Psychological Assessment

© 2020 American Psychological Association
ISSN: 1040-3590

2020, Vol. 32, No. 6, 553–567
http://dx.doi.org/10.1037/pas0000810

Do Differences Between Individuals Who Are Healthy Weight or Overweight on Self-Report Measures of Disinhibited Eating and Restrained Eating Reflect Reality or Item "Bias"?

Kelsie T. Forbush
University of Kansas

Q. Chelsea Song and Louis Tay
Purdue University

Sara R. Gould
Children's Mercy Kansas City, Kansas City, Missouri, and
University of Missouri-Kansas City School of Medicine

Danielle A. N. Chapa, Christopher C. Cushing, and
Lauren T. Ptomey
University of Kansas

In light of increasing rates of overweight and obesity worldwide, there is a critical need for accurate self-report measures of disinhibited and restrained eating behaviors across the weight spectrum. Item response theory was used to determine whether differences in disinhibited and restrained eating between healthy weight and overweight or obese individuals were due to item bias (i.e., differential item functioning). Study 1 participants were healthy weight ($n = 510$) or overweight or obese ($n = 304$) adults recruited from the community. Study 2 participants were healthy weight ($n = 778$) or overweight or obese ($n = 320$) college students. Study 1 participants completed the Eating Disorder Examination–Questionnaire (EDE-Q), Eating Disorder Inventory–3, Dutch Eating Behaviors Questionnaire, Restraint Scale, and Three-Factor Eating Questionnaire. Study 2 participants completed the Eating Pathology Symptoms Inventory (EPSI). Items on the Restraint Scale demonstrated the most evidence for bias (60% of items), whereas the majority of other scales demonstrated low to moderate levels of item bias (17–38% of items). However, EDE-Q Restraint and EPSI Binge Eating, Cognitive Restraint, Excessive Exercise, Muscle Building, and Negative Attitudes Toward Obesity scales did not show any evidence of differential item functioning among weight groups. Participants with the same level of disordered eating responded differently to certain eating disorder self-report items due to weight-bias, rather than true between-groups differences. Nevertheless, EDE-Q Restraint, EPSI Cognitive Restraint, and EPSI Binge Eating did not exhibit any evidence of bias and are ideal for assessing restrained and disinhibited eating across the weight spectrum in both research and clinical settings.

Public Significance Statement

In light of increasing rates of overweight and obesity worldwide and the comorbidity of high body mass index and eating disorders, there is a critical need for accurate self-report measures of disinhibited eating (overeating) and restrained eating (dieting) across the weight spectrum, particularly given that most disordered-eating measures were created in healthy weight samples. Our results showed that most scales performed well across the weight spectrum; however, participants with the same level of disordered eating responded differently to certain Restraint Scale items due to weight-bias, rather than true differences in restrained eating.

Keywords: disordered eating, binge eating, restraint theory, item bias, item response theory

This article was published Online First March 19, 2020.

Kelsie T. Forbush, Department of Psychology, University of Kansas; Q. Chelsea Song and Louis Tay, Department of Psychological Sciences, Purdue University; Sara R. Gould, Department of Developmental and Behavioral Health, Children's Mercy Kansas City, Kansas City, Missouri, and Department of Pediatrics, University of Missouri-Kansas City School of Medicine; Danielle A. N. Chapa and Christopher C. Cushing, Department of Psychology, University of Kansas; Lauren T. Ptomey, Department of Internal Medicine, Division of Physical Activity and Weight Management, University of Kansas.

This research was supported by grants from the American Psychological Association, American Psychological Foundation, and Academy for Eating Disorders to Kelsie T. Forbush.

Correspondence concerning this article should be addressed to Kelsie T. Forbush, Department of Psychology, University of Kansas, 1415 Jayhawk Boulevard, Lawrence, KS 66045. E-mail: kforbush@ku.edu

Eating disorders (EDs) and overweight are major public health issues. EDs are associated with psychosocial disability and lower quality of life (Ágh et al., 2016), as well as substantial medical (Mitchell, 2016; Westmoreland, Krantz, & Mehler, 2016) and psychiatric morbidity (Hudson, Hiripi, Pope, & Kessler, 2007; Keski-Rahkonen & Mustelin, 2016). Overweight and obesity are significant health problems affecting approximately one third of children and adolescents in the United States who will likely continue to have obesity as adults, and will cost approximately $147 billion annually (Finkelstein, Trogdon, Cohen, & Dietz, 2009; Ogden, Carroll, Curtin, Lamb, & Flegal, 2010). Overweight and obesity can result in increased risk for all-cause mortality and engenders risk for numerous medical issues, including Type II diabetes mellitus, hyperlipidemia, hypertension, and renal disease (Crawford et al., 2010; Hall et al., 2014). Although eating and weight disorders are serious when they occur alone, research shows that EDs often co-occur with overweight or obesity, particularly in EDs that involve binge eating (Hay, Girosi, & Mond, 2015). Presence of lifetime obesity in persons with an ED is also associated with a worse prognosis and greater clinical severity (Villarejo et al., 2012).

Given the interplay between ED behaviors and weight management issues, it is increasingly important to assess EDs and overweight together, rather than separately. For example, preoperative evaluation for bariatric surgery typically includes assessment of ED behaviors because the presence of binge eating may be contraindicated, or lead to worse postoperative outcomes. Persons who engage in binge eating and undergo bariatric surgery tend to lose less weight and experience more weight regain (for a review, see Meany, Conceição, & Mitchell, 2014), as well as experience increased depression and lower quality-of-life (Colles, Dixon, & O'Brien, 2008). Binge eating also predicts poor weight loss and more weight regain in randomized behavioral weight loss trials (Masheb et al., 2015; Pacanowski, Senso, Oriogun, Crain, & Sherwood, 2014).

Finally, accurate measurement of restrictive behaviors is important for higher weight individuals. For example, atypical anorexia nervosa—defined as meeting full criteria for anorexia nervosa, including a significant weight loss, but without presenting as underweight—is increasing in prevalence (or identification). Whitelaw, Gilbertson, Lee, and Sawyer (2014) demonstrated a fivefold increase in hospitalizations in Australia due to atypical anorexia nervosa over the past 6 years. Thirty-four percent of patients admitted to 14 U.S. adolescent medicine ED programs had atypical anorexia nervosa (Forman et al., 2014), and 26% of patients presenting to an Australian outpatient center with anorexia nervosa had atypical anorexia nervosa (Sawyer, Whitelaw, Le Grange, Yeo, & Hughes, 2016). Typically, these individuals were considered overweight or obese prior to losing weight, whereas this was rare for those with "traditional" (i.e., low weight) anorexia nervosa (Sawyer et al., 2016; Whitelaw et al., 2014). Results demonstrated that people with atypical anorexia nervosa required similar rates of hospitalization (Sawyer et al., 2016; Whitelaw et al., 2014), types and rates of medical complications (Sawyer et al., 2016; Whitelaw et al., 2014), and amount of weight loss (Whitelaw et al., 2014) as people with "traditional" anorexia nervosa. Thus, as noted by Whitelaw et al. (2014), "It is interesting to contemplate the validity of psychometric measures when patients are not underweight. . ." (p. e736).

Although many clinicians and researchers administer measures of ED behaviors to persons with overweight or obesity, certain popular ED assessments—for example, the Eating Disorder Examination (Cooper & Fairburn, 1987) and Eating Disorder Inventory (Garner, Olmstead, & Polivy, 1983)—were developed and validated in persons who were healthy weight and may not provide accurate assessment across the weight spectrum. Given that restrained eating (dieting) and disinhibited eating (eating for reasons other than hunger) are common across the weight spectrum (Sares-Jäske, Knekt, Männistö, Lindfors, & Heliövaara, 2019; Takakura et al., 2019; Wang et al., 2019), it is important to understand the psychometric properties of restrained and disinhibited eating assessments across weight-status boundaries. The current article was the first to test whether individual items included in self-report measures of restrained and disinhibited eating are valid across weight-status boundaries. Below we discuss past research and theoretical models of dietary restraint and disinhibited eating across the weight spectrum that highlight the potential impact of differential item functioning (DIF) on substantive research findings and clinical practice.

Restraint Theory

Restraint theory was developed by Herman and Polivy (1975) to describe dietary restraint. Dietary restraint is defined as cognitive efforts to avoid or limit food intake to lose weight, independent of whether attempts to "diet" actually result in reduced food consumption. In other words, restraint theory is focused on cognitive efforts to restrict eating, rather than whether or not calorie reduction was achieved. Restraint theory proposed that individuals with the highest level of dietary restraint would be most susceptible to lapses in cognitive control over eating due to increased physiological and cognitive hunger drives, abstinence-violation effects, and/or depletion of self-regulation resources to monitor calorie intake (Herman & Mack, 1975; Herman & Polivy, 1975, 1983). Dieting shifts control of eating from physiologically based hunger cues to cognitive control. Stimuli (internal or external), such as the sight or smell of food, that disrupt dietary self-control would, therefore, result in overeating or binge eating because persons with elevated dietary restraint have more hunger due to reduced caloric intake.

Early research that tested restraint theory used designs that assessed the effects of a diet-disrupting manipulation on subsequent intake. Results showed that restrained eaters who consumed a preload milkshake ate more during a subsequent ice cream taste test. Unrestrained eaters showed the opposite pattern by eating less ice cream if they had already consumed a preload milkshake. In other words, when restrained eaters experienced a dietary disruption, they ate more, whereas unrestrained eaters regulated their eating by reducing caloric intake.

Research Support for Restraint Theory

Theoretical models posited that prolonged restraint led to overconsumption which, in turn, increased risk for the development of EDs (Fairburn, 2008) and weight gain (Herman & Polivy, 1975). Consistent with theoretical predictions, longitudinal studies showed high levels of restraint were associated with increased binge-eating behavior and predicted maintenance of bulimia-nervosa symptoms (Stice, 2001; Stice & Agras, 1998). Furthermore, Sares-Jäske et al. (2019) demonstrated at an 11-year follow-up that those who self-identified as dieters (i.e., restrained eaters) experienced a greater increase in body mass index (BMI) over the study period than those who did not endorse dieting behavior. Studies that experimentally induced short-term calorie deficits led to increased food intake in people without EDs and in persons with bulimia nervosa or binge eating disorder (BED; Agras & Telch, 1998; Hetherington, Stoner, Andersen, & Rolls, 2000). However, other experimental studies of short-term caloric deprivation contradicted the theoretical model that underlies restraint theory (Lowe, 1992, 1994; Ruderman & Christensen, 1983). Ruderman and Christensen (1983) showed that among overweight participants, unrestrained eaters ate more than restrained eaters, whereas the reverse was true in healthy weight participants. Restrained eaters, who had a healthy weight, consumed more food following a preload snack, suggesting that weight status moderates the effect of dietary restraint on overeating. Moreover, findings from randomized weight-loss trials found that caloric deprivation and greater restraint led to reduced disinhibited eating (Epstein, Paluch, Saelens, Ernst, & Wilfley, 2001; Presnell & Stice, 2003).

