



Evaluating different versions of the Body Shape Questionnaire and invariance across age in Mexican women

Karina Franco-Paredes^{1,5} · Felipe J. Díaz-Reséndiz^{1,2} · Fabiola González-Betanzos³ · Carlos Hidalgo-Rasmussen^{1,4}

Received: 20 March 2019 / Accepted: 11 February 2020 / Published online: 21 February 2020
© Springer Nature Switzerland AG 2020

Abstract

Purpose To analyze the factorial structure of the eight short versions of the Body Shape Questionnaire (BSQ) and analyze the factorial invariance of the best model, its convergent validity and its internal consistency in three age groups of Mexican women.

Methods Women attending middle school ($n = 261$), high school ($n = 245$) and university ($n = 296$) participated in the study; their average age was 16.52 years ($SD = 2.98$), and they completed the BSQ, the Eating Attitudes Test (EAT-40) and the Questionnaire on the Influences of the Aesthetic Body Ideal (CIMEC, its acronym in Spanish).

Results The BSQ-8D model showed the best goodness-of-fit indices and evidence for configural, metric, scalar and structural invariance. Both the Cronbach's alpha coefficient and the Omega coefficient for the BSQ-8D were .89. Likewise, the BSQ-8D was correlated with the EAT-40 ($r = .60$, $p < .001$) and the CIMEC ($r = .77$, $p < .001$), and both the total score and its items discriminated between women with and without abnormal eating behaviors ($p < .001$).

Conclusion The BSQ-8D has the best psychometric properties for the measurement of body dissatisfaction in the general population of Mexican women.

Level of evidence Level V, cross-sectional descriptive study.

Keywords BSQ · Reliability · Factorial invariance · Validity · Women

Introduction

Body dissatisfaction is considered to be one of the major risk factors for eating disorders and is a condition present in a significant number of people of all ages, from childhood

to adulthood; however, body perception changes in different stages of development [1], and there are critical moments for body dissatisfaction, for example, in adolescence [2].

The Body Shape Questionnaire (BSQ-34 [3]) is one of the most widely used self-report measures, both in clinical practice and in research; it was created to assess the phenomenological experience of feeling fat and included items to evaluate dissatisfaction with body size and shape. The original questionnaire consists of 34 items. In subsequent studies, eight different short versions have been proposed: four included eight items (BSQ-8A, BSQ-8B, BSQ-8C and BSQ-8D [4]), one included 10 items (BSQ-10 [5]), one included 14 items (BSQ-14 [6]) and two included 16 items (BSQ-16A and BSQ-16B [4]).

The BSQ has been translated or adapted in 13 countries: Germany, Brazil, Colombia, Spain, the United States, France, Japan, Korea, Malaysia, Mexico, Portugal, Switzerland and Turkey [7]; and its psychometric properties have been analyzed in different studies, finding excellent internal consistency for the BSQ-34 ($\alpha = .93-.98$) in both

✉ Karina Franco-Paredes
karina.franco@academicos.udg.mx

¹ Behavior, Health and Quality of Life Research Group, Research Center on Risks and Quality of Life, Universidad de Guadalajara, Ciudad Guzmán, Mexico

² Laboratorio de Análisis Conductual, Universidad de Guadalajara, Ciudad Guzmán, Mexico

³ Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mexico

⁴ Center for Advanced Studies, Universidad de Playa Ancha, Valparaíso, Chile

⁵ Centro Universitario del Sur, Universidad de Guadalajara, Av. Enrique Arreola Silva 883, Colonia Centro, CP 49000 Ciudad Guzmán, Jalisco, Mexico

clinical and nonclinical samples [4, 5, 7–15]. Moreover, temporal stability reliability at 2 weeks ($r_{xx} = .90$ [12, 16]), 3 weeks ($r_{xx} = .88$ [17]) and 1 month between applications ($r_{xx} = .81–.95$ [15, 18]) was adequate.

The factorial structure of the BSQ-34 is controversial due to the variety of structures identified using both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). One [4, 15], two [8, 9], three [18], four [11, 15] and up to five [14] factors have been identified using EFA; whereas, only three studies conducted CFA analyzing one-factor model, just one study found good fit indices [5] and in two studies no adequate goodness-of-fit indices were obtained [7, 19]. Likewise, one study showed that the BSQ-34 had factorial invariance in cross-cultural and cross-linguistic comparisons between women from two countries [7] and other study found that there is no invariance among gender [20].

Analyses of the short versions of the BSQ have found that the internal consistency is also good ($\alpha = .83–.98$); the version with 14 items and the two versions with 16 items showed greater internal consistency ($\alpha = .93–.97$ [4–7, 12, 13, 21, 22]) than did the versions with eight items ($\alpha = .83–.98$ [4, 5, 7, 13, 16]). Likewise, the factorial structure of the short versions has been examined with CFA and the unidimensionality was confirmed only for some models: the BSQ-14 [12, 21], the BSQ-10 (5), the BSQ-8B [7, 13, 20] and the BSQ-8C [13, 16].

