The Impact of Gender on the Assessment of Body Checking Behavior

Lauren Alfano, BA[1], **Tom Hildebrandt, PsyD**[1], **Katie Bannon, MA**[2], **Catherine Walker, MA**[3], and **Kate E. Walton, PhD**[4]

[1] Eating and Weight Disorders Program, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 120, New York, NY 10029. Phone: 212-659-8673. Fax: 212-849-2561

[2] Rutgers, The State University of New Jersey

[3] Albany, The State University of New York

[4] St. John's University, New York, NY

Abstract

Body checking includes any behavior aimed at global or specific evaluations of appearance characteristics. Men and women are believed to express these behaviors differently, possibly reflecting different socialization. However, there has been no empirical test of the impact of gender on body checking. A total of 1024 male and female college students completed two measures of body checking, the Body Checking Questionnaire and the Male Body Checking Questionnaire. Using multiple group confirmatory factor analysis, differential item functioning (DIF) was explored in a composite of these measures. Two global latent factors were identified (female and male body checking severity), and there were expected gender differences in these factors even after controlling for DIF. Ten items were found to be unbiased by gender and provide a suitable brief measure of body checking for mixed gender research. Practical applications for body checking assessment and theoretical implications are discussed.

Body checking can be broadly defined as any behavior, such as weighing oneself or comparing one's body size to others', which is aimed at global or specific evaluations of appearance characteristics, including body size, facial symmetry, and body composition. Although body checking is believed to be both a relatively common behavior in the general public and a core feature of certain types of psychopathology (e.g., eating disorders, body dysmorphic disorder; Cash, 2002; Kaye, Strober, & Rhodes, 2002; Olivardia, 2001), it remains understudied.

Body checking behavior can occur in the general experience of one's body image without significant psychopathology. However, when pathological, it is believed to be both excessive in quantity and compulsive or ritualized in nature. For example, compulsive body checking is commonly present in those with body dysmorphic disorder (BDD) and often occurs in the context of a body ritual, such as repeated checking of one's face during a ritual of makeup application to hide blemishes (e.g., Phillips & Castle, 2002). Similar types of body evaluation, e.g., pinching one's thighs to check if they have increased in size, have

Correspondence to: Tom Hildebrandt.

been described in women with eating disorders (Mountford, Haase, & Waller, 2006). Research on these types of behavior suggest a strong association between high levels of body checking and other forms of eating disorder pathology (Mountford et al., 2006; Mountford, Haase, & Waller, 2007; Shafran, Fairburn, Robinson, & Lask, 2004). Consequently, the reduction of checking and avoidance has become common target for body image interventions in bulimia nervosa (BN; Rosen, 1997), BDD (Veale et al., 2001), and women with high shape and weight concern (Delinsky & Wilson, 2006).

Reas and colleagues (2002) developed a self-report measure (Body Checking Questionnaire; BCQ) to assess specific checking behaviors, and basic psychometric properties have been established in undergraduate (De Berardis et al., 2007; Reas et al., 2002), overweight (Latner, 2008), and binge eating disorder (Reas, White, & Grilo, 2006) samples. The 23-item measure appears to have three reliable highly inter-correlated factors (overall appearance, specific body parts, and idiosyncratic checking) and short-term test-retest reliability. In addition, previous research has demonstrated significant associations with shape/weight concerns and behavioral indicators of eating disorder pathology, upholding the validity of its scores. One limitation to the BCQ, however, is that the content of the items largely reflect rituals and body parts most relevant to women, which has lead to the recent development of a companion measure (Male Body Checking Questionnaire; MBCQ) by Hildebrandt, Walker, Alfano, Delinsky, and Bannon (2010) that includes items more consistent with the lean muscularity ideals common among men.

Men and women appear to have divergent body image concerns and evaluate different aspects of their respective appearances. Research has demonstrated that male and female "ideal bodies" differ significantly, with men generally desiring a body that is both lean and muscular while women tend to idealize thin physiques (Fallon & Rozin, 1985; Olivardia, 2002). Because of these differences in appearance concerns, it follows that men and women might check their appearances in different ways. For example, men may be more likely to check the hardness of their biceps or to compare their muscle size to others (e.g., Olivardia, 2001) than to check their thighs for cellulite or to check how their bottom looks in the mirror items which are generally considered to be areas of higher concern for women (Phillips & Diaz, 1997) and are assessed in most body checking measures (e.g., the BCQ, Reas et al., 2002). Even when men and women exhibit identical checking behaviors (e.g., weighing themselves on a scale), their motivations for doing so might be very dissimilar. Men may check with the hope of noticing an increase in their body weight (from increased muscle mass), while women typically desire a reduction in their body weight (Phillips & Diaz, 1997). In either scenario, the checking behavior can act much like pulling the handle on a slot machine, with the individual hoping to find desirable changes, but being at the mercy of random fluctuations in shape and or weight. The variable reinforcement of this behavior leads to excessive checking and a behavior that is resistant to extinction.