In summary, although several studies supported the original predictions of restraint theory, many other studies produced results that were inconsistent with it. Reasons for discrepancies could be (a) differential effects of dieting in people with and without obesity that result from different reserves of adipose tissue to withstand acute and prolonged dieting, (b) third variable effects in which overconsumption leads to both self-reported increases in dietary restraint and the onset of bulimia nervosa or BED, or (c) psychometric issues. Consistent with the possibility of psychometric issues, past research suggested that associations between disruption of dietary goals and restrained eating depended on the instrument used. For example, the "disinhibition effect," which refers to overeating following a preload, emerges only when the Restraint Scale is used, but not when other measures of restrained eating are used (e.g., Three-Factor Eating Questionnaire and Dutch Eating Behavior Questionnaire). We therefore posit that the specificity of the "disinhibition effect" to the Restraint Scale may be because the scale combines several different constructs within the same instrument. For example, it measures dieting, weight fluctuations (e.g., the number of pounds a person was over their desired weight, when they were at their maximum weight), guilt after eating, and eating sensibly when with others but "splurging" when alone. As Ouwens, van Strien, and van der Staak (2003b) argued, the "disinhibition effect" that emerges with the Restraint Scale could be because weight fluctuations and past overweight status are more important for predicting overeating than dietary restraint. Given that the scale only predicts overeating in dieters who are already highly prone to overeating, it is possible that it has DIF across the weight spectrum.

Disinhibited and Binge Eating

Disinhibited eating refers to nonhomeostatic eating in response to internal (e.g., sadness or boredom) or external cues (e.g., the sight, smell, or taste of food), rather than physiological hunger (Stunkard & Messick, 1985). Binge eating is a closely related construct that is defined by eating a large amount of food and feeling a subjective sense of loss of control during the eating episode (APA, 2013). Measures of disinhibited and binge eating have several strengths, including significant positive correlations with BMI; obesity; consumption of foods high in sugar, sodium, and salt (for a review, see Bryant, King, & Blundell, 2008); and future weight gain (Hays & Roberts, 2008). However, it is important to note that disinhibited and binge eating can occur across the weight spectrum, including among individuals who have significantly low body weight (Takakura et al., 2019; Wang et al., 2019). Disinhibited eating is a strong predictor of food intake in preload studies, with higher scores on measures of disinhibition predicting greater consumption during laboratory feeding studies irrespective of participants' level of restrained eating (Ouwens, van Strien, & van der Staak, 2003a; Van Strien, Cleven, & Schippers, 2000; Westenhoefer, Broeckmann, Münch, & Pudel, 1994). Scores on measures of disinhibition or binge eating predict weight loss and weight regain during randomized controlled weight reduction programs (Cuntz, Leibbrand, Ehrig, Shaw, & Fichter, 2001; Fogelholm, Kukkonen-Harjula, & Oja, 1999; McGuire, Wing, Klem, Lang, & Hill, 1999; Teixeira et al., 2010). Persons with BED or bulimia nervosa (disorders that, by definition, involve binge eating) score significantly higher than healthy control participants on self-report measures of disinhibition (Ardovini, Caputo, Todisco, & Dalle Grave, 1999; Brown, Bryant, Naslund, King, & Blundell, 2006). Disinhibition is also elevated in persons with binge-purge anorexia nervosa compared to restricting anorexia nervosa (Kiezebrink, Campbell, Mann, & Blundell, 2009). Disinhibition scores in persons with EDs are associated with poor psychological health, including increased levels of depression, anxiety, and body dissatisfaction (Brown et al., 2006). Moreover, given that disinhibition scores do not change appreciably after ED treatment, Bryant et al. (2008) argued that disinhibition may represent a transdiagnostic maintenance process that underlies ED psychopathology.

Limitations of Measures of Disinhibited and Binge Eating

Although there are fewer inconsistent findings for measures of disinhibited eating (vs. restrained eating), certain measures were developed and validated in comparatively small samples that did not include participants with overweight or obesity. For example, the Eating Disorder Inventory, which includes scales that assess restraint (Drive for Thinness) and overeating (Bulimia) was developed in 113 patients with anorexia nervosa and 577 female college students. Although the measure has subsequently been validated in persons with overweight or obesity, initial item selection was focused on restrictive eating pathology in an underweight sample. The Eating Disorder Examination was developed and validated in 100 patients with anorexia nervosa or bulimia nervosa, and did not include obese participants without EDs. Given that the Eating Disorders Inventory and Eating Disorder Examination were developed prior to the recognition of BED as a diagnostic category and before the current obesity "epidemic," it is possible that assessment instruments contain weight bias.

IRT-Based Studies of Restrained and Disinhibited Eating

The majority of restrained and disinhibited eating scales were developed approximately 30 years ago, when item response theory

(IRT) computational packages were less readily available. As a result, few studies used IRT to evaluate weight-related item bias for ED scales. We identified two studies that used Rasch modeling of the Three-Factor Eating Questionnaire (Herrmann et al., 2014; Ismail, Noh, Ismail, & Tamil, 2015) and one study that used Rasch modeling to derive a short-form of the Eating Disorder Examination–Questionnaire (EDE-Q; Gideon et al., 2016). Of these past IRT studies, only Herrmann et al. (2014) compared restrained and disinhibited eating item responses across the weight spectrum (see description below).

Herrmann et al. (2014) used Rasch modeling of the Three-Factor Eating Questionnaire in a sample of 395 overweight or obese adults enrolled in a behavioral weight-loss trial. The authors evaluated the dimensionality, item-model fit, and item difficulty parameters for each Three-Factor Eating Questionnaire scale. Nineteen of 21 Restraint subscale items (difficulty ranged from −2.51 to 2.03 logits), 12 of 16 Disinhibition subscale items (difficulty −1.65 to 1.79 logits), and 12 of 14 Hunger subscale items (difficulty −2.41 to 1.40 logits) were identified as "good fitting" items. The separation index (an index of internal consistency) for item measures was 8.67 logits for the Restraint subscale, 6.58 logits for the Disinhibition subscale, and 7.23 logits for the Hunger subscale with a reliability of .98 to .99. The authors noted certain items on the Three-Factor Eating Questionnaire Restraint scale were harder to endorse in overweight and obese persons. Results from Herrmann et al. (2014) indicated that the Three-Factor Eating Questionnaire Restraint items may underestimate restrained eating in people with overweight or obesity.

Current Study

The current study is significant as the first to apply IRT to test whether item bias exists across weight categories for common measures of restrained, disinhibited, and binge eating. Most IRT studies in the field of EDs used simpler IRT models (e.g., the two-parameter model or Rasch models) and did not focus on DIF across weight groups (K. C. Allison et al., 2008; Gideon et al., 2016; Ismail et al., 2015). Thus, it is unclear from past research whether restrained and disinhibited eating questionnaire items over- or underestimate ED psychopathology in persons with overweight or obesity.

Past research theorized that associations between disruption of dietary goals and restrained eating depended on the instrument used. For example, the "disinhibition effect," which refers to overeating following a preload, emerges only when the Restraint Scale is used, but not when other measures of restrained eating are used (e.g., Three-Factor Eating Questionnaire and Dutch Eating Behavior Questionnaire; Ouwens et al., 2003b). Thus, the specificity of the "disinhibition effect" that emerges from the Restraint Scale may be because the Restraint Scale includes content that inadvertently measures overweight status, rather than true dietary restraint. For example, the Restraint Scale contains items that focus on past weight fluctuations (e.g., the number of pounds a person was over their desired weight, when they were at their maximum weight). We, therefore, hypothesized that items measuring weight fluctuations on the Restraint Scale would have lower difficulty parameters for people who are overweight or obese by virtue of their weight status, rather than true differences in dietary restraint.

Results from Herrmann et al. (2014) indicated the possibility that Three-Factor Eating Questionnaire Restraint items may be harder to endorse (have higher difficulty) for persons with overweight or obesity. Thus, we further hypothesized that Three-Factor Eating Questionnaire Restraint items would underestimate dietary restraint in persons with overweight or obesity. Finally, we posited that measures that included persons with overweight or obesity in their initial questionnaire development (Eating Pathology Symptoms Inventory, Dutch Eating Behaviors Questionnaire, and Three-Factor Eating Questionnaire) would show evidence for the least amount of DIF compared to other measures of similar constructs.

Method

Participants and Procedures

We carried out two studies to test whether DIF existed for common measures of restraint and disinhibited (or binge) eating. Procedures were approved by the local Institutional Review Board and participants provided online informed consent prior to engaging in study procedures. Permission was obtained by copyright holders to administer study measures online via Qualtrics. At the time the surveys were administered, Qualtrics did not provide an option for random presentation of questionnaires. Thus, questionnaires were presented in a fixed order across all participants. We used the Centers for Disease Control and Prevention definitions of healthy, overweight, and obesity (https://www.cdc.gov/healthyweight/assessing/bmi/index.html). The Centers for Disease Control and Prevention–defined weight groups are used by the American Medical Association, which classifies obesity as a medical disease. Study 1 participants were adults recruited from a Midwestern community ($N = 814$) who had a healthy BMI ($n = 510$) or had an overweight or obese BMI ($n = 304$). Study 1 participants were recruited from fliers, bus and newspaper ads, and a mass e-mail sent to faculty and staff at a large Midwestern university. Participants were included if they were 18 or older and did not have active psychosis. Study 2 participants ($N = 1{,}098$) were Midwestern college students recruited from group testing as part of their introductory psychology course. Study 2 participants had a healthy BMI ($n = 778$) or a BMI that was within the overweight or obese range ($n = 320$). Study 1 and 2 participants were excluded if they were <18 years of age, were a non-native English speaker, or were underweight. Table 1 presents demographic data.

