

Measurement Invariance of the Multidimensional Body-Self Relations Questionnaire: Can We Compare Across Age and Gender?

Shayna A. Rusticus · Anita M. Hubley

Published online: 1 December 2006
© Springer Science + Business Media, Inc. 2006

Abstract The Multidimensional Body-Self Relations Questionnaire (MBSRQ) is one of the most widely used body image measures and one of the few measures developed with men and women of a wide age range. To make age and gender comparisons, however, measures must exhibit cross-group equivalence. Whether the MBSRQ subscales can be used to make such comparisons was examined in a sample of 1,262 adults (422 men, 840 women) aged 18 to 98 years. The results showed that body image was perceived quite differently across the groups and that not all MBSRQ subscales may be used to make age and gender comparisons. The importance of examining measurement invariance in body image measures is highlighted and recommendations for use of the MBSRQ are offered.

Keywords Body image · Measurement invariance · Gender differences · Age differences

The vast majority of body image research has focused on adolescents and college-aged women. Comparatively little research has been conducted with men, and even less with older adults. This has created a gap in the literature in our understanding of body image issues among these understudied populations. However, before the various aspects of body image can be examined and compared across groups, evidence is needed to show that the measures used function

equally for all groups in question (i.e., are invariant). This is an issue that has yet to be addressed in the body image field. All too often, measures are used with the assumption that they are measuring the same concept(s) across groups. For example, measures of body image may be validated for use with college samples but then used to make comparisons with other age groups. There are two problems with this. First, validity evidence should be provided for each age group to demonstrate the appropriateness of the measure for each group. Second, even if the measure has been found to provide valid information in all age groups, this does not guarantee that the measure functions the same way across groups as required for comparison purposes (Horn & McArdle, 1992).

Measurement invariance refers to whether “under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute” (Horn & McArdle, 1992, p. 117). Only through the demonstration of measurement invariance can a scale be deemed to measure the same attribute across groups. If there is no evidence of the presence or absence of measurement invariance, or if invariance is not obtained, any differences found between groups cannot be interpreted unambiguously. For example, age differences on a scale to measure fitness importance might be due to true differences between age groups on the underlying latent variable, such that younger individuals rate their fitness levels as more important to them than older individuals do, or these differences might be due to systematic biases in the way people of different ages respond to certain items, such that an item about physical strength may be more salient than the other items to younger individuals, whereas such is not the case for older individuals. Findings of no differences between groups are also open to corresponding alternative

S. A. Rusticus · A. M. Hubley (✉)
Dept. of ECPS, University of British Columbia,
2125 Main Mall, Vancouver, BC, Canada V6T 1Z4
e-mail: anita.hubley@ubc.ca

interpretations. In short, as stated by Horn (1991), “Without evidence of measurement invariance, the conclusions of a study must be weak” (p. 119).

Evidence of measurement invariance is accumulated on an incremental basis. Configural invariance, the weakest form of invariance, assesses whether the configuration of the salient and nonsalient factor loadings is equivalent across groups (i.e., each group has the same factor pattern). Configural invariance is the minimum condition for factorial invariance (Horn & McArdle, 1992). When configural invariance is met for a measure, it indicates that the same basic concept is being measured in each of the groups. However, because this level of invariance alone does not guarantee that the unit of measurement of the latent variable is the same for each group, one cannot make comparisons across groups. As an example, if a measure of appearance importance is found to exhibit configural invariance across gender, then this measure can be given to a combined sample of men and women to assess appearance importance for the whole group; however, comparisons across gender should not be conducted. If configural invariance is *not* supported for a measure, it may indicate that the groups conceptualized the construct differently, the groups attached different meanings to the construct, or that an extraneous variable introduced into the study (e.g., in data collection) differentially affected the groups (Cheung & Rensvold, 2002).

The next level of invariance is metric (or weak) invariance. Metric invariance assesses whether the unstandardized factor pattern weights (i.e., factor loadings) are equal across the groups. In so doing, one determines whether item scores are scaled to the factor scores using the same unit of measurement across the groups. This level of invariance provides a strong basis for inference that individuals from each of the groups interpret and respond to the measure in a similar way (Horn & McArdle, 1992). When metric invariance has been shown for a measure, the measure may be used to examine structural relationships or correlations between the construct of interest and other constructs across groups (Steenkamp & Baumgartner, 1998). Because this level of invariance cannot rule out whether scores on items may be systematically biased upward or downward for some groups, it is not recommended that one examine mean differences across groups. If metric invariance is *not* supported for a measure, it could suggest that the latent variable is poorly operationalized or that cross-group differences exist in how the latent variable is conceptualized (Cheung & Rensvold, 2000).

Finally, scalar (or strong) invariance assesses whether the factor loadings and intercepts are equal across the groups. This indicates that individuals who have the same value on the latent variable would obtain the same value on the observed variable regardless of their group membership.

Evidence of scalar invariance is necessary to make mean comparisons across groups (Meredith, 1993; Steenkamp & Baumgartner, 1998). When scalar invariance has been shown for a measure, the measure may be used to assess cross-group mean differences on the observed scores on the measure. If scalar invariance is *not* supported for a measure, the measure should not be used for cross-group mean comparisons because bias exists in how the groups respond to the indicators. Two possible causes of this bias could be group differences in (a) levels of extreme response styles, whereby one group has a greater tendency to select the extreme points on a Likert-type scale, or (b) acquiescence response styles, whereby one group has a tendency to systematically give higher or lower responses (e.g., women always responding two points higher than men do on a Likert-type scale; Cheung & Rensvold, 2000). A third possible cause relates to the relevance of the items that define the construct, whereby an item may be endorsed at a much higher rate for one group than the other because it “is more salient as a marker” of the latent variable for that group (Chan, 2000, p. 177).

The Multidimensional Body-Self Relations Questionnaire (MBSRQ; Cash, Winstead, & Janda, 1986) was selected for the present study because it is one of the most widely used measures of body image. No researchers, however, have specifically examined whether the subscales of the MBSRQ function equivalently across age and/or gender groups, and, thus, whether age or gender comparisons on this measure are appropriate. Unlike most body image measures, the standardization sample for the MBSRQ was selected to be representative of the U.S. population in terms of gender and age; thus, it is one of the few measures not standardized only with college-aged students. However, Horn and McArdle (1992) pointed out that the fact that a measure has been standardized with a varied sample does not mean that meaningful comparisons can be made across subsamples. Evidence of measurement invariance is still needed to ensure that valid group comparisons can be made. The purpose of the present study was to examine the configural, metric, and scalar invariance of the MBSRQ across three age groups of men and women to determine if it is appropriate to use this measure to make age and gender comparisons.

