2020

**Male figural rating scales: A critical review of the literature**

Daniel Talbot  
*The University of Notre Dame Australia, daniel.talbot@nd.edu.au*

John Cass

Evelyn Smith

Follow this and additional works at: [https://researchonline.nd.edu.au/sci_article](https://researchonline.nd.edu.au/sci_article)

This article was originally published as:  

Original article available here:  
10.1017/bec.2020.5

This article is posted on ResearchOnline@ND at [https://researchonline.nd.edu.au/sci_article/74](https://researchonline.nd.edu.au/sci_article/74). For more information, please contact researchonline@nd.edu.au.
This is the author’s version of the following article, as accepted for publication.


This article was published in *Behaviour Change*, 17 April, 2020.

Published version available online at: -

https://doi.org/10.1017/bec.2020.5
Background

Body dissatisfaction is typically conceptualized as a female-centric issue, with male presentations inaccurately represented as atypical and rare (Murray et al., 2017). Contrary to this representation, research has shown consistently that body dissatisfaction in male populations is highly prevalent and increasing (del Mar Bibiloni, Coll, Pich, Pons, & Tur, 2017; Frederick et al., 2007; Murray et al., 2017; Watkins, Christie, & Chally, 2008). Body dissatisfaction can be defined as a negative subjective evaluation of one’s body as a whole, or relating to specific aspects of one’s body such as body size, shape, muscularity/muscle tone, and weight (Grogan, 2016). Prior research has found that body dissatisfaction is associated with elevated emotional distress, preoccupation with appearance, and cosmetic surgery (Hoffman & Brownell, 1997; J. K. Thompson, Heinberg, Altabe, & Tantleff-Dunn, 1999) and is a risk factor for developing disordered eating behaviors (Grilo, Masheb, & Wilson, 2001; Mayo & George, 2014), muscle dysmorphia (Murray, Rieger, Touyz, & De la Garza Garcia Lic, 2010), depression (McCreary & Sasse, 2000; Olivardia, Pope, Borowiecki III, & Cohane, 2004; Presnell, Bearman, & Stice, 2004). Additionally, body dissatisfaction has been associated with obesity (Mond, van den Berg, Boutelle, Hannan, & Neumark-Sztainer, 2011; Wardle & Cooke, 2005), poorer mental and physical health-related quality of life (Griffiths et al., 2016; Griffiths, Henshaw, McKay, & Dunn, 2017).

Although there is some commonality, body dissatisfaction manifests differently in males and females. This is largely due to the differences in shape and composition of body ideals between sexes. The majority of men tend to idealize what is known as a mesomorphic build. The mesomorphic build is defined by a low percentage of body fat, combined with a defined, visible, but not excessive muscle build. Specifically, well-developed muscles on the chest, shoulders, arms, and slim waist and hips, and a V-shaped torso (Cafri & Thompson, 2004; Grogan & Richards, 2002; Pope, Phillips, & Olivardia, 2000). In Western society most
women also idealize a low percentage of body fat and a thin waist, but strive for an hourglass shape (as opposed to a mesomorphic shape). Women also aspire to a toned, muscular definition, however their muscular preference tends to be less than their male counterparts (Bozsik, Whisenhunt, Hudson, Bennett, & Lundgren, 2018; Swami et al., 2010). In this way, both sexes may be dissatisfied with their amount of body fat, with the size of their waist, and with their muscle tone. However, men’s dissatisfaction is more characterized by muscle shape and size, whilst women’s dissatisfaction traditionally has a greater focus on a slim build (Grogan, 2016).

Figural rating scales are a prominent method used to evaluate body dissatisfaction. Figural rating scales are typically comprised of a series of frontal view body images graded thin to obese (in men and women), and from thin to muscular (in men; Grogan, 2016). Once presented with these images, an individual is to indicate which of the figures best represents their current body (their perceived body), and which best represents the body that they would like to have (their desired body), respectively. The difference between an individual’s perceived and desired body is referred to as perceived–desired body discrepancy. Perceived–desired body discrepancy has been conceptualized as an index of body dissatisfaction, with greater discrepancies indicating greater levels of body dissatisfaction (Fingeret, Gleaves, & Pearson, 2004; Gardner & Brown, 2010). However, this conceptualization is questionable. Conceivably, most individuals would desire a ‘better’ body (e.g., a skinnier or and/or more muscular body). However, it cannot be assumed for every individual that a failure to embody an idealized figure results in, or is directly related to, body dissatisfaction - it is possible for one to desire a particular body type without being clinically dissatisfied with one’s own body. That said, discrepancy scores have been found to correlate highly with a number of measures related to body dissatisfaction, including eating disorder symptoms (Smith, Hawkeswood, Bodell, & Joiner, 2011), drive for muscularity (Gillen & Markey, 2015; Hildebrandt,
Langenbcher, & Schlundt, 2004; Novella, Gosselin, & Danowski, 2015), body appreciation (Mutale, Dunn, Stiller, & Larkin, 2016; Novella et al., 2015; Swami, Salem, Furnham, & Tovee, 2008), body mass index (BMI) and body fat percentage, and with direct measures of body dissatisfaction (Talbot, Cass, & Smith, 2019). Therefore, discrepancy scores provided by figural rating scales are a valuable screening tool for indicating body dissatisfaction and related psychological and physiological variables. Additionally, figural rating scales offer advantages over self-report measures of body dissatisfaction in that they are quick to administer (administration time is typically under one minute), they can be used by adolescents, allow for a visual depiction of an individual’s perceived body image, and are not dependent on language or literacy skills (Grogan, 2016).