Despite apparent gender differences in body checking, each measure has its own global factor, which suggests that the severity of body checking appears to underlie patterns of body checking (Hildebrandt et al., 2010; Reas et al., 2002). The psychometric implications of gender in the measurement of these constructs remain unstudied. Item bias/measure bias and differential item functioning (DIF) have been used to study the impact of subgroups on measurement and ultimately to provide more accurate comparisons between subgroups. Differential item functioning occurs when individuals from different groups (including gender, race, ethnicity, and others) have nonequivalent item scores at the same level of the latent trait. Psychometrically, the presence of DIF can reduce the validity of a measure. When unrecognized or statistically controlled, DIF can lead to faulty conclusions (e.g., women have greater body image disturbance than men). In this study of gender differences, for a body checking item, DIF would exist when a woman with a specific degree of body

checking severity is more likely to score higher on a given item than a man with the same degree of body checking severity or when a certain item is more or less strongly related to this latent severity for women than men. The same logic would apply if the research examined DIF as a result of race or ethnicity. Gender-based differential item functioning has been found for measures such as the Anxiety Severity Index (Van Dam, Earleywine, & Forsyth, 2009), the Multidimensional Personality Questionnaire Stress Reaction Scale (Smith & Reise, 1998), and the diagnostic criteria for personality disorders (Jane, Oltmanns, South, & Turkheimer, 2007).

The purpose of this study is to explore gender-based DIF in the BCQ and MBCQ. Because the items, format, item response scale, and directions are similar in both measures, we will focus on identifying common factors among the entire item set and evaluating these items for DIF using a confirmatory factor analysis (CFA) model. Differential item functioning can be uniform or non-uniform. Non-uniform DIF occurs when groups have significantly different λ (factor loading/discrimination) parameters, and uniform DIF occurs when groups have significantly different τ (threshold/difficulty) parameters. Factor loadings can be conceptualized as the strength of the relationship between the underlying latent trait and the individual item, so non-uniform DIF indicates that the bias occurs differently across levels of the latent trait. Item thresholds can be conceptualized as the probability of endorsing a certain item category (e.g., "Always" in a 5-point ordinal scale) given a certain level of the latent trait, so uniform DIF occurs when there are the same group differences across all levels of the latent trait. A number of different methods exist for examining DIF, each with different advantages depending upon the specific research goal (see Teresi [2006] for a review). For example, multiple-indicator multiple-cause (MIMIC) and item response theory (IRT) models are also used to examine DIF, but CFA models allow for a more thorough testing of DIF than MIMIC models, including tests for both uniform and non-uniform DIF, and are better equipped to handle multidimensional data than the traditional unidimensional IRT models (Raju, Laffitte, & Byrne, 2002).

Method

Participants and Procedure

Men and women from several previous unpublished (observations from Hildebrandt and his colleagues) and published (Hildebrandt et al., 2009) studies of body checking were collapsed to yield an appropriate number of subjects for investigating DIF. All studies used college undergraduates recruited from the psychology participant pool and who completed the BCQ and MBCQ as part of their participation. There were no differences in inclusion or exclusion criteria.

A total of 1024 (n = 559 male; n = 465 female) college students were included and received course credit for their participation. All recruitment procedures were similar; participants responded to ads about a study examining body image. Participants completed paper and pencil questionnaires which included the BCQ and MBCQ. The racial and ethnic breakdown of the sample indicated a fair degree of diversity with the most common being White/Caucasian (44.4%, n = 455), followed in frequency by Asian/Pacific Islander (21.0%, n = 215), Hispanic/Latino(a) (12.2%, n = 125), and Black/African American (6.6%, n = 71). Participants were an average of 19.34 years old (SD = 2.42) with an average body mass index (BMI) of 23.65 (SD = 4.07) kg/m^2.

Measures

BCQ—The BCQ is a 23-item measure of body checking behavior that utilizes a 5-point ordinal scale from 1= "*never*" to 5 = "*always.*" Psychometric evaluations have reliably

produced a three-factor structure (overall appearance, specific body parts, and idiosyncratic checking) in both mixed gender and female-only samples (Calugi, Dalle Grave, Ghisi, & Sanavio, 2006; Grilo et al., 2005; Reas et al., 2002; Reas, White, & Grilo, 2006). The coefficient alphas among undergraduate populations range from .66–.92 with the idiosyncratic subscale showing the lowest internal consistency across studies (Calugi et al., 2006; Haase, Mountford, & Waller, 2007; Reas et al., 2002). The consistent replication of a three-factor solution using principal component analysis (PCA) and confirmatory factor analysis (CFA) suggests good factorial validity, although the factors are highly correlated (r = .70–.81) and there is evidence of a global higher order factor (Reas et al., 2002). The BCQ has reportedly good 1–2 week test-retest reliability in undergraduate populations for the subscales and overall sum score (r = .84; Calugi et al., 2006; Reas et al., 2006). Furthermore, the overall sum score and subfactor sum scores have moderate to high correlations with theoretically-related constructs including overvaluation of shape and weight, eating disorder symptoms, body checking cognitions, and physique anxiety (Calugi et al., 2006; Grilo et al., 2005; Haase, Mountford, & Waller, 2006, 2007; Reas et al., 2002, 2006) yielding evidence for its validity.

