3-2023

Checking the Scales: A Psychometric Evaluation of the Weight Concerns Scale in a Sample of College-Aged Cisgender Men from the United States

Kyle M. Brasil
University of South Alabama

Callie E. Mims
University of South Alabama

Ryon C. McDermott
University of South Alabama

Mary E. Pritchard
Boise State University

Checking the Scales: A Psychometric Evaluation of the Weight Concerns Scale in a Sample of College-Aged Cisgender Men from the U.S.

Kyle M. Brasil*
University of South Alabama
kb1926@jagmail.southalabama.edu

Callie E. Mims
University of South Alabama

Ryon C. McDermott
University of South Alabama

Mary E. Pritchard
Boise State University

Abstract

Historically, Western societies have considered body image issues to predominantly affect young, White women. While in recent years men's body image issues have been increasingly highlighted by researchers and the media alike, many instruments currently used to identify clinically significant body image disturbances were developed and validated with samples solely of women and/or girls. One such measure, Killen et al.'s (1994) Weight Concerns Scale (WCS), was initially validated in a sample of adolescent girls. The WCS has yet to be validated in samples of men, despite being used in large national surveys of college men and women (e.g., the National Healthy Minds Survey; HMS) used to inform resources on college campuses. Accordingly, we used structural equation modeling to conduct invariance testing between college student cisgender men's (n = 2,248) and women's (n = 4,733) responses on the WCS via the HMS. Through the use of two different approaches of invariance testing, evidence for metric non-invariance of two of the five items was identified, and all five items evidenced a response pattern that favored women over men. Additionally, removing non-invariant items on the WCS impacted the moderating effect of gender with indicators of depression, anxiety, and eating disorder symptomology. These findings suggest that the use of the WCS may not be appropriate for use in a cis-male sample without modification.

Keywords: body image, Weight Concerns Scale, psychometrics, college students, gender differences, men

Public Significance Statement: Accurate and meaningful measurement is important to inform clinical decisions. The Weight Concerns Scale (WCS) is a measure that was traditionally developed for use with adolescent girls, and little is known about is appropriateness with cisgender men. Our results indicate that two of the five WCS items differ on what is being measured between cisgender men and women and that the WCS is differentially related to other measures for male and female respondents. As such, the WCS does not appear to be appropriate to use with cisgender men without modification.

Body image and weight-related concerns are common and widespread struggles faced by college students (Bucchianeri et al., 2013; Haynos et al., 2018). It is estimated that 70%-80% of college-aged men (Hobza & Rochlen, 2009; Neighbors & Sobal, 2007) and 80-90% of college-aged women struggle with their body image and weight (De & Chakraborty, 2015; Neighbors & Sobal, 2007). Students with body dissatisfaction are at higher risk for muscle-building supplement and steroid use (Brower et al., 1994; Eik et al., 2018), binge drinking, dieting (Eik et al., 2018), depression (Eik et al., 2018; Paxton et al., 2006), anxiety (Barnes et al., 2020; Regis et al., 2018), eating disorders and muscle dysmorphia (Almeida et al., 2021; Palmeroni et al., 2020), as well as low self-esteem (Paxton et al., 2006). In addition, many collegiate women and men engage in subclinical binge eating and dieting, and nearly one-third engage in compensatory behaviors such as excessive exercise (Lavender et al., 2010; Luce et al. 2008). This is particularly concerning, considering that the health system, productivity costs, lost well-being, and other societal economic costs of eating disorders totaled $326.5 billion in the 2018-2019 fiscal year alone (Streatfeild et al., 2021). Notably, this estimate is only for diagnosed eating disorders and does not include the costs of subclinical presentations, which likely would increase the estimated cost.

Historically, body image issues were identified as a concern predominantly expressed by White, cisgendered women. However, in recent decades researchers have noted that body image issues also impact cisgender men (Tantleff-Dunn et al., 2011; note, from here on, we use the term, "men" to refer to cisgender men). Indeed, researchers recognized and argued for different presentations of body image issues for men, identifying that men tended to seek muscular bodies, whereas women pursued thinner figures (McCreary & Sasse, 2000; Schaefer et al., 2015; Tantleff-Dunn et al., 2011). Following these newer lines of research, men's body image related issues have also become more publicly recognized, as evident by news articles (e.g., Baggs, 2021) and popular media discussing men's body image struggles.

Given the historical focus of body image issues as predominantly experienced by women and girls, many instruments designed to conduct research, screen for risk, and identify clinically significant body image disturbances were developed, validated, and normed using samples consisting completely of women and/or girls. One prominent example, and the focus for the present study, is Killen et al.'s (1994) Weight Concerns Scale (WCS). The WCS is a popular screening and research instrument that was initially developed with a sample of adolescent girls. The WCS was then further validated in international samples of college-aged women (da Silva et al., 2017; Ribeiro Dias et al., 2015) and has since been used globally by researchers, colleges and universities, and policy makers with samples of *men* and *women*. The data collected is then used by researchers and college campuses to make inferences about weight concerns amongst men, which may ultimately influence policy and funding decisions regarding body image issues within this population. However, to our knowledge, no previous researchers have examined the psychometric properties of the WCS in samples of men. This is concerning given that the WCS is widely used. Moreover, body image concerns may present differentially in men compared to women, in part due to different cultural body ideals for both genders (e.g., muscular ideal versus thin ideal; Pope et al., 1999). This means that assumptions, policies, and funding decisions may be made based on misidentified rates of body image concerns. Accordingly, the present study examined the measurement and external validity equivalence between men and women of the WCS in a large, national sample of college students.

Weight Concerns Scale

The Weight Concerns Scale (WCS) was developed by Killen et al. (1994) as a brief measure to assess girls' concerns about their weight. The measure contains five items, with sample items including, "How afraid are you of gaining 3 pounds?" and "Do you ever feel fat?" The WCS was initially validated in a sample of 967 sixth and seventh-grade girls in a longitudinal study over three years. Killen et al. (1994) found that weight concerns were highly associated with later onset of eating disorder symptoms, using a principal component analysis (cf. Killen et al., 1994).

The WCS has expanded in use over the years, leading to its utilization in samples of U.S. college-aged women and men (e.g., Forbush et al., 2014). Within these samples, the WCS has been associated with weight and shape concerns (Ribeiro Dias et al., 2015; da Silva et al., 2017), anxiety disorders and depression in collegiate women (Aspen et al., 2014), as well as suicidal ideation in collegiate men and women (Lipson & Sonneville, 2020). Indeed, the WCS has become a popular measure with researchers evaluating college-aged and young adult samples. For example, the Healthy Minds Study (Healthy Minds Network, 2022), the German and Hungarian ProYouth programs (Kindermann et al., 2017; Szabó et al., 2015) and the Stanford-Washington University Eating Disorder screen (Graham et al., 2019), which has been utilized by the National Eating Disorders Association to screen individuals for eating disorder risk (Fitzsimmons-Craft et al., 2019), all incorporate the WCS as a part of their eating and body image blocks. The Healthy Minds Study, in particular, represents one of the largest annual studies of college and university students (Healthy Minds Network, 2022). The information gathered by Healthy Minds is used by countless researchers and campus administrators across the U.S., and ultimately informs decisions regarding resource allocation and funding.