Method

Participants

A total of 1,262 participants (422 men, 840 women) took part in this study. The men ranged in age from 18 to 98 years ($M=39.7$, $SD=19.1$), and the women ranged in age from 18 to 89 years ($M=39.4$, $SD=18.7$). Participants were grouped

into three age categories as follows: 18 to 29 years=young adulthood¹ (185 men, 364 women), 30 to 54 years=middle-age (131 men, 267 women), and 55 years and older=older adulthood (106 men, 209 women). For the overall sample, 74.7% of participants identified themselves as White, 12.5% as East Asian, 5.5% as South Asian or Middle Eastern, 1.3% as Hispanic, and 5.7% as Other. Nearly 45% of the participants indicated that they had never been married, 40.3% were married or in common-law partnerships, 9.4% were divorced or separated, and 4.3% were widowed. Participants tended to be well educated: less than high school=2.5%, high school=13.8%, some college or university=42.3%, Bachelor's degree=24.7%, and Master's or PhD=16.4%.

Measures

The MBSRQ is a 69 item self-report inventory comprised of ten subscales that assess cognitive, behavioral, and affective components of body image (Cash, 2000; Cash et al., 1986). Mean scores for each subscale are calculated by taking the mean of its corresponding items (Cash, 2000). The Appearance Evaluation (AE) subscale (seven items) assesses feelings about physical appearance; higher scores indicate greater satisfaction with appearance. The Appearance Orientation (AO) subscale (12 items) assesses investment in appearance; higher scores indicate more importance and attention placed on looks and more engagement in grooming activities. The Fitness Orientation (FO) subscale (13 items) assesses investment in fitness level; higher scores indicate more value placed on fitness and more involvement in fitness activities. The Health Evaluation (HE) subscale (six items) assesses feelings about health; higher scores indicate perceptions of a healthy body. The Health Orientation (HO) subscale (eight items) assesses investment in a healthy lifestyle; higher scores indicate a more "health conscious" lifestyle. The Illness Orientation (IO) subscale (five items) assesses reactivity to being or becoming ill; higher scores indicate a greater awareness of personal symptoms of physical illness and a greater likelihood to seek medical attention. The Body Areas Satisfaction (BAS) subscale (nine items) assesses satisfaction with discrete aspects of appearance; higher scores indicate contentment with more areas of one's body. The Overweight Preoccupation (OP) subscale (four items) assesses fat anxiety, weight vigilance, dieting, and eating restraint; higher scores indicate greater weight preoccupa-

tion. Because the remaining two subscales (Fitness Evaluation and Self-Classified Weight) have only three and two items, respectively, invariance testing could not be conducted on these subscales. Thus, only eight of the ten subscales were investigated.

Procedure

Recruitment for this survey took place by means of "snowball sampling" via emails sent to student listserves asking students to take part and/or forward the survey to other adults they know, posters distributed throughout the community, and oral announcements made in psychology and human kinetics classrooms, community centers, senior citizens centers, and shopping malls. The data was collected in two forms: a web-based survey ($n=819$) and a paper and pencil survey ($n=443$). The content of the two surveys was identical. For the web survey, individuals were provided with a link to the survey materials and interested individuals completed the survey at a time and location of their choice. All information was collected on a secure server. Once data collection was complete, all survey materials were removed from the Internet. For the paper survey, an envelope that contained the research materials was provided to interested individuals. Once the survey was complete, participants were instructed to place the materials in a sealed envelope and to return the package by a specified time to a place set by one of the researchers or research assistants.

Model evaluation

All tests of measurement invariance were investigated using multi-group confirmatory factor analyses (MG-CFA) in LISREL (Joreskog & Sorbom, 1993). The maximum likelihood (ML) estimation method with a Pearson product moment covariance matrix was used to analyze the data. For each level of invariance, up to a total of six models were tested. For each model, we fixed the scale of the latent variable by fixing its variance to 1.0. The first model that was tested in all cases, herein called the full model, was the model that tested the three age groups (young adult, middle-aged adult, older adult) by two gender groups (men, women). If the full model did not meet invariance requirements, then five age and gender subgroups were tested. The women's model tested each female group across the three age groups, whereas the men's model tested each male group across the three age groups. The young adults' model tested the men and women at the youngest age group. Similarly, the middle-aged and older adults' models tested the men and women within their respective age groups.

Configural invariance was evaluated by using the commonly used chi-square test as well as four goodness-of-fit

¹ Young adults were categorized as 18–29 because the majority of research in the body image field has focused on young adults in their late teens and early 20s. It was felt that this age grouping would make the group of young adults in the present study more comparable to those in previous research.

indices that were recommended by Steenkamp and Baumgartner (1998): the root mean square error of approximation (RMSEA; values less than 0.08 indicate acceptable fit; Browne & Cudeck, 1993), the Comparative Fit Index (CFI; values of 0.90 or greater indicate acceptable fit; Vandenberg & Lance, 2000), the non-normed fit index (NNFI, also called the Tucker–Lewis Index; values of 0.90 or greater indicate acceptable fit; Vandenberg & Lance, 2000), and the consistent Akaike information criteria (CAIC; if the model CAIC is less than both the independence and saturated CAIC, it indicates acceptable fit; Diamantopoulos & Siguaw, 2000).

To test metric and scalar invariance, all models were placed in a hierarchical sequence of nested models so that systematic comparison tests could be conducted (Joreskog, 1971). Although the degree of invariance across nested models is most frequently assessed by chi-square difference tests (a critical value of less than 0.01 was used here), researchers have shown that differences in chi-square values are dependent on sample size (Brannick, 1995; Kelloway, 1995). Thus, Cheung and Rensvold (2002) recommended using change in CFI (Δ CFI) to assess differences between the models; values between 0 and -0.01 indicate model invariance. In cases where there was disagreement between the conclusions of the chi-square difference test and Δ CFI, the latter index was given more weight.

Figure 1 outlines the steps that were used to test for measurement invariance. The first step was to test configural invariance for the full model. This was tested by examining whether the items of each of the scales exhibited significant nonzero loadings on salient factors and zero loadings on non-salient factors. If this model was found to fit well, it could serve as a baseline model for comparisons with more restricted models. If configural invariance was not supported for the full model, configural invariance testing was examined for the five subgroups (e.g., women's model, young adults' model). For those models that met the requirements of configural invariance, further tests of metric invariance were conducted. For those models that did not meet configural invariance, principal components analyses were conducted on the individual groups to explore whether a different factor structure than the one specified in the literature emerged.

The second test of invariance that was examined was metric invariance. This was tested by constraining the matrix of factor loadings to be invariant across groups. If metric invariance was not supported for the full model, metric invariance testing was conducted for the five subgroups. For those models that met the requirements of metric invariance, tests of scalar invariance were conducted. If metric invariance was not found for any subgroup, testing was stopped.

