong internal reliability in a sample of male U. S. veterans, and scores were uniquely predicted by military trauma, controlling for other potential confounds (Arditte Hall, Bartlett, Iverson, & Mitchell, 2017). The EDDS has been validated in international samples and diverse cultures, including male and female Hong Kong adolescents (S. W. Lee et al., 2007) and in clinical and non-clinical Dutch adult women (Krabbenborg et al., 2012) and non-clinical samples of adult and adolescent Chilean females (Silva et al., 2012). Questions relating to height and weight were omitted from the original 22-item scale, as this information was collected elsewhere. Thus, the final questionnaire included 20 questions.

2.2.7. Anthropometrics and demographics

Self-report height and weight were collected during the eligibility screening questions for the study, and used to calculate BMI (weight in kg/height in meters squared). Participants were asked to provide age, gender, ethnicity, level of education, and current profession. Participants had the option to decline to answer any of these questions.

2.3. Handling of missing data

Missing values analyses of questionnaire items indicated only a small number of missing responses (0–1.1%). Little's MCAR test was used to assess the pattern of missingness. A non-significant *p* value on this test indicates that data are missing completely at random (MCAR). In this case, Little's MCAR test $\chi^2(5{,}675) = 5{,}825$, $p = .081$, indicated no pattern of data missingness. Given the very small number[2] and randomness of missing data, missing items were deleted pairwise. One participant had missing data on eight of the ten items on the DEBQ External Eating subscale, and their data were excluded for that measure.

Twenty participants (5.3%) did not provide their weight and/or height, thus making it impossible to calculate their BMI. Missing values analysis suggested that cases with missing BMI data just failed to meet the non-significance criteria for being MCAR (Little's MCAR test $\chi^2(31) = 45.2$, $p = .048$). Independent *t*-tests were conducted to identify differences between participants with BMI data available versus missing. Participants without BMI data were older, had lower self-esteem, worse body image, and had experienced more weight stigma from others, although effect sizes were small (all Cohen's *d*s ≤ 0.27). Based on these characteristics, it seems likely that individuals with missing BMI might tend to be heavier,

[2] Only two items (WBIS Q9 and RSE Q6) had more than one data point missing, both $n = 2$ missing.

and were either unwilling to convey this information, or chose not to weigh themselves and were unable to provide it. As BMI is strongly correlated with experienced weight stigma, and as the structural equation model used in this study does not allow missing values in covariates, non-random missing BMI data could lead to misrepresentation of the relationships in the model. Thus, a decision was made to impute missing BMI values. Linear regression analysis was conducted to see whether BMI could be predicted using the remaining independent variables. The model was significant, $F(12,338) = 18.7$, $p < .001$), and explained 39.9% of the variance in BMI. Thus, missing BMI values were imputed using the expectation maximization estimation. The EM method is an iterative procedure that estimates the means, covariance matrix, and correlation of scale variables with missing values based on the likelihood under the distribution of the variable. Each iteration is conducted in two steps: first, an E step uses log-likelihood to produce a conditional expectation of the missing data given the observed values and current estimate of the parameters, e.g., correlations; the M step then performs full information maximum likelihood estimation as though the missing data had been filled in, to compute parameters that maximize the expected log-likelihood from the E step. These parameter estimates are used in the subsequent E step, and the process repeats until convergence is achieved. Missing data on demographic variables (age 0.5%, gender 0.8%, race/ethnicity 19.5%) were not imputed.

3. Data analysis and results

All confirmatory factor analyses (CFA) and mediation analyses were conducted in Mplus version 8 (Muthén & Muthén, 2017). Other analyses were conducted using SPSS for Mac (IBM Software Group, Chicago, IL), version 23.

3.1 Sample characteristics

The sample was predominantly female (88%), and, although drawn from a reasonable wide geographic base (16 countries)[3], predominantly White. (71%; 19.5% did not provide an ethnicity and no other ethnic group made up more than 4% of the sample), with a mean age of 37.6 years (*SD* = 12.1; range 18–69). Participants were generally well educated and in white-collar professions. A good range of body sizes were represented in the sample. Average BMI was 36.8 kg/m^2 (*SD* = 8.9; range 25.0–76.2). Additionally, 32.2% of participants were weight-loss dieting, 27.7% were watching what they ate so as not to gain weight, and 39.8% were not dieting. Overall, 8.2% of the sample met the DSM-5 diagnostic criteria for BED and 7.9% met the criteria for BN. Neither BED nor BN correlated with either experienced (BED point biserial correlation r_{pb} = .09, p = .084; BN r_{pb} = -.01, p = .920) or internalised weight stigma (BED r_{pb} = -.06, p = .740; BN r_{pb} = .06, p = .560). As such, neither BED nor BN were included in subsequent analyses. Means, standard deviations, internal reliability statistics, and correlations for study variables are displayed in Table 1.

3.2. Experienced, internalized, and explicit weight stigma

The SSI, AFAQ, and BMI displayed mild levels of skewness. Log-transformations were performed but did not alter the results of subsequent analyses; thus, raw scores are presented throughout.

Experiences of weight stigma were ubiquitous, being reported by almost every participant. Levels of experienced stigma were consistent with those reported in other community samples (Puhl & Brownell, 2006; Vartanian & Novak, 2011), with frequent experiences of stigma recorded in most domains (see Supplementary Materials). There was a

[3] UK 45.4%, USA 34.0%, Oceania 7.1%, Canada 5.8%, Continental Europe 5.5%, Other 2.2%.

small but significant positive relationship between experienced and internalized weight stigma (Spearman's $\rho = .20$, $p < .001$). In addition, with the exception of physical assault, all domains of the Stigmatizing Situations Inventory were significantly correlated with higher levels of internalised weight stigma and lower self-esteem. However, after controlling for BMI, only nasty comments (from all sources), embarrassment of loved ones, negative assumptions, being excluded or ignored, and job discrimination remained significantly associated with internalisation, and only interpersonal sources of stigma remained significantly associated with self-esteem. Participants' own anti-fat attitudes were similar to those reported in other higher-weight samples (Burmeister et al., 2013; Durso, Latner, & Ciao, 2016) and had small to moderate correlations with internalised weight stigma, body satisfaction, and global self-esteem, in the expected directions.

Differences by demographic characteristics were explored using independent *t*-tests, one-way analysis of variance with robust test of equality of means, and χ^2 tests. Levels of experienced and internalised weight stigma did not differ by age, profession, or education, although there was a tendency for individuals with higher degrees to report less internalised weight stigma. No statistically significant gender differences emerged for the majority of study variables; however, consistent with the extant literature (Dutton et al., 2014; Hatzenbuehler, Keyes, & Hasin, 2009), women experienced around twice as much stigma as men overall, and women reported higher scores on all subscales of the Stigmatizing Situations Inventory with the exception of being physically attacked. These gender differences remained significant after controlling for BMI. Thus, subsequent regression analyses were controlled for gender.

3.3. Confirmatory factor analysis of WBIS, RSE, and Appearance Evaluation scales

Some previous studies using the WBIS have found that the first item on the questionnaire, “As an overweight person, I feel that I am just as competent as anyone,” loaded poorly onto the one-factor structure and had very low (Durso et al., 2016), or even negative (Hilbert, Baldofski, et al., 2014; M. S. Lee & Dedrick, 2016), item-total score correlation; however, other analyses have not found this item to be problematic (e.g., Durso & Latner, 2008; Gomez & Baile, 2015). Thus, prior to conducting the bifactor analysis, separate CFA models were tested for each individual scale to ensure that the items adequately represented the constructs of interest. All three models were an acceptable fit to the data (Table 2). For the WBIS, the first item had a lower factor loading (standardized $\lambda = .482$) compared with the other 10 items in the scale (standardized $\lambda = .675 – .906$, median .720). However, this loading was above the generally accepted cut-off of .3 for meaningful factor loading in a sample of this size (Stevens, 2002; cited in Field, 2013), and was statistically significant ($p < .001$), indicating that the item adequately captured the target construct in the present sample. Re-running the model with item 1 excluded resulted in only a small improvement in model fit (CFI = .93, SRMR = .04).[4] Thus, all eleven items were included in subsequent analyses. Factor loadings on the Appearance Evaluation scale ranged from .565 to .842, median = .755. Factor loadings on the RSE ranged from .607 to .835, median = .694.

Table 2. *Model fit indices for individual scale confirmatory factor analysis*

Measure	χ^2	*df*	CFI	SRMR
Weight Bias Internalization Scale	272	44	.92	.05
Appearance Evaluation	143	14	.91	.05
Rosenberg Self-Esteem scale	246	35	.90	.05

Note. All χ^2 $p < .001$. CFI = Comparative Fit Index; *df* = degrees of freedom; SRMR = standardized root mean squared residual.

3.4. Bifactor analysis

[4] Analysis conducted in SPSS indicated the item-total correlation for item 1 was .48.

Bifactor analysis is a form of CFA that allows for the parsing of variance into unique and joint components (Brunner, Nagy, & Wilhelm, 2012a; Chen, West, & Sousa, 2006; Rindskopf & Rose, 1988a). It takes its name from the fact that scale items are allowed to load onto two factors – their own specific construct and a general factor that captures any underlying commonality between the specific components (see Figure 1).