Men's Weight Concerns

The WCS has been used in samples of adolescent boys (Hochgraf et al., 2018; 2019; Jones et al., 2014), collegiate men (Forbush et al., 2014; Kindermann et al., 2017; Lipson & Sonneville, 2020; Szabó et al., 2015), and adult men (Markey et al., 2008; Markey & Markey, 2011). Despite the wide use of the WCS in men, some researchers have questioned its ability to measure eating disorder risk among men, as the WCS is missing key components of factors influencing said risk (Pearson et al., 2013). Specifically, the Weight Concerns Scale was designed to measure preoccupation with thinness and body shape in adolescent girls (Killen et al., 1994). Previously, researchers have found that the WCS correlates positively with measures of body dissatisfaction in adolescent females (Killen et al., 1994) and collegiate women (Ribeiro Dias et al., 2015), and to date, it has been used to measure weight-related body image concerns in collegiate women (Ribeiro Dias et al., 2015) and male and female adolescents (Hochgraf et al., 2019). While the WCS has been used to successfully predict the onset of eating disorders in adolescent girls (Killen et al., 1994), it has never been used in that capacity for men. Further, a recent study by Chua et al. (2021) found that in their sample of Singaporean men and women, women scored significantly higher than men on the WCS, but that men and women did not differ in frequency of disordered weight control behaviors or eating disorder risk status. Such findings suggest that the WCS may not be a psychometrically useful screening tool for eating disorder risk when applied to samples of men.

One reason the WCS has not been used to predict the onset of eating disorders in men may be that the items are less appropriate for male populations. The WCS focuses on five weight-related concerns: fear of weight gain, worry over weight and body shape, importance of weight, diet history, and perceived fatness (Killen et al., 1994). Although men who exhibit a preoccupation with thinness do indeed display higher levels of eating disorder symptoms (Grossbard et al., 2013), so do men who exhibit a preoccupation with muscularity (Arellano-Pérez et al., 2019; Brown et al., 2017; Grossbard et al., 2013). In fact, this focus on assessing preoccupation with thinness as a hallmark of eating disorder risk is one reason researchers believe men have been under-represented in eating disorder research (Brown et al., 2017; McCreary et al., 2007), as problematic eating behaviors in men may be the result of either trying to lose weight or gain weight (McCreary et al., 2007). Moreover, many men who exhibit signs of disordered eating do both in an attempt to increase muscularity while simultaneously increasing leanness through certain dieting behaviors (McCreary et al., 2007). Thus, multiple studies have suggested that assessments designed to assess body image and eating disorder risk in women are not appropriate for use in men, given that men often experience unique body concerns and risk factors for disordered eating behaviors (Kaminski et al., 2005; McCreary & Sasse, 2000).

Invariance Testing

The findings of studies using the WCS in samples of men should thus be interpreted with caution until this measure has been validated and supported in such samples. Some researchers (e.g., Pearson et al., 2013) have echoed this concern with the instrument but did not examine the psychometric properties of the WCS. Accordingly, there is a noteworthy gap in the literature regarding the measurement equivalence of the WCS between men and women. Specifically, researchers have yet to examine the WCS for measurement invariance (i.e., statistical equivalence) between men and women. Using structural equation modeling, researchers can examine several levels of invariance between different populations on a particular instrument (Cheung & Lau, 2012; Milfont & Fischer, 2010). If invariance is supported, researchers can assume that the latent construct is statistically equivalent and therefore make comparisons between groups with confidence. However, if invariance is not supported, this indicates that the instrument may be functioning differently for certain groups. Therefore, comparisons and interpretations between groups are inappropriate (Milfont & Fischer, 2010). As such, invariance testing is a powerful statistical tool and means for researchers to psychometrically evaluate instruments when they are used with a group on which the instrument was not originally validated.

Furthermore, invariance testing can be conducted at several levels, each with different implications in the interpretation of results. At the basic level (i.e., configural invariance), the instrument demonstrates the same pattern and number of factors per group. If the WCS, which is believed to be a unidimensional (i.e., one-factor explaining all the variation in the items) instrument were to fail at this level of invariance, then this would mean that the instrument may have a different factor structure in men vs women. If configural invariance holds, then more advanced levels of invariance can be tested to determine whether the latent construct assessed by the WCS has the same conceptual meaning among men and women (i.e., metric invariance), if the rating scale for each item has the same zero point (i.e., scalar invariance) for men and women, and whether the instrument measures the construct with the same degree of reliability and precision between men and women (i.e., residuals invariance). Finally, invariance analyses can be used to determine is the direct effects between an instrument and other instruments are equivalent (i.e., direct effects invariance). Thus, invariance provides information about the internal *and* external validity of an instrument between different groups (Kline, 2016).

Present Study

Although the WCS is a popular and widely used screening measure of body dissatisfaction (da Silva et al., 2017; Ribeiro Dias et al., 2015; Killen et al., 1994; Kindermann et al., 2017), researchers have yet to examine the measurement equivalence of the WCS between men and women to determine either the internal or external validity of the WCS when used with men. Accordingly, in the present study, we tested the WCS for configural, metric, scalar, residual, and direct effects invariance between collegiate men and women. Validity evaluations of an instrument require a sequential process, that is internal validity (e.g., measurement invariance) must be supported before examining external validity. Given that invariance testing is an exploratory process, we did not put forth any hypothesis regarding configural, metric, scalar, or residual levels of invariance. Instead, we hypothesized that the WCS would be positively correlated with measures of depressive, anxiety, and eating disorder symptoms in both the cisgender male and cisgender female samples. Specifically, previous research has identified positive associations between WCS scores and measures of disordered eating (Killen et al., 1994), depression, and anxiety (Aspen et al., 2014), and we expected scores on these measures to positively correlate with scores on the WCS. However, we did not advance any specific hypotheses as to whether gender would moderate these associations (i.e., whether the direct effects would be invariant or non-invariant between men and women).

Method

Procedures and Participants

This study was not preregistered. Data were drawn from the National Healthy Minds Study (HMS), an annual web-based survey initially launched in 2007 and consisting of a random sample of undergraduate and graduate students from over 400 universities across the U.S (Healthy Minds Network, 2022). After receiving institutional review board approval for archival data analysis and permission to use the data from the HMS organizers, the initial sample consisted of 14,078 university students who participated in the HMS between fall of 2015 and spring of 2019 and responded to variables of interest. We selected these data sets as they contained the WCS scale and were conducted prior to the COVID-19 pandemic. As literature using the WCS predominantly focuses on traditionally college-aged students or emerging adults, we removed the 2,508 participants outside the age range of 18-29. We further removed 211 participants who did not identify as cisgender men or cisgender women because there were not enough participants in any gender minority groups to run structural equation modeling, as well as 10 participants who opted not to state their gender identity as gender is a primary variable of interest in this study. Notably, there were also 4,368 participants whose reported gender identity and sex assigned at birth did not match, but did not indicate that they identified as transgender, genderqueer, or another gender minority. These participants were removed from the sample, as well. The final sample for invariance testing totaled 6,981 participants, including 4,733 cisgender women and 2,248 cisgender men with a mean age of 21.05 (SD = 2.60). See Table 1 for all demographic information.

Measures

Demographics

Participants were asked to self-report their gender presented with the following categories: Male, Female, Trans male/Trans man, Trans female/Trans woman, Genderqueer/Gender non-conforming, and Self-Identify/Other (please specify), with the latter including a free response option. Participants were asked to self-report their sex assigned at birth using the following categories: Male, Female, or Intersex.