Figural rating scales have also been utilized in eating disorders samples to measure body image disturbance. Typically, results indicate that there is a significant difference in participants’ selected perceived body, and their actual body composition (Cafri & Thompson, 2004). Moreover, eating disorder symptoms are significantly positively associated with figural rating scale index scores, meaning that greater differences between selected perceived and desired body selections are associated with greater eating disorder symptomatology (Baranowksi, Jorga, Djordjevic, Marinkovic, & Hetherington, 2003; Garner, Olmstead, & Polivy, 1983; Jung, Forbes, & Lee, 2009).

Historically, figural rating scales were produced depicting female bodies, and varying in terms of body fat percentage (Stunkard, 1983). However, due to evidence indicating a significant presence of body image issues and related disordered eating behaviors amongst the male population (Murray et al., 2017), there has been an increase in the development of figural rating scales for men (for review of hand-drawn silhouette figural rating scales, see Gardner & Brown, 2010). Initially, male figural rating scales mirrored female scales, in that they typically depicted a series of male bodies systematically varying in body fat percentage.
(e.g., Stunkard’s male figural rating scale; Stunkard, 1983). Although useful, these scales fail to include muscularity as a factor. Given that muscularity is an important consideration in the manifestation of body dissatisfaction in men (Ridgeway & Tylka, 2005), the inclusion of variations in muscularity may be essential for capturing a true reflection of male body image (Drewnowski & Yee, 1987; Stunkard, Sorensen, & Schulsinger, 1983; M. A. Thompson & Gray, 1995). Lynch and Zellner (1999) produced the first figural rating scale which held body fat constant and graded increasing muscle mass. Both Stunkard (1983) and Lynch and Zellner (1999) employed unidimensional variation in their figural rating scales based on a single biometric variable (body fat and muscularity, respectively). Since the publication of Lynch and Zellner’s scale, there have been many different male figural rating scales developed and published. However, many of these scales are still limited in psychometric evidence, design, and image quality.

Various male figural rating scales either fail to provide, or return poor validity and/or reliability, and as such may not provide valid or consistent indices of body dissatisfaction. In relation to scale design, the number of images comprising the figural rating scale is an important consideration. Too few images (i.e. representing discrete increments along the body dimension) results in a coarse scale – if there are not enough body type options for respondents to choose from then the scale cannot be as sensitive in determining differences in perceived and desired body selections between respondents. Too many images may cause problems for reliability (Cafri & Thompson, 2004). Ambrosi-Randic, Pokrajac-Buljan, and Taksic (2005) investigated the optimal number of figural rating scale stimuli for body size assessment, comparing unidimensional scales containing three, five, seven, and nine figures, respectively. The perceived-ideal discrepancy score for each of these four scales was calculated ($n = 320$), and then correlated with two measures of self-report body dissatisfaction. The authors found that discrepancy scores for scales containing five, seven,
and nine figures were highly, and equivalently correlated with each measure of body dissatisfaction, and thus concluded that the optimal number of images for a unidimensional figural rating scale was seven, plus or minus two. To the authors’ knowledge, no study has examined the optimal number of figural rating scale stimuli for bi-dimensional figural rating scales.

Likewise, the representation of the images themselves is another important consideration. Prior studies have evidenced the significance of muscular shape and tone in male body image (Cho & Lee, 2013; Ridgeway & Tylka, 2005). Therefore, visual access to muscle groups such as the abdominal and pectoral muscles may be essential for making an informed decision about one’s perceived and desired body. Many of the existing figural rating scales fail to grant access to these body areas, instead representing bodies covered by clothing.

Another important design consideration lies in whether figural rating scales are ordinal or interval measures. An ordinal figural rating scale presents with no regular change in body size between adjacent body images, as opposed to an interval scale in which body composition increases in regular intervals between figures. Interval figural rating scales are ideal as this allows for perceived, desired and discrepancy scores to be used in parametric statistical analyses (Gardner, 1975).