Both studies included data-quality checks to identify participants with inconsistent or invalid responses. For example, participants who initially reported that they never purged in their life, but later endorsed purging in the past month were removed from the dataset. Next, we examined responses to identify outliers using visual inspection of the data. In Study 1, some participants reported extreme weight gains or losses on the Restraint Scale. Participants with outlier responses on the Restraint Scale were contacted by the Principal Investigator to clarify their answers. Based on the participant's response, the item(s) were either updated for accuracy or kept the same. If the participant with outlier data on the Restraint Scale did not respond to our query, the item(s) were recoded as missing. Missing data were later imputed on a measure-by-measure basis using maximum likelihood multi-

Table 1
Study Demographics

Variable	Healthy weight	Overweight/obese	F or χ^2	p
Study 1				
Age	26.56 (12.32)	32.39 (14.06)	−5.99	<.001
Body mass index	22.02 (1.65)	30.00 (5.40)	−25.08	<.001
% Female	63.33%	39.47%	42.76	<.001
% Caucasian	90.78%	90.13%	.03	.854
Study 2				
Age	19.16 (1.33)	19.49 (1.87)	−2.89	.004
Body mass index	21.78 (1.68)	29.30 (4.39)	−29.78	<.001
% Female	57.97%	46.56%	11.45	.001
% Caucasian	87.15%	81.25%	5.87	.015

Note. We used multivariate analysis of variance to test for differences between groups for continuous variables, age and body mass index. Chi-Square values are provided for categorical variables.

ple imputation—averaged over 11 imputations—if 10% or less of their total responses for the questionnaire were missing.

Measures

All subscales chosen for this study were designed to represent homogeneous, unidimensional content. Study 1 measures included scales from the EDE-Q, Eating Disorder Inventory–3, Dutch Eating Behaviors Questionnaire, Restraint Scale, and Three-Factor Eating Questionnaire. The Eating Pathology Symptoms Inventory was administered to Study 2 participants. A total of 158 items were evaluated across both studies. Coefficient alpha, IRT reliability, means, and standard deviations are reported in Table 2. Below we

Table 2
Internal Consistency and Item Response Theory (IRT) Reliability for Study Measures and Mean Differences

Scale name	Healthy weight M (SD)	Overweight/obese	F	p	Cronbach's α	IRT reliability
Study 1 measures						
DEBQ						
Restrained Eating	2.41 (0.90)	2.64 (0.81)	12.41	<.01	.93	.92
Emotional Eating	2.00 (0.88)	2.21 (0.93)	9.25	<.01	.97	.95
External Eating	2.89 (0.60)	2.97 (0.57)	3.15	.08	.83	.91
TFEQ						
Cognitive Restraint	8.37 (5.19)	8.72 (5.05)	.81	.37	.87	.86
Disinhibition	4.81 (3.26)	6.84 (3.74)	60.28	<.01	.81	.80
Hunger	4.93 (3.36)	6.02 (3.56)	17.56	<.01	.81	.75
RS	11.30 (5.69)	15.15 (5.31)	79.23	<.01	.78	.84
EDE-Q						
Restraint	1.13 (0.57)	1.52 (0.53)	7.50	.01	.82	.81
EDI-3						
Drive for Thinness	4.98 (5.84)	6.39 (5.90)	10.18	<.01	.88	.83
Bulimia	2.69 (4.19)	3.38 (3.97)	4.87	.03	.84	.78
Study 2 measures						
EPSI						
Body Dissatisfaction	16.50 (7.03)	20.45 (7.46)	68.14	<.01	.92	.90
Binge Eating	18.53 (5.36)	19.62 (5.51)	9.14	<.01	.86	.85
Cognitive Restraint	7.47 (2.91)	8.13 (2.93)	11.25	<.01	.81	.81
Purging	6.52 (1.75)	6.97 (2.26)	12.12	<.01	.83	.66
Restricting	10.91 (4.30)	10.75 (4.13)	.32	.57	.85	.81
Excessive Exercise	12.47 (5.24)	12.53 (4.97)	.03	.85	.89	.84
Negative Attitudes Toward Obesity	15.07 (4.66)	14.27 (4.68)	6.55	.01	.76	.81
Muscle Building	8.02 (3.32)	8.79 (3.71)	11.33	<.01	.89	.88

Note. DEBQ = Dutch Eating Behavior Questionnaire; TFEQ = Three-Factor Eating Questionnaire; RS = Restraint Scale; EDE-Q = Eating Disorder Examination–Questionnaire; EDI-3 = Eating Disorder Inventory–3; EPSI = Eating Pathology Symptoms Inventory; AIC = average interitem correlation. The mean and standard deviations in this table were calculated using a multivariate analysis of variance that did not control for any variables. We also reran analyses controlling for demographic differences (data available upon request). For Study 1 measures, a multivariate analysis of covariance was used to test group differences controlling for age, body mass index, and gender. Only DEBQ Restrained Eating and EDI-3 Bulimia were significantly different between groups. For Study 2 measures, an analysis of variance was used to test group differences controlling for age, body mass index, gender, and race (Caucasian or non-Caucasian). Only Body Dissatisfaction and Muscle Building were significantly different between groups. The scoring varied across scales (i.e., some were averaged, some were summed based on scoring instructions provided for each measure), which led to different ranges for mean values across measures.

describe relevant features and psychometric properties of our study measures:

Eating Disorder Examination–Questionnaire. The EDE-Q (Fairburn & Beglin, 1994) is a 28-item self-report questionnaire that assesses ED behaviors and cognitions over the past 28 days. The EDE-Q contains four subscales, including Restraint, Eating Concern, Shape Concern, and Weight Concern. The EDE-Q contains three behavioral items that assess the frequency of binge eating and three behavioral items that measure the frequency of compensatory behavior. For the current study, only the five-item EDE-Q Restraint subscale was analyzed. Past studies show that the EDE-Q demonstrates evidence for strong internal consistency and convergent validity with other measures of ED symptoms (Berg, Peterson, Frazier, & Crow, 2012).

Eating Disorder Inventory–3. The Eating Disorders Inventory–3 (Garner, 2004) is a 91-item measure that is organized into 12 scales. For the purposes of this study, we only analyzed the six-item Drive for Thinness and eight-item Bulimia scales. Studies support the internal consistency, convergent validity, and criterion-related validity for distinguishing among persons with and without EDs (Salbach-Andrae et al., 2010).

Eating Pathology Symptoms Inventory. The Eating Pathology Symptoms Inventory (Forbush et al., 2013) is a 45-item measure that includes eight scales, which are named Body Dissatisfaction (seven items), Binge Eating (eight items), Excessive Exercise (five items), Purging (six items), Cognitive Restraint (three items), Restricting (six items), Muscle Building (five items), and Negative Attitudes Toward Obesity (five items). Eating Pathology Symptoms Inventory scores show evidence for strong construct and criterion-related validity (Forbush, Wildes, & Hunt, 2014; Forbush et al., 2013; Tang, Forbush, & Lui, 2015) and a robust factor structure (Coniglio et al., 2018).

Dutch Eating Behavior Questionnaire. The Dutch Eating Behavior Questionnaire (Van Strien, Frijters, Bergers, & Defares, 1986) has 33 items that assess: Restrained Eating (10 items), Emotional Eating (13 items), and External Eating (10 items). Dutch Eating Behavior Questionnaire scores show convergent validity with Eating Pathology Symptoms Inventory, Restraint Scale, and Three-Factor Eating Questionnaire scores (D. B. Allison, Kalinsky, & Gorman, 1992; Bohrer, Forbush, & Hunt, 2015). Dutch Eating Behavior Questionnaire scores show evidence for test–retest reliability and internal consistency (D. B. Allison et al., 1992; Forbush, Hilderbrand, Bohrer, & Chapa, 2019).

Revised Restraint Scale. The Restraint Scale (Polivy, Heatherton, & Herman, 1988) was developed to test restraint theory. The Restraint Scale contains 10 items that were designed to differentiate the eating behaviors of dieters (vs.) nondieters. Past studies show Restraint Scale scores have high internal consistency reliabilities (Klem, Klesges, Bene, & Mellon, 1990) and test–retest reliability (D. B. Allison et al., 1992), although internal consistency is slightly worse among overweight and obese persons (D. B. Allison et al., 1992; Bohrer et al., 2015). Past studies found that the Restraint Scale scores have low-to-modest discriminant validity from social desirability, but strong convergent validity with the Dutch Eating Behavior Questionnaire and Three-Factor Eating Questionnaire (D. B. Allison et al., 1992).

Three-Factor Eating Questionnaire. The Three-Factor Eating Questionnaire (Stunkard & Messick, 1985) is a 51-item measure that includes scales that assess Cognitive Restraint Over Eating (21 items), Disinhibition (16 items), and Hunger (14 items). The Three-Factor Eating Questionnaire is commonly used to assess psychological contraindications for bariatric surgery (Marek, Heinberg, Lavery, Merrell Rish, & Ashton, 2016), as well as to measure self-reported restrained and disinhibited eating in adult populations. Studies support the internal consistency and test–retest reliability of Three-Factor Eating Questionnaire scores (D. B. Allison et al., 1992; Forbush et al., 2019). The factor structure of the Three-Factor Eating Questionnaire Restraint scale is replicable in participants with obesity (Karlsson, Persson, Sjöström, & Sullivan, 2000).

Statistical Analyses

We used Item response theory differential item functioning with covariates (IRT-C-DIF) to evaluate the extent to which weight-based group membership affected the probability of endorsing restrained or disinhibited eating items. DIF occurs when people with the same latent level of psychopathology have unequal probabilities of giving a response to a certain item. To evaluate DIF, we used the bivariate residual (BVR; Vermunt & Magidson, 2005), which focuses on identifying and testing potential DIF items using IRT with covariates (IRT-C). The BVR uses a generalized latent variable modeling framework to identify whether there is DIF due to more than one group characteristic (e.g., obesity status and gender simultaneously).

We used IRT-C-DIF to evaluate DIF, using the methods and procedures that are recommended by Tay, Meade, and Cao (2015). IRT-C-DIF analyses were conducted at the subscale or scale level. Our analysis proceeded in four overall steps: (a) determine whether IRT-C-DIF assumptions were met, (b) estimate IRT-C-DIF model parameters, (c) assess for DIF, and (d) determine if DIF affected mean-level score differences. Below we describe our analytical steps in detail.

IRT-C-DIF assumptions. IRT-C-DIF requires that certain assumptions are met for DIF analyses to yield meaningful, valid results. For example, if subscales are multidimensional (assess more than one latent construct), interdependent, or if the model is not correctly specified (i.e., the model does not demonstrate a good fit to the data), then the assumptions of IRT are violated. Testing IRT-C-DIF assumptions is important because if assumptions are not met, it could impact the validity of IRT-C-DIF (Embretson & Reise, 2013).

Unidimensionality and model fit were evaluated through parallel analysis, confirmatory factor analysis (CFA), and exploratory bifactor analysis, which are recommended techniques for testing IRT-C-DIF assumptions (Tay et al., 2015). First, we conducted parallel analysis, in which eigenvalues from principal components analysis were compared to the 95th percentile eigenvalues from 500 sets of permuted raw data. The numbers of components with higher actual eigenvalues than permuted eigenvalues suggested substantive factors (O'Connor, 2000).