The bifactor model allows for the commonality among the scale items explained by the general self-judgment factor to be partitioned out, with the scale-specific factors representing only the unique shared variance among the items on each scale. This is achieved by specifying the inter-factor correlations to equal zero, thus forcing the common variance in the model (i.e., excluding the item-specific residual variance) to be split between four orthogonal factors. Variance in item scores are explained by the direct influence of the general factor, the influence of specific constructs, independent of the general factor, plus item-level residual variance not accounted for by either the underlying negative self-judgment factors or the scale-specific factors.[5]

Model-based scale reliabilities were calculated following Rodriguez et al. (2016). The overall reliability statistic ω is defined as the total amount of variance attributable to *all* the constructs underlying the total scale score (i.e., general plus scale-specific factors), divided by the total observed variance (i.e., the variance attributable to all the latent factors in the model

---

[5] Three alternative models were also tested: a unidimensional model with all items from the three scales loading onto a single common factor; a first-order model with each of the three scales represented by its own factor, which were allowed to correlate; and a higher-order model where the three first-order factors loaded onto a second-order general factor. The bifactor model was a superior fit to the data than these alternative factor structures (see Supplementary Material).

plus item-level residual unexplained variance).[6] The reliability statistic ω-H represents the amount of total variance attributed to variance on the general factor, ω-S refers to the proportion of total variance explained by specific factors and the general factor associated with the items of that specific factor (so, for example, the loadings of the 11 items in the WBIS scale onto the general factor), and ω-HS refers to the proportion of total variance attributable to specific factors after partitioning out the variance explained by the general factor. As with other reliability statistics, the value of ω can range from 0 to 1.

The bifactor model was a good fit for the data, $\chi^2(322) = 1049$, $p < .001$, CFI = .90, SRMR = .06.[7] Items on the WBIS loaded negatively onto the general construct, whereas items on the Appearance Evaluation and RSE scales loaded positively. Thus, the general construct appears to represent an underlying *positive* self-judgment factor. Given than approximately 80% of the variance in both the WBIS and the Appearance Evaluation scales was attributable to this general factor, compared with just under half of the variance in the RSE (Table 2), it would seem that the common underlying factor is capturing a more appearance-specific construct – a body-related positive self-judgment, rather than self-judgment more generally.

Consistent with the hypothesized underlying general construct, items on the WBIS and Appearance Evaluation scale loaded strongly onto the general factor and only weakly onto their specific factors once shared variance was partitioned out. As noted above, over

---

[6] This statistic is the latent-variable equivalent of Cronbach's α. Note, nomenclature for omega statistics varies across different authors.

[7] As an added precaution, the bifactor model was also tested without item 1 on the WBIS. This model was a worse fit for the data, $\Delta \chi^2 = 114$ Δ df = 25.

three-quarters of the variance in WBIS scores was explained by the underlying Positive Body-related Self-judgment factor, with just 15% accounted for by factor-specific commonality. The remainder was attributed to unique item-level variance. For six of the 11 items of the WBIS scale items – namely, questions 1, 2, 4, 9, 10, and 11– less than 10% of the variance was attributable to the construct-specific factor. Given that the general factor appears to be capturing variance in body-related self-judgment, it is unsurprising that only two of the seven items on the Appearance Evaluation scale demonstrated more than 10% of variance attributable to the factor-specific construct; over 80% of the variance in scale scores was attributable to variation in the underlying general factor.

Only three items on the IWS factor, and two on the Body Satisfaction factor had a loading above .4. Scale reliability statistics confirmed the poor internal validity of the construct-specific factors, with extremely low reliable score variance explained by the construct-specific IWS and Body Satisfaction factors after controlling for the general factor (ω-HS = .14 and .08, respectively). Thus, the apparent reliability of the individual scales (both ω-S > .90) is inflated by the presence of reliable variance that is actually explained by a more general underlying factor.

Additionally, two of the eleven items on the WBIS and three of the seven items on the Appearance Evaluation scale did not load significantly onto their specific factors after controlling for the underlying Positive Body-related Self-judgment general factor. These items represented the only two reverse-scored items on the WBIS (i.e., higher scores indicated lower IWS) and three of the five positively worded items on the Appearance Evaluation scale (i.e., higher scores indicated greater appearance satisfaction). The non-significant loadings of these items onto the construct-specific factors when controlling for the general factor, suggests that almost all of the reliable variance in these items was explained by the general Positive Body-related Self-judgment factor.

The remaining nine items loaded onto the IWS factor with loadings ranging from .196 to .463, median .331. An average of 12.0% of the variance in positively scored items (i.e., higher IWS) was explained by the IWS construct after partitioning out the general factor and the item-level residual variance, whereas the corresponding figure for reverse-scored items was only 0.7%. Thus, the IWS factor clearly represented a negative weight-based attitude toward the self.

Also of note was the content validity of the items that did and did not remain strong predictors of the construct-specific factors. For the IWS factor, excluding the reverse-scored items, one of which pertained to self-worth and the other to desire to change, five items loaded onto the factor with standardized loadings above .3. Two related to self-worth, one to concern about others' attitudes, and two to psychological distress. Items pertaining to desire to change, fat identity, and body image, did not fare so well. For the Appearance Evaluation scale, two of the seven items are reverse scored (i.e., indicating body dissatisfaction) and five are scored such that higher item responses indicate greater body satisfaction. Unsurprisingly, controlling for shared variance on the general Positive Body-related Self-judgment, three of these positive items no longer significantly loaded onto the construct-specific Body Satisfaction factor, and a fourth had the next lowest loading (.214). Further, an average of 6.0% of the construct-specific variance in positive items was accounted for by the specific factor, compared with 21.5% of the variance of reverse-scored items. Thus, the items representing the construct-specific factor, after partitioning out Positive Body-Related self-judgment, largely represent body *dissatisfaction*, with the valence of the factor being reversed from that of the original scale.

Items on the RSE loaded evenly onto the general and scale-specific factors, with variance on nine out of the ten items being attributable to factor-specific commonality after controlling for general positive body-related self-judgment. Further, half of the items loaded > .5 onto the Self-Esteem factor, and the Self-Esteem factor had reasonable scale-specific reliability, even when controlling for the general factor. Reliability of the construct-specific Self-Esteem factor (ω-HS = .44) was considerably higher than that for the other two construct-specific factors. Examination of the RSE item factor loadings on the construct-specific factor, once common variance with the other measures had been removed, indicated a small effect of item valence, such that mean loading of positively scored items was .429, whereas the mean loading of reverse-scored items was .551, suggesting, in line with the hypothesized interpretation of the underlying construct, that partitioning out the positive self-judgment variance strengthened the relationship between the reverse-scored items and the remaining Self-Esteem factor. Similarly, on average, only 19.8% of the variance in the positive items on the RSE was explained in the construct-specific Self-Esteem factor, compared with 31.4% of the variance on reverse-scored items. Thus, as with the other-construct specific factors, the Self-Esteem factor appears to be capturing slightly more negative self-judgments.

3.5. Mediation analysis

3.5.1 Eating behavior outcome

Exploratory (EFA) and confirmatory factor analysis (CFA) were used to define the latent disordered eating factor structure to be used as the dependent variable in the mediation analysis. First, principal axis factoring was run with Direct Oblimin rotation with Kaiser normalization on a random 50% of the sample. Oblique rotation was chosen as it was expected that subscales would be correlated with each other to some extent. It was stipulated that item factor loadings should be > .5 on the primary factor and < .3 on any other factors. As recommended by Stevens (2002; cited in Field, 2013) for samples greater than 250, factor extraction decisions were based on the scree plot, rather than eigenvalues. The resultant factor structure was confirmed on the other half of the sample. Goodness of model fit to the observed data was assessed with the two-index reporting method recommended by Hu and Bentler (1999) for maximum likelihood-based estimation, using the absolute fit index Standardized Root Mean Squared Residual (SRMR) and the incremental fit index Comparative Fit Index (CFI). Cut-off values close to .95 for the CFI and .08 for the SRMR generally indicate a good fit between the hypothesized model and the observed data (Hu & Bentler, 1999).

Exploratory factor analysis suggested a three-factor solution was the best fit for the data. As the subscales were allowed to correlate, unique variance explained by each factor in the rotated factor solution could not be determined; however, prior to rotation, the three factors explained 69.5% of the variance in eating behavior scores. The first factor, which was labelled Binge Behavior, comprised the two binge eating frequency scores and the EDDS sum score. The second factor, labelled Restraint, contained the DEBQ Restraint subscale and current dieting behavior. The final factor comprised the DEBQ External and Emotional Eating subscales, and was labelled Disinhibition. Confirmatory factor analysis indicated that this model was an acceptable fit for the data: $\chi^2(11) = 55.6$, $p < .001$, CFI = .91, SRMR = .07.[8]

---

[8] Two possible alternative models were also tested: a unidimensional model with all eating behavior measures loading onto a single "disordered eating" factor, and a second-order two-factor model, where the first-order factors represented more generic non-physiological eating patterns (made up of the DEBQ subscales and current dieting behavior) and eating pathology (made up of the EDDS symptom score and binge eating behavior), and these first-order factors then loaded onto a second-order "disordered eating" factor. Neither model was a good fit for the data.

3.5.2 Full mediation model

Structural equation modelling was conducted with experienced weight stigma as the predictor variable, the components of the bifactor measurement model as the mediators, and the disordered eating latent construct as the dependent variable. This approach allowed the relative contribution of the construct-specific IWS, Body Satisfaction, and Self-Esteem factors and the general underlying Positive Body-related Self-judgment factor to be assessed independently as mediators, whilst controlling for variance in the other factors. To this end, a parallel mediation model was tested – that is, the significance of indirect pathways between experienced weight stigma and disordered eating were explored for all possible mediators simultaneously. Thus, each specific indirect pathway controlled for all other pathways in the model.

Analyses were conducted using maximum likelihood (ML) estimation. Although ML estimation assumes multivariate normal distribution of the measured variables in the population, it is robust to mild to moderate normality violations (Curran, West, & Finch, 1996; Fan & Wang, 1998). Descriptive statistics indicated approximately normal distribution of indicators, with only mild values of skew and kurtosis (absolute values ≤ 1.0 and ≤ 1.4, respectively). Starting values were allocated to factor loadings and the variance of latent factors was fixed to unity to facilitate model identification. Goodness of model fit to the observed data was assessed using the same two-index reporting method outlined above. However, CFI tends to decline with increasing number of indicators in the model (Kenny & McCoach, 2003). In the present analysis, the maximum number of variables per factor was 28, thus, following Chen et al. (2012), a less stringent cut-off of .90 was used for the CFI to indicate goodness of fit.