Weight Concerns Scale

The WCS (Killen et al., 1994) is a five-item measure designed to assess body weight concerns. Sample items include “How afraid are you of gaining 3 pounds?” and “Do you ever feel fat?” Responses to items are reported on a Likert-type scale ranging from 4 to 7-point continuums. Higher scores are indicative of higher weight concerns. While initially validated in a sample of adolescent American girls, the WCS has been expanded and psychometrically validated in samples of college-aged and international adult women. The WCS has been found to have adequate construct validity, reliability, and strong invariance within these samples (da Silva et al., 2017; Ribeiro Dias et al., 2015). In the original WCS, item one is worded for girls and later women (e.g., “compared to other girls/women”; Killen et al., 1994). In the HMS survey, item one was reworded so that the stem reflected the participants reported gender (e.g., if the participant selected “Male” they were presented with “compared to other men”). In the present study, we combined cisgender women and men’s responses to analyze if there was a difference between the word stems and to be able to compare the items. Our internal consistency for the combined WCS was .76 for the total sample, .77 for women, and .71 for men.

Eating Disorder Symptomatology

The Sick Control One Fat Food (SCOFF; Morgan et al., 1999) is a five-item test designed to detect the presence of an eating disorder. Sample items include "Do you ever make yourself sick because you feel uncomfortably full?" and "Do you believe yourself to be fat when others say you are too thin?" Because this measure was developed in the U.K., and the Healthy Minds Study is disseminated to U.S. college students, the unit of weight referenced on one item was altered from the original scale's usage of stones to the more commonly used American unit of pounds. The "yes/no" response items are scored by assigning one point for every "yes" answer and zero points for every "no." A score of two or greater indicates the presence of an eating disorder. In our analyses, we used the total score to more align with standard usage of the SCOFF. A recent meta-analysis on 25 studies validating the SCOFF in a variety of settings, ages, and across both women and men concluded that validity was high across samples, with the highest sensitivity for young women with AN or BN, and the lowest sensitivity for men and community samples (Kutz et al., 2020).

Anxiety Symptomology

The General Anxiety Disorder-7 (Spitzer et al., 2006) is a seven-item measure used to assess generalized anxiety disorder (GAD). Item responses are reported on a four-point scale from 0-3, with 0 = "not at all" and 3 = "nearly every day". Scores range from 0-21, with higher scores indicative of increased likelihood of GAD and increased symptom severity. The GAD-7 has been found to be valid and reliable in both male and female U.S. college students (Byrd-Bredbenner et al., 2021).

Depressive Symptomology

The Patient Health Questionnaire-9 (Kroenke et al., 2001) is a nine-item scale used to measure depression severity. Responses are recorded on a four-point scale from 0-3, "not at all" and 3 = "nearly every day." Scores range from 0-27, with higher scores indicating increased depression severity. The PHQ-9 has been found to be equivalently reliable and valid across gender and racial groups in college students (Keum et al., 2018).

<u>Analysis Plan</u>

Invariance Testing

We tested the WCS for measurement invariance and external validity using multigroup structural equation modeling (SEM). Specifically, we used a combination of nested-model and direct-modeling approaches. Nested modeling approaches involved testing models with various parameters constrained to be equal between men and women compared to models in which those parameters were freely estimated. Specifically, we tested the following forms of invariance in this order: (a) configural invariance (i.e., determination of whether the model provided an acceptable fit while not imposing any cross-group equality constraints), (b) metric invariance (i.e., testing for equivalence between groups on the latent factor by adding cross-group equality constraints to item factor loadings), (c) scalar invariance (i.e., adding cross-group equality constraints to item intercepts), and (d) residuals invariance (i.e., adding cross-group equality constraints to item residuals). Because levels of invariance are hierarchical, with configural at the bottom and residuals at the top, a failure to support invariance at a lower level precluded the testing of invariance at a higher level. To test external validity, we examined direct effects invariance to determine if the regression paths between the WCS and depressive, anxiety, and eating disorder symptomology were significantly different between men and women.

To determine support for any form of invariance using the nested model approach, we used a scaled chi-square difference test. A significant increase in the scaled chi-square indicated that the model with cross-group equality constraints was a worse fit compared to a model without those equality constraints (i.e., invariance is not supported). However, the chi-square difference test is extremely sensitive to sample size (Kline, 2016). Therefore, we also used a direct model approach and calculated the bias-corrected bootstrapped confidence intervals (CIs) of the between-groups difference on the parameters in question (e.g., factor loadings, intercepts, residuals, or direct effects; Cheung & Lau, 2012). Within this approach, a 99% CI of the between-group difference that contains zero indicates that the difference between groups was not statistically significant (i.e., invariance is supported).

We also included three additional analysis procedures to provide an in-depth and nuanced comparison of the WCS between men and women. First, in SEM, one item is constrained to one and used to mathematically identify the latent construct. As such, that item is automatically assumed to be invariant in multigroup SEM – an assumption that is rarely tested in practice (Cheung & Lau, 2012). Thus, to thoroughly test invariance, we examined each level of invariance using all five items of the WCS as the identifying variable (see the supplementary file for results by referent item). This iterative process allowed for the identification of specific items that may be non-invariant that could otherwise be missed in multigroup SEM. Second, invariance historically yields a dichotomous (yes = invariant, no = non-invariant) outcome. Recent investigators have developed a method of calculating the effect size of metric and scalar non-invariance. Calculating the difference in means and covariance structures (d_{macs}; Nye & Drasgow, 2011) provides researchers a way to move beyond the yes or no framework to more of an ordinal perspective by describing the magnitude of non-invariance. Thus, we contextualized our findings using d_{macs} criteria for small (.40), medium (.60), and large (.80) non-invariance effect sizes (Nye et al., 2019) using open-sourced software developed by Nye and Drasgow (2011; available at https://psychology.psy.msu.edu/pers_nye/). Finally, regarding our external validity analyses, we planned to examine direct effects invariance with and without any items on the WCS identified as being non-invariant. This process allowed for a more nuanced understanding of whether the inclusion of problematic items impacted external validity.

For all models, we used the following established cutoff and fit indices to determine if our data provided acceptable fit at any level of invariance (Hu & Bentler, 1999; Kline, 2016) to evaluate each model: Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI) (values close to .95 are indicative of a good fit for both CFI and TLI); the Root Mean Square Error of Approximation (RMSEA) with 90% confidence intervals [CI] (where low values .06 and high values less than .10 are indicative of a good fit), and the Standardized Root-Mean-square Residual (SRMR; where values .08 are indicative of a good fit). We also reported the chi-square test statistic (where a non-significant value is indicative of a perfect fit to the data). However, we interpreted the chi-square test statistic with caution, due to its sensitivity to sample size (Kline, 2016). Data is available upon request and once the requesting induvial(s) have received permission from the HMS organizers.

Results

Preliminary Analyses

Prior to conducting our primary analyses, we screened the data for missing data, univariate outliers, and assumptions of normality. Missing values for the WCS items were minimal (i.e., less than .2%) to where missingness is considered to meet the assumptions of at least missing at random (Meyers et al., 2017) and there were no univariate outliers (i.e., z-scores greater than the absolute value of three). The WCS evidenced moderate positive skewness (i.e., approaching but not exceeding the absolute value of one) for men and small positive skewness for women. Due to these departures from normality and the size of our sample, we used a full information maximum likelihood estimator with robust standard errors for model fit.