Second, we applied a unidimensional CFA for ordinal data using Multivariate weighted least squares estimation using Laavan in R (Rosseel, 2012). CFA model-fit indices included (a) the Tucker–Lewis index and the comparative fit index (Bentler, 1990; Bentler & Bonnett, 1980; values closer to 1.00 indicated better fit and values above 0.95 provided evidence suggestive of good model-data fit, Hu & Bentler, 1999) and (b) the root mean squared error of approximation (RMSEA; values close to zero indicated better fit and values smaller than .06 suggestive of good fit, Hu & Bentler, 1999); however, it is important to note that RMSEA may not be trustworthy for assessing unidimensionality under certain conditions (see Rigdon, 1996). Kenny (2015) also cautioned against calculating the RMSEA, if the RMSEA of the null model was smaller than .158.

Finally, we estimated an exploratory bifactor model (Reise, Scheines, Widaman, & Haviland, 2013; Schmid & Leiman, 1957) and obtained values of omega hierarchical ($omega_h$) and explained common variance (ECV). The R package "psych" (Revelle, 2017) was used to conduct bifactor analyses. ECV and $omega_h$ both reflect the degree of unidimensionality in a scale. Specifically, $omega_h$ reflects how much variance in summed (standardized) scores can be attributed to a single general factor (McDonald, 2013). ECV reflects the amount of common variance beyond the general trait, regardless of the size of the item loadings on the general trait. High values of $omega_h$ (>.70) and ECV (>.60) suggest unidimensionality (Reise et al., 2013).

Estimation of IRT-C-DIF parameters. After determining whether data were sufficiently unidimensional and that IRT-C-DIF assumptions were met, we proceeded to estimate IRT item parameters using a graded response model (GRM; Samejima, 1969) in Latent GOLD 4.0 (J. K. Vermunt & Magidson, 2008). The GRM is commonly applied to self-reported personality or attitudinal data that are polytomously scored. Using a cumulative logit parameterization, the total number of response options was m and the probability of endorsing response category k ($k = 1 \ldots, m$) on item i was formulated as $P(u_i = k|\theta_j) = P^*_{i,k-1}(\theta_j) - P^*_{i,k}(\theta_j)$, $k = 1, \ldots, m$ where $P^*_{i,k}(\theta_j)$ was defined as the probability of endorsing response option k formulated as $P^*_{i,k}(\theta_j) = [1 + \exp(-[a_i\theta_j + b_{ik}])]^{-1}$, $k = 1, \ldots, m - 1$, and $P^*_{i,0}(\theta_j = 1$ and $P^*_{i,\delta}(\theta_j) = 0$. The parameters a_i and b_{ik} represented the item discrimination and item category boundaries, respectively.

Adjusted chi-square degrees of freedom ratios (adjusted χ^2/df) were computed for all individual (singles), pairs of (doubles) and triples sets of (triples) items to infer model fit. Adjusted chi-square statistics are less sensitive to sample size and more sensitive for detecting certain forms of model-data misfit, compared to examining individual items alone (Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001; Tay & Drasgow, 2012). Larger values of adjusted χ^2/df indicated worse fit. According to Drasgow, Levine, Tsien, Williams, and Mead (1995), when the ratio of chi-square to the degrees of freedom (i.e., χ^2/df) is smaller than 3.0, it indicates that the IRT model has an acceptable fit, a ratio <2.0 indicates good fit, and a ratio <1.0 indicates excellent fit.

IRT-C-DIF analyses. To assess DIF, we used the IRT-C method (Tay, Newman, & Vermunt, 2011; Tay, Vermunt, & Wang, 2013). In the current study, we tested DIF between weight groups. Specifically, we assessed whether the coefficient d_i was significantly different from zero, indicating DIF. $P^*_{i,k}(\theta_j) = [1 + \exp(-[a_i\theta_j + b_{ik} + d_iZ])]^{-1}$. The IRT-C procedure starts with an initial baseline model in which all items are constrained to be equal across the groups. The following steps are then undertaken iteratively: (1) a potential DIF item is identified based on the BVR. The BVR represents the standardized residual between the item (e.g., healthy weight vs. overweight or obese) and a covariate (gender) that is not accounted for by the model (i.e., the amount of residual variance that is likely due to DIF). The magnitude of the BVR indicates the extent to which there was local dependency between the item and the covariate, indicating DIF. The sequence is determined by inspecting the covariate x item BVR matrix to identify the largest BVR value for an item-covariate pair. The corresponding item is flagged for possible DIF. (2) To statistically test whether this potential DIF item exhibits significant DIF, a subsequent model is estimated with item parameters freely estimated between the groups. Significant differences on item parameters indicated that the item has DIF. (3) If there are significant differences in Step 2, we proceeded to update the baseline model and allowed the previously significant DIF item to be freely estimated. This excludes the item from BVR estimation, as there is no residual term given that the item is freely estimated. Then, the next potential DIF item can be examined using Steps 1 and 2. If there are no significant differences in Step 2, the procedure ends. This is because the remaining item with the largest potential for DIF does not exhibit statistically significant DIF. Using this procedure, we determined the proportion of items that had DIF.

Impact of IRT-C-DIF on mean-level score differences. To determine whether DIF affected mean-level score differences, we tested the extent to which standardized weight-group differences of a specific scale changed after accounting for DIF. For example, standardized group differences might have been 0.30 between participants with a healthy weight and those with overweight or obesity, but after accounting for DIF, the difference dropped to 0.15. This would indicate that DIF accounted for 0.15 of the mean-score difference, and only 0.15 was attributable to "true" differences in mean scores.

Results

IRT-C Assumptions

As presented in Table 3, we found that the majority of the scales were sufficiently unidimensional. Parallel analysis showed only one dominant dimension for most scales, and CFA fit statistics demonstrated moderate-to-good data fit. The only two exceptions were the Three-Factor Eating Questionnaire Cognitive Restraint and the Restraint Scale scales; parallel analysis pointed to two dimensions. The Three-Factor Eating Questionnaire Cognitive Restraint scale had raw eigenvalues of 6.15 and 1.46 for the first two factors compared to the 95th percentile values of 1.36 and 1.30, respectively. The difference between the raw and simulated second eigenvalues was small. Further, the CFA fit for the Three-Factor Eating Questionnaire Cognitive Restraint scale was good. Based on these findings, we deemed the data to be sufficiently unidimensional to proceed with IRT analyses. By contrast, the Restraint Scale had a second dominant dimension, and low $omega_h$ and ECV, indicating that applying IRT was not suitable.

Estimation of IRT-C-DIF Parameters

The doubles and triples χ^2/df statistic for each scale was about three and many scales had values smaller than three. This demonstrated that the estimated GRM fit the data well.

IRT-C-DIF Analyses

The proportion of items on a scale with DIF ranged from .00 to .60 for measures of restrained eating. The Restraint Scale had the

Table 3
Test of Unidimensionality and Item Response Theory (IRT) Model-Data Fit for Study 1 and Study 2

Scale name	Parallel analysis	CFA model-fit statistics					Bifactor model-fit statistics		IRT χ^2/df statistic		
		χ^2	*df*	CFI	TLI	RMSEA	$omega_h$	ECV	Singles	Doubles	Triples
DEBQ Restrained Eating	1	320.68	35	0.94	0.92	0.10	0.89	0.84	0.39	1.87	1.68
DEBQ Emotional Eating	1	1,358.58	65	0.88	0.85	0.16	0.63	0.60	1.78	3.51	3.93
DEBQ External Eating	1	880.17	35	0.78	0.71	0.18	0.73	0.63	0.25	2.27	2.42
TFEQ Cognitive Restraint	3	776.14	189	0.84	0.83	0.06	0.72	0.63	0.01	0.74	1.21
TFEQ Disinhibition	2	538.53	104	0.85	0.82	0.07	0.77	0.68	0.05	0.80	1.32
TFEQ Hunger	1	244.65	77	0.91	0.90	0.05	0.75	0.67	0.04	0.56	1.00
RS	2	612.63	35	0.72	0.64	0.15	0.70	0.58	0.03	1.62	1.75
EDE-Q Restraint	1	111.05	5	0.93	0.87	0.17	0.82	0.78	0.13	2.78	3.39
EDI-3 Drive for Thinness	1	94.04	9	0.96	0.94	0.11	0.87	0.82	0.27	0.82	1.06
EDI-3 Bulimia	1	379.83	20	0.85	0.78	0.15	0.78	0.66	0.04	0.58	0.83
EPSI Body Dissatisfaction	1	571.17	14	0.91	0.86	0.19	0.87	0.81	0.20	3.11	3.29
EPSI Binge Eating	1	440.28	20	0.87	0.82	0.14	0.75	0.65	0.42	2.06	2.08
EPSI Cognitive Restraint[a]	1	0.00	0	—	—	—	0.81	0.87	—	—	—
EPSI Purging	1	256.29	9	0.89	0.82	0.16	0.91	0.85	0.14	1.28	0.53
EPSI Restricting	1	617.56	9	0.80	0.66	0.25	0.77	0.61	0.19	2.55	3.13
EPSI Excessive Exercise	1	171.48	5	0.95	0.89	0.17	0.84	0.83	0.16	1.84	1.76
EPSI Muscle Building	1	41.36	5	0.99	0.97	0.08	0.84	0.80	1.42	2.20	2.09
EPSI Negative Attitudes Toward Obesity	1	215.98	5	0.93	0.87	0.20	0.78	0.75	0.39	2.54	2.68

Note. CFA = confirmatory factor analysis; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean squared error of approximation; DEBQ = Dutch Eating Behavior Questionnaire; TFEQ = Three-Factor Eating Questionnaire; RS = Restraint Scale; EDE-Q = Eating Disorder Examination–Questionnaire; EDI-3 = Eating Disorder Inventory–3; EPSI = Eating Pathology Symptoms Inventory; ECV = explained common variance.
[a] No CFA model-fit statistics are available for a three-item scale because it is a fully saturated model.

largest proportion of DIF (see Table 4). Consistent with high item bias for the Restraint Scale, score differences became smaller between healthy-weight and overweight-obese groups after accounting for DIF. In contrast to results for the Restraint Scale, Eating Pathology Symptoms Inventory Cognitive Restraint and EDE-Q Restraint showed no evidence for item bias between weight groups. For the other scales, although DIF was present, score difference changes were smaller than for the Restraint Scale.