A bootstrapping resampling procedure (DiCiccio & Efron, 1996; Fritz & MacKinnon,

2007) with 1000 bootstrap samples was used to construct confidence intervals for the path estimates and indirect effects. An effect was considered statistically significant if the 95% confidence interval did not include zero.

Gender and BMI were entered as covariates for all steps in the mediation pathway; however, only the relationship between gender and experienced weight stigma was significant. The model was then re-run with the non-significant covariate paths removed. Results of the parallel mediation analysis are displayed in Figure 2. Experienced weight stigma was not a significant predictor of global Positive Body-related Self-judgment when controlling for the construct-specific factors, but did significantly predict each of the construct-specific factors, controlling for all other factors in the model. The positive relationship between experienced weight stigma and the Body Satisfaction construct reflects that the construct-specific factor appears to represent body dissatisfaction.

The general and construct-specific factors all significantly predicted disordered eating controlling for each other. The Positive Body-related Self-judgment and IWS factors were positive predictors of disordered eating, as expected. Similarly, the self-esteem factor also positively predicted disordered eating, suggesting that higher self-esteem was associated with *more* disordered eating; this is consistent with the residual variance in the construct-specific factor representing more negative self-judgment. However, the Appearance Evaluation factor was a negative predictor of disordered eating, which would not be expected if this factor is capturing body dissatisfaction.

Standardized indirect effects are shown in Table 3. That the indirect pathways from experienced weight stigma to disordered eating via Body Satisfaction and Self-Esteem were negative is simply due to the shifted valence of the construct-specific factors and the multiplicative effect of the proximal and distal pathways, whereby a positive regression coefficient to the mediator and a negative regression coefficient from the mediator to the dependent variable are multiplied to provide a negative indirect effect, and vice versa. Both indirect pathways from experienced weight stigma to disordered eating via Positive Body-related Self-judgment and IWS were non-significant. These findings may have more to do with splitting out the total effect in this manner, when all pathways appear to be equally relevant.

Based on the model-derived standardized coefficients for the respective paths in these mediation models, a sample size of approximately 400 would be needed to detect a significant mediation effect with .8 power at a significance level of .05 using bootstrap resampling procedures (Fritz & MacKinnon, 2007). Thus, a sample size of 379 may be slightly underpowered to detect these significant indirect effects when split across four mediators. In the present analysis, the 95% confidence intervals for both the Positive Self-judgment and the IWS mediation effects were close to not including zero, and therefore a slightly larger sample size may have produced significant findings for these mediators.

3.6. Post-hoc analyses

Possible issues arising from splitting of total indirect effects between several mediators in the parallel mediation analysis may be overcome by considering one mediator at a time, but comparing the construct-specific bifactor mediator with total construct (i.e., without partitioning out common variance) mediator. This technique allows for the impact of ignoring multidimensionality to be explored. Following Gonzalez and MacKinnon (2016), two simple mediation models were run, the first using the construct-specific IWS factor derived from the bifactor model as the sole mediator, and the second using a unidimensional IWS factor based solely on the items of the WBIS as the mediator, ignoring other possible sources of common variance. If IWS, as delineated by partitioning out common variance using bifactor analysis, is the *true* mediator of the relationship between experienced weight stigma and disordered eating behavior, then treating IWS as unidimensional and running the mediation model with the original WBIS scale items effectively contaminates the *pure* variance of the construct-specific factor with unnecessary variance shared with other constructs. Under these circumstances, this would produce a poorer model fit, increase bias within the parameters, weaken the indirect effect, and reduce the statistical power of the analysis (Gonzalez & MacKinnon, 2016). The path diagram for the bifactor mediation model is displayed in Figure 3. The positive and negative items of the RSE were *a priori* allowed to load onto separate correlated factors in the bifactor model. The bifactor mediation model was a generally poor fit to the data, $\chi^2(595) = 1867$, $p < .001$ CFI = .85, SRMR = .15.

The unidimensional model (Figure 4) was an adequate fit for the data on most fit indices, $\chi^2(165) = 803$, $p < .001$ CFI = .85, SRMR = .08, and a significantly improved fit compared with the bifactor model, $\Delta\chi^2(430) = 1064$; critical value = 479.

Additionally, the standardized path coefficient from the mediator to the dependent variable was almost double in size in the unidimensional compared with the bifactor mediator model, and almost two-thirds of the variance in disordered eating was explained by the model with the unidimensional mediator, compared with under 20% in the model with the bifactor mediator. The total, direct, and indirect effects for the two models are shown in Table 4, and indicate that the indirect effect of experienced weight stigma on disordered eating was slightly stronger for the unidimensional than the bifactor mediator.

4. Discussion

The present study is the first to attempt to separate out the distinct aspects of internalized weight stigma, as measured by the WBIS, from those that share variance with the related constructs of body satisfaction and global self-esteem. Using bifactor analysis to identify shared and unique variance among these constructs, findings, suggest that a distilled IWS construct is not the *true* mediator of the relationship between experienced weight stigma and disordered eating. First, a bifactor model including the WBIS, RSE, and Appearance Evaluation scales was a good fit for the data, supporting the existence of a common underlying factor that accounted for shared variance between the measures, and which explained the data better than when considering the scales as capturing distinct constructs.[9] Given the extremely high correlations between these scales, this was unsurprising, and we had predicted that once this generic self-judgment variance had been removed, variance attributable to body- and weight-related self-judgment would be explained by the construct-specific factors. However, the data suggest that the common factor primarily represented body-related self-judgment, rather than a more generic self-evaluation. The fact that approximately half of the variance in RSE scores was explained by this body-related self-judgment factor, despite no reference to body or appearance in the items comprising this measure, indicates the substantial influence that low body-related esteem has on global self-esteem.

A body-related self-judgment factor would account for the exceptionally low residual variance in the Appearance Evaluation scores explained by the construct-specific factor. However, the very low construct-specific reliability for the WBIS does raise the question of whether the WBIS provides interpretable unique variance after controlling for the general underlying body-related self-judgment factor (Reise, Bonifay, & Haviland, 2013; Reise,

---

[9] Note, the good fit of a bifactor model speaks to the existence of this general underlying factor. It does not *per se* indicate that the domain-specific constructs provide useful additional information. This question is determined by examining residual variance, scale reliabilities, factor loadings, and so on.

Moore, & Haviland, 2010). Attempts to test this by assessing them simultaneously as mediators of the relationship between experienced weight stigma and disordered eating behavior confirmed that they cannot be reliably separated. Specifically, controlling for the indirect pathways between experienced weight stigma and eating behavior via body satisfaction, global self-esteem, and positive body-related self-judgment, the construct-specific IWS factor was no longer a significant mediator of the relationship. This finding was confirmed by simple mediation analyses in which a model containing a unidimensional WBIS mediator was a better fit to the data than one with a construct-specific IWS mediator with shared variance partitioned out.

The construct-specific IWS factor clearly represented a negative weight-based attitude toward the self, evidenced by the fact that the positively worded items of the WBIS, reflecting higher IWS, had more reliable variance explained by the construct-specific IWS factor once common variance had been partitioned out, whereas almost none of the variance in reverse-scored items was attributable to the construct-specific IWS factor. As such, this factor may be, in some ways, more distinctively related to the theorized meaning of the construct than was the case for the other residualized construct-specific factors; therefore, its reduced ability to predict such a key behavioral outcome raises questions about its utility in future research.

It is possible that findings from existing studies using the WBIS that have identified IWS as a significant mediator of the relationship between experienced weight stigma and disordered eating, may actually be measuring a construct that is at least partially confounded by the related constructs of body satisfaction and global self-esteem. However, the question arises, “Does it matter?” The WBIS has good internal validity (Durso & Latner, 2008), has been shown to be linked with several key outcomes of interest (Carels et al., 2010; Durso & Latner, 2008; Hilbert, Baldofski, et al., 2014; Webb & Hardin, 2016), and provides a potentially useful target for interventions aimed at improving wellbeing and quality of life in higher-weight individuals (Mensinger & Meadows, 2017; Pearl, Hopkins, Berkowitz, & Wadden, 2016). Even in the present analysis, the construct-specific IWS factor was significantly predicted by experienced weight stigma, and demonstrated a statistically significant positive relationship with disordered eating, even when controlling for body satisfaction, global self-esteem, and general positive body-related self-judgment. This is likely explained by the nature of the items that retained construct-specific variance. The 11-item WBIS scale was distilled from an original pool of 19 questions, covering a rather wide range of cognitive and affective responses to one's own weight status. The extent to which these items specifically related to the proposed definition of "self-devaluation due to weight" is questionable (Meadows & Higgs, 2019), and items were removed from the scale based on an *a priori* assumption of unidimensionality. A recent CFA of the original 19-item pool used to develop the WBIS, but without the *a priori* assumption of unidimensionality, identified two factors – seven items representing Weight-related distress and six representing actual Self-devaluation (WBIS-2F; Meadows & Higgs, 2019). Looking at the five items on the standard WBIS that had over 10% variance explained by the construct-specific factor in the present analysis, four of them were items that loaded onto the Weight-related distress subscale of the WBIS-2F. Thus, the statistical significance of the above relationships indicate that experienced weight stigma was positively associated with weight-related distress, and that weight-related distress was positively associated with disordered eating. In contrast, experienced weight stigma did not predict general positive body-related self-judgment, which may be considered to represent a more trait-like construct.