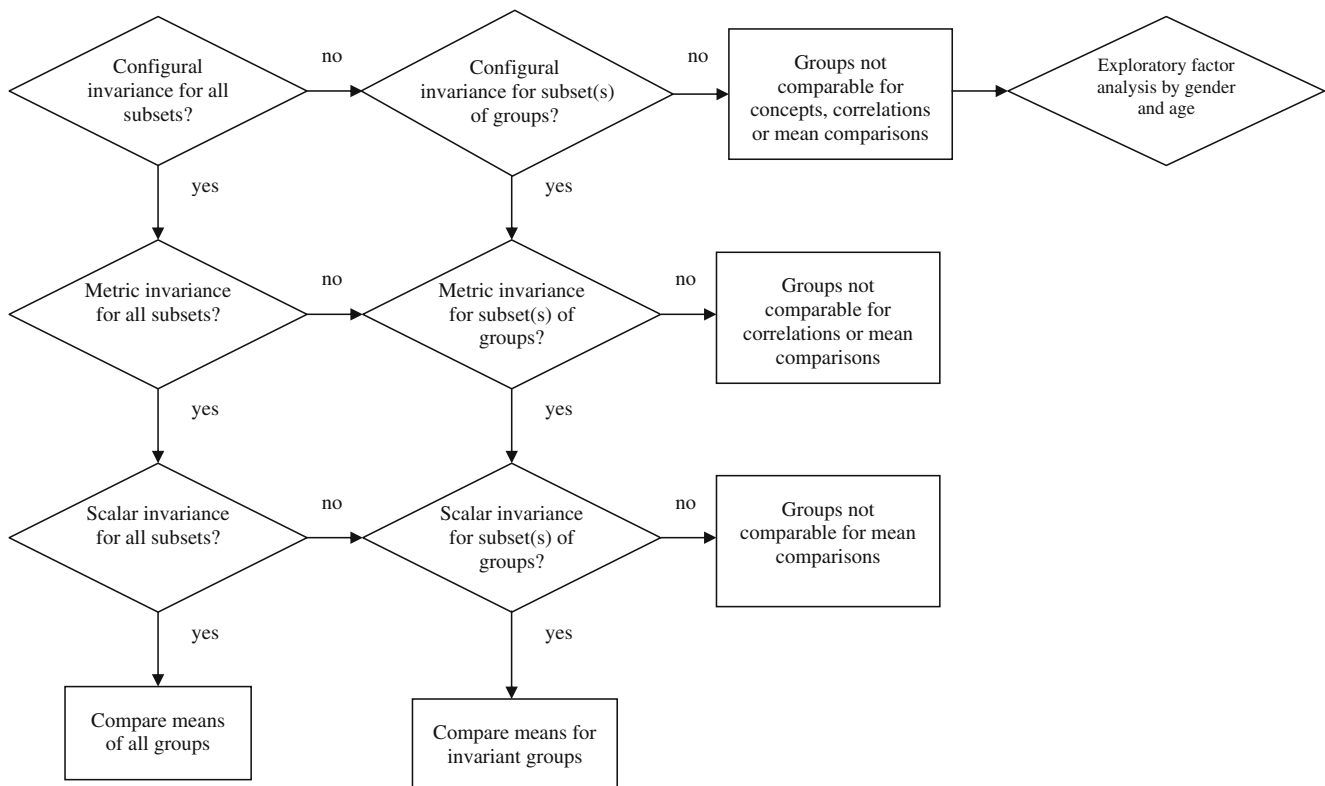


Fig. 1 Steps for assessing measurement invariance.

Table 1 Goodness-of-fit indices for the MBSRQ—appearance evaluation.

Model	χ^2	<i>df</i>	RMSEA	NNFI	CFI	Model CAIC	Independence CAIC	Saturated CAIC	$\Delta\chi^2$	Δdf	ΔCFI	Invariance?
Configural—full model	454.28*	84	0.15	0.91	0.94	1,163.14 ^a	6,989.35	1,366.53	–	–	–	Yes
Metric—full model	630.75*	119	0.15	0.92	0.92	1,039.29 ^a	6,989.35	1,366.53	176.47*	35	–0.02	No
Women’s model	296.29*	56	0.13	0.95	0.95	533.10 ^a	5,359.36	649.10	14.71	14	0	Yes
Men’s model	290.46*	56	0.18	0.82	0.84	508.10 ^a	1,650.44	591.18	117.76*	14	–0.06	No
Young adults’ model	225.70*	35	0.15	0.93	0.94	404.92 ^a	3,577.22	409.05	11.82	7	0	Yes
Middle-aged adults’ model	226.70*	35	0.16	0.89	0.91	366.87 ^a	2,181.94	391.24	98.22*	7	–0.04	No
Older adults’ model	138.61*	35	0.13	0.89	0.91	272.58 ^a	1,234.81	377.07	26.68*	7	–0.02	No
Scalar—women’s model	391.56*	66	0.14	0.94	0.94	735.18	5,454.40	649.20	109.98*	24	–0.01	Yes
Young adults’ model	316.91*	40	0.18	0.91	0.92	598.65	3,432.72	409.05	103.03*	12	–0.02	No

* $p < 0.001$ ^aModel CAIC < than both independence and saturated CAICs

The final test of invariance that was examined was scalar invariance. This was tested by constraining the vector of item intercepts across groups and then examining model fit. If this model was found to fit well, then mean differences of observed scores could be compared and such differences can be considered reflections of true differences between the groups on the latent variable. If scalar invariance was not supported for the full model, scalar invariance testing was conducted for the five subgroups. For those subgroups that met the requirements of scalar invariance, mean score comparisons were conducted on that scale for those groups. If scalar invariance was not found for any subgroup, testing was stopped.

² Overall model fit, using the chi-square test and goodness of fit indices, was only used to appraise configural invariance. Change in chi-square and change in CFI were used for metric and scalar invariance tests. However, for completeness of the tables, the results of the chi-square test and goodness-of-fit indices are presented for each level of invariance testing.

Results

The MBSRQ consists of separate subscales that can be used jointly or independently. As these subscales do not sum to an overall total score, and the entire measure is not always used within a single study, all measurement invariance tests were conducted separately for each subscale. Each subscale was assumed to be a unidimensional scale.

Appearance evaluation

Table 1 summarizes the results of the measurement invariance tests for the AE subscale². The first test of invariance across the six age and gender groups was for configural invariance for the full model. As three of the fit statistics indicated an acceptable fit of the model (i.e., NNFI, CFI, CAIC), it was concluded that configural invariance was supported for the full model.

Next, metric invariance for the full model was tested. Both the significant increase in chi-square from the configural invariance model and the ΔCFI indicate that

the hypothesis of metric invariance was *not* tenable for the full model. The next step was to assess whether metric invariance was established for any of the five age and gender subgroups. As summarized in Table 1, metric invariance was not met for the men's model (i.e., young, middle-aged, and older men), middle-aged adults' model (i.e., men and women within this age group), or older adults' model (i.e., men and women within this age group). Metric invariance was shown for the women's model (i.e., young, middle-aged, and older women) and for the young adults' model (i.e., men and women within this age group), which indicates that tests of scalar invariance could be conducted for these two models.

In the test for scalar invariance for the women's model, the significant increase in chi-square from the configural invariance model indicates that the hypothesis of scalar invariance was *not* tenable. However, the Δ CFI indicates that the hypothesis of scalar invariance was tenable. Because differences in chi-square values have been found to be dependent on sample size (Brannick, 1995; Kelloway, 1995), more weight was given to Δ CFI, and it was concluded that scalar invariance was shown for the women's model. When we tested scalar invariance for the young

adults' model, both the significant increase in chi-square from the configural invariance model and the Δ CFI indicate that the hypothesis of scalar invariance was *not* tenable.