The medium for the body stimuli included in scales presents as another significant aspect in scale design. Hand-drawn body stimuli may pose problems for validity and reliability. The majority of hand-drawn body stimuli are asymmetric to varying, but not non-specified degrees. Previous research has demonstrated that bilateral body asymmetry can influence the ratings of bodies and faces, with asymmetrical bodies and faces being rated as less attractive (Rhodes & Simmons, 2007; Tovee, Tasker, & Benson, 2000). Therefore, it is essential that bodies included in figural rating scales do not vary in terms of extent of
symmetry, else this may influence perceived and desired body selections. Based on this premise, symmetrical figures such as computer-rendered figures and photograph figures should be employed. Additionally, these figures typically provide greater and more accurate anatomical detail (e.g., specific muscle groups), which is essential in order for respondents to make an accurate selection of their perceived and ideal bodies. The inclusion of muscular shape and tone is essential as prior studies have found for men, the muscular definition and leanness of the abdominal region, arms, and chest is key in defining the ideal male body (Grogan & Richards, 2002; Ridgeway & Tylka, 2005), and therefore would be key in selecting their ideal body from those presented in figural rating scales.

Given the parallel increase in recognition of male body image concerns and the development of figural rating scales for men, combined with the various limitations of these scales described above, this review aims to (i) examine the design of each figural rating scale, including whether body fat and/or muscularity is measured, the nature of the stimulus, number of dimensions, number of images, whether chest, arms and legs are displayed, and whether the scale is interval or ordinal. Additionally, the present study aims to: (ii) examine the psychometric properties of established figural rating scales for men; (iii) examine the ecological validity, that is, the stimuli quality of rating scales; and (iv) examine the relationship between figural rating scale scores and eating disorder symptoms.

A set of criteria were designed to review figural rating scales. These criteria were based on established threshold standards for scale psychometric properties, established design standards for self-report scales, and critiques and recommendations of figural rating scale design outlined in the extant literature (Gardner & Brown, 2010; Gardner, Friedman, & Jackson, 1998). Based on these criteria, it is recommended that a figural rating scale should be a bi-axial scale (i.e., representing variations in both body fat and muscularity in uni- or bi-dimensional format), with a minimum of seven realistic computer or photo generated images
of bare skinned bodies, on an interval scale. Psychometric properties evaluated include concurrent and convergent validity, and test-retest and internal reliability. Concurrent and convergent validity should demonstrate correlation coefficients greater than .50 with biometric measures of body composition (such as Body Mass Index and Fat Free Mass Index), and self-report measures of body dissatisfaction and related constructs such eating disorder symptoms, respectively. Test-retest and internal reliability should demonstrate correlations of .80 or greater (Carmines, 1990).

**Procedure**

A systematic review of the literature on male figural rating scales was conducted using PsycInfo, PubMed, and Google Scholar. This search was conducted on articles published from 1983 (Stunkard’s first figural rating scale) until December 2018. The following search terms were used: “male figural rating scales” and “male figure rating scales”, “male silhouette rating scales”, or “male body dissatisfaction scales”. The literature searches were restricted to articles focusing on adults that were written in English and published in peer-reviewed journals. After culling irrelevant articles from the initial search, 43 relevant articles were examined (for extensive literature search details, see Appendix A). For each article identified, the reference list was subsequently examined for relevant articles that were potentially omitted by the initial search. This resulted in the identification of four additional articles. After excluding seven duplicates, 40 articles were identified. Each article was then examined in line with the following criteria: (i) the article included the development and validation of an adult male figural rating scale; and (ii) psychometric properties for reliability and/or validity evidence were examined. Consequently, a total of 20 scales were excluded including 14 scales that utilized female-only stimuli, two scales utilizing child body stimuli, and four scales that did not report psychometric properties of reliability or validity.
This resulted in 20 articles being selected to include in the review. Figure 1 displays the process of article selection.

[Insert Figure 1]

From the articles included in the review, the following data were collected: (i) the author(s) names, and year of publication of the study (the original author(s) of the scale are shown in bold, psychometric values provided from studies succeeding the original publication are shown in the row below the original study); (ii) the name of the figural rating scale; (iii) the number of participants included in the study; (iv) age range or mean of the validation sample in years; (v) the type of sample used when examining validity and/or reliability evidence of the measure; (vi) body dimensions of body fat and/or muscularity represented in the figural rating scale; (vii) stimulus type of figural rating scale, indicating whether the stimuli are hand-drawn, or computer rendered, silhouettes or detailed figures; (viii) whether the figural rating scales are uni- or bi-dimensional; (ix) the test-retest reliability values (test-retest timespan displayed in parentheses); (x) whether bodies in the figural rating scale have bare chest, arms, and legs displayed; (xi) whether the scales are ordinal or interval; (xii) the concurrent and convergent validity values; and (xiii) the correlations between the figural rating scale index score (index score = perceived body minus desired body) and eating disorder symptoms. Convergent validity was generally taken as the correlation between a figural rating scale index scores and other tools that measure body dissatisfaction. Concurrent validity was taken as the association between perceived body figural rating scale selection and body mass index (BMI), or fat free mass index (FFMI). In this way, concurrent validity described the ability of a responder to use a given figural rating scale to select a body that approximated their actual body composition (in terms of body fat and/or muscularity). Ecological validity is defined as the extent to which the materials used in a scale approximate the real-world material (Reis & Judd, 2000). For the present study, ecological validity
concerns the degree to which the male body stimuli used to construct a given figural rating scale accurately represents a human male body.