Despite the assumption of unidimensionality underpinning the development of the standard WBIS, the resulting scale appears to consist of numerous different concepts, including concern about others' attitudes, body image, and desire for weight change – which could arguably be categorized as body image, or as a pragmatic response to a pervasively anti-fat environment, even in the absence of self-devaluation, that appear to have little to do with self-devaluation *per se*. This analysis highlights that internal reliability of a construct (e.g., Cronbach's α), which tends to be high for the WBIS, may be misleading in terms of construct unidimensionality – items that are highly interrelated are not necessarily all measuring the same thing (McNeish, 2017). Indeed, examining the factor loadings of items on the WBIS following bifactor analysis in the present study revealed a pattern in the nature of items that loaded onto the construct-specific IWS factor above the 0.3 cut-off after partitioning out common variance with global self-esteem and body satisfaction. Excluding reverse-scored items, the items with satisfactory loadings onto the construct-specific factor were those relating to self-devaluation, psychological distress, and fear of others' attitudes. This appears to be more similar to the conceptualisation of weight self-stigma used in the development of the Weight Self-Stigma Questionnaire, another validated measure of IWS, which distinguishes between self-stigma and fear of being stigmatized by others (Lillis et al., 2010). Interestingly, a validation study of a Spanish version of the WBIS in 59 higher-weight, mostly treatment-seeking individuals, identified a two-factor structure (Gomez & Baile, 2015). The first factor explained 51% of the variance in WBIS scores in this sample, and included all of the items classified in the present study as describing self-devaluation, two mixed-concept items, and the item described herein as "fat identity" (see Table 2). The second factor explained a further 14% of the variance and included items about desire for change, fear of others' attitudes, aspects of body image, and one item relating to distress. Thus, despite its widespread use in current weight stigma research as a unidimensional measure of

IWS, the 11-item version of the WBIS appears to be capturing several related constructs in addition to weight-related self-devaluation.

While bifactor analysis provides a useful technique for addressing a vexing issue with conceptually overlapping variables, it is not without methodological challenges. In particular, some have questioned the feasibility of trying to interpret residualized construct-specific factors when there is a strong underlying general factor explaining the majority of the variance (Chen et al., 2012). This issue may help to explain the unexpected negative relationship between the construct-specific Body Satisfaction factor (body dissatisfaction) and disordered eating. Given that only 8.4% of the variance in Appearance Evaluation scores was attributable to this factor, with the majority being partitioned out onto the general Positive Body-related Self-Judgment factor, it would be unwise to over-interpret this finding. While the difficulty of interpreting the essence of construct-specific factors with low attributable variance certainly raises practical difficulties for the researcher, it does not present a reason to avoid the technique – rather it highlights the need for better construct clarity and specificity, and ought to be used more often during the early stages of measure development. Numerous studies in educational, health, and personality research have successfully identified differential direct and indirect pathways via construct-specific and general factors based on bifactor analyses (e.g., Király et al., 2015; Lac & Donaldson, 2017; Lauriola & Iani, 2016). Thus, it is not solely the fact that these constructs are conceptually related that is causing the statistical issue in the present analysis. Rather, given endemic and pervasive weight stigma in both Western and non-Western societies (Brewis, SturtzSreetharan, & Wutich, 2018), lack of differentiation between the constructs is likely due to considerable overlap between them in higher-weight individuals who reside in these fat-hostile environments (Rogge, Greenwald, & Golden, 2004).

The present study has a number of limitations. First, as noted above, the study may be slightly underpowered to detect significant effects on the mediation model. Secondly, the relatively small number of men in the sample makes it impossible to test for any gender differences in how experiences of weight stigma indirectly influence downstream eating behavior. A small number of studies (e.g., Boswell & White, 2015; M. E. Eisenberg, Ward, Linde, Gollust, & Neumark-Sztainer, 2017; Sattler, Deane, Tapsell, & Kelly, 2018) have identified gender differences in coping strategies for experienced weight stigma, but to our knowledge, nobody has looked at gender differences across mediation models. Future studies should aim for more equal recruitment across gender (including non-binary) identities.

Future research should also consider whether ethnic, racial, or cultural differences influence these findings. While weight stigma is a now nearly universal phenomenon (Brewis et al., 2018), some cultural differences remain in attitudes to higher-weight bodies (e.g., Aryeetey, 2016; although, note, most recent studies find few differences by race/ethnicity or cultural background), and differences in internalisation and coping strategies have also been identified by race, ethnicity, or nationality (Brewis & Wutich, 2014; M. H. Eisenberg, Street, & Persky, 2017; Himmelstein, Puhl, & Quinn, 2017). It is possible that in some samples poor body satisfaction is not necessarily accompanied by weight-related concerns, or at least, not to the same extent. It would also be worth replicating this study using alternative measures of IWS, such as the WBIS-2F or the Weight Self-Stigma Questionnaire, that operationalize different aspects of IWS, separating out self-devaluation from other concerns. Findings from such studies could help inform development of interventions aimed at minimizing the harms associated with experienced and internalized weight stigma.

In conclusion, IWS, as measured by the WBIS, does mediate the relationship between experienced weight stigma and maladaptive eating behaviors. However, much of the indirect effect appears to be transmitted via a more general underlying body-related self-judgment factor, rather than via a conceptually "pure" construct of weight-related self-devaluation. Greater conceptual clarity in the study of IWS is needed to fully understand the true mechanisms via which societal weight stigma impacts on individuals' self-directed judgments, how this impacts on downstream health-related behaviors, and how best to address the problem clinically or at the societal level. It remains to be established whether the relationship between experienced weight stigma and maladaptive eating behaviors can be explained via an indirect effect whereby experienced stigma reduces positive self-regard, without recourse to internalization of societal attitudes toward weight.

Acknowledgments

Special thanks to Ed Rigdon and Martin Brunner for technical advice regarding bifactor analysis.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest: None

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: APA.

Arditte Hall, K. A., Bartlett, B. A., Iverson, K. M., & Mitchell, K. S. (2017). Military-related trauma is associated with eating disorder symptoms in male veterans. *International Journal of Eating Disorders*, *50*(11), 1328–1331. https://doi.org/10.1002/eat.22782

Aryeetey, R. N. O. (2016). Perceptions and experiences of overweight among women in the Ga East District, Ghana. *Frontiers in Nutrition, 3*, Article 13. https://doi.org/10.3389/fnut.2016.00013

Barnes, R. D., Ivezaj, V., & Grilo, C. M. (2014). An examination of weight bias among treatment-seeking obese patients with and without binge eating disorder. *General Hospital Psychiatry*, *36*(2), 177–180. https://doi.org/10.1016/j.genhosppsych.2013.10.011

Boswell, R. G., & White, M. A. (2015). Gender differences in weight bias internalisation and eating pathology in overweight individuals. *Advances in Eating Disorders, 3*(3), 259–268. https://doi.org/10.1080/21662630.2015.1047881

Brauhardt, A., Rudolph, A., & Hilbert, A. (2014). Implicit cognitive processes in binge-eating disorder and obesity. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*(2), 285–290. https://doi.org/10.1016/j.jbtep.2014.01.001

Brewis, A., SturtzSreetharan, C., & Wutich, A. (2018). Obesity stigma as a globalizing health challenge. *Globalization and Health, 14*(1), 20. https://doi.org/10.1186/s12992-018-0337-x

Brewis, A. A., & Wutich, A. (2014). A world of suffering? Biocultural approaches to fat stigma in the global contexts of the obesity epidemic. *Annals of Anthropological Practice, 38*(2), 269–283. https://doi.org/10.1111/napa.12056

Brown, T. A., Cash, T. F., & Mikulka, P. J. (1990). Attitudinal body image assessment: Factor analysis of the Body-Self Relations Questionnaire. *Journal of Personality Assessment*, *55*(1–2), 135–144. https://doi.org/10.1080/00223891.1990.9674053

Brunault, P., Rabemampianina, I., Apfeldorfer, G., Ballon, N., Couet, C., Réveillère, C., Gaillard, P., & El-Hage, W. (2015). The Dutch Eating Behavior Questionnaire: Further psychometric validation and clinical implications of the French version in normal weight and obese persons. *La Presse Médicale*, *44*(12), e363–e372. https://doi.org/10.1016/j.lpm.2015.03.028

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Burmeister, J. M., Hinman, N., Koball, A., Hoffmann, D. A., & Carels, R. A. (2013). Food addiction in adults seeking weight loss treatment. Implications for psychosocial health and weight loss. *Appetite*, *60*, 103–110. https://doi.org/10.1016/j.appet.2012.09.013

Carels, R. A., Hlavka, R., Selensky, J. C., Solar, C., Rossi, J., & Miller, J. C. (2019). A daily diary study of internalised weight bias and its psychological, eating and exercise correlates. *Psychology & Health, 34*(3), 306–320. https://doi.org/10.1080/08870446.2018.1525491

Carels, R. A., Wott, C. B., Young, K. M., Gumble, A., Koball, A., & Oehlhof, M. W. (2010). Implicit, explicit, and internalized weight bias and psychosocial maladjustment among treatment-seeking adults. *Eating Behaviors*, *11*(3), 180–185. https://doi.org/10.1016/j.eatbeh.2010.03.002

Cash, T. F. (2000). MBSRQ Users' Manual. Retrieved from http://www.body-images.com/assessments/mbsrq.html

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, *80*(1), 219–251. https://doi.org/10.1111/j.1467-6494.2011.00739.x

Chen, F. F., West, S., & Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Crandall, C. S. (1994). Prejudice against fat people: Ideology and self-interest. *Journal of Personality and Social Psychology*, *66*(5), 882–894. https://doi.org/10.1037//0022-3514.66.5.882

Crandall, C. S., D'Anello, S., Sakalli, N., Lazarus, E., Wieczorkowska Nejtardt, G., & Feather, N. T. (2001). An attribution-value model of prejudice: Anti-fat attitudes in six nations. *Personality and Social Psychology Bulletin, 27*(1), 30–37. https://doi.org/ 10.1177/0146167201271003

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16

Dakanalis, A., Zanetti, M. A., Clerici, M., Madeddu, F., Riva, G., & Caccialanza, R. (2013). Italian version of the Dutch Eating Behavior Questionnaire. Psychometric proprieties and measurement invariance across sex, BMI-status and age. *Appetite*, *71*, 187–195. https://doi.org/10.1016/J.APPET.2013.08.010