Appearance orientation

Table 2 summarizes the results of the measurement invariance tests for the AO subscale. As three of the fit indices (i.e., NNFI, CFI, CAIC) indicate an acceptable fit of the model, it was concluded that configural invariance was supported for the full model. When we tested metric invariance for the full model, the significant increase in chi-square from the configural invariance model indicates that the hypothesis of invariance was *not* tenable. However, the Δ CFI indicates that the hypothesis of invariance was tenable. We gave more weight to the Δ CFI results, and concluded that metric invariance was observed for the full model. When we tested scalar invariance for the full model, both the significant increase in chi-square from the configural invariance model and the Δ CFI indicate that the hypothesis of scalar invariance was *not* tenable for the full model. Thus, tests were conducted to determine whether scalar invariance could be shown for any of the

Table 2 Goodness-of-fit indices for the MBSRQ—appearance orientation.

Model	χ^2	<i>df</i>	RMSEA	NNFI	CFI	Model CAIC	Independence CAIC	Saturated CAIC	$\Delta\chi^2$	Δ <i>df</i>	Δ CFI	Invariance?
Configural—full model	819.51**	324	0.09	0.94	0.95	2,022.71 ^a	11,153.07	3,806.76	—	—	—	Yes
Metric—full model	999.69**	384	0.09	0.94	0.94	1,727.67 ^a	11,153.07	3,806.76	180.18*	60	−0.01	Yes
Scalar—full model	1,542.20**	434	0.12	0.90	0.89	2,547.76 ^a	11,159.06	3,807.13	722.69**	110	−0.06	No
Women's model	835.83**	206	0.11	0.91	0.91	1,390.26 ^a	7,515.09	1,808.50	296.62**	44	−0.04	No
Men's model	523.94**	206	0.11	0.90	0.90	985.90 ^a	3,589.85	1,646.86	242.84**	44	−0.06	No
Young adults' model	474.10**	130	0.10	0.91	0.91	880.50 ^a	4,073.33	1,139.49	137.74**	22	−0.03	No
Middle-aged adults' model	352.30**	130	0.09	0.94	0.91	701.13 ^a	3,793.63	1,089.89	109.27**	22	−0.01	Yes
Older adults' model	312.86**	130	0.10	0.94	0.94	686.91 ^a	3,211.06	1,050.91	71.95**	22	−0.02	No

* $p < 0.01$

** $p < 0.001$

^a Model CAIC < than both independence and saturated CAICs

Table 3 Goodness-of-fit indices for the MBSRQ—fitness orientation.

Model	χ^2	<i>df</i>	RMSEA	NNFI	CFI	Model CAIC	Independence CAIC	Saturated CAIC	$\Delta\chi^2$	Δdf	ΔCFI	Invariance?
Configural—full model	1,372.06*	395	0.11	0.93	0.94	1,625.76 ^a	17,662.22	4,441.22	–	–	–	Yes
Metric—full model	1,534.49*	460	0.11	0.93	0.94	2,256.94 ^a	17,662.22	4,441.22	207.43*	65	0	Yes
Scalar—full model	1,866.08*	510	0.12	0.92	0.92	2,945.70 ^a	17,689.19	4,441.65	493.62*	115	–0.02	No
Women's model	1,112.47*	243	0.12	0.93	0.93	1,768.37 ^a	12,701.51	2,109.92	285.39*	46	–0.02	No
Men's model	638.72*	243	0.10	0.91	0.91	1,096.63 ^a	4,929.03	1,921.34	95.96*	46	–0.01	Yes
Young adults' model	548.38*	154	0.10	0.95	0.95	945.77 ^a	8,035.81	1,329.41	70.31*	23	0	Yes
Middle-aged adults' model	515.83*	154	0.11	0.94	0.94	921.14 ^a	6,065.63	1,271.53	63.69*	23	0	Yes
Older adults' model	483.04*	154	0.12	0.89	0.90	874.76 ^a	3,499.91	1,226.06	40.07	23	0	Yes

* $p < 0.001$ ^a Model CAIC < than both independence and saturated CAICs

five age and gender subgroups. As summarized in Table 2, scalar invariance was not met for the women's, men's, young adults', or older adults' models. Scalar invariance was only supported for the middle-aged adult model.

Fitness orientation

Table 3 summarizes the results of the measurement invariance tests for the FO subscale. Three of the fit indices (i.e., NNFI, CFI, CAIC) indicate an acceptable fit of the model, so it was concluded that configural invariance was supported for the full model. With respect to metric invariance for the full model, the significant increase in chi-square from the configural invariance model indicates that the hypothesis of invariance was *not* tenable, but the ΔCFI indicates differently. Following the arguments in previous analyses, it was concluded that metric invariance was shown for the full model. Scalar invariance for the full model was tested next. Both the significant increase in chi-square from the configural invariance model and the ΔCFI indicate that the hypothesis of scalar invariance was *not* tenable for the full model. The final step was to assess whether scalar invariance was established for any of the five age and gender subgroups. As summarized in Table 3, scalar invariance was not shown for the women's model, but was found for the men's, young adults', middle-aged adults', and older adults' models.

Health evaluation

Table 4 summarizes the results of the measurement invariance tests for the HE subscale. The chi-square test and two of the fit indices (i.e., RMSEA, NNFI) did *not* support the fit of this model, so it was concluded that configural invariance was not supported for the full model. The next step was to test whether configural invariance was present for any of the five age and gender subgroups. As summarized in Table 4, configural invariance was not supported for the women's model, young adults' model, or middle-aged adults' model. Configural invariance was supported, however, for the men's model and older adults' model and thus metric invariance for these models was tested next. For both models, the significant increase in chi-square from the configural invariance model and the ΔCFI indicate that the hypothesis of metric invariance was *not* tenable.

To explore why configural invariance did not hold for the women's, young adults' or middle-aged adults' model, exploratory principal components analyses (PCAs) were conducted on this subscale to determine whether a non-unidimensional factor pattern emerged for these groups. The number of factors to extract for each group was determined by conducting a parallel analysis (Reise, Waller, & Comrey, 2000). The results of the PCA for the three women's age groups indicated the presence of one factor

for all three groups. Configural invariance held for the men's model, which indicates that a one factor model holds for each of the men's age groups; thus, it was not necessary to run a PCA for the young adult or middle-aged adult men.

Health orientation

Table 5 summarizes the results of the measurement invariance tests for the HO subscale. The chi-square test and two of the fit indices (i.e., RMSEA, NNFI) did *not* support the fit of this model; thus it was concluded that configural invariance was not supported for the full model. The next step in the analysis was to test whether configural invariance could be found for any of the five age and gender subgroups. As summarized in Table 5, configural invariance was not met for the women's, young adults', or middle-aged adults' models. Configural invariance was supported for the men's and older adults' model and thus the next step was to test for metric invariance for these groups. For the men's model, both the non-significant chi-square and the Δ CFI indicate that the hypothesis of metric invariance was tenable. For the older adults' model, both the significant increase in chi-square from the configural invariance model and the Δ CFI indicate that the hypothesis of metric

invariance was *not* tenable. Because metric invariance was supported for the men's model, the final test of measurement invariance examined for this group was for scalar invariance. Both the significant increase in chi-square from the configural invariance model and the Δ CFI indicate that the hypothesis of scalar invariance was *not* tenable.