MBCQ—Hildebrandt et al. (2010) developed a companion measure to the BCQ that initially included 19 items, but after evaluation in three separate samples, a 16-item scale appeared to have the best psychometric properties. The response scale mirrors that of the BCQ. Hildebrandt et al. reported evidence of acceptable internal consistency (α = .72–.86) for the subscales (Global Muscle Checking, Chest and Shoulder Checking, Other Comparative Checking, and Behavioral Testing) and full scale as well as significant correlations with associated measures of muscle dysmorphia and eating disorder psychopathology in undergraduates. One week test-retest reliability of the sum scale also proved acceptable (r = .84) with similar results for the individual subscales (r = .68–.79). As with the BCQ, a series of exploratory and confirmatory factor analyses revealed evidence of a higher order global body checking factor with highly intercorrelated subfactors (r = .66–.88) yielding evidence for its validity.

Statistical Analyses

The first step was to establish the dimensionality and factor structure of the combined BCQ and MBCQ in order to establish an item set to be evaluated for DIF. Based on the existing psychometric investigations of both the BCQ and MBCQ, we sought to establish a more parsimonious factor structure than those previously described (Hildebrandt et al., 2010; Reas et al., 2002). Using previously published data, we identified 19 items between the two measures that represented global body checking constructs related to the female thinness ideal and the male muscularity ideal. These items were chosen based on three criteria. First, they were part of the measures' subscales with the highest eigenvalue. Second, these items were part of the measures' subscales with the highest loading on the second order factor. Third, the content of these items reflected global or overall body checking behaviors that would be expected in a general population. Items not included were nuanced and more closely tied to the psychopathology of eating disorders and muscle dysmorphia. The 19 items chosen included the items from the Overall Appearance (OA) subscale of the BCQ (items 3, 5, 8, 11–13, 15, 17, 20, and 21) and the Global Checking (GC) scale of the MBCQ (items 1–5, and 15) from the MBCQ. The items of the OA have held unidimensionality in previous mixed gender samples (Reas et al., 2002; Reas et al., 2006). The MBCQ items of the GC scale have shown some evidence of multidimensionality in women (Hildebrandt et al.). However, the PCA conducted by Hildebrandt and colleagues likely generated nuisance factors (see Thompson [2002]) as factors 2–5 all had eigenvalues close to 1.0, with negligible differences between them. BCQ items 11–13 also loaded highly on the GC factor among women and were part of a highly correlated additional factor among men (r = .82),

so we decided to include these items in our initial investigations of dimensionality and factor structure, assuming a more parsimonious grouping of items would reflect a single global checking factor.

Once suitable dimensionality was established, we used a multiple group CFA model to test for uniform and non-uniform DIF by evaluating parameter invariance. This procedure involved the calculation of a chi-square difference test ($\Delta\chi^2$) comparing a model where the parameter under investigation is held equivalent across groups and comparing it to a model where the same parameter is estimated freely in both groups. Based on recommendations for multiple-group CFA (Brown, 2006), we tested group differences following five steps: (1) factor structure in each group separately; (2) equivalence of λ parameter across groups (i.e., uniform DIF); (3) equivalence of τ parameters (non-uniform DIF); (4) equivalence of factor variances/covariances; and (5) equivalence of factor means. These steps are progressive, building upon the previous step. Because both measures use a 5-point ordinal scale, there are four τ parameters for each item. All modeling was conducted using *M*plus software version 4.2 (Muthén & Muthén, 1998–2007). There were no missing data on the measures included in analyses and model fit was assessed using root-mean square error of approximation (RMSEA ≤ .05 is good fit; Hu & Bentler, 1999) and comparative fit index (CFI; Bentler, 1990; scale 0–1.0, ≥ .95 is good fit, Hu & Bentler).

Results

The selected items were first subjected to CFAs in order to establish the appropriate factor structure. To establish dimensionality, we compared a series of competing models. First, a single factor model was estimated using all 19 items in the entire sample, but did not fit the data well, $\chi^2(95) = 635.10$, $p \le .001$; CFI = .73; RMSEA = .42. Results of a two-factor model, separating female (10 items) and male (9 items) body checking items into separate factors, indicated a better but still poor fit, $\chi^2(97) = 932.25$, $p \le .001$; CFI = .81; RMSEA = .17. The two latent variables were modeled based on the evidence for dimensionality from published studies of the BCQ and MBCQ. We labeled the first latent variable as *Female Body Checking Severity* (Female-BCS) and the second latent variable as *Male Body Checking Severity* (Male-BCS). We compared models with and without correlated factors and with and without cross-loading items. The best fitting model included correlated factors but no cross-loading items, $\chi^2(88) = 1,800.71$, $p \le .001$; CFI = .93; RMSEA = .09. The Chi-square difference test indicated that the correlated factor model provided a better fit to the data than a model with no cross loadings and uncorrelated factors, $\Delta \chi^2(1) = 521.60$, $p \le .001$. We ran the same series of tests independently with men and women. The correlated two-factor model fit the data well for men, $\chi^2(88) = 1,960.21$, $p \le .001$; CFI = .97; RMSEA = .04, and women, $\chi^2(88) = 1,899.95$, $p \le .001$; CFI = .96; RMSEA = .05.