Table 4
Weight-Group Differential Item Functioning (DIF) Analysis for Study 1 and Study 2

Scale	Score difference[a]		DIF proportion
	No DIF[b]	Accounting for DIF[c]	
Study 1			
Dutch Eating Behavior Questionnaire Restrained Eating	0.27	0.19	0.30
Dutch Eating Behavior Questionnaire Emotional Eating	0.24	0.23	0.08
Dutch Eating Behavior Questionnaire External Eating	0.19	0.22	0.10
Three-Factor Eating Questionnaire Cognitive Restraint	0.11 (*n.s.*)	0.14 (*n.s.*)	0.24
Three-Factor Eating Questionnaire Disinhibition	0.53	0.49	0.19
Three-Factor Eating Questionnaire Hunger	0.37	0.39	0.07
Restraint Scale	0.54	0.29	0.60
Eating Disorder Examination–Questionnaire Restraint	0.29	—	0.00
Eating Disorder Inventory–3 Drive for Thinness	0.32	0.22	0.17
Eating Disorder Inventory–3 Bulimia	0.29	0.32	0.38
Study 2			
Body Dissatisfaction	0.61	0.70	0.43
Binge Eating	0.21	—	0.00
Cognitive Restraint	0.23	—	0.00
Purging	0.40	—	0.00
Restricting	−0.08 (*n.s.*)	−0.18	0.33
Excessive Exercise	0.00 (*n.s.*)	—	0.00
Muscle Building	0.34	—	0.00
Negative Attitude Toward Obesity	−0.18	—	0.00

Note. *n.s.* = not significant; DIF proportion = proportion of items with DIF. Blank cells did not have DIF.
[a] Standardized score difference between overweight – healthy weight score. [b] Score difference assuming the scale does not comprise any DIF. [c] Score difference after accounting for DIF.

This suggested that item DIF may cancel out to produce scale scores for persons of healthy weight versus overweight or obesity that are fairly equivalent and not biased. See Table 5 for a list of restrained eating items that showed evidence for bias.

Scales that assessed disinhibited eating had lower DIF (vs. measures of restrained eating) with the proportion of DIF ranging from .00 to .38. Eating Disorders Inventory Bulimia had the highest proportion of DIF between weight categories, whereas the Eating Pathology Symptoms Inventory Binge Eating had no item bias. Other measures had very low levels of proportion of DIF that ranged from .08 to .10. Taken together, measures of disinhibited eating (other than Eating Disorders Inventory Bulimia) showed little evidence for item bias. See Table 5 for a list of disinhibited eating items that showed evidence for bias.

Table 5
List of Biased Items in Study 1 and Study 2

Studies
Study 1
DEBQ Restrained Eating
Trying to eat less at mealtimes (+)
Trying not to eat between meals (+)
Trying not to eat in the evenings (+)
DEBQ Emotional Eating
Wanting to eat when anxious, worried, or tense (+)
DEBQ External Eating
Wanting to eat after seeing others eat (−)
TFEQ Cognitive Restraint
Not eating after eating quota of calories (−)
Life is too short to worry about dieting (+)
Eating less after eating a food that is typically not allowed (+)
Paying attention to changes in body figure (−)
Amount of consciousness about foods eaten (−)
TFEQ Disinhibition
Going on a diet more than one time (+)
Minimal weight change in past ten years (+)
Taking a long time to eat unconsciously (+)
TFEQ Hunger
Eating more than three times per day due to hunger levels (−)
RS
Pounds over desired weight (+)
Maximum weight loss in one month (+)
Maximum weight gain in one week (+)
Typical weight fluctuation in one week (+)
Frequency of dieting (+)
Changes in life due to five-pound weight fluctuation (−)
EDI-3 Drive for Thinness
Thoughts about dieting (+)
EDI-3 Bulimia
Stuffing self with food (+)
Eating binges without ability to stop eating (−)
Thoughts about bingeing or overeating (−)
Study 2
Eating Pathology Symptoms Inventory
Body Dissatisfaction
Trying on different clothes due to body dissatisfaction (−)
Dissatisfaction with hip size (−)
Dissatisfaction with thigh size (−)
Eating Pathology Symptoms Inventory
Restricting
Others surprised if they knew how little was eaten (+)
Skipping two meals in a row (+)

Note. DEBQ = Dutch Eating Behavior Questionnaire; TFEQ = Three-Factor Eating Questionnaire; RS = Restraint Scale; EDI-3 = Eating Disorder Inventory–3; (+) = easier for persons with overweight to endorse; (−) = more difficult for persons with overweight to endorse.

As shown in Table 1, demographic characteristics (e.g., age, gender and race) varied across healthy weight and overweight/obese groups for both Study 1 and 2 samples. To test whether DIF was influenced by demographic characteristics, we conducted an additional covariate analysis using the restricted mixed-measurement item response model with covariates method (MM-IRT-C; Tay et al., 2011; Tay et al., 2013). The advantage of restricted MM-IRT-C is that it can simultaneously test for DIF across multiple groups. For example, restricted MM-IRT-C could test to what extent DIF occurred on specific Eating Disorders Inventory–3 and Eating Pathology Symptoms Inventory dimensions for weight groups, while controlling for demographic characteristics such as gender and race. Measurement items were entered as observed indicators, whereas gender and race were entered as dichotomous and categorical variables, respectively. BIC was used to determine model fit, whereas BVR was used to evaluate local misfit. We did not control for age; although differences were significant, they were small and unlikely to reflect meaningful developmental differences (e.g., the difference in age in Studies 1 and 2 were 5.83 and 0.33 years, respectively).

Results of covariate analysis suggested that gender and race did not influence the detection of DIF (see Table 6). Note that after accounting for gender and race, the standardized score difference between healthy weight and overweight/obese groups slightly increased for Eating Disorders Inventory–3 Drive for Thinness, Eating Pathology Symptoms Inventory Body Dissatisfaction, and Eating Pathology Symptoms Inventory Cognitive Restraint both when not controlling for DIF items (e.g., Eating Disorders Inventory Drive for Thinness: $d = .32$, $d_{cov} = .37$) and when controlling for DIF items (e.g., Eating Disorders Inventory Drive for Thinness: $d = .22$, $d_{cov} = .25$). Further analyses showed that the increase was mainly due to gender effects. For example, the standardized group difference between female and male, after controlling for weight and race was $d_{cov} = .94$ when not controlling for DIF items, and $d_{cov} = .96$ when controlling for DIF items, was larger than the difference between healthy weight and overweight/obese weight groups. That is, Eating Disorders Inventory Drive for Thinness, Eating Pathology Symptoms Inventory Body Dissatisfaction, and Eating Pathology Symptoms Inventory Cognitive Restraint exhibit large gender differences, which are reflected in the results for healthy weight and overweight/obese groups.

Discussion

This was the first study to test whether item bias exists across the weight spectrum for common measures of restrained, disinhibited, and binge eating. We hypothesized that the Restraint Scale and other measures that were developed to test restraint theory would show the most evidence for DIF. Results of our study supported our hypotheses. Specifically, we posited that the Restraint Scale would overestimate restrained eating among persons with overweight or obesity because the Restraint Scale included content that is dependent upon a person's history of weight fluctuations and past overweight status. In other words, even if a person had low current dietary restraint, they may still score highly on the Restraint Scale due to past a history of overweight or obesity. Given that past research posited that the Restraint Scale only predicts overeating in dieters who are already highly prone to

Table 6
Weight-Based Differential Item Functioning (DIF) Analysis for Study 1 and Study 2: Controlling for Gender and Race as Covariates

Scale	Score difference[a]		DIF proportion
	No DIF[b]	Accounting for DIF[c]	
Study 1			
Dutch Eating Behavior Questionnaire Restrained Eating	0.30	0.21	0.30
Dutch Eating Behavior Questionnaire Emotional Eating	0.24	0.24	0.08
Dutch Eating Behavior Questionnaire External Eating	0.19	0.22	0.10
Three-Factor Eating Questionnaire Cognitive Restraint	0.12 (*n.s.*)	0.14	0.24
Three-Factor Eating Questionnaire Disinhibition	0.55	0.54	0.19
Three-Factor Eating Questionnaire Hunger	0.37	0.39	0.07
Restraint Scale	0.56	0.57	0.60
EDE-Q Restraint	0.29	—	0.00
Eating Disorder Inventory–3 Drive for Thinness	0.37	0.25	0.17
Eating Disorder Inventory–3 Bulimia	0.28	0.32	0.38
Study 2			
Body Dissatisfaction	0.73	0.83	0.43
Binge Eating	0.21	—	0.00
Cognitive Restraint	0.30	—	0.00
Purging	0.44	—	0.00
Restricting	−0.09 (*n.s.*)	−0.19	0.33
Excessive Exercise	0.00	—	0.00
Muscle Building	0.32	—	0.00
Negative Attitude Toward Obesity	−0.19	—	0.00

Note. EDE-Q = Eating Disorder Examination–Questionnaire; *n.s.* = not significant; DIF proportion = proportion of items with DIF. Blank cells did not have DIF.
[a] Standardized score difference between overweight – healthy weight score. [b] Score difference assuming the scale does not comprise any DIF. [c] Score difference after accounting for DIF.

overeating (Ouwens et al., 2003a; Ouwens et al., 2003b), we hypothesized that the Restraint Scale would show DIF across the weight spectrum. We further hypothesized that: 1) Three-Factor Eating Questionnaire Restraint scale items would underestimate restrained eating in persons with overweight or obesity and 2) measures that included persons with overweight or obesity in the initial questionnaire development and validation process (Dutch Eating Behavior Questionnaire, Eating Pathology Symptoms Inventory, and Three-Factor Eating Questionnaire) would show evidence for the least amount of DIF compared to other measures of similar constructs.

Consistent with our hypotheses, the Restraint Scale performed the worst out of all the scales we tested. The one-factor model for the Restraint Scale had a poor fit to the data, which suggested that this measure may be tapping into more than one latent construct. Sixty percent of items on the Restraint Scale demonstrated DIF between healthy and overweight/obese participants, suggesting the presence of substantial bias related to a participant's past or current weight status. The majority of items that were biased pertained to a history of past weight fluctuations, which were endorsed at a lower threshold among overweight/obese participants (vs.) healthy weight participants who were matched on latent levels of restraint. Another Restraint Scale item assessed spending too much time thinking about eating or food, but this item was biased in the direction of healthy weight participants endorsing this item at a lower threshold than overweight or obese participants. Items that measured guilt over eating, eating in secret, and a person's highest ever maximum weight did not show evidence for bias between groups. However, given that six out of 10 Restraint Scale items showed evidence for bias, we suggest that caution be used in interpreting mean-level differences between weight groups, particularly in patients or participants who are overweight or obese. In support of our hypothesis, we found that measures that were developed and validated in samples that included overweight or obese persons (Dutch Eating Behavior Questionnaire, Eating Pathology Symptoms Inventory, and Three-Factor Eating Questionnaire) had less bias than measures developed in mostly healthy weight samples (Eating Disorders Inventory–3 and EDE-Q). However, contrary to our expectation, the EDE-Q Restraint scale had no evidence of bias between weight groups.