Primary Analyses

Configural Invariance

As an initial step to configural invariance, we first tested models for men and women separately and iteratively with the five different referent items to identify the model. For men, the models across all referents were equivalent and provided acceptable fit to the data, (n = 2248) $\chi2$ (10) = 2428.44, p < .000 (CFI = .992, TLI = .985, RMSEA = .040 [90% CI = .025, .057], SRMR = .014). For women, the models were also equivalent and provided an acceptable fit to the data, (n = 4733) $\chi2$ (10) = 7958.94, p < .001 (CFI = .993, TLI = .986, RMSEA = .049 [90% CI = .039, .061], SRMR = .012). We then moved to a configural invariance model with both men and women combined but without any cross-group equality constraints. The fit of the configural invariance models were acceptable and equivalent regardless of the referent item, (N = 6981) $\chi2$ (10) = 10382.84, p < .001 (CFI = .993, TLI = .986, RMSEA = .046 [90% CI = .037, .056, SRMR = .013). Thus, these results supported the presence of configural invariance and warranted moving to more restrictive levels of invariance.

Metric Invariance

Having found support for configural invariance, we moved to test metric invariance by constraining men's and women's factor loading to be equal. Metric invariance models were, again, iteratively conducted using each item as a referent to identity the latent construct. The fit indices of each metric invariant model were equivalent across referent items: (N = 6981) $\chi 2$ (20) = 10382.84, p < .001 (CFI = .986, TLI = .98, RMSEA = .054 [90% CI = .046, .062, SRMR = .032). We then ran a series of scaled chi-square difference tests comparing the metric invariant models to the configural invariant models. All $\chi 2$ difference tests were statistically significant (see Table 2), however, an examination of the 99% CIs of the between-group differences suggested that non-invariance was relegated to two of the five items, thus indicating partial metric invariance. Specifically, items one, two, three, and four of the WCS evidenced statistically equivalent factor loadings for men and women. However, item five was non-invariant between women and men, and when item five was used to identify the latent construct, all of the items were non-invariant. See Table 3 for our results.

Scalar Invariance

Next, we examined the WCS for scalar invariance by constraining both the factor loadings and the item intercepts to be equal between men and women. Given that item 5 was non-invariant at the metric level, we kept this factor loading freely estimated. Regardless of referent item, the fit indices were equivalent across scalar models and evidenced poor fit: (N = 6981) $\chi 2$ (20) = 10382.84, p < .001 (CFI = .854, TLI = .846, RMSEA = .151 [90% CI = .145, .157, SRMR = .127). All $\chi 2$ difference tests were statistically significant and were further confirmed via bias-corrected bootstrapped confidence intervals of the between-group differences on each item's intercept. These results did not support scalar invariance and did not support moving to residual invariance testing.

Effect Sizes of Non-Invariance

Following the failure to support scalar invariance, we calculated the d_{macs} of the difference on WCS items between cisgender women and men at the metric and scalar levels. The d_{macs} of items 1, 3, and 4 of the WCS were .09, .18, and .12, respectively. These fall within the small to negligible range of effect sizes for d_{macs} (Nye et al., 2019). For items 2 and 5, the d_{macs} were 0.546 and 0.427, respectively, which is within the medium range of effect sizes for d_{macs} (Nye et al., 2019). The effect size for item 2 was of note considering that metric invariance was supported for item 2 via multigroup SEM.

External Validity

Given that our results supported the WCS being partially invariant between cisgender men and cisgender women, we explored the concurrent validity of the WCS between cisgender men and women. Specifically, we examined the relationships between the latent variable of the WCS with the observed outcomes on total scores of the PHQ-9, GAD-7, and SCOFF. Prior to examining the preliminary construct validity of the WCS, we confirmed metric invariance between cisgender men and cisgender women on the PHQ-9, GAD-7, and SCOFF. Specifically, our result supported at least partial invariance for these instruments. This process was not the focus of our study, however the information is available upon request.

To provide initial concurrent validity, we created correlational configural models to examine the associations between the WCS with the PHQ-9, GAD-7, and SCOFF total scores in men and women without any cross-group constraints. We used a bootstrap of 1000 iterations. We began by examining the model with all WCS items. Given there were two items that were identified to be non-invariant (i.e., item 2 and 5) we also examined the models when item 2 was removed, when item 5 was removed, and when both items were removed. The initial model (i.e., with all WCS items) evidenced adequate fit: (n = 6760) $\chi 2$ (34) = 290.978, p < .001 (CFI = .988, TLI = .980, RMSEA = .047 [90% CI = .042, .052, SRMR = .02). The correlations for cisgender men and cisgender women were all statistically significant (see Table 4). We also examined the bootstrapped difference between the unstandardized correlations for cisgender men and cisgender women. Gender differences were significant for all paths between cisgender men and cisgender women (see Table 4). In each instance, the associations between WCS and the concurrent validity variables were stronger for cisgender women compared to cisgender men.

We then examined the model without item 2, which evidenced adequate fit: (n = 6760) $\chi 2$ (22) = 217.419, $p < .001$ (CFI = .988, TLI = .977, RMSEA = .051 [90% CI = .045, .058, SRMR = .02). Again, all correlations for cisgender men and cisgender women were statistically significant (see Table 4). The differential tests were significantly different between men and women on the associations between the WCS and the PHQ-9 and SCOFF but not for GAD-7 scores. Women again had stronger associations for these paths (see Table 4).

We re-examined the model adding item 2 back and removing item 5, which evidenced adequate fit: (n = 6760) $\chi 2$ (22) = 99.433, $p < .001$ (CFI = .996, TLI = .992, RMSEA = .031 [90% CI = .024, .037, SRMR = .014). Again, all correlations for cisgender men and cisgender women were statistically significant (see Table 4). The differential tests were significantly different between men and women on the associations between the WCS and the PHQ-9, SCOFF, and GAD-7. Women again had stronger associations for these paths (see Table 4). However, the model evidenced even better fit indices when compared to the model with all WCS items.

In the final iteration, we removed both items 2 and 5, which evidenced adequate fit: (n = 6760) $\chi 2$ (12) = 65.957, $p < .001$ (CFI = .995, TLI = .988, RMSEA = .036 [90% CI = .028, .045, SRMR = .015). Again, all correlations for cisgender men and cisgender women were statistically significant (see Table 4). The differential tests were similar to the results when removing only item 2, and there were significantly different associations between men and women on the WCS and the PHQ-9 and SCOFF. Women again had stronger associations for these paths. Similar to the results of the model without item 2, the differential test was not statistically significantly between men and women on the path between the WCS and GAD-7 (see Table 4). The final model also had slightly better fit indices than the model with only item 2 removed. These results provide support for a moderation effect by gender when examining the relationships between the WCS and the PHQ-9 and SCOFF. However, when removing item 2 and items 2 and 5 together, the paths this moderation effect disappears and the relationships are now equal between men and women.

Discussion

Researchers have used the WCS in samples of collegiate men, however, the WCS has not been psychometrically validated with samples of collegiate men. We addressed this gap in the literature by testing for configural, metric, scalar, and residuals invariance between men and women on the WCS. Additionally, we examined the external validity of the WCS for men by comparing external correlates empirically and conceptually linked to weight concerns.