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*(3), 189–228. https://doi.org/10.1214/ss/1032280214

Donnellan, M. B., Ackerman, R. A., & Brecheen, C. (2016). Extending structural analyses of the Rosenberg Self-Esteem Scale to consider criterion-related validity: Can composite self-esteem scores be good enough? *Journal of Personality Assessment*, *98*(2), 169–177. https://doi.org/10.1080/00223891.2015.1058268

Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2015). Measures of self-esteem. In G. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 131–157). London: Elsevier. https://doi.org/10.1016/B978-0-12-386915-9.00006-1

Durso, L. E., & Latner, J. D. (2008). Understanding self-directed stigma: Development of the Weight Bias Internalization Scale. *Obesity*, *16 Suppl 2*, S80–S86. https://doi.org/10.1038/oby.2008.448

Durso, L. E., Latner, J. D., & Ciao, A. C. (2016). Weight bias internalization in treatment-seeking overweight adults: Psychometric validation and associations with self-esteem, body image, and mood symptoms. *Eating Behaviors*, *21*, 104–108. https://doi.org/10.1016/j.eatbeh.2016.01.011

Durso, L. E., Latner, J. D., & Hayashi, K. (2012). Perceived discrimination is associated with binge eating in a community sample of non-overweight, overweight, and obese adults. *Obesity Facts*, *5*(6), 869–880. https://doi.org/10.1159/000345931

Durso, L. E., Latner, J. D., White, M. A., Masheb, R. M., Blomquist, K. K., Morgan, P. T., & Grilo, C. M. (2012). Internalized weight bias in obese patients with binge eating disorder: A ssociations with eating disturbances and psychological functioning. *International Journal of Eating Disorders*, *45*(3), 423–427. https://doi.org/10.1002/eat.20933

Dutton, G. R., Lewis, T. T., Durant, N., Halanych, J., Kiefe, C. I., Sidney, S., Kim, Y., & Lewis, C. E. (2014). Perceived weight discrimination in the CARDIA study: Differences by race, sex, and weight status. *Obesity, 22*(2), 530–536. https://doi.org/10.1002/oby.20438

Eisenberg, M. H., Street, R. L., & Persky, S. (2017). “It runs in my family …”: The association of perceived family history with body dissatisfaction and weight bias internalization among overweight women. *Women & Health, 57*(4), 478–493. https://doi.org/10.1080/03630242.2016.1170095

Eisenberg, M. E., Ward, E., Linde, J. A., Gollust, S. E., & Neumark-Sztainer (2017). Exposure to teasing on popular television shows and associations with adolescent body satisfaction. *Journal of Psychosomatic Research, 103*, 15–21. https://doi.org/10.1016/j.jpsychores.2017.09.014

Fan, X., & Wang, L. (1998). Effects of potential confounding factors on fit indices and parameter estimates for true and misspecified SEM models. *Educational and Psychological Measurement*, *58*(5), 701–735. https://doi.org/10.1177/0013164498058005001

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). London: Sage Publications (UK).

Friedman, K. E., Reichmann, S. K., Costanzo, P. R., Zelli, A., Ashmore, J. A., & Musante, G. J. (2005). Weight stigmatization and ideological beliefs: Relation to psychological functioning in obese adults. *Obesity Research*, *13*(5), 907–916. https://doi.org/10.1038/oby.2005.105

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*(3), 233–239. https://doi.org/10.1111/j.1467- 9280.2007.01882.x

Gomez, S. S., & Baile, J. I. (2015). Psychometric properties of a Spanish adaptation of the Weight Bias Internalization Scale (WBIS). *Nutrición Hospitalaria*, *32*(4), 1510–1515. https://doi.org/10.3305/nh.2015.32.4.9343

Gonzalez, O., & MacKinnon, D. P. (2018). A bifactor approach to model multifaceted constructs in statistical mediation analysis. *Educational and Psychological Measurement, 78*(1), 5–31. https://doi.org/10.1177/0013164416673689

Griffiths, R. A., Beumont, P. J. V., Giannakopoulos, E., Russell, J., Schotte, D., Thornton, C., Touyz, S. W., & Varano, P. (1999). Measuring self-esteem in dieting disordered patients: The validity of the Rosenberg and Coopersmith contrasted. *International Journal of Eating Disorders*, *25*(2), 227–231. https://doi.org/10.1002/(SICI)1098-108X(199903)25:2<227::AID-EAT13>3.0.CO;2-4

Hatzenbuehler, M. L., Keyes, K. M., & Hasin, D. S. (2009). Associations between perceived weight discrimination and the prevalence of psychiatric disorders in the general population. *Obesity, 17*(11), 2033–2039. https://doi.org/10.1038/oby.2009.131

Hilbert, A., Baldofski, S., Zenger, M., Löwe, B., Kersting, A., & Braehler, E. (2014). Weight Bias Internalization Scale: Psychometric properties and population norms. *PLOS ONE*, *9*(1), e86303. https://doi.org/10.1371/journal.pone.0086303

Hilbert, A., Braehler, E., Haeuser, W., & Zenger, M. (2014). Weight bias internalization, core self-evaluation, and health in overweight and obese persons. *Obesity*, *22*(1), 79–85. https://doi.org/10.1002/oby.20561

Himmelstein, M. S., Puhl, R. M., & Quinn, D. M. (2017). Intersectionality: An understudied framework addressing weight stigma. *American Journal of Preventive Medicine, 53*(4), 421–431. https://doi.org/10.1016/j.amepre.2017.04.003

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hübner, C., Baldofski, S., Zenger, M., Tigges, W., Herbig, B., Jurowich, C., Kaiser, S., Dietrich, A., & Hilbert, A. (2015). Influences of general self-efficacy and weight bias internalization on physical activity in bariatric surgery candidates. *Surgery for Obesity and Related Diseases*, *11*(6), 1371–1376. https://doi.org/10.1016/j.soard.2014.11.013

Hübner, C., Schmidt, R., Selle, J., Kohler, H., Muller, A., de Zwaan, M., & Hilbert, A. (2016). Comparing self-report measures of internalized weight stigma: The Weight Self-Stigma Questionnaire versus the Weight Bias Internalization Scale. *PLOS ONE*, *11*(10), e0165566. https://doi.org/10.1371/journal.pone.0165566

Innamorati, M., Imperatori, C., Lamis, D. A., Contardi, A., Castelnuovo, G., Tamburello, S., Manzoni, G. M., & Fabbricatore, M. (2017). Weight Bias Internalization Scale discriminates obese and overweight patients with different severity levels of depression: The Italian version of the WBIS. *Current Psychology*, *36*(2), 242–251. https://doi.org/10.1007/s12144-016-9406-6

Kelley, T. L. (1927). *Interpretation of educational measurements*. New York, NY: Macmillan.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, *10*(3), 333–351. https://doi.org/10.1207/S15328007SEM1003_1

Király, O., Urbán, R., Griffiths, M. D., Ágoston, C., Nagygyörgy, K., Kökönyei, G., & Demetrovics, Z. (2015). The mediating effect of gaming motivation between psychiatric symptoms and problematic online gaming: An online survey. *Journal of Medical Internet Research*, *17*(4), e88. https://doi.org/10.2196/jmir.3515

Krabbenborg, M. A. M., Danner, U. N., Larsen, J. K., van der Veer, N., van Elburg, A. A., de Ridder, D. T. D., Evers, C., Stice, E., & Engels, R. C. M. E. (2012). The Eating Disorder Diagnostic Scale: Psychometric features within a clinical population and a cut-off point to differentiate clinical patients from healthy controls. *European Eating Disorders Review*, *20*(4), 315–320. https://doi.org/10.1002/erv.1144

Lac, A., & Donaldson, C. D. (2017). Higher-order and bifactor models of the Drinking Motives Questionnaire: Examining competing structures using confirmatory bifactor analysis. *Assessment*, *24*(2), 222–231. https://doi.org/10.1177/1073191115603503

Latner, J. D., Barile, J. P., Durso, L. E., & O'Brien, K. S. (2014). Weight and health-related quality of life: The moderating role of weight discrimination and internalized weight bias. *Eating Behaviors*, *15*, 586–590. https://doi.org/10.1016/j.eatbeh.2014.08.014

Lauriola, M., & Iani, L. (2017). Personality, positivity and happiness: A mediation analysis using a bifactor model. *Journal of Happiness Studies*, *18*, 1659–1682. https://doi.org/10.1007/s10902-016-9792-3

Lee, M. S., & Dedrick, R. F. (2016). Weight Bias Internalization Scale: Psychometric properties using alternative weight status classification approaches. *Body Image*, *17*, 25–29. https://doi.org/10.1016/j.bodyim.2016.01.008

Lee, S. W., Stewart, S. M., Striegel-Moore, R. H., Lee, S., Ho, S., Lee, P. W. H., Katzman, M. A., & Lam, T. (2007). Validation of the Eating Disorder Diagnostic Scale for use with

Hong Kong adolescents. *International Journal of Eating Disorders*, *40*(6), 569–574. https://doi.org/10.1002/eat.20413

Lillis, J., Luoma, J. B., Levin, M. E., & Hayes, S. C. (2010). Measuring weight self-stigma: The Weight Self-Stigma Questionnaire. *Obesity*, *18*(5), 971–976. https://doi.org/10.1038/oby.2009.353

Lin, Y.-C., Latner, J. D., Fung, X. C. C., & Lin, C.-Y. (2018). Poor health and experiences of being bullied in adolescents: Self-perceived overweight and frustration with appearance matter. *Obesity*, *26*(2), 397–404. https://doi.org/10.1002/oby.22041

Magallares, A. (2014). Right wing autoritharism, social dominance orientation, controllability of the weight and their relationship with antifat attitudes. *Universitas Psychologica, 13*(2), 771–779. https://doi.org/10.11144/Javeriana.UPSY13-2. rwas