Where a range of values is presented for validity evidence, multiple measures of validity evidence were provided in the study. For example, if a study provided three correlational values for concurrent validity, \( r = .35, .45, .55 \), these values would be presented as .35-.55.

**Results**

Table 1 displays descriptive information relating to the validation and reliability assessment of included scales, and the psychometric properties of known male figural rating scales. Of note, each figural rating scale was assigned a number, displayed in the first column of Table 1.

**Scale Design**

Of the 20 figural rating scales, 10 measured body fat exclusively, one measured masculinity exclusively, and nine measured both body fat and masculinity. Seventeen of the scales were unidimensional, meaning that bodies were presented varying on one axis. Three scales were bi-dimensional, with bodies presented in grid form varying in different combinations of body fat and masculinity.

Figural rating scales differed in terms of number of images represented, ranging from 7-100. The average amount of images presented in figural rating scales (excluding one clear outlier - the Somatomorphic Matrix; (Pope, Borowiecki, & Cohane, 2000) for uni- and bi-dimensional scales was 9.81 and 29.5, respectively. Twelve scales included body stimuli that displayed exposed chest, arms and legs, whilst five scales included stimuli with chest, arms and legs covered by clothing, and three scales were hand-drawn silhouettes, and therefore displayed no body composition detail. Sixteen of the scales were interval (consistent distance
<table>
<thead>
<tr>
<th>Scale Name</th>
<th>Scale Description</th>
<th>Number of Images</th>
<th>Number of Dimensions</th>
<th>Interval or Ordinal</th>
<th>Number of Images</th>
<th>Age (years)</th>
<th>Type of Participant</th>
<th>Body Fat/Muscularity</th>
<th>Stimulus Description</th>
<th>Number of Images</th>
<th>Social Validation</th>
<th>Body Image Assessor (BMI)</th>
<th>Test-Retest Reliability</th>
<th>Sex of Scale</th>
<th>Descriptive and Psychometric Properties of Male Figural Rating Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Body Size Scale</td>
<td>Photograph Figures</td>
<td>9</td>
<td>Uni</td>
<td>Interval</td>
<td></td>
<td>30</td>
<td>Community Sample</td>
<td>Cameroonian</td>
<td>Photograph Figures</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td>Description and psychometric properties of male figural rating scales.</td>
</tr>
<tr>
<td>Brazilian Photographic Figure Rating Scale for Men (BPFRS-M)</td>
<td>Photograph Figures</td>
<td>8</td>
<td>Uni</td>
<td>Interval</td>
<td></td>
<td>Mean 22.4</td>
<td>Brazilian Undergraduate Students</td>
<td>Brazilian</td>
<td>Photograph Figures</td>
<td></td>
<td></td>
<td></td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-report Figural Drawing Scale</td>
<td>Hand-drawn Figures</td>
<td>9</td>
<td>Uni</td>
<td>Yes Ordinal</td>
<td></td>
<td>20-44</td>
<td>NR</td>
<td>BF</td>
<td>Hand-drawn Figures</td>
<td></td>
<td></td>
<td></td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle and Fat Silhouette Measure</td>
<td>Hand-drawn Figures</td>
<td></td>
<td>Uni</td>
<td>Yes Ordinal</td>
<td></td>
<td>18-23</td>
<td>Male College Volunteers</td>
<td>Muscle and Fat</td>
<td>Hand-drawn Figures</td>
<td></td>
<td></td>
<td></td>
<td>0.83 (28-31 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body Image Assessment Scale-BD</td>
<td>Hand-drawn Silhouettes</td>
<td>17</td>
<td>Uni</td>
<td>Interval</td>
<td></td>
<td></td>
<td>Undergraduate Students</td>
<td>BF</td>
<td>Hand-drawn Silhouettes</td>
<td></td>
<td></td>
<td></td>
<td>0.88 (2 weeks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-Card Discrete Scale</td>
<td>Hand-drawn Silhouettes</td>
<td>13</td>
<td>Uni</td>
<td>Interval</td>
<td></td>
<td></td>
<td>Body Image Assessor Scale</td>
<td>BF</td>
<td>Hand-drawn Silhouettes</td>
<td></td>
<td></td>
<td></td>
<td>0.75 (3 weeks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle Pictorial Measure</td>
<td>Hand-drawn Figures</td>
<td>9</td>
<td>Uni</td>
<td>Yes Ordinal</td>
<td></td>
<td>18-44</td>
<td>US Undergraduate Students</td>
<td>Muscle</td>
<td>Hand-drawn Figures</td>
<td></td>
<td></td>
<td></td>
<td>0.81 (2 weeks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body Size Guide (BSG)</td>
<td>Photograph Figures</td>
<td>10</td>
<td>Uni</td>
<td>No</td>
<td></td>
<td></td>
<td>Community Sample</td>
<td>BF</td>
<td>Photograph Figures</td>
<td></td>
<td></td>
<td></td>
<td>0.94 (3 weeks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bodybuilder Image Grid (BIG-O)</td>
<td>Hand-drawn Figures</td>
<td>30</td>
<td>Bi</td>
<td>Yes Ordinal</td>
<td></td>
<td>18-44</td>
<td>Weightlifters</td>