The results of our invariance analyses provided full support for configural invariance and some support for partial metric invariance. However, scalar invariance failed, and thus residuals invariance could not be tested. These findings suggest that, in our large national sample, the WCS has the same number of factors between men and women (i.e., configural invariance) but the overall latent construct of weight concern appears to be slightly different for men than women in ways that impact the amount of variation explained in one item (i.e., item 5 evidenced significant gender differences on how it loaded on the latent weight concerns construct), thus failing to support full metric invariance. Additionally, the rating scale used to assess men and women on all five of the WCS items were not statistically equivalent (i.e., scalar invariance was not supported). The combination of metric and scalar non-invariance is of concern, because it means that some items may not be as relevant to men and can thus artificially weight down men's scores relative to women.

Although the effect sizes of non-invariance for items 1, 3, and 4, were generally small in magnitude, suggesting that non-invariance would likely have little practical effect (Nye et al., 2019), the other two items evidenced medium effect sizes. Given that the WCS only contains five items, this means that 40% of the scale might not be accurately reflecting men's weight concerns. Item 5, "Do you ever feel fat?" emerged as particularly problematic. It was the only item to fail invariance testing at the metric level and the scalar level and produced a medium d_{macs} effect size. These results suggest that men and women may interpret this item differently. Specifically, non-invariance may be due to differences in fat and body talk between women and men. Researchers have suggested that while body talk among women may be more focused on being "fat," negative body talk among men may focus more on a lack of muscularity (Sladek et al., 2014), corresponding to the different cultural body ideals that pressure men and women (Pope et al., 1999). Thus, men may be more prone to expressing their weight concerns through rhetoric that is more salient to their experience of body image, such as feeling insufficiently muscular or toned, rather than feeling fat.

In addition to the phrase, "feeling fat" potentially being more relevant for women, the present non-invariance may signal important differences in what it means to feel "fat." For example, researchers have demonstrated that, while fat talk among women is prevalent across a wide variety of different situations, for men, fat talk is situation specific–-

most common at the gym or playing sports (Engeln et al., 2013). Likewise, while women's most common response to fat talk is to deny that the initiator of the fat talk is fat, men's most common response is to validate this concern (Engeln et al., 2013). In this way, it may be the case that fat talk, and "feeling fat" in particular, is inherently different in men and women, such that "feeling fat" for women is associated more so with body anxiety and shame manifested in or expressed as the physical feeling of fatness (Shannon & Mills, 2015), whereas for men, "feeling fat" may be representative of appraisal of their actual (verses their felt) physique, which they feel they can actively shape or change (SturtzStreetharan et al., 2020).

Item 2 of the WCS, "How afraid are you of gaining 3 pounds?" also was problematic, though it was shown to be invariant at the metric level. Further probing with d_{macs} indicated that this item exhibited a medium effect size for measurement non-invariance. Given that d_{macs} provide an assessment of both metric and scalar non-invariance, it is likely that item 2's medium effect size reflected issues with scalar non-invariance. Regardless, this item may also suffer from the same problems of relevance impacting item 5. Specifically, researchers have demonstrated that while the vast majority of women report wanting to lose weight, men tend to be evenly split between wanting to lose and wanting to gain weight (Davis & Cowles, 1991; McKinley, 2006). Of those who report wanting to gain weight, researchers are of the consensus that they hope to gain muscle mass (McCabe & Ricciardelli, 2004), corresponding to the muscular body ideal for men. Thus, it follows that being afraid to gain three pounds might mean something different for women and men; for example, while men may consider the muscle-to-adipose tissue ratio of that weight gain, women might assume this statement refers to gaining adipose tissue or weight gain in general. Thus, the utilization of Item 2 of the WCS may likewise be inappropriate in assessing men's weight concerns. In light of these findings that two of the five items of the WCS may not be suitable in evaluating men's weight concerns, researchers are encouraged to consider omitting these items if choosing to use the WCS to evaluate men's weight concerns, as well as incorporate other measures of body image that may be more salient to men's experiences, particularly those that include a muscularity component.

Furthermore, when examining the external validity of the WCS with men as compared to women, our associations were within the expected directions, thus supporting our main hypothesis. Additionally, direct effects invariance testing indicated a moderation effect by gender. However, categorical moderation effects cannot be supported if there are issues of metric non-invariance in SEM; invariance testing needs to be conducted prior to moderation analyses (Kline, 2016). That is, without assurances that the same latent construct is being assessed for men and women (i.e., metric invariance) it is impossible to determine if moderation is due to true gender differences or a measurement bias. For the WCS, this problem was evident in relation to items 2 and 5. Indeed, our results from the models with those items removed continued to support moderation by gender but only for the associations with PHQ-9 and SCOFF. The moderation between the WCS and GAD-7 disappeared. Thus, it is likely that there exists a true moderation effect by gender on depression and eating disorder symptomology, whereas any potential moderation effect with anxiety symptomology may be due measurement bias. This pattern of results further highlights the importance of measurement invariance testing prior to moderation in SEM (Kline, 2016) and suggests clinicians and researchers should consider omitting items 2 and 5 and/or incorporating other body-related measures when using the WCS with men.

Limitations and Future Directions

Although our study was the first to examine the psychometric properties of the WCS in a sample of men, the study had several limitations that should be noted. First, unlike Killen's (1994) study, our study was cross-sectional in nature and thus could not examine the predictive validity of the WCS on psychological outcomes (e.g., eating disorder symptoms) in men. Future studies may wish to examine the predictive validity of the WCS in men both with and without Items 2 and 5. Second, as the focus of this study was on cisgendered individuals, we cannot comment on the appropriateness of use of the WCS in gender minority populations. Additionally, there were not enough individuals who identified with a minority gender group to have the power to conduct invariance testing in our sample. To our knowledge, no studies have examined the psychometric properties of the WCS in individuals who identify as part of a minority gender group, constituting a significant gap in the current research base and an important area for future research. Finally, as we used the Healthy Minds Study sample, our population was limited to college students. Given the widespread use of the WCS in non-college populations (e.g., adolescents, community samples), future researchers may wish to examine the psychometric properties of the WCS in those populations, especially in men and minority gender samples.

Clinical Implications

To the authors' knowledge, this study marked the first to validate the psychometric properties of the WCS in cisgender men. The present study identified support for partial invariance. However, using two different approaches of invariance testing, evidence for metric non-invariance of Items 2 and 5 was identified. This indicates that these items may capture different latent constructs in men and women, may be interpreted differently by men and women, and potentially do not measure weight concerns in men as accurately as they do in women.

This finding possesses several key implications for measuring weight concerns in men for both researchers and clinicians alike. Firstly, the findings of previous research estimating weight concerns in men using the WCS must be interpreted with extreme caution, as they may not accurately reflect the prevalence of weight and shape preoccupation in men. Organizations that continue to rely on the WCS to measure weight and shape concerns as well as eating disorder risk in men may be underreporting such issues within this population. This may in turn contribute to both misconceptions about the rate of eating and body image disturbances in men as well as insufficient funding allocated towards prevention and treatment of these issues. Furthermore, organizations that use the WCS as part of a self, online, or at-home screening measure for eating disorders, such as the National Eating Disorder Association, may both under-identify eating disorders in at-risk men who participate in these screenings as well as further deter some men experiencing these issues from seeking help. For example, if a cisgender man with disordered eating tendencies and weight concerns that are more muscularity focused takes the WCS as part of a screening tool, he may doubt the validity of his disordered eating and body image concerns, especially after answering questions that may be more representative of eating disorder issues traditionally associated with women, such as gaining a small amount of weight or feeling fat. In this way, using screening measures for all genders that have only been developed and validated in samples of women, such as the WCS, may contribute to the continued stigmatization of eating disorders as "women's issues," and act as a barrier for those of other genders in receiving treatment. Moreover, individuals who are not aware of the differences in body image and weight concerns between men and women may report to a provider they completed a screener and the results indicated they were not at risk for disordered eating.