However, only four studies have tested the measurement invariance of short versions: the BSQ-10 showed metric invariance in cross-cultural and cross-linguistic comparisons between women from two countries [5], the BSQ-16A and BSQ-8B showed strong invariance when applied to independent samples of women [7]; the BSQ-8 (the author no specified which version) presented metric invariance between the Brazilian and the Portuguese women samples [19]; and the BSQ-8B was not invariant between women and men [20]. Therefore, it is clear that there are no solid and sufficient findings regarding the factor structure of BSQ short versions and behavior in populations that differ by age, sex or some other characteristic. Additionally, the short versions were also correlated with other instruments, such as the Eating Attitudes Test, the Bulimic Investigatory Test of Edinburgh [6], the Body Weight and Body Shape Concerns test [7] and the Body Dissatisfaction subscale of the Eating Disorders Inventory [12].

These findings showed that BSQ-34 scores are valid and reliable; however, the scores of some short versions show better psychometric properties than the original version. In spite of these findings, no studies have jointly analyzed which of the eight short versions shows better psychometric properties in women of different ages. In addition, studies have been conducted without considering the body perception changes that occur with age and the possibility that the

scale is not equivalent or invariant among women of different age groups.

Therefore, the aim of the present study was to analyze the factor structure of the eight short versions of the BSQ and to analyze the factorial equivalence of the best model, its convergent validity and its internal consistency in Mexican women in three age ranges: 12–15 years, between 15 and 18 years, and between 18 and 25 years.

Methods

Participants

The total sample comprised 802 women with three different educational levels. “*Sample A*” consisted of middle-school students ($n = 261$), “*sample B*” consisted of high-school students ($n = 245$), and “*sample C*” consisted of university students ($n = 296$), with an average age of 13.49 years ($SD = 1.03$ years), 16.09 years ($SD = 1.09$ years) and 19.55 years ($SD = 2.16$ years), respectively. These women attended public institutions of Ciudad Guzmán, a small city located in the southern region of the state of Jalisco, Mexico. For this study, educational levels corresponded to the age ranges indicated in the objective.

Instruments

The Body Shape Questionnaire (BSQ-34 [3]) was adapted to Spanish (Spain) by Raich et al. [14]. This instrument evaluates dissatisfaction with body image due to body weight and shape. The BSQ consists of 34 direct items with six response options on a Likert scale ranging from never = 1 to always = 6. The study for the adaptation and validation of the BSQ-34 in Mexican women showed excellent reliability ($\alpha = .98$ [8]). The eight short versions were analyzed in the present study. Two versions by Evans and Dolan [4] include 16 items, the BSQ-14 version by Dowson and Henderson [6]; the BSQ-10 version by Warren et al. [5] was also included, and finally the study analyzed the four 8-item versions by Evans and Dolan [4].

The Eating Attitudes Test (EAT-40 [23]) evaluates abnormal eating behaviors associated with eating disorders. The EAT-40 is a 40-item questionnaire with six response options on a Likert scale from never to always, and each item is rated on a scale from 0 to 3. A total score that fluctuates between 0 and 120 is calculated, with a higher score indicating a greater presence of abnormal eating behaviors. The EAT-40 was validated in Mexico and showed an internal consistency of $\alpha = .90$ in the clinical sample and $\alpha = .93$ in the total sample. The cut-off point for abnormal eating behaviors is 28 [24].

The Questionnaire on Influences of the Aesthetic Body Model (CIMEC, its acronym in Spanish [25]) evaluates the influence of social agents and situations in which the aesthetic model of thinness is promoted. The CIMEC consists of 40 items with three response options. The analysis of the psychometric properties of the CIMEC in Mexican women showed adequate internal consistency ($\alpha = .94$), and a four-factor structure was identified [26].

Procedure

The research was conducted in educational institutions, and the institutions' directors allowed the research team to collect data during class time. The objective of the research was explained, and the women's voluntary participation in the study was requested. Participants who provided written informed consent were asked to complete the questionnaires. For women in middle school and high school, their parents needed to provide informed consent. Those who agreed to participate needed approximately 15 min to complete questionnaires. At all times, one of the researchers in charge remained in the classroom to answer any questions and to ensure that participants do not share or discuss their response. Participants did not receive any remuneration or other form of inducement for their participation. Rates of refusal were calculated. For the total sample 4.74% ($n = 38$) of women refused to participate, 20 from middle school, 10 from high school and 8 from university. Reasons for denying were not asked.

The present research was carried out with strict adherence to the Code of Ethics for Psychologists [27] and the ethical principles of the American Psychological Association [28]. The protocol was approved by Bioethics Committee of the Centro Universitario del Sur of the Universidad de Guadalajara.

Statistical analysis

The analysis was performed in AMOS 21.0 (Analysis of Moment Structures). The internal structure of the scale was examined with confirmatory factor analysis (CFA), which was used to determine the following: (1) whether the original 34-item scale fits a one- or two-factor model; (2) how well the short versions of the BSQ-34 perform against the full version of the questionnaire; and (3) the factorial equivalence of the best version of the BSQ in the three age groups, based on analysis, to show that equal scores can lead to equal interpretations of results among women of different ages [29].