Except for the Restraint Scale, several measures of restrained eating performed well between groups. Eating Pathology Symptoms Inventory Cognitive Restraint and EDE-Q Restraint did not show any evidence of item bias. Thus, clinicians and researchers interested in measuring restrained eating could utilize these scales without modifications in persons who are healthy weight, overweight, or obese. The Dutch Eating Behavior Questionnaire Restrained Eating, Eating Disorders Inventory–3 Drive for Thinness, Eating Pathology Symptoms Inventory Restricting, and Three-Factor Eating Questionnaire Cognitive Control of Eating scales had moderate bias that ranged from 17% to 33% of items. Items within restrained eating measures that showed evidence for bias focused on diet-related cognitions (e.g., thinking about dieting and "making up" later if a nonallowed food was eaten) or items related to "watching one's weight" (e.g., trying to eat less due to watching weight).

Consistent with our hypotheses, the majority of bias was in the direction of items having a lower threshold of endorsement for restrained-eating items for persons who had overweight or obesity. The finding that overweight and obese people had lower thresholds for endorsement of item content focused on diet-related cognitions and "weight watching" was intriguing because it indicated that persons who are overweight or obese will score higher than their healthy weight counterparts on certain measures of restrained eating due to item bias, rather than due to true elevations in dietary restraint. Finally, items that showed low or no evidence for bias were more behavioral (vs. cognitive) and included items to assess issues such as excluding unhealthy food, avoiding food with high calorie content, working to limit dietary intake (whether or not one "succeeded"), going for eight or more hours without eating, and setting specific calorie limits. In other words, behaviors associated with dieting showed less evidence for bias, whereas items that measured general thoughts related to dieting or dieting due to unhappiness with weight or shape were endorsed at a lower threshold for persons with overweight or obesity.

Most measures of disinhibited eating had low to moderate levels of item bias. The Eating Pathology Symptoms Inventory Binge Eating scale performed the best with no items showing evidence for item bias. The Eating Disorders Inventory–3 Bulimia scale had the most item bias, with two items having higher thresholds and one item having a lower threshold for endorsement in persons with overweight or obesity. Other scales, including the Dutch Eating Behavior Questionnaire Emotional Eating, Dutch Eating Behavior Questionnaire External Eating, Three-Factor Eating Questionnaire Disinhibition, and Three-Factor Eating Questionnaire Hunger, had low levels of bias, ranging from 7% to 19% of items. Interestingly, biased Three-Factor Eating Questionnaire Disinhibition items were similar to items in the Restraint Scale; content included history of weight fluctuations. Dutch Eating Behavior Questionnaire items were related to a desire to eat when feeling anxiety and worry (which had a lower threshold for endorsement among overweight or obese persons) and desire to eat after seeing others eat (which had a higher threshold for overweight or obese persons). Interestingly, the content that was biased on the Eating Disorders Inventory–3 was related to binge eating, which was similar to content on the Eating Pathology Symptoms Inventory. Both Eating Disorders Inventory–3 Bulimia and Eating Pathology Symptoms Inventory Binge Eating ask about eating to the point of feeling sick or "stuffed," eating large amounts of food, and feeling it was hard to stop eating after starting. However, whereas the Eating Disorders Inventory–3 had the most evidence for bias, the Eating Pathology Symptoms Inventory had the least. One reason for this discrepancy could be due to the fact that all of the Eating Disorders Inventory–3 items that showed evidence for bias were reverse scored. Other studies found that reverse-scored items tend to have worse psychometric properties than items that were scored in the positive direction (Barnette, 2000; Conrad et al., 2004; Woods, 2006; Zhang, Noor, & Savalei, 2016). Thus, part of the reason for item bias for the Eating Disorders Inventory–3 may be due to problematic scale construction methods.

There were several limitations of the current study. First, although our study included common measures of disinhibited eating, newer measures of food reinforcement—for example, the Power of Food Scale (Lowe et al., 2009) were not assessed because it were not available at the time that data were collected. Second, the Restraint Scale did not show evidence for unidimensionality. To address this issue, we ran additional a supplementary multiple group analysis, which does not require the assumption of unidimensionality. Results of multiple group analysis also indicated that the Restraint Scale was variant between weight groups (data available upon request), suggesting that the Restraint Scale is indeed biased across weight categories. Third, height and weight were collected from online self-report, rather than measured objectively. Research has shown that persons who are overweight or obese are more likely to underestimate their weight, which means that some persons classified as having a healthy weight may have been overweight or obese (Spencer, Appleby, Davey, & Key, 2002). On the other hand, nationally representative studies show there is high agreement between self- and objectively measured height and weight with correlations ranging from $\geq .89$ for weight and $> .95$ for height (Kuczmarski, Kuczmarski, & Najjar, 2001; Spencer et al., 2002). Other research found high agreement (r = .99) among young adults aged 18–35 for BMI assessed via online self-report and objective measurement (Pursey, Burrows, Stanwell, & Collins, 2014). Thus, although some participants may have been placed into an incorrect weight category, the extent of misclassification was likely small. Finally, our study focused on one individual difference variable (weight) and the effect of weight classification on item bias for measures of restrained and disinhibited eating. However, due to the relative lack of ethnic and racial diversity in our samples, it was unclear how the intersection of race and gender may have interacted with weight status to influence responses to measures of restrained and disinhibited eating. Consideration of intersectionality is important given that certain racial and ethnic groups have higher rates of overweight and obesity than their Caucasian counterparts.

Implications and Conclusions

Clinical implications of the proposed work indicate that certain measures of restrained and disinhibited eating may lead to inaccurate decisions due to item bias. For example, when item bias was detected for the Three-Factor Eating Questionnaire Disinhibition scale, it tended to overpathologize persons with overweight or obesity, meaning that they would be more likely to score above clinical cut-points for disinhibited eating due to bias, rather than true problems with nonhomeostatic eating. The effect of this bias would be that persons may be disqualified from bariatric surgery or behavioral weight loss studies. Three-Factor Eating Questionnaire Cognitive Restraint items that showed evidence for bias underestimated restrained eating in overweight or obese persons which could potentially lead to a missed opportunity to help individuals at-risk for the development of atypical anorexia nervosa. Thus, to provide accurate assessment of restrained and disinhibited eating behaviors across the weight spectrum, we recommend using measures that were listed as bias free or adjusting scores to remove biased items to prevent misclassifications.

The current study has implications for the future assessment of restraint theory. Specifically, the "disinhibition effect," in which individuals who score high on restraint tend to eat more after a milkshake test, may be a psychometric artifact of the Restraint Scale, rather than a true psychological phenomenon. The Restraint Scale measures histories of weight fluctuation and past overweight status, which may be more important than dietary restraint for predicting overeating. We suggest that rather than using the Restraint Scale to measure dietary restraint, clinicians and researchers should directly measure histories of weight fluctuation or weight suppression (i.e., the difference between a person's current weight and highest nonpregnancy weight) or assess concrete behaviors that are associated with dietary restraint that underlie weight change (e.g., avoiding high calorie foods).

In conclusion, EDs and overweight are major public health issues. The current study is significant as the first to identify restrained and disinhibited eating items that are biased among overweight and obese persons, with important implications for improving both research and clinical assessment. Although the majority of items on the Restraint Scale showed evidence for biased responding between weight groups, the EDE-Q Restraint and Eating Pathology Symptoms Inventory Binge Eating, Cognitive Restraint, Excessive Exercise, Muscle Building, and Negative Attitudes Toward Obesity scales showed no evidence for item bias. Thus, there is an array of strong psychometric tools from which to choose in persons with and without overweight or obesity.

References

Ágh, T., Kovács, G., Supina, D., Pawaskar, M., Herman, B. K., Vokó, Z., & Sheehan, D. V. (2016). A systematic review of the health-related quality of life and economic burdens of anorexia nervosa, bulimia nervosa, and binge eating disorder. *Eating and Weight Disorders, 21,* 353–364. http://dx.doi.org/10.1007/s40519-016-0264-x

Agras, W. S., & Telch, C. F. (1998). The effects of caloric deprivation and negative affect on binge eating in obese binge-eating disordered women. *Behavior Therapy, 29,* 491–503. http://dx.doi.org/10.1016/S0005-7894(98)80045-2

Allison, D. B., Kalinsky, L. B., & Gorman, B. S. (1992). A comparison of the psychometric properties of three measures of dietary restraint. *Psychological Assessment, 4,* 391–398. http://dx.doi.org/10.1037/1040-3590.4.3.391

Allison, K. C., Engel, S. G., Crosby, R. D., de Zwaan, M., O'Reardon, J. P., Wonderlich, S. A., . . . Stunkard, A. J. (2008). Evaluation of diagnostic criteria for night eating syndrome using item response theory analysis. *Eating Behaviors, 9,* 398–407. http://dx.doi.org/10.1016/j.eatbeh.2008.04.004

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM–5).* Washington, DC: Author.

Ardovini, C., Caputo, G., Todisco, P., & Dalle Grave, R. (1999). Binge eating and restraint model: Psychometric analysis in binge eating disorder and normal weight bulimia. *European Eating Disorders Review, 7,* 293–299. http://dx.doi.org/10.1002/(SICI)1099-0968(199908)7:4<293::AID-ERV273>3.0.CO;2-S

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement, 60,* 361–370. http://dx.doi.org/10.1177/00131640021970592

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246. http://dx.doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonnett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606. http://dx.doi.org/10.1037/0033-2909.88.3.588

Berg, K. C., Peterson, C. B., Frazier, P., & Crow, S. J. (2012). Psychometric evaluation of the eating disorder examination and eating disorder examination-questionnaire: A systematic review of the literature. *International Journal of Eating Disorders, 45,* 428–438. http://dx.doi.org/10.1002/eat.20931

Bohrer, B. K., Forbush, K. T., & Hunt, T. K. (2015). Are common measures of dietary restraint and disinhibited eating reliable and valid in obese persons? *Appetite, 87,* 344–351. http://dx.doi.org/10.1016/j.appet.2014.12.226

Brown, K., Bryant, E., Naslund, E., King, N., & Blundell, J. (2006). Traits that promote weight gain in obesity, bulimia nervosa and EDNOS. *Obesity Reviews, 7,* 330.