Major, B., Hunger, J. M., Bunyan, D. P., & Miller, C. T. (2014). The ironic effects of weight stigma. *Journal of Experimental Social Psychology*, *51*, 74–80. https://doi.org/10.1016/j.jesp.2013.11.009

Mann, M., Hosman, C. M. H., Schaalma, H. P., & de Vries, N. K. (2004). Self-esteem in a broad-spectrum approach for mental health promotion. *Health Education Research*, *19*(4), 357–372. https://doi.org/10.1093/her/cyg041

Mautner, R. D., Owen, S. V., & Furnham, A. (2000). Cross-cultural explanations of body image disturbance in western cultural samples. *International Journal of Eating Disorders*, *28*(2), 165–172. https://doi.org/10.1002/1098- 108X(200009)28:2<165::AID-EAT5>3.0.CO;2-G

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

Meadows, A., & Higgs, S. (2019). The multifaceted nature of weight-related self-stigma: Validation of the Two-Factor Weight Bias Internalization Scale (WBIS-2F). *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2019.00808

Mehak, A., Friedman, A., & Cassin, S. E. (2018). Self-objectification, weight bias internalization, and binge eating in young women: Testing a mediational model. *Body Image, 24*, 111–115. https://doi.org/10.1016/j.bodyim.2018.01.002

Mensinger, J. L., & Meadows, A. (2017). Internalized weight stigma mediates and moderates physical activity outcomes during a healthy living program for women with high body mass index. *Psychology of Sport and Exercise*, *30*, 64–72. https://doi.org/10.1016/j.psychsport.2017.01.010

Mond, J., Mitchison, D., Latner, J., Hay, P., Owen, C., & Rodgers, B. (2013). Quality of life impairment associated with body dissatisfaction in a general population sample of women. *BMC Public Health*, *13*, 920. https://doi.org/10.1186/1471-2458-13-920

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén. Retrieved from www.statmodel.com

Myers, A., & Rosen, J. C. (1999). Obesity stigmatization and coping: Relation to mental health symptoms, body image, and self-esteem. *International Journal of Obesity*, *23*, 221–230. https://doi.org/10.1038/sj.ijo.0800765

O'Hara, L., Tahboub-Schulte, S., & Thomas, J. (2016). Weight-related teasing and internalized weight stigma predict abnormal eating attitudes and behaviours in Emirati female university students. *Appetite*, *102*, 44–50. https://doi.org/10.1016/j.appet.2016.01.019

Palmeira, L., Pinto-Gouveia, J., & Cunha, M. (2017). Exploring the efficacy of an acceptance, mindfulness & compassionate-based group intervention for women struggling with their weight (Kg-Free): A randomized controlled trial. *Appetite*, *112*, 107–116. https://doi.org/10.1016/j.appet.2017.01.027

Pearl, R. L., & Puhl, R. M. (2016). The distinct effects of internalizing weight bias: An experimental study. *Body Image*, *17*, 38–42. https://doi.org/10.1016/j.bodyim.2016.02.002

Pearl, R. L., Puhl, R. M., & Dovidio, J. F. (2015). Differential effects of weight bias experiences and internalization on exercise among women with overweight and obesity. *Journal of Health Psychology*, *20*(12), 1626–1632. https://doi.org/10.1177/1359105313520338

Pearl, R. L., Hopkins, C. H., Berkowitz, R. I., & Wadden, T. A. (2018). Group cognitive-behavioral treatment for internalized weight stigma: A pilot study. *Eating and Weight Disorders*, *23*, 357–362. https://doi.org/10.1007/s40519-016-0336-y

Pepper, A. C., & Ruiz, S. Y. (2007). Acculturation's influence on antifat attitudes, body image and eating behaviors. *Eating Disorders*, *15*, 427–447. https://doi.org/10.1080/10640260701667912

Puhl, R. M., & Brownell, K. D. (2006). Confronting and coping with weight stigma: An investigation of overweight and obese adults. *Obesity*, *14*(10), 1802–1815. https://doi.org/10.1038/oby.2006.208

Puhl, R. M., & Heuer, C. A. (2009). The stigma of obesity: A review and update. *Obesity*, *17*(5), 941–964. https://doi.org/10.1038/oby.2008.636

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, *23*(1), 51–67. https://doi.org/10.1207/s15327906mbr2301_3

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rogge, M. M., Greenwald, M., & Golden, A. (2004). Obesity, stigma, and civilized oppression. *Advances in Nursing Science, 27*(4), 301–315. https://doi.org/10.1097/00012272-200410000-00006

Rosenberg, M. (1965). *Society and the adolescent self-image*. New Jersey: Princeton University Press.

Sani, S. H. Z., Fathirezaie, Z., Brand, S., Pühse, U., Holsboer-Trachsler, E., Gerber, M., & Talepasand, S. (2016). Physical activity and self-esteem: Testing direct and indirect relationships associated with psychological and physical mechanisms. *Neuropsychiatric Disease and Treatment*, *12*, 2617–2625. https://doi.org/10.2147/NDT.S116811

Sattler, K. M., Deane, F. P., Tapsell, L., & Kelly, P. J. (2018). Gender differences in the relationship of weight-based stigmatisation with motivation to exercise and physical activity in overweight individuals. *Health Psychology Open, 5*(1), Article 2055102918759691. https://doi.org/10.1177/2055102918759691

Schvey, N. A., Roberto, C. A., & White, M. A. (2013). Clinical correlates of the Weight Bias Internalization Scale in overweight adults with binge and purge behaviours. *Advances in Eating Disorders: Theory, Research and Practice*, *1*(3), 213–223. https://doi.org/10.1080/21662630.2013.794523

Silva, J. R., Behar, R., Cordella, P., Ortiz, M., Jaramillo, K., Alvarado, R., & Jorquera, M. J. (2012). Validation of the Spanish version of the Eating Disorders Diagnostic Scale. *Revista Médica de Chile*, *140*(12), 1562–1570. https://doi.org/10.4067/S0034- 98872012001200007

Stice, E., Fisher, M., & Martinez, E. (2004). Eating Disorder Diagnostic Scale: Additional evidence of reliability and validity. *Psychological Assessment*, *16*(1), 60–71. https://doi.org/10.1037/1040-3590.16.1.60

Stice, E., Telch, C. F., & Rizvi, S. L. (2000). Development and validation of the Eating Disorder Diagnostic Scale: A brief self-report measure of anorexia, bulimia, and binge-eating disorder. *Psychological Assessment*, *12*(2), 123–131. https://doi.org/10.1037/1040-3590.12.2.123

Stice, E., & Shaw, H. E. (2002). Role of body dissatisfaction in the onset and maintenance of eating pathology. *Journal of Psychosomatic Research*, *53*(5), 985–993. https://doi.org/10.1016/S0022-3999(02)00488-9 van Strien, T., Frijters, J. E. R., Bergers, G. P. A., & Defares, P. B. (1986). The Dutch Eating Behavior Questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behavior. *International Journal of Eating Disorders*, *5*(2), 295–315. https://doi.org/10.1002/1098-108X(198602)5:2<295::AID- EAT2260050209>3.0.CO;2-T

Vartanian, L. R. (2015). Development and validation of a brief version of the Stigmatizing Situations Inventory. *Obesity Science & Practice*, *1*(2), 119–125. https://doi.org/10.1002/osp4.11

Vartanian, L. R., & Novak, S. A. (2011). Internalized societal attitudes moderate the impact of weight stigma on avoidance of exercise. *Obesity*, *19*(4), 757–762. https://doi.org/10.1038/oby.2010.234

Wang, Y.-F., Ha, S., Zauszniewski, J. A., & Ross, R. (2018). Psychometric properties of the Chinese version of the Dutch Eating Behavior Questionnaire in a sample of Taiwanese parents. *Obesity Research & Clinical Practice*, *12*(1), 129–132. https://doi.org/10.1016/j.orcp.2017.11.005

Wardle, J. (1987). Eating style: A validation study of the Dutch Eating Behaviour Questionnaire in normal subjects and women with eating disorders. *Journal of Psychosomatic Research, 31*(2), 161–169. https://doi.org/10.1016/0022-3999(87)90072-9

Webb, J. B., & Hardin, A. S. (2016). An integrative affect regulation process model of internalized weight bias and intuitive eating in college women. *Appetite*, *102*, 60–69. https://doi.org/10.1016/j.appet.2016.02.024

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLOS ONE, 11*(3), e0152719. https://doi.org/10.1371/journal.pone.0152719

Table 1. Means, Standard Deviations, Internal Reliability, and Correlations Between Study Variables

Variable	1	2	3	4	5	6	7	8	9	10
1. BMI	–	.54***	.07	-.13*	-.10	-.13	-.18*	.01	-.03	.04
2. EWS[a]			.20***	.01	-.18***	-.26**	.01	.11*	.13*	.23***
3. IWS[b]				.43***	-.80***	-.73***	.37*	.41***	.51***	.70***
4. Anti-fat attitudes[c]					-.32***	-.24***	.21*	.30***	.23***	.42***
5. Body satisfaction[d]						.63***	-.30*	-.32***	-.40***	-.56***
6. Self-esteem[e]							-.17**	-.24***	-.30***	-.53***
7. Dietary restraint[f]								.13*	.19***	.34***
8. External eating[f]									.57***	.50***
9. Emotional eating[f]										.49***
10. Eating pathology[g]										–
Possible range	≥ 25.0	0–3	1–7	0–9	1–5	0–30	1–5	1–5	1–5	0–113
Mean	36.8	0.9	4.2	1.7	2.4	17.3	2.9	3.2	3.2	25.1
Standard deviation	8.9	0.6	1.4	1.7	0.9	6.0	0.8	0.7	1.0	12.2
Actual range	25.0–76.2	0–2.8	1–7	0–8	1–4.7	0–30	1–5	1.4–4.9	1–5	3–75
α		.96	.93	.86	.89	.91	.90	.87	.95	.81