The CFAs were performed using a random subsample of 50% of the data from groups "A", "B" and "C" and employed the maximum likelihood estimation (ML) method. A critical assumption running ML estimation is that data are

multivariate normal. We reviewed univariate kurtosis and its critical ratio as prerequisite, values range from -4.17 to 1.3 . Kline (2005) considers absolute values equal to or greater than 7 are indicative of departure from normality. At the same time, the Mardia's coefficient evaluates multivariate normality and it was equal to 4.76; thus, it is considered as multivariate normality. The ML method provides consistent, efficient, and unbiased parameter estimates as well as an omnibus test of model fit [30]. The Chi-square statistic (χ^2) is presented to assess the fit of each model. Since this indicator is sensitive to sample size, other model fit indicators are used as well [31], such as the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the root mean square error of approximation (RMSEA) and its corresponding 90% confidence interval (90% CI). Acceptable model fit is defined according to the following criteria: $RMSEA < .08$ (90% CI), $CFI > .90$, and $TLI > .90$; however, a good fit of the model is attained when $RMSEA < .05$, $CFI > .95$, and $TLI > .95$ [31, 32]. We also used the Akaike information criterion (AIC) in the comparison of models, with smaller values representing a better fit of the hypothesized model [33].

Factorial invariance

The invariance or equivalence of a factorial model assesses the psychometric equivalence of a construct across groups. Invariance test implies the evaluation, by multigroup CFA (MG-CFA), of three basic levels of invariance: (1) Configural invariance, (2) Measurement invariance and (3) Structural invariance [34, 35]. Configural invariance considers that the factorial structure is similar across the different groups. The invariance in the measure, refers to the degree to which the measurement parameters of each item are similar across groups [36]; this analysis focuses on the invariance of the factorial loadings (metric or weak invariance), the invariance to which the intercepts are added (scalar or strong invariance) and the invariance to which the residuals are added (residual or strict invariance). Finally, structural invariance focuses on the equality of latent or unobserved variables, thus adding constraints to the matrix of factorial variances and covariances.

Testing for factorial equivalence encompasses a series of hierarchical steps that begins with the separate determination of a baseline model for each group. This model represents the one that best fits the data, considering both parsimony and substantive meaningfulness [37]. The importance of this model lies in the fact that it serves as a baseline model (unconstrained model) with which the other models are compared to evaluate equivalence. If the configural model (unconstrained model) has an adequate fit, it can be concluded that both the number of factors and the factorial loadings pattern are similar across the groups,

and the evaluation of more restrictive invariance models is justified [38].

To evaluate the weak invariance, the baseline model is taken, and the factorial loadings of all items are constrained (made equal) for the three groups. If the overall model fit is significantly worse in the weak invariance model compared to the configural invariance model, it indicates that at least one loading is not equivalent across the groups, and therefore each item loading should be tested. To evaluate strong invariance and strict invariance, global tests are also performed in which the intercepts and the residuals, respectively, were constrained, in addition to the restriction of the factorial weights. If the global tests do not meet the criteria of invariance, an analysis of the corresponding parameter by item must be performed; those procedures are also known as Differential Item Functioning (DIF) analysis. Finally, to evaluate the structural invariance, the common factor variances are equalized.

The likelihood ratio (LR) test and an increase in CFI (Δ CFI) are used as equivalence criteria. For the LR test, the invariant (constrained) models are considered to be nested in the baseline (unconstrained) model, and it is, therefore, possible to estimate a significant loss of fit in the restricted model. This analysis uses the likelihood ratio (LR) test, which determines the difference in the Chi-square ($\Delta\chi^2$) statistic between the baseline model and the constrained model. This difference follows a Chi-square distribution, with degrees of freedom equal to the difference between the degrees of freedom of the models that are compared (Δdf), and if this value is significant, then the models are not equivalent. However, since χ^2 is sensitive to sample size, the criterion proposed by Cheung and Rensvold [39] was used in addition to the LR test. This criterion refers to the change in the CFI between the models, so the authors note that if Δ CFI > .01, then the model in which the parameters are restricted does not hold up; in this study, the fulfillment of both criteria is considered as a proof of lack of invariance.

Reliability and validity evidence

To analyze reliability, Cronbach's alpha coefficient (α) and the omega coefficient (ω) were calculated, regarding that alpha coefficient estimation is affected by number of response alternatives, omega coefficient is a better estimator given that considers factorial weights. Thus, this is a more stable measure of reliability [40]. For the latter, the factorial loads of the model of best fit were used, according to the process described by Viladrich et al. [41]. The construct validity for the best-fitting model was analyzed, and Pearson's correlation was calculated with the EAT-40 and the CIMEC. Furthermore, the scores for the women with and without abnormal eating behaviors were compared.

Results

Normative data

Table 1 shows the mean and standard deviation of the original BSQ and each of the short versions that were evaluated for each group and the total sample, as well as the statistical tests of differences between groups and their level of significance.