To explore why configural invariance did not hold for the women's, young adults', or middle-aged adults' models, exploratory PCAs were conducted on this subscale separately for each of the three women's age groups to determine whether a non-unidimensional factor pattern emerged for these groups. The results of the PCA indicated that two factors accounted for the variance of the items for the young adult and middle-aged adult women, whereas one factor accounted for the variance of the items for the older adult women. For the young adult and middle-aged adult women, items 8 (health knowledge), 29 (reading health literature), and 52 (fitness knowledge) loaded on one factor, and items 18 (health importance), 19 (avoid health threats), 28 (health taken for granted), and 38 (no nutrition effort) loaded on a second factor. Item 9 (healthy lifestyle) loaded on both factors for the young adult women and on the second factor for the middle-aged adult women.

Table 4 Goodness-of-fit indices for the MBSRQ—health evaluation.

Model	χ^2	<i>df</i>	RMSEA	NNFI	CFI	Model CAIC	Independence CAIC	Saturated CAIC	$\Delta\chi^2$	Δdf	Δ CFI	Invariance?
Configural—full model	270.63**	54	0.14	0.89	0.93	865.18 ^a	3,654.48	1,024.90	—	—	—	No
Women's model	203.81**	27	0.16	0.88	0.93	490.28	2,686.18	486.83	—	—	—	No
Men's model	66.82**	27	0.10	0.91	0.95	320.80 ^a	941.25	443.39	—	—	—	Yes
Young adults' model	114.22**	18	0.14	0.88	0.93	294.55 ^a	1,489.30	306.79	—	—	—	No
Middle-aged adults' model	97.83**	18	0.15	0.86	0.92	271.16 ^a	1,042.39	293.43	—	—	—	No
Older adults' model	58.58**	18	0.12	0.93	0.96	218.40 ^a	1,082.26	282.80	—	—	—	Yes
Metric—men's model	101.69**	39	0.10	0.91	0.92	266.35 ^a	941.25	443.39	34.87**	12	-0.03	No
Older adults' model	79.75**	24	0.12	0.93	0.94	194.79 ^a	1,085.70	282.94	21.17*	6	-0.02	No

* $p < 0.01$

** $p < 0.001$

^aModel CAIC < than both independence and saturated CAICs

Table 5 Goodness-of-fit indices for the MBSRQ—health orientation.

Model	χ^2	<i>df</i>	RMSEA	NNFI	CFI	Model CAIC	Independence CAIC	Saturated CAIC	$\Delta\chi^2$	Δdf	ΔCFI	Invariance?
Configural—full model	435.69**	120	0.12	0.88	0.92	1,247.23 ^a	4,318.04	1,756.96	–	–	–	No
Women's model	289.61**	60	0.12	0.87	0.92	687.89 ^a	2,768.43	834.56	–	–	–	No
Men's model	146.07**	60	0.10	0.90	0.93	487.20 ^a	1,513.54	760.09	–	–	–	Yes
Young adults' model	193.22**	40	0.13	0.87	0.91	450.21 ^a	1,869.61	525.92	–	–	–	No
Middle-aged adults' model	141.48**	40	0.12	0.87	0.91	371.06 ^a	1,260.27	503.02	–	–	–	No
Older adults' model	100.98**	40	0.10	0.91	0.94	317.87 ^a	1,134.11	484.80	–	–	–	Yes
Metric—men's model	177.05**	76	0.10	0.92	0.93	408.75 ^a	1,623.26	760.09	30.99	16	0	Yes
Older adults' model	123.02**	48	0.10	0.89	0.91	286.33 ^a	953.61	484.80	22.04*	8	–0.03	No
Scalar—men's model	255.66**	88	0.12	0.87	0.87	578.58 ^a	1,513.54	760.09	109.59**	12	–0.06	No

* $p < 0.01$ ** $p < 0.001$ ^a Model CAIC < than both independence and saturated CAICs

Illness orientation

Table 6 summarizes the results of the measurement invariance tests for the IO subscale. Three of the fit indices (NNFI, CFI, CAIC) indicated an acceptable fit of the model, so it was concluded that configural invariance was supported for the full model. The next step in the analysis was to test for metric invariance for the full model. Both the significant increase in chi-square from the configural invariance model and the ΔCFI indicate that the hypothesis of metric invariance was *not* tenable. Next, metric invariance was examined for the five age and gender subgroups. As summarized in Table 6, metric invariance was not shown for the men's model or middle-aged adults' model. Metric invariance was met for the women's, young adults', and older adults' model, and thus the final step was to examine scalar invariance for these models. As indicated in Table 6, scalar invariance was not met for the young adults' model or older adults' model. Scalar invariance was shown for the women's model.

Body areas satisfaction subscale

Table 7 summarizes the results of the measurement invariance tests for the BAS subscale. The chi-square test and two of the fit indices (i.e., RMSEA, NNFI) did *not* support the fit of this model, so it was concluded that configural invariance was not supported for the full model.

The next step was to test whether configural invariance could be shown for any of the five age and gender subgroups. As summarized in Table 7, configural invariance was not met for the women's, men's, middle-aged adults', or older adults' models. Configural invariance was met, however, for the young adults' model, and so the next step was to test for metric invariance for this model. Both the non-significant increase in chi-square from the configural invariance model and the ΔCFI indicate that the hypothesis of metric invariance was tenable. The final test of measurement invariance for this model was for scalar invariance. Both the significant increase in chi-square from the configural invariance model and the ΔCFI indicate that the hypothesis of scalar invariance was *not* tenable.

To explore why configural invariance did not hold for the women's, men's, middle-aged adults', or older adults' models, exploratory PCAs were conducted on this subscale separately for each group to determine whether a non-unidimensional factor pattern emerged for any group. The results of the PCAs indicated that two factors accounted for the variance of the items for the middle-aged adult group of women and that one factor accounted for the variance of the items for the young and older adult groups of women and for all three groups of men. For the middle-aged adult group of women, items 61 (face), 62 (hair), and 68 (height) loaded on one factor, and items 63 (lower torso), 64 (mid torso), 65 (upper torso), 66 (muscle tone), 67 (weight), and 69 (overall) loaded on a second factor.

Table 6 Goodness-of-fit indices for the MBSRQ—illness orientation.