We then estimated a two-factor multiple group CFA with all λ parameters, τ parameters, covariances (Ψ), and factor means (μ) freely estimated. To identify the model, we constrained the factor variances in males to 1.0 and the means to zero. Once we identified an item without significant DIF for each factor, we fixed that parameter to 1.0 to free up the variances and means. The multiple group CFA fit the data well, $\chi^2(195) = 2302.34$, $p \le .001$; CFI = .97; RMSEA = .05, and served as the baseline model for DIF tests. Table 1 reports the invariance tests for λ and τ parameters. Three items had significantly different λ between groups for the Female-BCS factor. Two additional items had significantly different λ and τ parameters between groups suggesting a mixture of uniform and non-uniform DIF in the Female-BCS factor. A similar pattern emerged for the Male-BCS factor. Two items had significantly different λ parameters and two additional items had significantly different λ and τ parameters. Ten items (five Female-BCS factor items and four Male-BCS items) were

unbiased measures of the latent factors indicating partial invariance in the measurement of both types of latent body checking severity.

A closer examination of the pattern and types of DIF for the Female-BCS items reveals that 'checking one's bottom in the mirror' and lying down to 'feel my bones touch the floor' are more strongly related to college women's Female-BCS. Furthermore, college women have a higher likelihood of engaging in these behaviors (i.e., lower τ parameters). For college men, Female-BCS is more strongly related to looking at others 'to see how my body size compares with their body size' and the using specific clothing to 'try on to make sure they still fit'. A similar examination of DIF for the Male-BCS factor indicates college men's latent severity is more strongly related to comparing 'the size of my chest muscles with others' and the comparison of 'overall leanness or muscle definition to others' than for college women. These behaviors are also more likely to occur among college men. Latent Male-BCS was also more strongly related to certain checking behaviors. Specifically, evaluating the hardness of one's biceps or the visibility of one's abdominal muscles were more strongly related to Male-BCS in college men than women.

Constraining the items without DIF to be equal between groups and freeing the identified DIF parameters, we then tested between-group differences in factor variances and covariances. Factor variance was not significantly different between groups for the Female-BCS factor, $\chi^2(1) = 3.72$, $p = .05$, or Male-BCS factor, $\chi^2(1) = 2.08$, $p = .15$. The covariance between the Female and Male BCS factors was significantly different between genders, $\chi^2(1) = 37.56$, $p \leq .001$. The relationship between Female and Male BCS among college women was significantly higher ($\Psi = .79$, $SE = .25$) than college men ($\Psi = .65$, $SE = .21$). Finally, we tested between-gender differences in each latent factor and found both to be significantly different for Female BCS, $\chi^2(1) = 129.72$, $p \leq .001$, and Male BCS, $\chi^2(1) = 203.72$, $p \leq .001$. After controlling for DIF, college men had a latent mean of $\mu = -0.35$, $SE = .14$ and college women had a latent mean of $\mu = 1.11$, $SE = .17$ for the Female-BCS factor. The standardized effect for this difference ($ES = -1.68$) indicates that the latent mean for females was approximately one and a half standard deviations higher than for males. The opposite pattern of results was found for Male-BCS factor means; college women had a latent mean of $\mu = -0.11$, $SE = .19$ and college men had a mean of $\mu = 1.00$, $SE = .12$. The standardized difference ($ES = 1.25$) indicates that men scored about one and a fourth standard deviations higher on the Male-BCS than females.

The multiple group CFA model provided a marginal fit to the data, but after controlling for DIF and differences between factor covariance and factor means, the final model provided a good fit to the data, $\chi^2(128) = 2,899.57$, $p \leq .001$; CFI = .99; RMSEA = .02. This finding suggests that the gender bias in specific items will contribute to poor fitting models when estimated in mixed gender samples. The related finding - that 10 items displayed no evidence of DIF - is also significant. The 10 items without significant DIF could be considered a purified item set that could be used in mixed gender samples of undergraduates. Coefficient alphas for the neutral item Male-BCS and Female-BCS were $\alpha = .96$ and $\alpha = .94$, respectively. The Female-BCS was significantly correlated with the BCQ total score ($r = .81$, $p \leq .001$) and the Male-BCS significantly correlated with the MBCQ total score ($r = .83$, $p \leq .001$). A single group CFA model with two latent factors (Female-BCS and Male-BCS) including only these 10 items assuming no DIF, provided an acceptable overall fit to the data, $\chi^2(97) = 773.87$, $p \leq .001$; CFI = .97; RMSEA = .05.

Discussion

Body checking describes an individual's attempt to assess aspects of his or her appearance through means such as looking in a mirror, weighing himself or herself, checking the

hardness and symmetry of one's biceps, and checking to see if certain clothes fit. The BCQ (Reas et al., 2002) is the most commonly used measure of body checking in both clinical and non-clinical samples of men and women (e.g., Reas et al., 2002; Reas et al., 2006), despite the potential for biased measurement of male body checking. This study is the first to identify DIF in the BCQ and the MBCQ, and the first, to our knowledge, to evaluate DIF in any body image measure. Analysis of the baseline model demonstrated differences between men and women's body image concerns consistent with those described in the literature, but raised questions about the validity of the BCQ or MBCQ in mixed gender samples. Whereas the items displaying DIF could be conceptualized as more specific to male or female body image, the final CFA model revealed 10 items that provide a satisfactory measure of gender-neutral body checking severity. The recommendations based on these findings are to avoid using the BCQ and MBCQ in mixed gender samples unless measurement bias is controlled. If the specific study aims require a mixed gender sample, these findings suggest that utilizing the 10 neutral items are likely to provide the most parsimonious measure of global female and male body checking severity.