Likewise, clinicians that use the WCS as a screening tool for weight and shape concerns or disordered eating in all clients regardless of gender may underdiagnose these issues in their cisgender male-identifying clients. Because of this, if researchers or clinicians decide to use the WCS to identify weight concerns in men going forward, they are encouraged to consider removing Items 2 and 5, as well as include body image measures that have been validated in men and that more so reflect men's experiences of body image, in particular, muscularity dissatisfaction.

Conclusion

This study assessed the internal and external validity of the WCS in a large national sample of college students using the Healthy Minds data set. Our findings suggest that although the WCS exhibited configural invariance, the latent construct of weight concerns may be different in men than in women. In particular, two items were cause for major concern in our male sample (items 2 and 5), and all five items evidenced a response pattern that favored women over men. This questions the validity of using the WCS in its entirety in a cis-male sample. Furthermore, given that researchers have questioned relying on scales that assess only thinness as a measure of body image concern in men (Kaminski et al., 2005; McCreary & Sasse, 2000), our findings suggest that in addition to perhaps using a modified version of the WCS in men (removing items 2 and 5), researchers and clinicians assessing body image concerns in men might be better served to utilize additional measures of body image dissatisfaction, such as those that incorporate a muscularity component (e.g., Sladek et al., 2014).

References

Aspen, V., Weisman, H., Vannucci, A., Nafiz, N., Gredysa, D., Kass, A. E., Trockel, M., Jacobi, C., Wilfley, D. E., & Taylor, C. B. (2014). Psychiatric co-morbidity in women presenting across the continuum of disordered eating. *Eating Behaviors*, *15*(4), 686–693. https://doi.org/10.1016/j.eatbeh.2014.08.023

Almeida, M., Brown, T. A., Campos, P. F., Amaral, A. C. S., & Carvalho, P. H. B. (2021). Dissonance-based eating disorder prevention delivered in-person after an online training: A randomized controlled trial for Brazilian men with body dissatisfaction. *International Journal of Eating Disorders, 54*(3), 293–304. https://doi.org/10.1002/eat.23423

Arellano-Pérez, O. G., Vázquez-Cervantes, M. J., Fernández Cortés, T. L., & Saucedo-Molina, T. de J. (2019). Drive for muscularity and disordered eating behaviors and its relationship with anthropometric indicators and physical activity in Mexican adolescent men. *Eating and Weight Disorders, 24*(4), 661–670. https://doi.org/10.1007/s40519-019-00736-2

Baggs, M. (2021, April 28). Body image affects half of men's mental health, new study shows. *BBC*. https://www.bbc.com/news/newsbeat-56904488

Barnes, M., Abhyankar, P., Dimova, E., & Best, C. (2020). Associations between body dissatisfaction and self-reported anxiety and depression in otherwise healthy men: A systematic review and meta-analysis. *PLoS ONE, 15*(2). https://doi.org/10.1371/journal.pone.0229268

Brower, K. J., Blow, F. C., & Hill, E. M. (1994). Risk factors for anabolic-androgenic steroid use in men. *Journal of Psychiatric Research, 28*(4), 369–380.

Brown, T. A., Forney, K. J., Pinner, D., & Keel, P. K. (2017). A randomized controlled trial of The Body Project: More Than Muscles for men with body dissatisfaction. *International Journal of Eating Disorders, 50*(8), 873–883. https://doi.org/10.1002/eat.22724

Bucchianeri, M. M., Arikian, A. J., Hannan, P. J., Eisenberg, M. E., & Neumark-Sztainer, D. (2013). Body dissatisfaction from adolescence to young adulthood: Findings from a 10-year longitudinal study. *Body Image, 10*(1), 1-7. https://doi/org10.1016/j.bodyim.2012.09.001

Byrd-Bredbenner, C., Eck, K., & Quick, V. (2021). GAD-7, GAD-2, and GAD-mini: Psychometric properties and norms of university students in the United States. *General Hospital Psychiatry, 69*, 61-66. https://doi.org/10.1016/j.genhosppsych.2021.01.002

Cheung, G. W., & Lau, R. S. (2012). A direct comparison approach for testing measurement invariance. *Organizational Research Methods, 15*(2), 167-198. https://doi.org/10.1177/1094428111421987

Chua, S. N., Fitzsimmons-Craft, E. E., Austin, S. B., Wilfley, D. E., & Taylor, C. B. (2021). Estimated prevalence of eating disorders in Singapore. *International Journal of Eating Disorders, 54*(1), 7–18. https://doi.org/10.1002/eat.23440

da Silva, W. R., Santana, M. de S., Maroco, J., Maloa, B. F. S., & Campos, J. A. D. B. (2017). Body weight concerns: Cross-national study and identification of factors related to eating disorders. *PLoS ONE, 12*(7). https://doi.org/10.1371/journal.pone.0180125

Davis, C., & Cowles, M. (1991). Body image and exercise: A study of relationships and comparisons between physically active men and women. *Sex Roles, 25*(1–2), 33–44. https://doi.org/10.1007/BF00289315

De, S., & Chakraborty, R. (2015). Body dissatisfaction among female university students: Metacognitive awareness. *Psychological Studies, 60*(3), 257-264. https://doi.org/10.1007/s12646-015-0326-6

Eik, N. T. T., Austin, S. B., Blashill, A. J., Murray, S. B., & Calzo, J. P. (2018). Prospective health associations of drive for muscularity in young adult males. *International Journal of Eating Disorders, 51*(10), 1185–1193. https://doi.org/10.1002/eat.22943

Engeln, R., Sladek, M. R., & Waldron, H. (2013). Body talk among college men: Content, correlates, and effects. *Body Image, 10*(3), 300–308. https://doi.org/10.1016/j.bodyim.2013.02.001

Fitzsimmons-Craft, E. E., Balantekin, K. N., Graham, A. K., Smolar, L., Park, D., Mysko, C., Funk, B., Taylor, C. B., & Wilfley, D. E. (2019). Results of disseminating an online screen for eating disorders across the U.S.: Reach, respondent characteristics, and unmet treatment need. *International Journal of Eating Disorders, 52*(6), 721–729. https://doi.org/10.1002/eat.23043

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*(1), 85–91. https://doi.org/10.1002/eat.22180

Graham, A. K., Trockel, M., Weisman, H., Fitzsimmons-Craft, E. E., Balantekin, K. N., Wilfley, D. E., & Taylor, C. B. (2019). A screening tool for detecting eating disorder risk and diagnostic symptoms among college-age women. *Journal of American College Health, 67*(4), 357–366. https://doi.org/10.1080/07448481.2018.1483936

Grossbard, J. R., Atkins, D. C., Geisner, I. M., & Larimer, M. E. (2013). Does depressed mood moderate the influence of drive for thinness and muscularity on eating disorder symptoms among college men? *Psychology of Men & Masculinity, 14*(3), 281–287. https://doi.org/10.1037/a0028913