<td>BF</td>
<td>Hand-drawn Figures</td>
<td></td>
<td></td>
<td></td>
<td>0.86 (2 weeks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arm Fat and Muscle CV = .34 - .38</td>
<td>10-Card Discrete Scale</td>
<td>10</td>
<td>Uni</td>
<td>Ordinal</td>
<td></td>
<td></td>
<td>US Master’s Students</td>
<td>Muscle and Fat</td>
<td>Hand-drawn Figures</td>
<td></td>
<td></td>
<td></td>
<td>0.34 (2 weeks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body Fat and Muscle CV = .42 - .57</td>
<td>10-Card Discrete Scale</td>
<td>10</td>
<td>Uni</td>
<td>Ordinal</td>
<td></td>
<td></td>
<td>US Master’s Students</td>
<td>Muscle and Fat</td>
<td>Hand-drawn Figures</td>
<td></td>
<td></td>
<td></td>
<td>0.42 (2 weeks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle Pictorial Measure</td>
<td>Hand-drawn Figures</td>
<td>9</td>
<td>Uni</td>
<td>Yes Ordinal</td>
<td></td>
<td>18-44</td>
<td>US Undergraduate Students</td>
<td>Muscle</td>
<td>Hand-drawn Figures</td>
<td></td>
<td></td>
<td></td>
<td>0.81 (2 weeks)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1**
<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Instrument/Acronym</th>
<th>Study Population</th>
<th>Follow-up Duration</th>
<th>Method of Score Computation</th>
<th>Method of Drawing Figures</th>
<th>Method of Presentation of Images</th>
<th>Interval</th>
<th>Pearson's Correlation</th>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutale et al., 2016</td>
<td>Body Dissatisfaction Scale</td>
<td>100 Undergraduate Students</td>
<td>BF Per. = .96, Des. = .88</td>
<td>Computer-Rendered Figures</td>
<td>Computer</td>
<td><em>mean</em></td>
<td>1 week</td>
<td>NA</td>
<td>BF CC = .39, CV = .81</td>
</tr>
<tr>
<td>Novella et al., 2015</td>
<td>Presentation of Images on a Continuum Scale (PICS)</td>
<td>180 Undergraduate Students</td>
<td>M and BF</td>
<td>Computer-Rendered Figures</td>
<td>Hand-Drawn</td>
<td>Yes</td>
<td>NA</td>
<td>NA</td>
<td>M CC = .44, CV = .51</td>
</tr>
<tr>
<td>Pope et al., 2000</td>
<td>Cafri et al., 2004</td>
<td>Somatomorphic Matrix</td>
<td>BF Per. = .64, Des. = .79</td>
<td>Hand-Drawn Figures</td>
<td>Hand-Drawn</td>
<td>Yes</td>
<td>Interval</td>
<td>M Index = .34, BF Index = .57</td>
<td></td>
</tr>
<tr>
<td>Pulvers et al., 2004</td>
<td>Male Body Image Instrument</td>
<td>200 African-American Public Housing Residents</td>
<td>BF</td>
<td>Hand-Drawn Figures</td>
<td>Hand-Drawn</td>
<td>Yes</td>
<td>Interval</td>
<td>BF CV = .51, - .61</td>
<td></td>
</tr>
<tr>
<td>Ralph Nearman &amp; Filik, 2018</td>
<td>Male Body Scale (MBS) and Male Fit Body Scale (MFBS)</td>
<td>200 Community Sample</td>
<td>BF Per. = .77, Des. = .54</td>
<td>Hand-Drawn Figures</td>
<td>Hand-Drawn</td>
<td>Yes</td>
<td>Interval</td>
<td>BF CC = .39, CV = .47</td>
<td></td>
</tr>
<tr>
<td>Stunkard et al., 1983</td>
<td>Ralph Nearman &amp; Filik, 2018</td>
<td>Visual Body Scale for Men (VBSM)</td>
<td>BF</td>
<td>Hand-Drawn Figures</td>
<td>Hand-Drawn</td>
<td>Yes</td>
<td>Interval</td>
<td>BF Des. = &lt; .01, Index = .39</td>
<td></td>
</tr>
<tr>
<td>Talbot et al., 2018</td>
<td>Visual Body Scale for Men (VBSM)</td>
<td>200 Undergraduate Students</td>
<td>BF</td>
<td>Computer-Rendered Figures</td>
<td>Computer</td>
<td>Yes</td>
<td>2-6 weeks</td>
<td>BF CC = .39, CV = .81, M CC = .39, CV = .41</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Instrument/Scale</td>
<td>Participants</td>
<td>Figures</td>
<td>Rating Scale</td>
<td>Method</td>
<td>Interval</td>
<td>Per.</td>
<td>Des.</td>
<td>Index</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------</td>
<td>--------------</td>
<td>---------</td>
<td>--------------</td>
<td>--------</td>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>Thompson &amp; Gray, 1995</td>
<td>Contour Drawing Scale</td>
<td>Undergraduate students</td>
<td>BF Hand-drawn Figures</td>
<td>Uni-dimensioonal figural rating scale</td>
<td>Bi-dimensional figural rating scale</td>
<td>(1 week)</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Tucker, 1982</td>
<td>Perceived Somatotype Scale</td>
<td>Undergraduate students</td>
<td>M Hand-drawn Figures</td>
<td>Uni-dimensional figure scale</td>
<td>Bi-dimensional figure scale</td>
<td>(2 weeks)</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williamson et al., 2000</td>
<td>Body Image Assessment for Obesity (BIO-O)</td>
<td>US Community sample</td>
<td>BF Hand-drawn Silhouettes</td>
<td>Uni-dimensional figure scale</td>
<td>Bi-dimensional figure scale</td>
<td>(2 weeks)</td>
<td>No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
between each body reported) in nature, whilst the remaining four were ordinal scales (bodies are shown in order of magnitude but there is no standard of measurement of differences).