The items with significant DIF suggest interesting patterns. The practices of lying on the floor to evaluate whether one's bones touch or checking one's bottom in the mirror appear to be strong indicators of female body checking and occur more frequently among women than men. The measurement of these behaviors, in addition to checking one's body in reflective surfaces, trying on specific clothes to evaluate body size, and comparing one's overall size to others appeared to be the sources of measurement bias in female body checking severity. These differences may reflect the divergent ways men and women evaluate thinness or overall appearance. Evaluating chest muscles and overall muscle leanness appear to be stronger indicators of body checking in men than women. Not surprisingly, these behaviors were more common among men as well, even after controlling for latent checking severity. The items with DIF in each scale appear to reflect stereotypically male and female body checking motivated by evaluating either the female thinness ideal or the male lean muscular ideal.

The identification of items that do not show DIF, however, suggests certain behaviors may be measured without gender bias and this is especially important to consider in context of measure development and for subsequent theory about group differences in body image constructs. These Female–BCS behaviors included evaluating oneself from different angles or different positions, pulling one's clothes tight to see how they look, asking others' about their weight or size, checking for body fat, and flattening one's stomach. The Male-BCS behaviors included flexing bicep muscles, evaluating shoulder broadness, chest leanness, and identifying muscle definition. After controlling for latent severity, the measurement properties of these items are equivalent, which may reflect their ability to tap into checking of male or female versions of thinness and lean muscularity ideals. This interpretation assumes that there is some flexibility in the desired bodies of both men and women. For instance, some women may prefer or idealize an athletic body where muscle definition and size are important indicators. Alternatively, some men may prefer or idealize a thinner body that is marked by a lack of body fat and flat stomach. This potential flexibility in ideals is evident in weightlifting populations of men (Hildebrandt, Alfano, & Langenbucher, 2010; Hildebrandt, Schlundt, Langenbucher, & Chung, 2006), but little research on the flexibility of the thinness ideal has been done with women. However, some researchers have noted the drive for muscularity among women (e.g., McCreary & Saucier, 2009), but the link to a separate ideal is uncertain. This same flexibility may also reflect gender-influenced expressions of certain types of psychopathology, such as eating disorders (Hildebrandt & Alfano, 2009).

There were several limitations to this study. First, the sample consists only of male and female undergraduates, which may limit the generalizability of these findings, although psychometric evaluations of body checking measures suggest comparable measurement properties (e.g., factor structure) of body checking in clinical and non-clinical groups. Second, we evaluated DIF in a select group of body checking items. We sought parsimony in our model; however, it is possible that different patterns of DIF would emerge when evaluated within a more complex multidimensional structure. Finally, the current study approached the investigation of DIF from a multiple group CFA approach. It is possible that other statistical approaches to studying DIF would provide different results or provide other relevant information about these latent traits. In particular, the use of item response theory (IRT) can provide important information about the value of each item that is not inherent in CFA approaches (Teresi, 2006).

Nevertheless, it is clear that (a) in general, body checking appears to manifest itself differently between men and women at varying levels of body checking severity and (b) the BCQ and MBDQ may not accurately capture frequency and severity of body checking as standalone measures of body checking. This is a particularly relevant issue when measuring evidence of body image concern and dissatisfaction in certain clinical populations, such as those with eating disorders or muscle dysmorphia (Phillips, O'Sullivan, & Pope, 1997). Thus, data from the current study demonstrate that gender is a factor that cannot be ignored when assessing body checking behavior and is suggestive of the possibility that body image measures may have gender-based DIF. Psychometric evaluations of existing measures could identify DIF which would offer options for increased validity. In addition, research reporting on new measures should use DIF studies as part of the standard development and validation process. The existence of a purified item-set, however, does provide an opportunity to measure body checking in mixed-gender samples. These 10 items could be used for building alternative assessment methods designed to better measure body checking severity such as in a computer adaptive format where unbiased item sets are essential to evaluate severity in heterogeneous samples (Gibbons, Rush, & Immekus, 2008). Finally, the DIF identified in the current study provides an opportunity for theoretical exploration of the body checking construct. It is possible that a number of different factors explain why men and women differ in their pattern of body checking, and the study of these patterns could help explain gender differences in certain types of psychopathology where body checking occurs.

Research Highlights

- Body checking is a symptom of body image disturbance that when measured via self-report questionnaire demonstrates measurement bias

- The bias found in body checking measures reflects underestimation of checking severity for opposite sex

- This measurement bias can be removed by using a 10-item subset of existing measures of body checking

Acknowledgments

References

Ackerman, TA. Multidimensional item response theory modeling. In: McArdle, JJ.; Maydeu-Olivares, A., editors. Contemporary psychometrics. New York: Routledge Press; 2005. p. 3-26.