Model	χ^2	<i>df</i>	RMSEA	NNFI	CFI	Model CAIC	Independence CAIC	Saturated CAIC	$\Delta\chi^2$	Δdf	ΔCFI	Invariance?
Configural—full model	110.58*	30	0.12	0.90	0.95	604.08 ^a	1,980.45	732.07	–	–	–	Yes
Metric—full model	189.87*	54	0.11	0.91	0.92	482.27 ^a	1,989.44	732.14	207.43*	24	–0.03	No
Women's model	101.91*	24	0.14	0.92	0.93	270.32 ^a	1,330.83	347.79	15.24	9	–0.01	Yes
Men's model	79.69*	24	0.13	0.86	0.89	225.45 ^a	636.07	316.70	55.82*	9	–0.09	No
Young adults' model	72.88*	14	0.13	0.89	0.92	196.41 ^a	837.82	219.13	7.43	4	–0.01	Yes
Middle-aged adults' model	45.20*	14	0.14	0.91	0.93	155.60 ^a	567.78	209.59	21.28*	4	–0.04	No
Older adults' model	22.51	14	0.06	0.97	0.98	129.52 ^a	550.08	202.10	1.34	4	0	Yes
Scalar—women's model	143.80*	31	0.12	0.91	0.90	371.04	1,330.83	347.79	57.13*	16	–0.04	No
Young adults' model	74.86*	18	0.11	0.92	0.92	241.90	837.82	219.13	9.41	8	–0.01	Yes
Older adults' model	105.32*	18	0.15	0.80	0.82	246.82	567.78	209.59	84.15*	8	–0.16	No

* $p < 0.001$ ^aModel CAIC < than both independence and saturated CAICs

Overweight preoccupation

Table 8 summarizes the results of the measurement invariance tests for the OP subscale. Three of the fit indices (i.e., NNFI, CFI, CAIC) indicated an acceptable fit of the model, so it was concluded that configural invariance was supported for the full model. The next step was to test for metric invariance for the full model. Both the non-significant increase in chi-square from the configural invariance model and the ΔCFI indicate that the hypothesis of metric invariance was tenable. This was followed by a test for scalar invariance for the full model. Both the significant increase in chi-square from the configural invariance model and the ΔCFI indicate that the hypothesis of scalar invariance was *not* tenable. The final step was to assess whether scalar invariance was established for any of the five age and gender subgroups. As summarized in Table 8, scalar

invariance was not met for the women's or men's models. Scalar invariance was shown, however, for the young, middle, and older adults' models.

Gender and age differences in mean scores on the MBSRQ subscales

Table 9 presents the means and standard deviations for each subscale of the MBSRQ. For those models that met the requirements for scalar invariance, univariate one-way analyses of variance (ANOVAs) were conducted with gender or age group as the independent variable and the subscale mean score as the dependent variable. Two-way ANOVAs with gender and age group could not be conducted because scalar invariance was not achieved for the full model for any subscale, and was only achieved for select subgroups.

Table 7 Goodness-of-fit indices for the MBSRQ—body areas satisfaction.

Model	χ^2	<i>df</i>	RMSEA	NNFI	CFI	Model CAIC	Independence CAIC	Saturated CAIC	$\Delta\chi^2$	Δdf	ΔCFI	Invariance?
Configural—full model	745.59*	162	0.14	0.87	0.90	1,684.81 ^a	6,601.92	2,196.42	–	–	–	No
Women's model	448.58*	81	0.14	0.88	0.91	925.16 ^a	4,242.87	1,043.36	–	–	–	No
Men's model	297.01*	81	0.14	0.86	0.89	678.46 ^a	2,318.46	950.11	–	–	–	No
Young adults' model	262.28*	54	0.13	0.90	0.92	565.59 ^a	2,888.28	657.40	–	–	–	Yes
Middle-aged adults' model	263.33*	54	0.15	0.85	0.88	533.82 ^a	2,005.91	628.78	–	–	–	No
Older adults' model	219.98*	54	0.14	0.85	0.89	463.82 ^a	1,646.91	606.29	–	–	–	No
Metric—young adults' model	280.24*	62	0.10	0.92	0.93	522.93 ^a	2,792.80	657.40	17.96	8	0	Yes
Scalar—young adults' model	355.88*	70	0.13	0.89	0.89	687.13	2,888.28	657.40	93.60*	8	–0.09	No

* $p < 0.001$ ^a Model CAIC < than both independence and saturated CAICs**Table 8** Goodness-of-fit indices for the MBSRQ—overweight preoccupation.

Model	χ^2	<i>df</i>	RMSEA	NNFI	CFI	Model CAIC	Independence CAIC	Saturated CAIC	$\Delta\chi^2$	Δdf	ΔCFI	Invariance?
Configural—full model	41.77**	12	0.11	0.92	0.97	432.47 ^a	1,312.27	488.09	–	–	–	Yes
Metric—full model	76.20**	31	0.08	0.95	0.96	311.41 ^a	1,312.27	488.09	34.43	19	–0.01	Yes
Scalar—full model	155.72**	42	0.11	0.91	0.89	491.68	1,312.27	488.09	113.95**	30	–0.08	No
Women's model	101.80**	18	0.13	0.89	0.89	285.37	874.25	231.86	73.53**	12	–0.08	No
Men's model	42.89**	18	0.11	0.92	0.92	209.80 ^a	419.97	211.14	29.38*	12	–0.06	No
Young adults' model	18.25	10	0.06	0.98	0.98	149.96	548.88	146.09	10.71	6	–0.01	Yes
Middle-aged adults' model	12.82	10	0.04	0.99	0.99	138.37 ^a	414.66	139.73	12.27	6	0	Yes
Older adults' model	35.64**	10	0.13	0.88	0.90	158.27	321.71	134.73	7.39	6	–0.01	Yes

* $p < 0.01$ ** $p < 0.001$ ^a Model CAIC < than both independence and saturated CAICs

Table 9 Means (standard deviations) for the subscales of the MBSRQ.

	Young M (SD)	Middle-aged M (SD)	Older M (SD)
Appearance evaluation			
Men	3.69 (0.70)	3.48 (0.70)	3.49 (0.50)
Women	3.43 (0.82) ^a	3.33 (0.89) ^a	3.23 (0.77) ^a
Appearance orientation			
Men	3.23 (0.62)	3.20 (0.66) ^a	3.49 (0.50)
Women	3.48 (0.60)	3.46 (0.63) ^a	3.49 (0.63)
Fitness orientation			
Men	3.76 (0.74) ^{a, b}	3.45 (0.66) ^{a, c}	3.43 (0.67) ^{a, d}
Women	3.47 (0.71) ^b	3.48 (0.72) ^c	3.40 (0.70) ^d
Health evaluation			
Men	3.90 (0.73)	3.77 (0.60)	3.84 (0.62)
Women	3.72 (0.72)	3.86 (0.76)	3.81 (0.82)
Health orientation			
Men	3.46 (0.70)	3.56 (0.66)	3.72 (0.60)
Women	–	–	3.95 (0.60)
Illness orientation			
Men	2.99 (0.75) ^a	3.18 (0.78)	3.34 (0.80)
Women	3.26 (0.77) ^a	3.28 (0.76)	3.27 (0.79)
Body areas satisfaction			
Men	3.60 (0.68)	3.50 (0.61)	3.62 (0.49)
Women	3.39 (0.64)	–	3.36 (0.65)
Overweight preoccupation			
Men	1.89 (0.74) ^a	2.34 (0.82) ^b	2.13 (0.72) ^c
Women	2.60 (0.89) ^a	2.69 (0.91) ^b	2.50 (0.89) ^c

Means with the same superscript within each subscale exhibit scalar invariance and can be compared; missing means and standard deviations represent subscales that did not exhibit unidimensionality, and thus the mean subscale score should not be used.