Ecological Validity

Ecological validity was conceptualized as the extent to which the figural rating scale stimuli represents the human male form, as appraised by the authors of the present study. Of all the figural rating scales included in the present study, 10 included hand-drawn figures (scales from studies 3, 4, 7, 9, 12, 13, 14, 15, 18, 19), four included hand-drawn silhouettes (scales from studies 5, 6, 16, and 20) four included computer rendered figures (scales from studies 10, 11, 16, and 17), and three included photograph figures (scales 1, 2, and 8; Table 1). Photograph figures present as the most representative of real male bodies, followed by computer rendered figures, hand-drawn figures, and then finally hand-drawn silhouette figures.

Test-Retest Reliability

Fifteen out of 20 studies reported test-retest reliability evidence, which included 48 individual tests of reliability across all included studies. However, of these 48 individual tests, only 27 (57%) of these measures returned acceptable test-retest reliability as decreed by 0.80 recommended standard for test–retest reliability (Carmines, 1990). Only five studies reported that both perceived and desired body ratings were above the recommended standard (scales from studies 4, 9, 10, 15, and 19; Table 1).

Concurrent and Convergent Validity

Out of the 20 figural rating scales included in this study, 17 reported some measure of convergent or concurrent validity evidence for male samples. A range of self-report body
dissatisfaction questionnaires were used to provide convergent validity, including the Male Body Attitudes Scale (Tylka, Bergeron, & Schwartz, 2005), the revised Male Body Attitudes Scale (Ryan, Morrison, Roddy, & McCutcheon, 2011), the Drive for Muscularity Scale (McCreary & Sasse, 2000), the Body Dissatisfaction subscale of the Eating Disorder Inventory (Garner et al., 1983), and the Multidimensional Body-Self Relations Questionnaire–Appearance Scales (Brown, Cash, & Mikulka, 1990). Convergent validity was assessed through Body Mass Index, body fat percentage, and/or Fat Free Mass Index. Of note, 11 studies reported convergent validity evidence (scales from studies 2, 7, 8, 9, 11, 13, 14, 16, 17, 19, and 20) and 11 reported concurrent validity evidence (scales from studies 1, 3, 5, 6, 11, 13, 14, 15, 16, 17, and 18; Table 1).

**Correlations with Eating Disorder Symptoms**

Four of the 20 scales (scales from studies 14, 15, 16, and 17) included a correlational analysis between perceived, desired, and/or index scores (index scores = difference score between perceived and desired body selection), and eating disorder symptoms. Results showed moderate to high positive correlations between eating disorder symptoms (measured via the Eating Disorders Examination Questionnaire; EDE-Q; (Fairburn & Beglin, 1994), or the EDE-Q short; EDE-QS; (Gideon et al., 2016) and body perceived and index scores. No significant correlations were found between eating disorder symptomatology and muscularity index scores. Only Ralph-Nearman and Filik (2018) reported a significant association between participants’ desired body (as selected on the Male Fit Body Scale, a unidimensional muscularity rating scale) and the EDE-Q Global score.

**Discussion**
From this review it is evident that there is a vast number of male figural rating scales available in the extant literature. Scales vary in terms of scale design, stimulus quality and medium, and psychometric soundness.