Haynos, A. F., Wall, M. M., Chen, C., Wang, S. B., Loth, K., & Neumark-Sztainer, D. (2018). Patterns of weight control behavior persisting beyond young adulthood: Results from a 15-year longitudinal study. *International Journal of Eating Disorders, 51*(9), 1090-1097. https://doi.org/10.1002/eat.22963

Healthy Minds Network (2022). *Research on adolescent and young adult mental health*. https://healthymindsnetwork.org/

Hobza, C. L., & Rochlen, A. B. (2009). Gender role conflict, drive for muscularity, and the impact of ideal media portrayals on men. *Psychology of Men & Masculinity, 10*(2), 120-130. https://doi.org/10.1037/a0015040

Hochgraf, A. K., McHale, S. M., & Fosco, G. M. (2018). Parent responsiveness and gender moderate bidirectional links between self-esteem and weight concerns during adolescence. *Journal of Family Psychology, 32*(6), 828–834. https://doi.org/10.1037/fam0000434.supp

Hochgraf, A. K., McHale, S. M., & Fosco, G. M. (2019). Interparental conflict and gender moderate the prospective link between parents' perceptions of adolescents' weight and weight concerns. *International Journal of Eating Disorders, 52*(8), 904–913. https://doi.org/10.1002/eat.23093

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

Jones, M., Lynch, K. T., Kass, A. E., Burrows, A., Williams, J., Wilfley, D. E., & Taylor, C. B. (2014). Healthy weight regulation and eating disorder prevention in high school students: A universal and targeted web-based intervention. J*ournal of Medical Internet Research, 16*(2), 28–39. https://doi.org/10.2196/jmir.2995

Kaminski, P. L., Chapman, B. P., Haynes, S. D., & Own, L. (2005). Body image, eating behaviors, and attitudes toward exercise among gay and straight men. *Eating Behaviors, 6*(3), 179–187. https://doi.org/j.eatbeh.2004.11.003

Keum, B. T., Miller, M. J., & Inkelas, K. K. (2018). Testing the factor structure and measurement invariance of the PHQ-9 across racially divers U.S. college students. *Psychological Assessment, 30*(8), 1096-1106. https://doi.org/10.1037/pas0000550

Killen, J. D., Taylor, C. B., Hayward, C., Wilson, D. M., Haydel, K. F., Hammer, L. D., Simmonds, B., Robinson, T. N., Litt, I., Varady, A., & Kramer, H. (1994). Pursuit of thinness and onset of eating disorder symptoms in a community sample of adolescent girls: A three-year prospective analysis. *International Journal of Eating Disorders, 16*(3), 227–238. https://doi.org/10.1002/1098-108X(199411)16:3<227::AID-EAT2260160303>3.0.CO;2-L

Kindermann, S., Moessner, M., Ozer, F., & Bauer, S. (2017). Associations between eating disorder related symptoms and participants' utilization of an individualized internet-based prevention and early intervention program*. International Journal of Eating Disorders, 50*(10), 1215–1221. https://doi-org/10.1002/eat.22774

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kutz, A. M., Marsh, A. G., Gunderson, C. G., Maguen, S., Masheb, R. M. (2020). Eating disorder screening: A systematic review and meta-analysis of diagnostic test characteristics of the SCOFF. *Journal of General Internal Medicine, 35*(2020), 885-893. https://doi.org/10.1007/s11606-019-05478-6

Lavender, J. M., De Young, K. P., & Anderson, D. A. (2010). Eating Disorder Examination Questionnaire (EDE-Q): Norms for undergraduate men. *Eating Behaviors, 11*(2), 119-121. https://doi.org/10.1016/j.eatbeh.2009.09.005

Lipson, S. K., & Sonneville, K. R. (2020). Understanding suicide risk and eating disorders in college student populations: Results from a national study. *International Journal of Eating Disorders, 53*(2), 229–238. https://doi-org/10.1002/eat.23188

Luce, K. H., Crowther, J. H., & Pole, M. (2008). Eating Disorder Examination Questionnaire (EDE-Q): Norms for undergraduate women. *International Journal of Eating Disorders, 41*(3), 273-276. https://doi.org/10.1002/eat.20504

Markey, C. N., Gomel, J. N., & Markey, P. M. (2008). Romantic relationships and eating regulation: An investigation of partners' attempts to control each other's eating behaviors. *Journal of Health Psychology, 13*(3), 422–432. https://doi.org/10.1177/1359105307088145

Markey, C., & Markey, P. (2011). Romantic partners, weight status, and weight concerns: An examination using the actor-partner interdependence model. *Journal of Health Psychology, 16*(2), 217–225. https://doi.org/10.1177/1359105310375636

McCabe, M. P., & Ricciardelli, L. A. (2004). Body image dissatisfaction among males across the lifespan. *Journal of Psychosomatic Research, 56*(6), 675–685. https://doi.org/10.1016/S0022-3999(03)00129-6

McCreary, D. R., Hildebrandt, T. B., Heinberg, L. J., Boroughs, M., & Thompson, J. K. (2007). A review of body image influences on men's fitness goals and supplement use. *American Journal of Men's Health, 1*(4), 307–316. https://doi.org/10.1177/1557988306309408

McCreary, D. R., & Sasse, D. K. (2000). An exploration of the drive for muscularity in adolescent boys and girls. *Journal of American College Health, 48*(6), 297-304. https://doi.org/10.1080/07448480009596271

McKinley, N. M. (2006). Longitudinal gender differences in objectified body consciousness and weight-related attitudes and behaviors: Cultural and developmental contexts in the transition from college. *Sex Roles, 54*(3–4), 159–173. https://doi.org/10.1007/s11199-006-9335-1

Meyers, L.S., Gamst, G., & Guarino, A. J. (2017). *Applied multivariate research: Design and interpretation* (3rd ed.). Sage.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research, 3*(1), 111-121. https://doi.org/10.21500/20112084.857

Morgan, J. F., Reid, F., & Lacey, J. H. (1999). The SCOFF questionnaire: Assessment of a new screening tool for eating disorders. *British Medical Journal, 319*(7223), 1467-1468. https://doi.org/10.1136/bmj.319.7223.1467

Neighbors, L. A., & Sobal, J. (2007). Prevalence and magnitude of body weight and shape dissatisfaction among university students. *Eating Behaviors, 8*(4), 429-439. https://doi.org/10.1016/j.eatbeh.2007.03.003

Nye, C. D., Bradburn, J., Olenick, J., Bialko, C., & Drasgow, F. (2019). How big are my effects? Examining the magnitude of effect sizes in studies of measurement equivalence. *Organization Research Methods, 22*(3), 678-709. https://doi.org/10.1177/1094428118761122

Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. *Journal of Applied Psychology, 96*(5), 966-980. https://doi.org/10.1037/a0022955

Palmeroni, N., Luyckx, K., Verschueren, M., & Claes, L. (2020). Body dissatisfaction as a mediator between identity formation and eating disorder symptomatology in adolescents and emerging adults. *Psihološka Obzorja / Horizons of Psychology, 60*(1), 328-346. https://doi.org/10.5334/pb.564

Paxton, S. J., Neumark-Sztainer, D., Hannan, P. J., & Eisenberg, M. E. (2006). Body dissatisfaction prospectively predicts depressive mood and low self-esteem in adolescent girls and boys. *Journal of Clinical Child and Adolescent Psychology, 35*(4), 539–549. https://doi.org/10.1207/s15374424jccp3504_5