CFA and comparison models

Table 2 shows the evaluation of the two-factor solution and the one-factor solution of the original BSQ-34 and for each short version of the BSQ-34 (the two versions with 16 items, the version with 14 items, the version with 10 items and the four 8-item versions) using CFA. Goodness-of-fit indices were analyzed for the nine models evaluated. In the original 34-item scale, none of the models showed a good fit, even when the two-factor model performed better. In the 16-item versions, the BSQ-16B showed better fit indices, and version A did not have a good model fit. All other short versions of

Table 1 Normative data for each version of the BSQ

Version	Sample A ($n=261$)	Sample B ($n=245$)	Sample C ($n=296$)	Total ($n=802$)	$F(2, 801)$	p
BSQ-34	75.60 (33.35)	79.23 (38.07)	72.25 (31.28)	75.47 (34.22)	2.80	.061
BSQ_16A	37.01 (16.38)	38.50 (18.34)	35.41 (15.12)	36.87 (16.59)	2.34	.096
BSQ-16B	36.13 (17.03)	38.13 (19.41)	34.44 (16.01)	36.11 (17.48)	3.00	.050
BSQ_14	33.75 (16.89)	35.85 (19.03)	32.53 (15.94)	33.94 (17.27)	2.50	.082
BSQ_10	19.20 (9.09)	20.14 (9.94)	18.58 (8.22)	19.26 (9.06)	1.99	.137
BSQ_8A	18.99 (8.65)	19.43 (9.34)	17.47 (7.70)	18.56 (8.57)	3.99	.019
BSQ_8B	18.02 (8.47)	19.07 (9.39)	17.94 (7.93)	18.31 (8.58)	1.40	.246
BSQ_8C	18.20 (8.71)	19.10 (9.86)	17.22 (8.16)	18.11 (8.91)	3.02	.049
BSQ_8D	17.93 (8.79)	19.03 (9.81)	17.22 (8.15)	18.00 (8.91)	2.79	.062

Numbers outside the parentheses are the mean, and those in parentheses are the standard deviation

Table 2 Values of goodness-of-fit indices for confirmatory factor analysis models

	χ^2	<i>df</i>	RMSEA (90% CI)	NFI	CFI	GFI	AGFI	AIC
1. BSQ-34 (two factors)	1718.02	526	.08 (.07, .08)	.82	.87	.78	.75	1856.02
2. BSQ-34 (one factor)	1844.36	527	.08 (.08, .08)	.81	.75	.76	.72	1980.37
3. BSQ-16A	462.11	104	.09 (.08, .10)	.87	.89	.86	.82	526.11
4. BSQ-16B	394.98	104	.08 (.07, .09)	.92	.94	.89	.86	410.98
5. BSQ-14	400.36	77	.10 (.09, .11)	.91	.93	.86	.81	456.35
6. BSQ-10	97.95	35	.07 (.05, .08)	.94	.93	.95	.92	137.95
7. BSQ-8A	131.48	20	.12 (.10, .14)	.91	.92	.92	.85	163.48
8. BSQ-8B	73.38	20	.08 (.06, .10)	.94	.95	.96	.92	105.38
9. BSQ-8C	78.40	20	.08 (.06, .11)	.95	.96	.95	.92	110.39
10. BSQ-8D	39.01	20	.04 (.02, .07)	.98	.99	.98	.96	71.01

df, degrees of freedom, *RMSEA* root of the mean square error of approximation, *90% CI* 90% confidence interval, *NFI* normal fit index, *CFI* comparative fit index, *GFI* goodness-of-fit index, *AGFI* adjusted goodness-of-fit index, *AIC* Akaike information criterion

the BSQ showed good model fit indices, with better indices for the 8-item versions, except for the BSQ-8A. For the model comparison, the BSQ-8D showed better overall fit indicators and the lowest AIC; thus, it was retained as a final model.

For the BSQ-8D version, the factorial weights for each item in each of the samples and in the total sample are shown in Table 3. All the standardized factorial weights were $> .50$, ranged from .60 to .84 for sample A, .61 to .88 for sample B, .54 to .86 for sample C, and .61 to .84 for total sample.

Analysis of the factorial invariance of the BSQ-8D by age

The MG-CFA was performed to test the validity of the BSQ-8D in the different age groups. Table 4 shows the Chi-square and degrees of freedom values for each model, as

well as, the LR test and the Δ CFI to assess the invariance. The unconstrained model fit examines configural invariance of the BSQ-8D; current conventions for evaluating model fit consider the Chi-square test, and two more alternative fit indices. Although the Chi-square test does not show a good fit ($\chi^2 = 175.69$, $df = 60$, $p < .05$), the rest of the indices contradict this conclusion: the Tucker–Lewis Index (TLI = .96) and the mean square root (RMSEA = .04) allow us to accept the base model of the configural invariance. For testing (1) Measurement invariance and (2) Structural invariance, the models that do not cover the invariance criteria (a significant LR test and a Δ CFI $> .01$), are Italic. The existence of metric invariance and scalar invariance (weak and strong) has been proven in the global tests, a. Measurement loadings ($p = .045$, Δ CFI = .003) and b. Measurement intercepts ($p = .001$, Δ CFI = .009), but not strict c. Measurement residuals ($p = .001$ and Δ CFI = .02). In that case, the tests for the