**Scale Design**

Over half of the reviewed figural rating scales failed to measure both body fat and muscularity. As discussed above, it is essential to examine both body fat and muscular bulk to adequately capture two critical aspects of male body image (Ridgeway & Tylka, 2005). Additionally, a number of scales presented clothed bodies, or bodies without detail of muscular tone and shape. If body stimuli are presented wearing clothes certain features of the male body will be not visible to the individual making the selection (Cho & Lee, 2013; Ridgeway & Tylka, 2005). Effectively, this means less visual information to inform body selection, possibly resulting in a less reliable decision. This is particularly problematic when considering decisions around muscularity.

The number of images comprising reviewed scales varied significantly, with the number of images on a single dimension ranging between seven and 100. All scales presented at least seven figures, thus complying to the minimum optimal number of figures outlined by Ambrosi-Randić and colleagues (2005). However, it is difficult to apply the results of this study to the present review given that Ambrosi-Randić and colleagues (2005) did not examine figural rating scales with greater than nine figures or bi-dimensional scales.

Results showed that four of the existing figural rating scales are ordinal measures, in that there is not a regular change in body size between adjacent body images. Thus, the validity of the perceived-desired discrepancy index as a measure is compromised. Ideally, figural rating scales should be interval scales, with a regular change in body size between adjacent body images. This would also ensure that *perceived, desired*, and index scores could
be included in parametric statistical analyses. Based only on the above, we recommend scales from studies 9, 11, 12, 14, 16, and 17, as they best fit the design criteria described above (i.e. measure both body fat and muscularity, allow for visual access to chest, arm, and leg regions, and are interval scales; Table 1).

**Ecological Validity**

Ecological validity refers to the extent to which the figural rating scale stimuli accurately represents the human male form. Results show that the majority of figural rating scales are constructed using hand-drawn stimuli. Only scales from studies 1, 2, and 8 included photograph figures, and scales from studies 10, 11, 16, and 17 included computer rendered figures (Table 1). Due to the risk of asymmetry, the use of hand-drawn body stimuli is not ideal as it may compromise validity and reliability of the scale (Rhodes & Simmons, 2007; Tovee et al., 2000).

**Test-Retest Reliability**

Results showed that approximately three-quarters of studies reviewed reported test-retest reliability evidence, which included 48 individual tests of reliability across all included studies. However, around half of these measures returned acceptable test-retest reliability as decreed by 0.80 recommended standard for test-retest reliability (Carmines, 1990). Only scales from studies 4, 9, 10, 15, and 19 reported that both *perceived* and *desired* body ratings were above the recommended standard (Table 1). It is essential for figural rating scales to demonstrate reliability, particularly in a clinical setting (e.g., to track the progress of body perception disturbances).

**Concurrent and Convergent Validity**
Approximately one quarter of studies failed to report concurrent and convergent validity evidence. Concurrent validity indicates the extent to which figural rating scale index scores indicate body dissatisfaction. Without an assessment of concurrent validity, index scores cannot be interpreted beyond an indication of discrepancy in perceived and desired body image. Therefore, it is essential that index scores are assessed for their sensitivity to indicate body dissatisfaction, as this is a primary use of figural rating scales.

Amongst all articles examined in the present review, convergent validity was typically assessed by examining associations between figural rating scales perceived body and biometric measures of body composition. Although this approach appears to be theoretically sound, many of the studies used BMI, a problematic measure of body composition in males as muscular bulk (as opposed to body fat) can often be the driving factor behind BMI (Burkhauser & Cawley, 2008; Rothman, 2008). Further, many studies reported biometric measures relied were self-report estimates of height and weight. This is problematic as prior studies indicate that participants are relatively poor at estimating these constructs, and many partake in deception (Bowman & Delucia, 1992). Based on these criteria, scales from studies 9, 11, 13, 14, 16, and 17 provided the best validity evidence (Table 1).

**Eating Disorder Symptoms**

The consideration of the utility of figural rating scales to predict eating disorder symptoms is important given that, despite a historical underrepresentation in the peer-reviewed literature, a significant proportion of men suffer with eating disorders or cope with a significant portion of eating disorder symptoms (Coffino, Udo, & Grilo, 2019; Duncan, Ziobrowski, & Nicol, 2017; Limbers, Cohen, & Gray, 2018). Results showed that only scales from studies 14, 15, 16, and 17 examined the relationship between eating disorder symptoms and figural rating scale scores (Table 1). Amongst these studies, participants’ perceived body
fat percentage and body fat index scores positively correlated with eating disorder symptoms. However, apart from the Male Fit Body Scale (Ralph-Nearman & Filik, 2018), there were no significant associations between muscle figural rating scales and eating disorder symptoms. This could potentially be explained by considering the sensitivity (or lack thereof) of the EDE-Q for detecting masculinity-related body concern in males. Prior research comparing male and female norms on the EDE-Q have found that males have significantly lower scores across all EDE-Q subscales compared to females (Lavender, De Young, & Anderson, 2010; Reas, Overas, & Ro, 2012), which could be accounted for by a lack of muscle-related items.