Note. In addition to bivariate correlations, partial correlations controlling for BMI were also calculated. Partial correlations controlling for BMI had little effect on correlation coefficients, with the following exceptions: controlling for BMI, experienced weight stigma was significantly correlated with explicit anti-fat attitudes and dietary restraint (both $r = .11$, $p < .05$), and more strongly correlated with

external eating (r = .19, p < .001), and emotional eating (r = .23, p < .05). BMI = body mass index; EWS = experienced weight stigma; IWS = internalised weight stigma. EWS measured with Stigmatizing Situations Inventory; IWS measured with WBIS;
[a]Stigmatizing Situations Inventory; Cronbach's α for individual subscales ranged between .52 and .89; seven of the ten multi-item subscales had αs >.70. No reliability statistic was obtained for the subscale 'Being attacked' as this consists of a single item; [b]Weight Bias Internalization Scale; [c]Anti-Fat Attitudes Questionnaire; [d]MBSRQ-Appearance Evaluation scale; [e]Rosenberg Self-Esteem scale; [f]Dutch Eating Behavior Questionnaire subscale; [g]Eating Disorders Diagnostic Scale.
$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 2. Standardized factor loadings, proportion of variance associated with general and specific factors, scale reliabilities, explained common variance and percent uncontaminated correlations for bifactor model

Item numbers[a]	Valence[b]	Standardized factor loadings: POS	IWS	BS	SE	% Variance explained: % General	% Specific	% Residual
Weight Bias Internalization Scale								
1. As an overweight person, I feel that I am just as competent as anyone.[c]	–	-.484	.099†			23.4%	1.0%	75.6%
2. I am less attractive than most other people because of my weight.[d]	+	-.685	.232			47.0%	5.4%	47.6%
3. I feel anxious about being overweight because of what people might think of me.[e]	+	-.543	.436			29.5%	19.0%	51.5%
4. I wish I could drastically change my weight.[f]	+	-.757	.225			57.3%	5.1%	37.6%
5. Whenever I think a lot about being overweight, I feel depressed.[g]	+	-.718	.452			51.6%	20.4%	28.0%
6. I hate myself for being overweight.[cg]	+	-.786	.463			61.8%	21.4%	16.8%
7. My weight is a major way that I judge my value as a person.[c]	+	-.715	.383			51.1%	14.7%	34.2%
8. I don't feel that I deserve to have a really fulfilling social life as long as I'm overweight.[c]	+	-.580	.342			33.7%	11.7%	54.6%
9. I am OK being the weight that I am.[d]	–	-.768	.066†			59.0%	0.4%	40.6%
10. Because I'm overweight, I don't feel like my true self.[h]	+	-.699	.196			48.9%	3.8%	47.3%
11. Because of my weight, I don't understand how anyone attractive would want to date me.[cde]	+	-.706	.246			49.8%	6.1%	44.1%
Scale total						78.1%	15.1%	6.8%
Appearance Evaluation								
1. My body is sexually appealing. [self/other][i]	+	.728		.214		53.1%	4.6%	42.4%
2. I like my looks just the way they are. [self]	+	.850		.038†		72.3%	0.1%	27.5%
3. Most people would consider me good-looking. [other]	+	.474		.491		22.5%	24.1%	53.4%
4. I like the way I look without my clothes on. [self]	+	.817		.029†		66.7%	0.1%	33.2%
5. I like the way my clothes fit me. [self]	+	.778		.101†		60.6%	1.0%	38.4%
6. I dislike my physique. [self]	–	.584		.289		34.1%	8.4%	57.5%
7. I am physically unattractive. [self/other]	–	.699		.588		48.8%	34.6%	16.6%
Scale total						82.4%	8.4%	9.2%
Rosenberg Self-Esteem Scale								
1. On the whole, I am satisfied with myself.	+	.756			.220	57.2%	4.8%	38.0%
2. At times, I think I am no good at all.	–	.432			.537	18.6%	28.8%	52.5%
3. I feel that I have a number of good qualities.	+	.435			.471	18.9%	22.2%	58.9%
4. I am able to do things as well as most other people.	+	.406			.444	16.5%	19.7%	63.8%
5. I feel I do not have much to be proud of.	–	.447			.601	20.0%	36.2%	43.8%
6. I certainly feel useless at times.	–	.402			.605	16.2%	36.7%	47.2%
7. I feel that I am a person of worth, at least on an equal plane with others.	+	.500			.573	25.0%	32.9%	42.1%
8. I wish I could have more respect for myself.	–	.502			.372	25.2%	13.8%	61.0%
9. All in all, I am inclined to feel that I am a failure.	–	.546			.642	29.8%	41.2%	29.0%
10. I take a positive attitude toward myself.	+	.698			.438	48.7%	19.2%	32.2%
Scale total						47.8%	43.7%	8.5%
ω / ω-S		.966	.932	.908	.915			
ω-H / ω-HS		.864	.141	.076	.437			

Note. The copyrighted content of the MBSRQ Appearance Evaluation scale is provided with its author's permission. Research use of the licensed material requires permission of the author at www.body-images.com.
BS = body satisfaction factor; IWS = internalized weight stigma factor; POS = positive body-related self-judgment general factor; SE = global self-esteem factor.
[a]Numbers represent item numbers on each scale, as originally published. [b] The valence column indicates items that are scored 'as is' (positively valenced) and those that are reverse-scored (negatively valenced). [c]Item content appears to pertain to weight-related self-worth. [d]Item content appears to pertain to aspects of body image. [e]Item content appears to pertain to others' attitudes toward higher-weight individuals. [f]Item content appears to pertain to desire to change weight. [g]Item content appears to pertain to distress at weight status; [h]Item content appears to pertain to fat identity. [i] Self/other indicates whether the item refers to the respondent's own views (self) or their impressions of other people's views (other).
[†]Non-significant factor loading ($p \geq .05$).

Table 3. Indirect Effects of Experienced Weight Stigma on Disordered Eating via General and Construct-Specific Factors From Bifactor Analysis

Pathway	Coefficient	*SE*	95% CI	*p*
Total effect	.20	.07	[.05, .34]	.007
Indirect effects				
EWS → POS → Disordered eating	.06	.05	[-.02, .19]	.229
EWS → IWS → Disordered	.06	.04	[-.03, .14]	.134
EWS → BS → Disordered eating	-.05	.02	[-.10, -.00]	.029
EWS → SE → Disordered eating	-.05	.03	[-.10, .01]	.045

Note. Standardized coefficients shown. 95% confidence intervals based on 1000 bootstrapped samples. The direct effect is usually reported to provide an indication of the existence of other potential mechanisms that are not included in the model, and that, by definition, must be responsible for any residual significant effect of the predictor on the outcome when indirect effects have been taken into account. It is calculated as the difference between the total effect and the sum of the indirect effects, and is therefore not meaningful when the indirect effects have opposite valences as in the present case; hence this effect is not reported.
BS = body satisfaction factor; EWS = experienced weight stigma (Stigmatizing Situations Inventory); IWS = internalized weight stigma factor; POS = positive body-related self-judgment general factor.

Table 4. *Indirect effects of experienced weight stigma on disordered eating via bifactor and unidimensional internalized weight stigma factors*

Pathway	Coefficient	*SE*	95% CI	*p*
Bifactor IWS				
Total effect	.17	.07	[.04, .32]	.018
Direct effect	.07	.09	[-.08, .26]	.445
Indirect effect	.10	.05	[-.02, .20]	.057
Unidimensional WBIS				
Total effect	.20	.08	[.05, .34]	.008
Direct effect	.04	.07	[-.10, .16]	.574
Indirect effect	.16	.04	[.08, .25]	<.001

Note. Standardized coefficients shown. 95% confidence intervals based on 1000 bootstrapped samples. IWS = Internalized weight stigma factor; WBIS = Weight Bias Internalization Scale.

A bifactor analysis of the Weight Bias Internalization Scale: What are we really measuring?

Supplementary material.

1. Frequency of experienced weight stigma

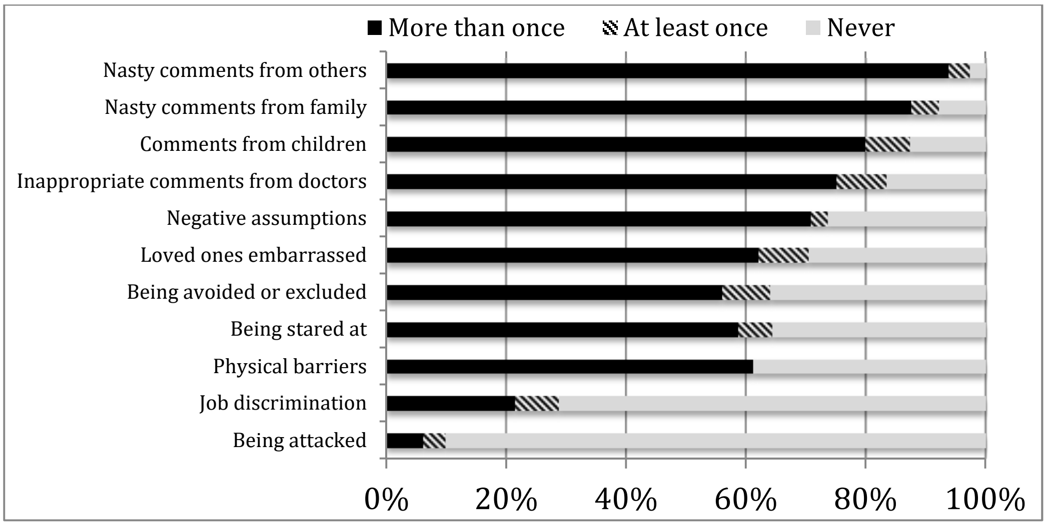

Figure S1. Frequency of experienced weight stigma across eleven domains.

2. Bifactor analysis of measures of internalized weight stigma, body satisfaction, and global self-esteem: Alternative models

Four competing models were tested using confirmatory factor analysis following the guidelines outlined in Brunner et al (2012b). The first model was a one-factor model in which a proposed negative self-judgment construct was regressed on all items of the WBIS, RSE, and Appearance Evaluation scales simultaneously (Figure S2).