Table 3 Standardized factorial loadings (λ) of the BSQ-8D in each of the samples and in the overall sample

Item	Factorial weights (λ)			
	Sample A	Sample B	Sample C	Total
1. Have you been so worried about your shape that you have been feeling that you ought to diet?	.67	.84	.78	.77
2. Have you noticed the shape of other women and felt that your own shape compared unfavorably?	.78	.88	.74	.81
3. Has being naked, such as when taking a bath, made you feel fat?	.84	.84	.86	.84
4. Have you not gone out to social occasions (e.g., parties) because you have felt bad about your shape?	.64	.67	.67	.65
5. Have you worried about other people seeing rolls of flesh around your waist or stomach?	.72	.85	.80	.79
6. When in company have you worried about taking up too much room (e.g., sitting on a sofa or a bus seat)?	.64	.61	.62	.62
7. Have you pinched areas of your body to see how much fat there is?	.62	.68	.54	.61
8. Have you avoided situations where people could see your body (e.g., communal changing rooms or swimming baths)?	.60	.78	.66	.68

Table 4 Invariance models test for three age groups (sample A: 12–15, sample B: 15–18 and sample C: 18–25 years old)

Model	χ^2	df	LR Test			Δ CFI
			$\Delta\chi^2$	Δ df	p	
1. Configural invariance						
Unconstrained model	175.69	60				
2. Measurement invariance						
a. Measurement loadings (weak)	199.72	74	24.03	14	.045	.003
b. Measurement intercepts (strong)	235.70	90	60.01	30	.001	.009
c. Measurement residuals (strict)	<i>314.59</i>	<i>104</i>	<i>138.9</i>	<i>44</i>	<i>.001</i>	<i>.020</i>
Residual_Item2	234.39	90	58.70	30	.001	.009
Residual_Item12	<i>243.67</i>	<i>92</i>	<i>67.98</i>	<i>32</i>	<i><.001</i>	<i>.011</i>
Residual_Item14	<i>241.97</i>	<i>92</i>	<i>66.28</i>	<i>32</i>	<i><.001</i>	<i>.010</i>
Residual_Item18	<i>249.04</i>	<i>92</i>	<i>73.35</i>	<i>32</i>	<i><.001</i>	<i>.013</i>
Residual_Item24	<i>245.89</i>	<i>92</i>	<i>70.20</i>	<i>32</i>	<i><.001</i>	<i>.012</i>
Residual_Item27	<i>248.87</i>	<i>92</i>	<i>73.18</i>	<i>32</i>	<i><.001</i>	<i>.012</i>
Residual_Item30	235.93	92	60.24	32	.002	.008
Residual_Item31	<i>253.71</i>	<i>94</i>	<i>78.02</i>	<i>34</i>	<i><.001</i>	<i>.013</i>
3. Structural invariance						
	<i>244.44</i>	<i>94</i>	<i>68.75</i>	<i>34</i>	<i>.016</i>	<i>.010</i>

Italicized values indicate those models that do not meet the criteria of invariance [invariance criteria: a non-significant LR test ($p > .05$) and Δ CFI $\leq .01$]

Chi square = χ^2 , df = degrees of freedom

residual of each item were carried out; results show that only items 2 and 30 are equivalent in the three age groups. (3) Structural invariance holds up because Δ CFI = .01, which suggests that the variances of the latent variable are the same for the three groups.

Reliability

The BSQ-8D showed adequate reliability indicators, as evaluated by Cronbach's alpha ($\alpha = .89$) and by the Omega coefficient ($\omega = .89$).

Convergent validity

The correlations between the total scores of the BSQ-8D and the EAT-40 ($r = .60$, $p < .001$) and between those of the BSQ-8D and the CIMEC ($r = .77$, $p < .001$) were calculated.

Discriminant validity

Considering the cut-off point of the EAT-40, two groups were identified: women with abnormal eating behaviors ($n = 94$) and women without abnormal eating behaviors ($n = 100$). The total score of the BSQ ($t = 15.99$, $p < .001$, $r = .75$) and its eight items discriminated between both groups, with significantly higher scores in women with abnormal eating behaviors.

Discussion

The objective of the present study was to analyze the factorial structure of the eight short versions of the BSQ and to analyze the factorial equivalence of the best model, its convergent validity and its internal consistency in Mexican women in three age ranges: 12–15 years, between 15 and 18 years, and between 18 and 25 years. First, evidence was generated that did not confirm the two-factor structure for the BSQ-34 previously generated with EFA in Mexican women from the general population and from a clinical sample [8]. In the present study, the two-factor model demonstrated better performance than the one-factor model; however, the goodness-of-fit indices were inadequate.

The models in the different short versions were analyzed. The model with the best fit was the BSQ-8D, and the unidimensionality of this scale was confirmed. This finding is in line with previous studies where unidimensionality of different short versions of the BSQ was analyzed [5, 7, 12, 16, 19–21].