Pearson, C. M., Guller, L., McPherson, L., Lejuez, C. W., & Smith, G. T. (2013). Validation of an existing measure of eating disorder risk for use with early adolescents. *Eating Behaviors, 14*(2), 113–118. https://doi.org/10.1016/j.eatbeh.2013.01.006

Pope, H. G., Olivardia, R., Gruber, A., & Borowiecki, J. (1999). Evolving ideals of male body image as seen through action toys. *International Journal of Eating Disorders, 26*(1), 65–72. https://doi.org/10.1002/(SICI)1098-108X(199907)26:1<65::AID-EAT8>3.0.CO;2-D

Regis, J. M. O., Ramos-Cerqueira, A. T. A., Lima, M. C. P., & Torres, A. R. (2018). Social anxiety symptoms and body image dissatisfaction in medical students: Prevalence and correlates. *Jornal Brasileiro de Psiquiatria, 67*(2), 65-73. https://doi.org/10.1590/0047-2085000000187

Ribeiro Dias, J. C., da Silva, W. R., Maroco, J., & Campos, J. A. D. B. (2015). Validation of the Weight Concerns Scale applied to Brazilian university students. *Body Image, 14*, 72–76. https://doi.org/10.1016/j.bodyim.2015.03.012

Ribeiro Dias, J. C., Maroco, J., & Bonini Campos, J. A. D. (2015). Weight Concerns Scale applied to college students: Comparison between pencil-and-paper and online formats. *Cyberpsychology, Behavior, and Social Networking, 18*(3), 188–192. https://doi.org/10.1089/cyber.2014.0392

Schaefer, L. M., Burke, N. L., Thompson, J. K., Dedrick, R. F., Heinberg, L. J., Calogero, R. M., Bardone-Cone, A. M., Higgins, M. K., Frederick, D. A., Kelly, M., Anderson, D. A., Schaumberg, K., Nerini, A., Stefanile, C., Dittmar, H., Clark, E., Adams, Z., Macwana, S., Klump, K. L., Vercellone, A. C., Paxton, S. J., & Swami, V. (2015). Development and validation of the Sociocultural Attitudes Towards Appearance Questionnaire-4 (SATAQ-4). *Psychological Assessment, 27*(1), 54–67. https://doi.org/10.1037/a0037917

Shannon, A., & Mills, J. S. (2015). Correlates, causes, and consequences of fat talk: A review. *Body Image, 15*, 158–172. https://doi.org/10.1016/j.bodyim.2015.09.003

Sladek, M. R., Engeln, R., & Miller, S. A. (2014). Development and validation of the Male Body Talk Scale: A psychometric investigation. *Body Image, 11*(3), 233–244. https://doi.org/10.1016/j.bodyim.2014.02.005

Spitzer, R. L., Kroenke, K., & Williams, J. B. W. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Streatfeild, J., Hickson, J., Austin, S. B., Hutcheson, R., Kandel, J. S., Lampert, J. G., Myers, E. M., Richmond, T. K., Samnaliev, M., Velasquez, K., Weissman, R. S., & Pezzullo, L. (2021). Social and economic cost of eating disorders in the United States: Evidence to inform policy action. *International Journal of Eating Disorders*, *54*(5), 851–868. https://doi.org/10.1002/eat.23486

SturtzSreetharan, C., Agostini, G., Wutich, A., Mitchell, C., Rines, O., Romanello, B., & Brewis, A. (2020). “I need to lose some weight”: Masculinity and body image as negotiated through fat talk. *Psychology of Men & Masculinities*, *21*(1), 148–161. https://doi.org/10.1037/men0000219

Szabó, K., Czeglédi, E., Babusa, B., Szumska, I., Túry, F., Sándor, I., & Bauer, S. (2015). The European initiative ProYouth for the promotion of mental health and the prevention of eating disorders screening results in Hungary. *European Eating Disorders Review*, *23*(2), 139–146. https://doi.org/10.1002/erv.2345

Tantleff-Dunn, S., Barnes, R. D., & Larose, J. G. (2011). It’s not just a “woman thing:” The current state of normative discontent. *Eating Disorders, 19*(5), 392–402. https://doi.org/10.1080/10640266.2011.609088

Table 1

Demographic Characteristics of the Sample (N = 6981)

Demographics	***M***		***SD***		**Range**	**% of Sample**	
	Women	**Men**	**Women**	**Men**	**Both**	**Women**	**Men**
Age	21.00	21.14	2.57	2.68	18-29		
Race/Ethnicity							
White						73.7	73.3
Black/African-American						7.1	4.7
American Indian or Alaskan Native						1.2	1.0
Asian American/Asian						15.4	16.3
Native Hawaiian or Pacific Islander						0.9	0.8
Middle Eastern, Arab, or Arab American						1.7	3.0
Other (Self-Identify)						0.8	2.1

Note. Cisgender women (n = 4733) and cisgender men (n = 2248).

Table 2

χ2 Difference Tests at the Metric and Scalar Levels

Model Level	**Difference Scaling Correction**	**Sattora-Bentler Chi-Square Difference**	**Difference in Degrees of Freedom**	***p*-value for difference**
Metric	0.979	76.095	4	<.001
Scalar	1.113	1369.934	5	<.001

Note. This information reflects the Weight Concern Scale when item one (WCS1) is held constant as the referent item.

Table 3

Metric invariance results of factor loadings and Scalar Invariance results of Intercepts per item on the Weight Concerns Scale

Item	Standardized Factor loading (Women)	Standardized Factor Loading (Men)	99% Bias Corrected CI of the Difference between men and women	Intercepts for women	Intercepts for men	99% Bias Corrected CI of the Difference between men and women	d_{macs}
WCS1	.655***	.812***	[-.59, .16][a]	3.15***	3.04***	[-.18, -.04][a]	.064
WCS2	.75***	.798***	[-.15, .06]	2.57***	1.72***	[-.91, -.77]	.546
WCS3	.505***	.555***	[-.15, .31]	3.0***	2.41***	[-.74, -.47]	.161
WCS4	.562***	.685***	[-.08, .05]	1.9***	1.73***	[-.21, -.12]	.17
WCS5	.771***	.784***	[.18, .37]	3.25***	2.39***	[-.93, -.80]	.427

Notes. WCS1 = item one of the Weight Concerns Scale (WCS), WCS2 = item two of the WCS, WCS3 = item three of the WCS, WCS4 = item four of the WCS, and WCS5 = item five of the WCS. This information reflects the WCS when WCS1 is held constant as the referent item, [a] denotes results for WCS1 when WCS2 is held constant as the referent item, and *** denotes significant < .001.

Table 4

Correlations of WCS versions with External Validity Measures

Measure	PHQ-9		GAD-7		SCOFF	
	Men	Women	Men	Women	Men	Women
WCS_T	.326***	.36***	.294***	.307***	.558***	.611***
WCS_2	.342***	.37***	.311	.309	.576***	.612***
WCS_5	.297***	.33***	.288***	.289***	.532***	.606***
WCS_2_5	.296***	.331***	.303	.282	.533***	.619***

Note. WCS_T = WCS with all items, WCS_2 = WCS without item 2, WCS_5 = WCS without item 5, WCS_2_5 = WCS without items 2 and 5,*** = difference between men and women, and all correlations were significant < .001.