Appearance evaluation As scalar invariance was established for the women's model, a univariate ANOVA was conducted to test for any age-related differences. Results indicated a significant small effect for age group, $F(2, 837) = 4.27, p = 0.014, \eta^2 = 0.014$.³ Follow-up post hoc analyses indicated that young adult women reported significantly greater satisfaction with appearance than older adult women did. There were no significant differences between the middle-aged women and either the younger women or older women.

Appearance orientation As scalar invariance was established for the middle-aged adults, a univariate ANOVA was conducted to test for gender differences. Results indicated a significant small effect for gender, such that women reported greater investment in appearance than men did, $F(1, 396) = 14.77, p < 0.001, \eta^2 = 0.036$.

³ Effect sizes are reported in addition to the statistical test results to indicate whether an effect is non-trivial or not (Zumbo & Hubley, 1998). Kirk's (1996) criteria for interpreting effect size are as follows: small effect = 0.010 to 0.058, medium effect = 0.059 to 0.137, and large effect = > 0.137. Kirk's criteria are for omega-sq.; however, these criteria may be appropriately applied to interpreting partial eta-squared, which is a similar measure of strength of association.

Fitness orientation As scalar invariance was established for four of the five subgroups, a series of univariate ANOVAs were conducted to test for age differences for men and gender differences for young, middle-aged, and older adults. Results indicated a small significant effect for age group for the men, $F(2, 419) = 10.94, p < 0.001, \eta^2 = 0.050$. Follow-up post hoc analyses indicated that young men reported significantly higher investment in fitness than did both middle-aged men and older adult men. There were no significant differences between the middle-aged and older men. Results for the gender differences indicated a significant small effect for gender for young adults, such that young men reported significantly more investment in fitness than young women did, $F(1, 547) = 19.66, p < 0.001, \eta^2 = 0.035$. There were no significant main effects for gender for middle-aged adults, $F(1, 396) = 0.19, n.s., \eta^2 < 0.001$, or older adults, $F(1, 313) = 0.134, n.s., \eta^2 < 0.001$.

Illness orientation As scalar invariance was established only for young adults, a univariate ANOVA was conducted to test for gender differences. Results indicated a significant small effect for gender, such that young women reported a greater awareness of physical symptoms and a greater

Table 10 Levels of invariance attained for the subscales of the MBSRQ.

Subscale	Configural	Metric	Scalar
Appearance evaluation	Full model	Women's model Young adults' model	Women's model
Appearance orientation	Full model	Full model	Middle-aged adults' model
Fitness orientation	Full model	Full model	Men's model Young adults' model Middle-aged adults' model Older adults' model
Health evaluation	Men's model Older adults' model		
Health orientation	Men's model Older adults' model	Men's model	
Illness orientation	Full model	Women's model Young adults' model Older adults' model	Young adults' model
Body areas satisfaction	Young adults' model	Young adults' model	
Overweight preoccupation	Full model	Full model	Young adults' model Middle-aged adults' model Older adults' model

likelihood of seeking medical attention than young men did, $F(1, 549)=15.61$, $p<0.001$, $\eta^2=0.028$.

Overweight preoccupation As scalar invariance was established for the young, middle-aged, and older adult models, a series of univariate ANOVAs were conducted to test for gender differences in each of the three age groups. Results indicated significant effects for gender for each of the age groups, such that women in the young, $F(1, 549)=87.57$, $p<0.001$, $\eta^2=0.138$, middle-aged, $F(1, 396)=14.47$, $p<0.001$, $\eta^2=0.035$, and older, $F(1, 313)=13.39$, $p<0.001$, $\eta^2=0.04$, groups reported significantly higher scores than their male counterparts did. There was a large effect size for the group of young adults and small effect sizes for the middle-aged and older adults.

Scalar invariance was not established for any subgroups for the HE, HO, and BAS subscales, so no group comparisons were made on these subscales.

Discussion

Although the importance of measurement invariance in group comparisons has received increased attention in recent years and more research is being conducted to examine the equivalence of various measures across groups, there appears to be no published research to date on measurement invariance in the body image field. Evidence of measurement invariance is important for scientific inference, and lack of measurement invariance casts suspicion and doubt on both the conclusions drawn

from the data and theory developed from research studies (Horn & McArdle, 1992).

As the purpose of the present study was to assess whether the MBSRQ ultimately could be used to make gender and age group comparisons, the levels of invariance that were examined were configural, metric, and scalar invariance. The results of the measurement invariance tests for the subscales of the MBSRQ clearly illustrate that the multidimensional nature of body image is perceived quite differently across the age and gender groups, as evidenced by no two subscales demonstrating the same level of invariance to the same degree. Table 10 provides a summary of the levels of invariance achieved for each subscale in the present study.

Based on the varied findings for the MBSRQ subscales, it is important for researchers to consider the goals of their research as they decide which of these subscales to use. Because the AE, AO, FO, IO, and OP subscales met the requirements for configural invariance for the full model, the results support the use of these five subscales to assess their respective latent variables across the age and gender groups. Therefore, for example, if a researcher wanted to use the AE subscale as a measure of appearance satisfaction in a sample of women, the results of the present study would support the use of this subscale for that purpose.

For the HE and HO subscales, configural invariance was only supported for men and for older adults, which suggests that these subscales should only be used to assess these constructs for adult men of all ages and for older adults. Configural invariance was not found for the HE subscale for the women's model, although the PCAs for this

subscale indicated the presence of only one factor for each group of women. We suggest that this subscale may be used with women to assess the construct of health evaluation, but researchers should use some caution in interpreting the findings. On the one hand, the decision for invariance is based on a binary decision-making scheme using selected criteria. In the case of HE, had the NNFI result been 0.90 instead of 0.89, we would have concluded that configural invariance for the full model was achieved. Thus, there is the possibility that there is measurement error in the decision for the lack of invariance of the HE subscale. On the other hand, relying on the criteria selected, it is possible that there might be one or more minor secondary factors that are not detected by the less strict PCA that may have a slight impact on the results. The findings for the HO subscale, however, are clearer. Because the PCAs for the HO subscale showed the presence of two factors for the young and middle-aged women, a single subscale score should not be used with these two groups of women to assess investment in a healthy lifestyle.

For the BAS subscale, configural invariance was only supported for young adults, which suggests that this subscale should only be used with this group. However, as the results of PCAs indicated the presence of one factor for all groups except for the middle-aged women, this subscale may be used with men of all ages and with young and older adult women, but, as with the HE subscale, caution should be used in interpreting the findings for middle-aged and older adult men and older adult women. This subscale should not be used with middle-aged women to assess body areas satisfaction.

Only three of the subscales (AO, FO, and OP) met the requirements of metric invariance for the full model. Thus, for these subscales, the results provide support for the use of these subscales to examine structural relationships, or correlations, between these subscales and other latent variables across all age and gender groups. Four of the subscales (AE, HO, IO, BAS) received support for metric invariance for one or more subgroups. For these subscales, we can only recommend them to examine relationships among correlations for those subgroups that show metric invariance. For example, for the AE subscale, correlations can be examined between scores on AE and other measures across the different age groups of women and between young men and women. For the HO subscale, correlations can be examined for men. For the IO subscale, correlations can be examined for women, young adults, and older adults. For the BAS subscale, correlations can be examined for young adults. For the remaining subgroups, and for the HE subscale, we cannot recommend them to examine correlations across groups.