Once the best model of the different short versions of the BSQ was identified, the MG-CFA was conducted. The main objective of generating evidence for the invariance of an instrument is the establishment of a multigroup reference model that acceptably fits the data. In other words, this reference model allows the researcher to determine whether the same elements are indicators of the same latent factor in each group [42]. The results obtained showed configural invariance, metric (weak) and scalar (strong) invariance, but no residual (strict) was observed. Although residual

invariance is a component for full measurement invariance, it is not a prerequisite for testing mean differences because the residuals are not part of the factor. For this reason, this test is often omitted in invariance research [38]. Finally, structural invariance has been proven which suggests that the variances of the latent variable are the same for the three groups.

In the validation study of the original BSQ and in subsequent validations, it has been shown that both the full version and the short version show evidence of convergent validity because their scores are correlated positively with those of other tests validated in the field of eating disorders [6, 7, 12]. In the present study, the score on the BSQ-8D was positively correlated with the score on the EAT-40 ($r = .60$), one of the most widely used instruments to assess specific characteristics of eating disorders. The BSQ-8D score was also correlated with the CIMEC score ($r = .77$). That is, the relationship between body dissatisfaction and the presence of abnormal eating behaviors was confirmed [43], as well as between body dissatisfaction and the influence of the aesthetic model of thinness.

As expected, women with abnormal eating behaviors showed significantly higher mean scores both in the total BSQ and in the eight items compared than those who did not show abnormal eating behaviors. These findings demonstrate that the BSQ-8D clearly discriminates between unhealthy and healthy samples, which is consistent with the findings of other studies for both the original scale [4, 5, 7, 13, 16] and for the short versions using clinical samples [3, 8, 14, 22], women considered to have probable cases of bulimia nervosa [3], women with a history of vomiting [6], women with different levels of body dissatisfaction [10], and obese women who diet [17].

The internal consistency of the BSQ-8D was considered to be excellent ($\alpha = .89$) and was within the range obtained in previous studies for both the short versions ($\alpha = .83$ – $.98$ [4, 5, 7, 13, 16, 19, 20] and for the original version ($\alpha = .93$ – $.98$ [4, 5, 7–15, 19]. Furthermore, this value was confirmed by calculating the Omega coefficient. This abbreviated version of the BSQ represents less than 25% of application time without detectable losses in the metric quality of the instrument. Model comparison is a technique to identify which version of the instrument measures best the concept and, therefore, contains the most appropriate items for it.

One limitation of this study is that it did not include a clinical sample diagnosed with an eating disorder. Furthermore, on the original scale (BSQ-34), some items are not appropriate for men; an advantage of the BSQ-8D is that its items do not show gender bias. Thus, it is suggested that future research incorporates samples of males; thus far, only five studies have analyzed the validity of the BSQ scores in samples of males [10, 12, 16, 17, 22], and three had samples of fewer than 60 participants [12, 16, 17], which is

considered an insufficient number of studies for performing certain statistical analyses, such as factor analysis.

Given that the BSQ is frequently used in studies to evaluate body dissatisfaction in Mexican women, the main contribution of this study is to provide evidence that the scores of BSQ-8D show adequate psychometric properties. Specifically, the BSQ-8D was found to have a one-factor structure, and it was also found to be a multigroup reference model that acceptably fits the data. Likewise, the scores on this instrument correlate with those of other tests, discriminate between women with and without abnormal eating behaviors and present excellent internal consistency. Therefore, given the low number of items and their psychometric properties, the BSQ-8D is emerging as a useful tool for evaluating body dissatisfaction in different age groups. However, it is important to note that since reliability and validity are not characteristics of the instrument but rather of the instrument scores obtained in a particular sample, it is necessary to evaluate the psychometric properties of the BSQ-8D when it is used with other samples or in other contexts.

What is already known on this subject?

The Body Shape Questionnaire (BSQ) was designed to measure body shape concerns and it has been demonstrated to be a psychometrically sound measure. However, there is evidence that its items could be redundant and several short versions of the BSQ have been proposed and it is necessary to evaluate which of the eight short versions shows better psychometric properties in women of different ages.

What does this study add?

The BSQ-8D short version showed good psychometric properties in the current study and structural invariance has been proven on women in three age ranges. This measure could serve a useful tool to assess body concerns among Mexican women of different ages.

Acknowledgements The authors thank their research assistants, Mayra G. Solórzano-Barragán and Jessica Reynoso-Sánchez, for their support with the data collection.

Compliance with ethical standards

Conflict of interest The authors declare that they have not found conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants or their parents.