Bryant, E. J., King, N. A., & Blundell, J. E. (2008). Disinhibition: Its effects on appetite and weight regulation. *Obesity Reviews, 9,* 409–419. http://dx.doi.org/10.1111/j.1467-789X.2007.00426.x

Chernyshenko, O. S., Stark, S., Chan, K.-Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36,* 523–562. http://dx.doi.org/10.1207/S15327906MBR3604_03

Colles, S. L., Dixon, J. B., & O'Brien, P. E. (2008). Grazing and loss of control related to eating: Two high-risk factors following bariatric surgery. *Obesity, 16,* 615–622. http://dx.doi.org/10.1038/oby.2007.101

Coniglio, K. A., Becker, K. R., Tabri, N., Keshishian, A. C., Miller, J. D., Eddy, K. T., & Thomas, J. J. (2018). Factorial integrity and validation of the Eating Pathology Symptoms Inventory (EPSI). *Eating Behaviors, 31,* 1–7. http://dx.doi.org/10.1016/j.eatbeh.2018.07.004

Conrad, K. J., Wright, B. D., McKnight, P., McFall, M., Fontana, A., & Rosenheck, R. (2004). Comparing traditional and Rasch analyses of the Mississippi PTSD Scale: Revealing limitations of reverse-scored items. *Journal of Applied Measurement, 5,* 15–30.

Cooper, Z., & Fairburn, C. (1987). The eating disorder examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders. *International Journal of Eating Disorders, 6,* 1–8. http://dx.doi.org/10.1002/1098-108X(198701)6:1<1::AID-EAT2260060102>3.0.CO;2-9

Crawford, A. G., Cote, C., Couto, J., Daskiran, M., Gunnarsson, C., Haas, K., . . . Schuette, R. (2010). Prevalence of obesity, Type II diabetes mellitus, hyperlipidemia, and hypertension in the United States: Findings from the GE Centricity Electronic Medical Record database. *Population Health Management, 13,* 151–161. http://dx.doi.org/10.1089/pop.2009.0039

Cuntz, U., Leibbrand, R., Ehrig, C., Shaw, R., & Fichter, M. M. (2001). Predictors of post-treatment weight reduction after in-patient behavioral therapy. *International Journal of Obesity and Related Metabolic Disorders, 25,* S99–S101. http://dx.doi.org/10.1038/sj.ijo.0801710

Drasgow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement, 19,* 143–166. http://dx.doi.org/10.1177/014662169501900203

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Florence, KY: Psychology Press. http://dx.doi.org/10.4324/9781410605269

Epstein, L. H., Paluch, R. A., Saelens, B. E., Ernst, M. M., & Wilfley, D. E. (2001). Changes in eating disorder symptoms with pediatric obesity treatment. *The Journal of Pediatrics, 139,* 58–65. http://dx.doi.org/10.1067/mpd.2001.115022

Fairburn, C. G. (2008). Eating disorders: The transdiagnostic view and the cognitive behavioral theory. *Cognitive behavior therapy and eating disorders* (pp. 2–22). New York, NY: Guilford Press.

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16,* 363–370.

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., & Dietz, W. (2009). Annual medical spending attributable to obesity: Payer-and service-specific estimates. *Health Affairs, 28,* w822–w831. http://dx.doi.org/10.1377/hlthaff.28.5.w822

Fogelholm, M., Kukkonen-Harjula, K., & Oja, P. (1999). Eating control and physical activity as determinants of short-term weight maintenance after a very-low-calorie diet among obese women. *International Journal of Obesity and Related Metabolic Disorders, 23,* 203–210. http://dx.doi.org/10.1038/sj.ijo.0800825

Forbush, K. T., Hilderbrand, L. A., Bohrer, B. K., & Chapa, D. A. N. (2019). Test-retest reliability of common measures of eating disorder symptoms in men versus women. *Assessment, 26,* 419–431. http://dx.doi.org/10.1177/1073191117700267

Forbush, K. T., Wildes, J. E., & Hunt, T. K. (2014). Gender norms, psychometric properties, and validity for the Eating Pathology Symptoms Inventory. *International Journal of Eating Disorders, 47,* 85–91. http://dx.doi.org/10.1002/eat.22180

Forbush, K. T., Wildes, J. E., Pollack, L. O., Dunbar, D., Luo, J., Patterson, K., . . . Watson, D. (2013). Development and validation of the Eating Pathology Symptoms Inventory (EPSI). *Psychological Assessment, 25,* 859–878. http://dx.doi.org/10.1037/a0032639

Forman, S. F., McKenzie, N., Hehn, R., Monge, M. C., Kapphahn, C. J., Mammel, K. A., . . . Woods, E. R. (2014). Predictors of outcome at 1 year in adolescents with *DSM–5* restrictive eating disorders: Report of the national eating disorders quality improvement collaborative. *The Journal of Adolescent Health, 55,* 750–756. http://dx.doi.org/10.1016/j.jadohealth.2014.06.014

Garner, D. M. (2004). *EDI-3 (Eating Disorder Inventory–3) professional manual*. Lutz, FL: Psychological Assessment Resources.

Garner, D. M., Olmstead, M. P., & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders, 2,* 15–34. http://dx.doi.org/10.1002/1098-108X(198321)2:2<15::AID-EAT2260020203>3.0.CO;2-6

Gideon, N., Hawkes, N., Mond, J., Saunders, R., Tchanturia, K., & Serpell, L. (2016). Development and psychometric validation of the Eating Disorder Examination–Questionnaire S, a 12 item short form of the Eating Disorder Examination–Questionnaire (EDE-Q). *PLoS ONE, 11,* e0152744. http://dx.doi.org/10.1371/journal.pone.0152744

Hall, M. E., do Carmo, J. M., da Silva, A. A., Juncos, L. A., Wang, Z., & Hall, J. E. (2014). Obesity, hypertension, and chronic kidney disease. *International Journal of Nephrology and Renovascular Disease, 7,* 75–88. http://dx.doi.org/10.2147/IJNRD.S39739

Hay, P., Girosi, F., & Mond, J. (2015). Prevalence and sociodemographic correlates of *DSM–5* eating disorders in the Australian population. *Journal of Eating Disorders, 3,* 19. http://dx.doi.org/10.1186/s40337-015-0056-0

Hays, N. P., & Roberts, S. B. (2008). Aspects of eating behaviors “disinhibition” and “restraint” are related to weight gain and BMI in women. *Obesity, 16,* 52–58. http://dx.doi.org/10.1038/oby.2007.12

Herman, C. P., & Mack, D. (1975). Restrained and unrestrained eating. *Journal of Personality, 43,* 647–660. http://dx.doi.org/10.1111/j.1467-6494.1975.tb00727.x

Herman, C. P., & Polivy, J. (1975). Anxiety, restraint, and eating behavior. *Journal of Abnormal Psychology, 84,* 66–72. http://dx.doi.org/10.1037/0021-843X.84.6.666

Herman, C. P., & Polivy, J. (1983). A boundary model for the regulation of eating. *Psychiatric Annals, 13,* 918–927. http://dx.doi.org/10.3928/0048-5713-19831201-03

Herrmann, S. D., Gao, Y., Lambourne, K., Willis, E. A., Paxton, R. J., & Donnelly, J. E. (2014). Calibrating the Three-Factor Eating Questionnaire using Rasch modeling. *Research Quarterly for Exercise and Sport, 85,* A68.

Hetherington, M. M., Stoner, S. A., Andersen, A. E., & Rolls, B. J. (2000). Effects of acute food deprivation on eating behavior in eating disorders. *International Journal of Eating Disorders, 28,* 272–283. http://dx.doi.org/10.1002/1098-108X(200011)28:3<272::AID-EAT4>3.0.CO;2-Q

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

Hudson, J. I., Hiripi, E., Pope, H. G., Jr., & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological Psychiatry, 61,* 348–358. http://dx.doi.org/10.1016/j.biopsych.2006.03.040

Ismail, R., Noh, M., Ismail, N. H., & Tamil, A. M. (2015). A tale of two construct validation analysis: Rasch model and exploratory factor analysis approach for three-factor eating questionnaire (TFEQ-R21) among Malaysian male workers. *The Medical Journal of Malaysia, 70,* 169–176.

Karlsson, J., Persson, L.-O., Sjöström, L., & Sullivan, M. (2000). Psychometric properties and factor structure of the Three-Factor Eating Questionnaire (TFEQ) in obese men and women. Results from the Swedish Obese Subjects (SOS) study. *International Journal of Obesity, 24,* 1715–1725. http://dx.doi.org/10.1038/sj.ijo.0801442

Kenny, D. A. (2015). *Measuring model fit*. Retrieved from http://davidakenny.net/cm/fit.htm

Keski-Rahkonen, A., & Mustelin, L. (2016). Epidemiology of eating disorders in Europe: Prevalence, incidence, comorbidity, course, consequences, and risk factors. *Current Opinion in Psychiatry, 29,* 340–345. http://dx.doi.org/10.1097/YCO.0000000000000278

Kiezebrink, K., Campbell, D., Mann, E., & Blundell, J. (2009). Similarities and differences between excessive exercising anorexia nervosa patients compared with *DSM–IV* defined anorexia nervosa subtypes. *Eating and Weight Disorders, 14,* e199–e204. http://dx.doi.org/10.1007/BF03325117

Klem, M. L., Klesges, R. C., Bene, C. R., & Mellon, M. W. (1990). A psychometric study of restraint: The impact of race, gender, weight and marital status. *Addictive Behaviors, 15,* 147–152. http://dx.doi.org/10.1016/0306-4603(90)90018-S

Kuczmarski, M. F., Kuczmarski, R. J., & Najjar, M. (2001). Effects of age on validity of self-reported height, weight, and body mass index: Findings from the Third National Health and Nutrition Examination Survey, 1988–1994. *Journal of the American Dietetic Association, 101,* 28–34. http://dx.doi.org/10.1016/S0002-8223(01)00008-6

Lowe, M. R. (1992). Staying on versus going off a diet: Effects on eating in normal weight and overweight individuals. *International Journal of Eating Disorders, 12,* 417–424. http://dx.doi.org/10.1002/1098-108X(199212)12:4<417::AID-EAT2260120409>3.0.CO;2-J

Lowe, M. R. (1994). Putting restrained and unrestrained nondieters on short-term diets: Effects on eating. *Addictive Behaviors, 19,* 349–356. http://dx.doi.org/10.1016/0306-4603(94)90057-4

Lowe, M. R., Butryn, M. L., Didie, E. R., Annunziato, R. A., Thomas, J. G., Crerand, C. E., . . . Halford, J. (2009). The Power of Food Scale: A new measure of the psychological influence of the food environment. *Appetite, 53,* 114–118. http://dx.doi.org/10.1016/j.appet.2009.05.016