Bentler PM. Comparative fit indices in structural models. Psychological Bulletin 1990;107:238–246. [PubMed: 2320703]

Brown, T. Confirmatory factor analysis for applied research. New York: Guilford Press; 2006.

Calugi S, Dalle Grave R, Ghisi M, Sanavio E. Validation of the Body Checking Questionnaire (BCQ) in an eating disorders population. Behavioural and Cognitive Psychotherapy 2006;34:233–242.

Cash, TF. Cognitive-behavioral perspectives on body image. In: Cash, TF.; Pruzinsky, T., editors. Body image: A handbook of theory, research, and clinical practice. New York, NY: The Guilford Press; 2002. p. 38-46.

De Berardis D, Carano A, Gambi F, Campanella D, Giannetti P, Mancini E, …Ferro FM. Alexithymia and its relationships with body checking and body image in a non-clinical female sample. Eating Behavior 2007;8:296–304.

Dagne GA, Howe GW, Brown CH, Muthén BO. Hierarchical modeling of sequential behavioral data: An empirical Bayesian approach. Psychological Methods 2002;7:262–280. [PubMed: 12090414]

Delinsky SS, Wilson GT. Mirror exposure for the treatment of body image disturbance. International Journal of Eating Disorders 2006;39:108–116. [PubMed: 16231342]

Fallon AE, Rozin P. Sex differences in perceptions of desirable body shape. Journal of Abnormal Psychology 1985;91:102–105. [PubMed: 3980849]

Gibbons RD, Rush AJ, Immekus JC. On the psychometric validity of the domains of the PDSQ: An illustration of the bi-factor item response theory model. Journal of Psychiatric Research 2008;43:401–410. [PubMed: 18554611]

Gregorich S. Do self report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. Medical Care 2006;44:S78–S94. [PubMed: 17060839]

Grilo CM, Reas DL, Brody ML, Burke-Martindale CH, Rothschild BS, Masheb RM. Body checking and avoidance and the core features of eating disorders among obese men and women seeking bariatric surgery. Behaviour Research and Therapy 2005;43:629–637. [PubMed: 15865917]

Haase AM, Mountford V, Waller G. Understanding the link between body checking cognitions and behaviors: The role of social physique anxiety. International Journal of Eating Disorders 2007;40:241–246. [PubMed: 17167755]

Hildebrandt T, Alfano L. A review of men and boys with eating disorders: Working towards an empirically derived diagnostic system. International Journal of Child and Adolescent Health 2009;2:185–196.

Hildebrandt T, Alfano L, Langenbucher JW. Body image disturbance in 1000 male appearance and performance enhancing drug users. Journal of Psychiatric Research 2010;44:841–846. [PubMed: 20110092]

Hildebrandt T, Langenbucher JW, Carr SJ, Sanjuan P. Modeling population heterogeneity in appearance-and performance-enhancing (APED) use: Applications of mixture modeling in 400 regular APED users. Journal of Abnormal Psychology 2007;116:717–733. [PubMed: 18020718]

Hildebrandt T, Schlundt D, Langenbucher JW, Chung T. Presence of muscle dysmorphia symptoms among male weightlifters. Comprehensive Psychiatry 2006;47:127–135. [PubMed: 16490571]

Hildebrandt T, Walker DC, Alfano L, Delinsky S, Bannon K. Development and validation of a male specific body checking questionnaire. International Journal of Eating Disorders 2010;43:77– 87. [PubMed: 19247988]

Hu L, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus alternatives. Structural Equation Modeling 1999;6:1–55.

Jane JS, Oltmanns TF, South SC, Turkheimer E. Gender bias in diagnostic criteria for personality disorders: An item response theory analysis. Journal of Abnormal Psychology 2007;116:116–175. [PubMed: 17324022]

Jöreskog KG, Goldberger AS. Estimation of a model with multiple indicators and multiple causes of a single latent variable. Journal of the American Statistical Assocation 1975;70:631–639.

Latner JD. Body checking and avoidance among behavioral weight-loss participants. Body Image 2008;5:91–98. [PubMed: 18405867]

McCreary DR, Saucier DM. Drive for muscularity, body comparison, and social physique anxiety among men and women. Body Image: An International Journal of Research 2009;6:24–30.

Mislevy R. Bayes modal estimation in item response models. Psychometrika 1986;51:177–195.

Mountford V, Haase A, Waller G. Body checking in the eating disorders: Associations between cognitions and behaviors. International Journal of Eating Disorders 2006;39:708–715. [PubMed: 16868998]

Mountford V, Haase A, Waller G. Is body checking in the eating disorders more closely related to diagnosis or to symptom presentation? Behaviour Research and Therapy 2007;45:2704–2711. [PubMed: 17765868]

Muthén, B. Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In: Wainer, H.; Braun, H., editors. Test validity. Hillsdale NJ: Erlbaum Associates; 1988. p. 213-238.

Muthén B. Beyond SEM: General latent variable modeling. Behaviormetrika 2002;29:81–117.

Muthén B, Kao CF, Burstein L. Instructional sensitivity in mathematics achievement test items: Applications of a new IRT-based detection technique. Journal of Educational Measurement 1991;28:1–22.