**Recommendations**

It is recommended that this review be utilized by researchers and clinicians to aide in their decision when selecting a male figural rating scale. Second, as a general rule, figural rating scales with no validity and/or reliability evidence should not be used. Ideally, validity evidence should return correlations of greater than .50, denoting a large effect size for a correlation coefficient, and reliability evidence should return interclass correlations of .80 or greater (Carmines, 1990). Third, figural rating scales that encompass variations in body fat and masculinity should be prioritized so that both dimensions of male body image can be assessed, unless the researcher/clinician only requires one of the two dimensions. Fourth, the quality of the body image stimuli comprising the figural rating scale should be considered. Figural rating scales that present photographic or realistic computer-rendered figures, such as the Presentation of Images on a Continuum Scale (Novella et al., 2015) and the New Somatomorphic Matrix-Male (Talbot, Smith, Cass, & Griffiths, 2019), should be utilized. Fifth, figural rating scales should be constructed as interval scales with regular differences between adjacent bodies. This will ensure that perceived body scores, desired body scores,
perceived-desired index scores can be included in parametric statistical analysis. Consequently, no single study met all four recommendations.

Figural rating scales from studies 9, 11, 14, 16, and 17 were the closest to satisfying criteria relating to scale design, validity, and reliability. The scales in these studies satisfied all criteria, except what is described below. Study 9 provided a validation for the Bodybuilder Image Grid (Hildebrandt et al., 2004). This scale was only limited by the quality of the stimuli (hand-drawn figures) and moderate convergent validity for body fat selection scores. Study 11 used the Presentation of Images on a Continuum Scale (Novella et al., 2015), which only failed to assess test-retest reliability evidence, and reported moderate to low concurrent validity evidence for body fat rating scores. Study 14 used the Male Body Scale and Male Fit Body Scale (Ralph-Nearman & Filik, 2018). These scales were limited in the use of hand-drawn body stimuli, and low test-retest reliability scores across all measures. Study 16 utilized the Visual Body Scale for Men (Talbot et al., 2018), which was limited in that two out of six measures returned test-retest reliability evidence below the recommended cut-off, and moderate concurrent validity evidence. Study 17 provided a validation for the New Somatomorphic Matrix-Male (Talbot et al., 2019). This scale was only limited by one score returning low test-retest reliability evidence, small convergent validity evidence for muscularity, and moderate concurrent validity evidence for some measures.

Limitations of this review are noted. First, the generalizability of the validity and reliability evidence provided is limited to men aged 17 and older. Notably, there were no child or adolescent samples included in studies within this review. The utility of included figural rating scales for these populations is, therefore unknown. Second, a large portion of studies utilized samples of undergraduate students, non-clinical samples, and/or Western samples. This further limits the generalizability of presented psychometrics.
In sum, the majority of existing male figural rating scales fell short of our established criteria. Generally, scales were limited by a lack of reliability and validity evidence, poor quality of body stimuli, a failure to represent both dimensions of male body image, poor visual access to key body areas, and/or non-interval scale designs. However, there are several male figural rating scales that meet the majority of our criteria. This included three uni-dimensional ratings scales: the Presentation of Images on a Continuum Scale (Novella et al., 2015), the Male Body Scale and Male Fit Body Scale (Ralph-Nearman & Filik, 2018), and Visual Body Scale for Men (Talbot et al., 2018), and two bi-dimensional rating scales: the Bodybuilder Image Grid (Hildebrandt et al., 2004) and the New Somatomorphic Matrix-Male (Talbot et al., 2019).
References


Appendix A

<table>
<thead>
<tr>
<th>Name of Database</th>
<th>Search term</th>
<th>Initial search results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(17/12/2018)</td>
</tr>
<tr>
<td>PsychInfo</td>
<td>Male figural rating scales</td>
<td>724</td>
</tr>
<tr>
<td></td>
<td>Male figure rating scales</td>
<td>89,269</td>
</tr>
<tr>
<td></td>
<td>Male silhouette rating scales</td>
<td>922</td>
</tr>
<tr>
<td></td>
<td>Male body dissatisfaction scales</td>
<td>16,395</td>
</tr>
<tr>
<td>PubMed</td>
<td>Male figural rating scales</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Male figure rating scales</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>Male silhouette rating scales</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Male body dissatisfaction scales</td>
<td>163</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>Male figural rating scales</td>
<td>9,520</td>
</tr>
<tr>
<td></td>
<td>Male figure rating scales</td>
<td>17,800</td>
</tr>
<tr>
<td></td>
<td>Male silhouette rating scales</td>
<td>12,300</td>
</tr>
<tr>
<td></td>
<td>Male body dissatisfaction scales</td>
<td>19,900</td>
</tr>
</tbody>
</table>

*Note:* No articles beyond the first 20 pages of PsychInfo and Google Scholar searches were examined; Google Scholar search results are approximate values.