The one-factor model represents a situation where individual differences in scores on items across all three scales are influenced solely by differences in a single common latent factor – negative self-judgment, plus item-level residual error, with no scale-specific contribution (Figure S2A). The residual error for each item, which represents variance not accounted for by the negative self-judgment factor, comprises item-specific variance and random measurement error.

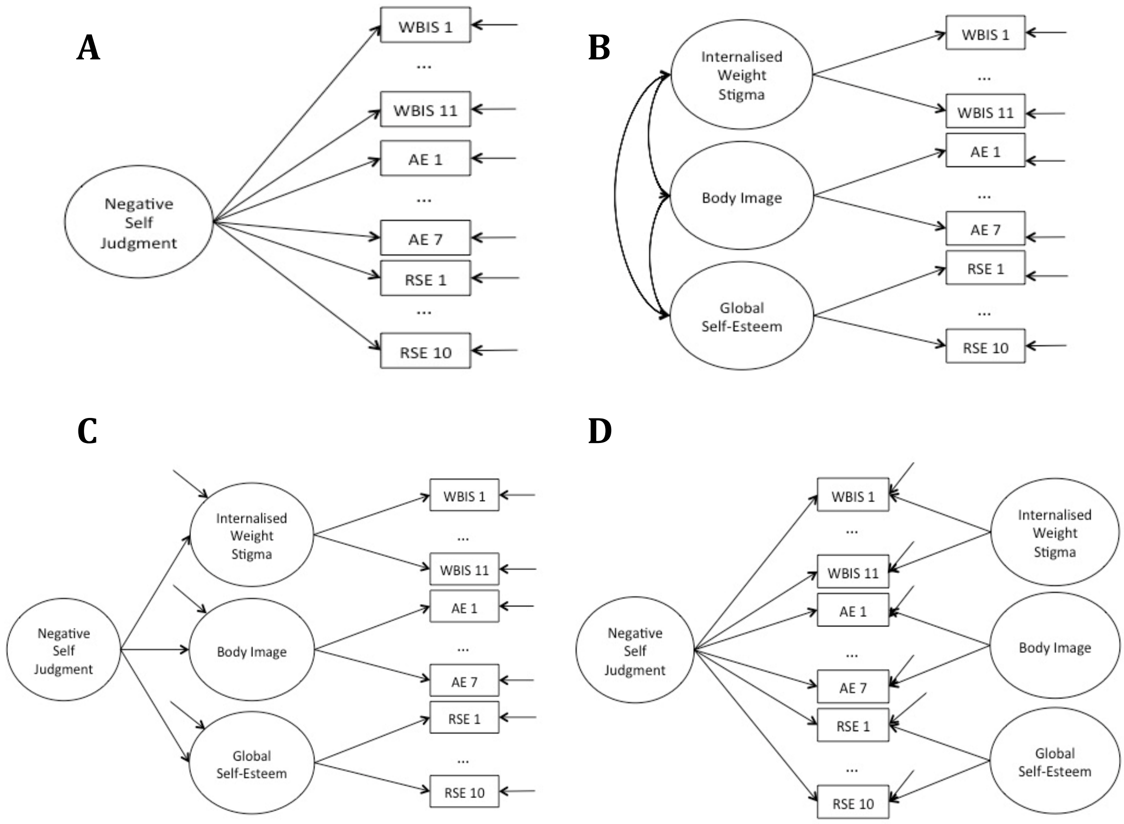

Figure S2. Schematic representation of the four alternative models. For clarity, only first and final scale items are shown for each measure, with remaining items represented by ellipses. A. One-factor model. B. First-order model. C. Second-order model. D. Bifactor model. AE = Appearance Evaluation scale; RSE = Rosenberg Self-Esteem Scale; WBIS = Weight Bias Internalization Scale.

The second model tested was a first-order CFA in which the items of each scale load only onto a scale-specific factor representing each individual construct. That is, each item is assumed to be influenced by a first-order, construct-specific factor – namely internalized weight stigma, body satisfaction, or global self-esteem, plus item-level residual error. This model does not include a general negative self-judgement factor and therefore does not account for shared variance between the three constructs. The three specific factors are allowed to correlate freely (Figure S2B).

The third model, a second-order CFA, allows the three construct-specific factors to load onto a higher-order negative self-judgment factor (Figure S2C). This model therefore includes both scale-specific variance and inter-scale shared variance attributable to the common latent factor. Variance of the first-order factors is now comprised partly of shared variance linked to the superordinate negative self-judgment construct and partly of scale-specific variance beyond the contribution of the higher-order factor. This model essentially imposes structural requirements onto the previously freely loading inter-factor correlations in the first-order model. That is, the second-order model constrains the correlations between the first-order factors to be zero and instead replaces them with factor loadings onto the second-order construct. Thus, the second-order general construct is theorized to explain the correlations (i.e., the common variance) between the scale-specific factors; the scales are correlated *because* they share a common cause. While the first-order and second-order models represent conceptually distinct models, they are statistically equivalent in this case, and therefore not discriminable (Rindskopf & Rose, 1988b). Note that the arrangement of factors in this hierarchical structure implies that the second-order factor influences the first-order variables, which, in turn, influence the individual scale items. That is, the second-order negative self-judgment factor has an *indirect* effect on item scores, via its influence on the first-order constructs. The influence of first-order constructs on the individual item scores is likewise partially accounted for by the influence of the second-order negative self-judgment factor on the first-order factors.

The final model tested was a bifactor model, in which both a general factor and three construct-specific factors were included – note all four factors are now first-order factors.

Distinguishing the bifactor model from the second-order model, the bifactor model allows for the commonality among the scale items explained by the general negative self-judgment factor to be partialled out, with the scale-specific factors representing only the unique shared variance among the respective scale items on each scale, beyond that accounted for by the common underlying factor (Figure S2D). This is achieved by specifying the inter-factor correlations to equal zero, thus forcing the common variance in the model (i.e., excluding the item-specific residual variance) to be split between four orthogonal factors. Note, in the bifactor model, variance in item scores are explained by the *direct* influence of the general factor, the influence of specific constructs, independent of the general factor, plus item-level residual variance not accounted for by either the underlying negative self-judgment factors or the scale-specific factors.

Model fit

Goodness of model fit to the observed data was assessed using the same two-index reporting method outlined above. However, CFI tends to decline with increasing number of indicators in the model (Kenny & McCoach, 2003). In the present analysis, the maximum number of variables per factor was 28, thus, following Chen (2012), a less stringent cut-off of .90 was used for the CFI to indicate goodness of fit.

Comparison of model fit (i.e., selection of superior models) was assessed using fit indices (CFI, SRMR) plus χ^2 difference tests for nested models (one-factor versus bifactor, first-order versus second-order, and second-order versus bifactor models). A reduction in χ^2 greater than the critical value for the change in degrees of freedom indicates a significantly better model fit. Information criteria (Akaike Information Criterion [AIC], Bayes Information Criterion [BIC], and sample size-adjusted Bayes Information Criterion [SSA-BIC]) were used to compare non-nested models (one-factor versus first-order and first-order versus bifactor models).[10] No specific guidelines are available for comparing information criteria values between models, other than lower values indicate improvement in model fit; however, differences of greater than 10 can be considered to correspond to "very strong evidence" of superior model fit (Burnham & Anderson, 2004; Raftery, 1995).

Results

One-factor model

Model fit statistics for the four models are displayed in Table S2. The one-factor model was only a moderate fit for the data. Additionally, reliability of the one-factor negative self-judgment construct was only .36. Thus, the observed association between the items was not adequately explained by the influence of a single negative self-judgment factor.

Table S2. Model Fit Indices and Information Criteria for Alternative Model Formulations

Model	χ^2	*df*	CFI	SR	AIC	BIC	SSA-
One-factor model	2041	35	.767	.078	28685	29016	28749
First-order factor model	1357	34	.860	.065	28007	28350	28074
Second-order factor	1357	34	.860	.065	28007	28350	28074
Bifactor model	1049	32	.900	.062	27749	28190	27835

Note. All χ^2 $p < .001$. AIC = Akaike's Information Criterion; BIC = Bayes Information Criterion; CFI = Comparative Fit Index; *df* = degrees of freedom; SRMR = standardized root mean squared residual; SSA-BIC = sample-size adjusted Bayes Information Criterion.

First-order factor model

The first-order factor model was a markedly better fit for the data than the one-factor model, having lower AIC (difference = 678), lower BIC (difference = 666), and lower sample-size adjusted BIC (difference = 675). Additionally, the model fit indices CFI and SRMR were improved compared with the one-factor model.

As established in preliminary CFA of individual constructs, the three constructs were well specified and model-based reliabilities for the three factors were high: IWS ω = .93, body satisfaction (BI) ω = .89, and global self-esteem (SE) ω = .91. However, inter-factor

[10] Information criteria can also be compared for nested models.

correlations were extremely high (IWS with BI and SE, $r = -.88$ and -.78, respectively; BI with SE, $r = .70$; all $p < .001$), suggesting a large amount of common variance between the three constructs.

Second-order factor model

As noted above, this model is statistically identical to the first-order model; model fit indices are therefore the same for both models, as are, by definition, the scale reliabilities of the first-order factors. However, the scale reliability for the second-order negative self-judgment factor was only moderate ($\omega = .697$), suggesting that this conceptualisation of the model only moderately explains the total variance in the observed data.

Bifactor model

The bifactor model was a good fit for the data, and significantly better than the second-order model ($\Delta\chi^2(25) = 308$, $p < .0001$). Model fit indices and information criteria also indicated superior model fit. However, model fit of the bifactor model was also significantly superior to the one-factor model ($\Delta\chi^2(28) = 902$, $p < .0001$), indicating that the individual scales do explain a significant proportion of individual scale variance, beyond that accounted for by an underlying general factor.