Muthén, LK.; Muthén, BO. Mplus user's guide. 5. Los Angeles, CA: Muthén and Muthén; 1998–2007.

Muthén B, Speckart G. Latent variable probit ANCOVA: Treatment effects in the California Civil Addict Programme. British Journal of Mathematical and Statistical Psychology 1985;38:161–170. [PubMed: 4074657]

Olivardia, R. Body image and muscularity. In: Cash, TF.; Pruzinsky, T., editors. Body image: A handbook of theory, research, and clinical practice. New York: Guilford Press; 2002. p. 210-218.

Olivardia R. Mirror, mirror on the wall, who's the largest of them all? The features and phenomenology of muscle dysmorphia. Harvard Review of Psychiatry 2001;9:254–259. [PubMed: 11553529]

Phillips, KA.; Castle, DJ. Body dysmorphic disorder. In: Castle, DJ.; Phillips, KA., editors. Disorders of body image. Philadelphia, PA: Wrightson Biomedical Publishing; 2002. p. 101-120.

Phillips KA, Diaz SF. Gender differences in body dysmorphic disorder. Journal of Nervous Mental Disorders 1997;185:570–577.

Phillips KA, O'Sullivan RL, Pope HG Jr. Muscle dysmorphia. Journal of Clinical Psychiatry 1997;58:361. [PubMed: 9515975]

Raju NS, Laffitte LJ, Byrne BM. Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. Journal of Applied Psychology 2002;87:517–529. [PubMed: 12090609]

Reas DL, Whisenhunt BL, Netemeyer R, Williamson DA. Development of the body checking questionnaire: A self-report measure of body checking behaviors. International Journal of Eating Disorders 2002;31:324–333. [PubMed: 11920995]

Reas DL, White MA, Grilo CM. Body checking questionnaire: Psychometric properties and clinical correlates in obese men and women with binge eating disorder. International Journal of Eating Disorders 2006;39:326–331. [PubMed: 16528695]

Rosen JC. Body image assessment and treatment in controlled studies of eating disorders. International Journal of Eating Disorders 1997;20:331–343. [PubMed: 8953321]

Samejima, F. Graded response model. In: van der Linden, W.; Hambleton, RK., editors. Handbook of modern item response theory. New York: Springer; 1997. p. 85-100.

Shafran R, Fairburn CG, Robinson P, Lask B. Body checking and its avoidance in eating disorders. International Journal of Eating Disorders 2004;35:93–101. [PubMed: 14705162]

Smith LL, Reise SP. Gender difference on negative affectivity: An IRT study of differential item functioning on the Multidimensional Personality Questionnaire Stress Reaction Scale. Journal of Personality and Social Psychology 1998;75:1350–1362. [PubMed: 9866192]

Teresi JA. Overview of quantitative measurement methods - Equivalence, invariance, and differential item functioning in health applications. Medical Care 2006;44:S39–249. [PubMed: 17060834]

Thompson, B. Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington DC: American Psychological Association; 2004.

Van Dam NT, Earleywine M, Forsyth JP. Gender bias in the sixteen-item anxiety sensitivity index: An application of polytomous differential item functioning. Journal of Anxiety Disorders 2009;23:256–259. [PubMed: 18786807]

Veale D, Riley S. Mirror, mirror on the wall, who is the ugliest of them all? The psychopathology of mirror gazing in body dysmorphic disorder. Behaviour Research and Therapy 2001;39:1381–1393. [PubMed: 11758697]

Table 1

Multiple Group Confirmatory Factor Analysis for Two-Factor Model of BCQ and MBCQ Items