Marek, R. J., Heinberg, L. J., Lavery, M., Merrell Rish, J., & Ashton, K. (2016). A review of psychological assessment instruments for use in bariatric surgery evaluations. *Psychological Assessment, 28,* 1142–1157. http://dx.doi.org/10.1037/pas0000286

Masheb, R. M., Lutes, L. D., Kim, H. M., Holleman, R. G., Goodrich, D. E., Janney, C. A., . . . Damschroder, L. J. (2015). High-frequency binge eating predicts weight gain among veterans receiving behavioral weight loss treatments. *Obesity, 23,* 54–61. http://dx.doi.org/10.1002/oby.20931

McDonald, R. P. (2013). *Test theory: A unified treatment*. New York, NY: Psychology Press. http://dx.doi.org/10.4324/9781410601087

McGuire, M. T., Wing, R. R., Klem, M. L., Lang, W., & Hill, J. O. (1999). What predicts weight regain in a group of successful weight losers? *Journal of Consulting and Clinical Psychology, 67,* 177–185. http://dx.doi.org/10.1037/0022-006X.67.2.177

Meany, G., Conceição, E., & Mitchell, J. E. (2014). Binge eating, binge eating disorder and loss of control eating: Effects on weight outcomes after bariatric surgery. *European Eating Disorders Review, 22,* 87–91. http://dx.doi.org/10.1002/erv.2273

Mitchell, J. E. (2016). Medical comorbidity and medical complications associated with binge-eating disorder. *International Journal of Eating Disorders, 49,* 319–323. http://dx.doi.org/10.1002/eat.22452

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32,* 396–402. http://dx.doi.org/10.3758/BF03200807

Ogden, C. L., Carroll, M. D., Curtin, L. R., Lamb, M. M., & Flegal, K. M. (2010). Prevalence of high body mass index in U.S. children and adolescents, 2007–2008. *Journal of the American Medical Association, 303,* 242–249. http://dx.doi.org/10.1001/jama.2009.2012

Ouwens, M. A., van Strien, T., & van der Staak, C. P. (2003a). Absence of a disinhibition effect of alcohol on food consumption. *Eating Behaviors, 4,* 323–332. http://dx.doi.org/10.1016/S1471-0153(03)00033-3

Ouwens, M. A., van Strien, T., & van der Staak, C. P. (2003b). Tendency toward overeating and restraint as predictors of food consumption. *Appetite, 40,* 291–298. http://dx.doi.org/10.1016/S0195-6663(03)00006-0

Pacanowski, C. R., Senso, M. M., Oriogun, K., Crain, A. L., & Sherwood, N. E. (2014). Binge eating behavior and weight loss maintenance over a 2-year period. *Journal of Obesity, 2014,* 249315. http://dx.doi.org/10.1155/2014/249315

Polivy, J., Heatherton, T. F., & Herman, C. P. (1988). Self-esteem, restraint, and eating behavior. *Journal of Abnormal Psychology, 97,* 354–356. http://dx.doi.org/10.1037/0021-843X.97.3.354

Presnell, K., & Stice, E. (2003). An experimental test of the effect of weight-loss dieting on bulimic pathology: Tipping the scales in a different direction. *Journal of Abnormal Psychology, 112,* 166–170. http://dx.doi.org/10.1037/0021-843X.112.1.166

Pursey, K., Burrows, T. L., Stanwell, P., & Collins, C. E. (2014). How accurate is web-based self-reported height, weight, and body mass index in young adults? *Journal of Medical Internet Research, 16,* e4. http://dx.doi.org/10.2196/jmir.2909

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73,* 5–26. http://dx.doi.org/10.1177/0013164412449831

Revelle, W. R. (2017). *psych: Procedures for personality and psychological research*. Evanston, IL: Northwestern University. Retrieved from https://cran.r-project.org/web/packages/psych/index.html

Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling, 3,* 369–379. http://dx.doi.org/10.1080/10705519609540052

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more: Version 0.5–12 (BETA). *Journal of Statistical Software, 48,* 1–36. http://dx.doi.org/10.18637/jss.v048.i02

Ruderman, A. J., & Christensen, H. (1983). Restraint theory and its applicability to overweight individuals. *Journal of Abnormal Psychology, 92,* 210–215. http://dx.doi.org/10.1037/0021-843X.92.2.210

Salbach-Andrae, H., Schneider, N., Bürger, A., Pfeiffer, E., Lehmkuhl, U., & Holzhausen, M. (2010). Psychometrische Gütekriterien des Eating Disorder Inventory (EDI-2) bei Jugendlichen [Psychometric properties of the Eating Disorder Inventory (EDI-2) in adolescents]. *Zeitschrift fur Kinder- und Jugendpsychiatrie und Psychotherapie, 38,* 219–228. http://dx.doi.org/10.1024/1422-4917/a000035

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph, 34,* 100–114. http://dx.doi.org/10.1007/BF03372160

Sares-Jäske, L., Knekt, P., Männistö, S., Lindfors, O., & Heliövaara, M. (2019). Self-report dieting and long-term changes in body mass index and waist circumference. *Obesity Science & Practice, 5,* 291–303. http://dx.doi.org/10.1002/osp4.336

Sawyer, S. M., Whitelaw, M., Le Grange, D., Yeo, M., & Hughes, E. K. (2016). Physical and psychological morbidity in adolescents with atypical anorexia nervosa. *Pediatrics, 137,* e20154080. http://dx.doi.org/10.1542/peds.2015-4080

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53–61. http://dx.doi.org/10.1007/BF02289209

Spencer, E. A., Appleby, P. N., Davey, G. K., & Key, T. J. (2002). Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutrition, 5,* 561–565. http://dx.doi.org/10.1079/PHN2001322

Stice, E. (2001). A prospective test of the dual-pathway model of bulimic pathology: Mediating effects of dieting and negative affect. *Journal of Abnormal Psychology, 110,* 124–135. http://dx.doi.org/10.1037/0021-843X.110.1.124

Stice, E., & Agras, W. S. (1998). Predicting onset and cessation of bulimic behaviors during adolescence: A longitudinal grouping analysis. *Behavior Therapy, 29,* 257–276. http://dx.doi.org/10.1016/S0005-7894(98)80006-3

Stunkard, A. J., & Messick, S. (1985). The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychosomatic Research, 29,* 71–83. http://dx.doi.org/10.1016/0022-3999(85)90010-8

Takakura, S., Aso, C. S., Toda, K., Hata, T., Yamashita, M., & Sudo, N. (2019). Physical and psychological aspects of anorexia nervosa based on duration of illness: A cross-sectional study. *BioPsychoSocial Medicine, 13,* 32. http://dx.doi.org/10.1186/s13030-019-0173-0

Tang, X., Forbush, K. T., & Lui, P. P. (2015). Development and validation of the Chinese-language version of the eating pathology symptoms inventory. *International Journal of Eating Disorders, 48,* 1016–1023. http://dx.doi.org/10.1002/eat.22423

Tay, L., & Drasgow, F. (2012). Adjusting the adjusted χ^2/df ratio statistic for dichotomous item response theory analyses: Does the model fit? *Educational and Psychological Measurement, 72,* 510–528. http://dx.doi.org/10.1177/0013164411416976

Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods, 18,* 3–46. http://dx.doi.org/10.1177/1094428114553062

Tay, L., Newman, D. A., & Vermunt, J. K. (2011). Using mixed-measurement item response theory with covariates (MM-IRT-C) to ascertain observed and unobserved measurement equivalence. *Organizational Research Methods, 14,* 147–176. http://dx.doi.org/10.1177/1094428110366037

Tay, L., Vermunt, J. K., & Wang, C. (2013). Assessing the item response theory with covariate (IRT-C) procedure for ascertaining differential item functioning. *International Journal of Testing, 13,* 201–222. http://dx.doi.org/10.1080/15305058.2012.692415

Teixeira, P. J., Silva, M. N., Coutinho, S. R., Palmeira, A. L., Mata, J., Vieira, P. N., . . . Sardinha, L. B. (2010). Mediators of weight loss and weight loss maintenance in middle-aged women. *Obesity, 18,* 725–735. http://dx.doi.org/10.1038/oby.2009.281

Van Strien, T., Cleven, A., & Schippers, G. (2000). Restraint, tendency toward overeating and ice cream consumption. *International Journal of Eating Disorders, 28,* 333–338. http://dx.doi.org/10.1002/1098-108X(200011)28:3<333::AID-EAT11>3.0.CO;2-%23

Van Strien, T., Frijters, J. E., Bergers, G. P., & 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,* 295–315. http://dx.doi.org/10.1002/1098-108X(198602)5:2<295::AID-EAT2260050209>3.0.CO;2-T

Vermunt, J. K., & Magidson, J. (2005). *Technical guide for Latent GOLD 4.0: Basic and advanced*. Belmont, MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. (2008). Latent GOLD 4.5 [Computer program]. Belmont, MA: Statistical Innovations.

Villarejo, C., Fernández-Aranda, F., Jiménez-Murcia, S., Peñas-Lledó, E., Granero, R., Penelo, E., . . . Menchón, J. M. (2012). Lifetime obesity in patients with eating disorders: Increasing prevalence, clinical and personality correlates. *European Eating Disorders Review, 20,* 250–254. http://dx.doi.org/10.1002/erv.2166

Wang, S. B., Mancuso, C. J., Jo, J., Keshishian, A. C., Becker, K. R., Plessow, F., . . . Eddy, K. T. (2019). Restrictive eating, but not binge eating or purging, predicts suicidal ideation in adolescents and young adults with low-weight eating disorders. *International Journal of Eating Disorders*. Advance online publication. http://dx.doi.org/10.1002/eat.23210

Westenhoefer, J., Broeckmann, P., Münch, A.-K., & Pudel, V. (1994). Cognitive control of eating behaviour and the disinhibition effect. *Appetite, 23,* 27–41. http://dx.doi.org/10.1006/appe.1994.1032

Westmoreland, P., Krantz, M. J., & Mehler, P. S. (2016). Medical complications of anorexia nervosa and bulimia. *The American Journal of Medicine, 129,* 30–37. http://dx.doi.org/10.1016/j.amjmed.2015.06.031

Whitelaw, M., Gilbertson, H., Lee, K. J., & Sawyer, S. M. (2014). Restrictive eating disorders among adolescent inpatients. *Pediatrics, 134,* e758–e764. http://dx.doi.org/10.1542/peds.2014-0070

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment, 28,* 186–191. http://dx.doi.org/10.1007/s10862-005-9004-7

Zhang, X., Noor, R., & Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PLoS ONE, 11,* e0157795. http://dx.doi.org/10.1371/journal.pone.0157795

Received February 13, 2019
Revision received January 18, 2020
Accepted January 22, 2020 ■