None of the subscales met the hypothesis of scalar invariance for the full model. Thus, our results suggest that

researchers should not use any of these subscales in their present forms to make comparisons across the six age and gender groups together. Five of the subscales (AE, AO, FO, IO, and OP), however, did meet scalar invariance requirements for one or more subgroups. For the AE subscale, comparisons may be conducted across age groups for women. For the AO subscale, gender comparisons may be conducted across middle-aged adults. For the FO subscale, comparisons may be conducted across all three age groups for men and across gender for all three age groups. For the IO subscale, gender comparisons may be conducted across young adults. Finally, for the OP subscale, gender comparisons may be conducted across all three age groups. For those subgroups and subscales that did not meet the requirements for scalar invariance, differences in item interpretation or measurement bias prevent accurate interpretation of observed mean differences among groups.

Examination of gender differences for those subscales that exhibit scalar invariance (i.e., AO, IO, and OP) indicated that, in most cases, women scored higher than men. The exception to this was the FO subscale, on which young men scored higher than young women and no significant differences were found for middle-aged and older adults. Examination of age differences for those subscales that exhibit scalar invariance (i.e., AE, FO) indicated that young adults tended to score higher than their older counterparts.

Conclusion

There are three important implications of this research that should be noted. First, the varied findings for the MBSRQ in this study highlight the importance of examining measurement invariance before *any* measure is used across different groups. Failure to do so may impact the validity of conclusions drawn and potentially distort ensuing theory.

Second, our results must also call into question many of the cross-group mean differences or correlational findings that have been reported in the past based on the MBSRQ. For instance, Paxton and Phythian (1999) used the MBSRQ in a sample of middle-aged and older men and women to examine gender and age effects. Several of their conclusions on gender differences are based on MBSRQ subscales, such as AO and HO, which have not been shown in this study to exhibit scalar invariance for these groups. Thus, the findings that women scored higher than men may be the result of differences in interpretation and measurement bias of the items for this subscale rather than true differences in the latent variable. Claims of gender differences on these subscales without evidence of measurement invariance may have negative social consequences by leading to incorrect gender-based expectations (Huble &

Zumbo, 1996; Messick, 1988) that can further impact policy and program decisions and distort theory.

Third, our findings should serve as a warning to researchers that caution should be used in deciding what subscales of the MBSRQ are the most appropriate for the purposes of their research. That is, if researchers are interested in gender differences, they should only use those subscales that show scalar invariance for both men and women, such as the FO subscale. If, for example, researchers are interested in relationships between body image and other measures of self-concept, then they need to choose subscales, such as the AO subscale, that have been found to be metrically invariant for the groups in which they are interested. Finally, if researchers are looking for a scale to measure a specific construct for a particular group of individuals, such as young men and women, they need to choose a subscale, such as the BAS subscale, that has demonstrated configural invariance for those groups.

An important point about current measurement invariance testing is that it is based on a binary decision-making scheme. Thus, some of the decisions for the different levels of invariance may be based on fit statistics that just met or did not meet the cutoff values for indicating invariance (i.e., with a cut-off criterion of 0.90, a CFI value of 0.89 does not support invariance, but a value of 0.90 does). This was the case for the HE subscale when configural invariance was evaluated for the full model. If the current study were to be replicated, there is a chance that some of the borderline results may change. The conclusions of the present study should be viewed with this in mind.

The results of the present study open several avenues for future research. First, these results highlight the need for further studies of measurement invariance on the MBSRQ subscales and other measures of body image. Our findings are limited to the age groupings used in the present study and need to be replicated. Second, more research is needed to examine why various levels of invariance have not been met for the subscales of the MBSRQ. As Cheung and Rensvold (2002) note, lack of invariance should not be seen merely as an obstacle to be overcome, but also as an indicator of potentially interesting information about how different groups view a construct. It may not be too surprising that scalar invariance, or even metric or configural invariance, was not met for men and women across the adult age range for the HE or HO subscales given that health concerns become much more important than appearance to one's body image as individuals get older (Clarke, 2002). However, the lack of invariance for other subscales (e.g., BAS) is not as easy to interpret. More research is needed to investigate possible sources of invariance as well as ways to overcome this issue if one wishes to make cross-group mean comparisons.

Acknowledgements We thank Dr. Thomas Cash for the use of his measures and his permission to exceed the number of copies normally allowed, Dr. Bruno D. Zumbo for his statistical advice, and our research assistants (Beth Chan, Sarah Chan, Kristine Hagen, Anne Muscat, and Martin Rusticus) for their assistance in participant recruitment and data collection.

We also acknowledge the support given to Shayna Rusticus through a Social Sciences and Humanities Research Council (SSHRC) Canada Graduate Master's Scholarship, the Norske Skog Canada Limited Fellowship, and the University of British Columbia Faculty of Education Graduate Student Research Grant.

References

- Brannick, M. T. (1995). Critical comment on applying covariance structure modeling. *Journal of Organizational Behavior*, *16*, 201–213.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Cash, T. F. (2000). *The multidimensional body-self relations questionnaire users' manual*. Available from the author at www.body-images.com.
- Cash, T. F., Winstead, B. A., & Janda, J. H. (1986). Body image survey report: The great American shape-up. *Psychology Today*, pp. 30–37, April.
- Chan, D. (2000). Detection of differential item functioning on the Kirton Adaptation—Innovation inventory using multiple-group mean and covariance structure analyses. *Multivariate Behavioral Research*, *35*, 169–199.
- Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. *Journal of Cross-Cultural Psychology*, *31*, 187–212.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.
- Clarke, L. H. (2002). Older women's bodies and the self: The construction of identity in later life. *Canadian Review of Sociology and Anthropology*, *38*, 441–464.
- Diamantopoulos, A., & Siguaw, J. A. (2000). *Introducing LISREL: A guide for the uninitiated*. London, England: Sage.
- Horn, J. L. (1991). Discussion of the issues of factorial invariance. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 114–125). Washington, DC: American Psychological Association.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Journal of Experimental Aging Research*, *18*, 117–144.
- Huble, A. M., & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. *Journal of General Psychology*, *123*, 207–215.
- Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426.
- Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.
- Kelloway, K. E. (1995). Structural equation modelling in perspective. *Journal of Organizational Behaviour*, *16*, 215–224.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746–759.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525–543.
- Messick, S. (1988). The once and future issues of validity: Assessing

- the meaning and consequences of measurement. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 33–45). Hillsdale, NJ: Erlbaum.
- Paxton, S. J., & Phythian, K. (1999). Body image, self-esteem, and health status in middle and later adulthood. *Australian Psychologist, 34*, 116–121.
- Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*, 287–297.
- Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78–90.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–69.
- Zumbo, B. D., & Hubley, A. M. (1998). A note on misconceptions concerning prospective and retrospective power. *Journal of the Royal Statistical Society, Series D: The Statistician, 47*, 385–388.