References

- Bucchianeri M, Arikian A, Hannan P, Eisenberg M (2013) Body dissatisfaction from adolescence to young adulthood: findings from a 10-year longitudinal study. *Body Image* 10(1):1–15. <https://doi.org/10.1016/j.bodyim.2012.09.001>
- Jiménez Flores P, Jiménez Cruz A, Bacardi Gascón M (2017) Insatisfacción con la imagen corporal en niños y adolescentes: revisión sistemática. [Dissatisfaction with body image in children and adolescents: a systematic review]. *Nutr Hosp* 34(2):479. <https://doi.org/10.20960/nh.455>
- Cooper PJ, Taylor MJ, Cooper Z, Fairburn CG (1987) The development and validation of the Body Shape Questionnaire. *Int J Eat Disord* 6(4):485–494. [https://doi.org/10.1002/1098-108X\(198707\)6:4%3c485::AID-EAT2260060405%3e3.0.CO;2-O](https://doi.org/10.1002/1098-108X(198707)6:4%3c485::AID-EAT2260060405%3e3.0.CO;2-O)
- Evans C, Dolan B (1993) Body Shape Questionnaire: derivation of shortened “alternate forms”. *Int J Eat Disord* 13(3):315–321. [https://doi.org/10.1002/1098-108x\(199304\)13:3%3c315::aid-eat2260130310%3e3.0.co;2-3](https://doi.org/10.1002/1098-108x(199304)13:3%3c315::aid-eat2260130310%3e3.0.co;2-3)
- Warren CS, Cepeda-Benito A, Gleaves DH, Moreno S, Rodriguez S, Fernandez MC et al (2008) English and Spanish versions of the Body Shape Questionnaire: measurement equivalence across ethnicity and clinical status. *Int J Eat Disord* 41(3):265–271. [https://doi.org/10.1002/1098-108x\(200803\)41:3%3c265::aid-eat2260410310%3e3.0.co;2-3](https://doi.org/10.1002/1098-108x(200803)41:3%3c265::aid-eat2260410310%3e3.0.co;2-3)
- Dowson J, Henderson L (2001) The validity of a short version of the Body Shape Questionnaire. *J Psychiatry Res* 102(3):263–271. [https://doi.org/10.1016/s0165-1781\(01\)00254-2](https://doi.org/10.1016/s0165-1781(01)00254-2)
- da Silva WR, Dias JCR, Maroco J, Campos JADB (2014) Confirmatory factor analysis of different versions of the Body Shape Questionnaire applied to Brazilian university students. *Body Image* 11(4):384–390. <https://doi.org/10.1016/j.bodyim.2014.06.001>
- Vázquez R, Galán J, López X, Alvarez G, Mancilla-Díaz JM, Caballero A et al (2011) Validez del Body Shape Questionnaire (BSQ) en mujeres mexicanas [Validity of the Body Shape Questionnaire (BSQ) in Mexican women]. *Rev Mex Trastor Aliment Mex J Eat Disord* 2:42–52
- Castrillon D, Montaña IL, Avendaño G (2007) Validación del Body Shape Questionnaire (Cuestionario de la Figura Corporal) BSQ para la población Colombiana [Validation of the Body Shape Questionnaire (BSQ) for the Colombian Population]. *Acta Colomb Psicol* 10(1):15–23
- Conti MA, Cordás TA, Dias de Oliveira MDR (2009) A study of the validity and reliability of the Brazilian version of the Body Shape Questionnaire (BSQ) among adolescents. *Rev Bras Saúde Matern Infant* 9(3):331–338. <https://doi.org/10.1590/S1519-38292009000300012>
- Di Pietro M, da Silveira DX (2009) Validade interna, dimensionalidade e desempenho da escala Body Shape Questionnaire em uma população de estudantes universitários [Internal validity, dimensionality and performance of the Body Shape Questionnaire in a group of Brazilian college students]. *Rev Bras Psiquiatr* 31(55):21–24. <https://doi.org/10.1590/S1516-44462008005000017>
- Ghaderi ATA, Scott B (2004) The reliability and validity of the Swedish version of the Body Shape Questionnaire. *Scand J Psychol* 45(4):319–324. <https://doi.org/10.1111/j.1467-9450.2004.00411.x>
- Pook M, Tuschen-Caffier B, Brähler E (2008) Evaluation and comparison of different versions of the Body Shape Questionnaire. *Psychiatry Res* 158(1):67–73. <https://doi.org/10.1016/j.psychres.2006.08.002>
- Raich RM, Mora M, Soler A, Avila C, Clos I, Zapater L (1996) Adaptación de un instrumento de evaluación de la insatisfacción corporal [Adaptation of a body dissatisfaction assessment instrument]. *Clínica y Salud* 7(1):51–66
- Rousseau A, Knotter A, Barbe P, Raich RM, Chabrol H (2005) Étude de validation de la version française du Body Shape Questionnaire [Validation study of the French version of the Body Shape Questionnaire]. *L'Encephale* 31:162–173. [https://doi.org/10.1016/s0013-7006\(05\)82383-8](https://doi.org/10.1016/s0013-7006(05)82383-8)
- Welch E, Lagerström M, Ghaderi A (2012) Body Shape Questionnaire: psychometric properties of the short version (BSQ-8C) and norms from the general Swedish population. *Body Image* 9(4):8–11. <https://doi.org/10.1016/j.bodyim.2012.04.009>
- Rosen JC, Jones A, Ramirez E, Waxman S (1996) Body Shape Questionnaire: studies of validity and reliability. *Int J Eat Disord* 20(3):315–319. [https://doi.org/10.1002/\(SICI\)1098-108X\(199611\)20:3%3c315::AID-EAT11%3e3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1098-108X(199611)20:3%3c315::AID-EAT11%3e3.0.CO;2-Z)
- Akdemir A, Inandi T, Akbas D, Kahilogullari AK (2012) Validity and reliability of a Turkish version of the Body Shape Questionnaire among female high school students: preliminary examination. *Eur Eat Disord Rev* 20(11):114–115. <https://doi.org/10.1002/erv.1106>
- Silva WR, Costa D, Pimenta F, Maroco J, Campos JA (2016) Psychometric evaluation of a unified Portuguese-language version of the Body Shape Questionnaire in female university students. *Cad Saude Publica* 32(7):1–12. <https://doi.org/10.1590/0102-311X0133715>
- da Silva WR, Swami V, Nogueira A, Marôco J, Ochner CN, Alvarez J (2019) The Body Shape Questionnaire is not invariant across sex: evidence from Portuguese-speaking university students. *Percept Mot Skills* 126(3):462–476
- Franco-Paredes K, Bautista-Díaz ML, Díaz-Reséndiz FJ, Arredondo M (2018) Propiedades psicométricas del Body Shape Questionnaire (BSQ-14) en mujeres universitarias mexicanas [Psychometric properties of the Body Shape Questionnaire (BSQ-14) in female Mexican university students]. *Rev Evaluar* 18(1):1–11
- Kapstad H, Nelson M, Overås M, Ro O (2015) Validation of the Norwegian short version of the Body Shape Questionnaire (BSQ-14). *Nord J Psychiatry* 69(7):509–514. <https://doi.org/10.3109/08039488.2015.1009486>
- Garner D, Garfinkel PE (1979) The eating attitudes test: an index of the symptoms of anorexia nervosa. *Psychol Med* 9:273–279. <https://doi.org/10.1017/S0033291700030762>
- Alvarez-Rayón G, Mancilla-Díaz JM, Vázquez-Arévalo R, Unikel-Santoncini C, Caballero-Romo A, Mercado-Corona D (2004) Validity of the eating attitudes test: a study of Mexican eating disorders patients. *Eat Weight Disord* 9:243–248. <https://doi.org/10.1007/bf03325077>
- Toro J, Salamero M, Martínez E (1994) Assessment of sociocultural influences on the aesthetic body shape model in anorexia nervosa. *Acta Psychiatr Scand*. <https://doi.org/10.1111/j.1600-0447.1994.tb08084.x>
- Vázquez R, Alvarez G, Mancilla JM (2000) Consistencia interna y estructura factorial del Cuestionario de Influencia de los Modelos Estéticos Corporales (CIMEC), en población mexicana [Internal consistency and factorial structure of the Questionnaire on the influences of the Aesthetic Body Ideal (CIMEC) in the Mexican population]. *Salud Ment* 23(6):18–24
- Sociedad Mexicana de Psicología (2010) Código ético del psicólogo [Ethical code for psychologists], 4a edn. Trillas, México
- American Psychological Association (2002) 2010 Amendments to the 2002 “Ethical principles of psychologists and code of conduct. American Psychological Association, Washington
- Barrada JR, van Strien T, Cebolla A (2016) Internal Structure and measurement invariance of the Dutch eating behavior questionnaire (DEBQ) in a (nearly) representative Dutch community sample. *Eur Eat Disord Rev* 24(6):503–509. <https://doi.org/10.1002/erv.2448>