BCQ/MBCQ Item	Male Model					Female Model				
	λ (SE)	τ_1 (SE)	τ_2 (SE)	τ_3 (SE)	τ_4 (SE)	λ (SE)	τ_1 (SE)	τ_2 (SE)	τ_3 (SE)	τ_4 (SE)
BCQ 3: I have special clothes which I try on to make sure they still fit.	.799 (.025)	.316 (.041)	.925 (.044)	1.289 (.056)	1.999 (.080)	.657 (.024)	.316 (.041)	.925 (.044)	1.289 (.056)	1.999 (.080)
BCQ 5: I check my reflection in glass doors or car windows to see how I look.	.661 (.036)	−.915 (.061)	−.338 (.049)	.392 (.049)	1.158 (.068)	.563 (.067)	−.915 (.061)	−.338 (.049)	.392 (.049)	1.158 (.068)
BCQ 8: I look at others to see how my body size compares with their body size.	.754 (.035)	−.899 (.062)	−.137 (.053)	.602 (.057)	1.578 (.086)	.654 (.052)	−.899 (.062)	−.137 (.053)	.602 (.057)	1.578 (.086)
BCQ 11: I ask others about their weight/clothing size so I can compare my weight/size	.612 (.029)	.608 (.048)	1.282 (.064)	1.853 (.096)	2.225 (.122)	.612 (.029)	.608 (.048)	1.282 (.064)	1.853 (.096)	2.225 (.122)
BCQ 12: I check to see how my bottom looks in the mirror.	.451 (.042)	.184 (.053)	.943 (.061)	1.898 (.099)	3.305 (.228)	.711 (.034)	−.172 (.053)	.041 (.063)	1.118 (.092)	2.112 (.109)
BCQ 13: I practice sitting/standing in various positions to see how I look in each position.	.698 (.022)	.159 (.039)	.790 (.041)	1.777 (.071)	2.966 (.103)	.698 (.022)	.159 (.039)	.790 (.041)	1.777 (.071)	2.966 (.103)
BCQ 15: I check to see if my fat jiggles.	0.511 (.034)	.604 (.049)	1.253 (.070)	1.820 (.101)	2.277 (.120)	0.511 (.034)	.604 (.049)	1.253 (.070)	1.820 (.101)	2.277 (.120)
BCQ 17: I suck in my gut to see what it is like when my stomach is completely flat.	.560 (.032)	−.045 (.050)	.600 (.050)	1.323 (.069)	1.919 (.110)	.560 (.032)	−.045 (.050)	.600 (.050)	1.323 (.069)	1.919 (.110)
BCQ 20: I lie down on the floor to see if I can feel my bones touch the floor.	.701 (.037)	1.834 (.048)	2.758 (.042)	2.998 (.071)	3.542 (.97)	.977 (.033)	−.049 (.031)	.575 (.042)	1.523 (.070)	2.005 (.97)
BCQ 21: I pull my clothes as tightly as possible around myself to see how I look.	.648 (.020)	.155 (.037)	.805 (.044)	1.767 (.071)	2.285 (.110)	.648 (.020)	.155 (.037)	.805 (.044)	1.767 (.071)	2.285 (.110)
MBCQ 1: I check the hardness of my biceps to ensure I have not lost any muscle.	.741 (.018)	−.778 (.055)	−.203 (.049)	.719 (.056)	1.637 (.088)	.666 (.020)	−.778 (.055)	−.203 (.049)	.719 (.056)	1.637 (.088)
MBCQ 2: I look at my abdominal muscles (6-pack) in the mirror.	.664 (.071)	−.795 (.053)	−.280 (.047)	.502 (.051)	1.186 (.068)	.464 (.050)	−.795 (.053)	−.280 (.047)	.502 (.051)	1.186 (.068)
MBCQ 3: I flex my biceps when looking in the mirror to ensure symmetry of my muscles.	.871 (.013)	−.622 (.052)	.034 (.049)	.626 (.056)	1.187 (.070)	.871 (.013)	−.622 (.052)	.034 (.049)	.626 (.056)	1.187 (.070)
MBCQ 4: I compare the size of my muscles to others	.791 (.017)	−.885 (.057)	−.200 (.049)	.626 (.055)	1.608 (.084)	.791 (.017)	−.885 (.057)	−.200 (.049)	.626 (.055)	1.608 (.084)

BCQ/MBCQ Item	Male Model					Female Model				
	λ (SE)	τ_1 (SE)	τ_2(SE)	τ_3 (SE)	τ_4 (SE)	λ (SE)	τ_1 (SE)	τ_2(SE)	τ_3 (SE)	τ_4 (SE)
MBCQ 5: I compare my overall leanness or muscle definition to others.	.803 (.015)	−1.190 (.061)	−.325 (.050)	.667 (.057)	1.664 (.089)	**.713 (.014)**	**−.260 (.055)**	**−.025 (.051)**	**1.310 (.052)**	**1.988 (.081)**
MBCQ 11: I compare the leanness or definition of my chest muscles with others.	.707 (.022)	−.233 (.047)	.439 (.051)	1.347 (.072)	2.039 (.119)	.707 (.022)	−.233 (.047)	.439 (.051)	1.347 (.072)	2.039 (.119)
MBCQ 12: I compare the size of my chest muscles with others.	.901 (.010)	.190 (.038)	.777 (.045)	1.430 (.051)	2.094 (.101)	**.787 (.011)**	**.782 (.034)**	**.977 (.041)**	**1.880 (.052)**	**3.004 (.91)**
MBCQ 13: I compare the broadness of my shoulders with others.	.704 (.022)	−.260 (.052)	.292 (.051)	1.109 (.063)	1.969 (.114)	.704 (.022)	−.260 (.052)	.292 (.051)	1.109 (.063)	1.969 (.114)
MBCQ 15: I flex my muscles when looking in the mirror to find lines or striations in the muscle.	.685 (.026)	−.106 (.050)	.408 (.050)	1.029 (.063)	1.678 (.090)	.685 (.026)	−.106 (.050)	.408 (.050)	1.029 (.063)	1.678 (.090)

Notes. CFA = confirmatory factor analysis. BCQ = Body Checking Questionnaire. MBCQ = Body Checking Questionnaire. **BOLD** = Item parameter was nonequivalent using chi-square difference test ($p <$ .001) and significantly different between genders. The remaining items were equivalent and showed no evidence of differential item functioning. BCQ measure latent Female-Body Checking Severity and MBCQ items measure latent Male-Body Checking Severity. λ = the factor loading or strength of the relationship between the item and male or female body checking severity. τ = the probability threshold for endorsing the ordinal

category of the item response over the categories. For example τ_1 = the logit parameter or log odds of endorsing "sometimes" over endorsing "rarely".