30. Bollen KA (1989) New incremental fit index for general structural equation models. *Sociol Methods Res* 17(3):303–316. <https://doi.org/10.1177/0049124189017003004>
31. Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model Multidiscip J* 6(1):1–55. <https://doi.org/10.1080/10705519909540118>
32. Bentler PM (1990) Comparative fit indexes in structural models. *Psychol Bull* 107(2):238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
33. Hu L-T, Bentler PM (1995) Evaluating model fit. In: Hoyle R (ed) *Structural equation modeling: concepts, issues, and applications*. Sage Publications Inc, Thousand Oaks, pp 76–99
34. Byrne BM, van de Vijver FJR (2010) Testing for measurement and structural equivalence in large-scale cross-cultural studies: addressing the issue of nonequivalence. *Int J Test* 10(2):107–132. <https://doi.org/10.1080/15305051003637306>
35. Meredith W (1993) Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58(4):525–543. <https://doi.org/10.1007/BF02294825>
36. Vandenberg RJ (2002) Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organ Res Methods* 5(2):139–158. <https://doi.org/10.1177/1094428102005002001>
37. Byrne BM (2008) Testing for multigroup equivalence of a measuring instrument: a walk through the process. *Psicothema* 20(4):872–882
38. Putnick DL, Bornstein MH (2016) Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev Rev* 4:71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
39. Cheung GW, Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Model Multidiscip J* 9(2):233–255. https://doi.org/10.1207/S15328007SEM0902_5
40. Gerbing DW, Anderson JC (1988) An update paradigm for scale development incorporating unidimensionality and its assessment. *J Mark Res* 25(2):186–192
41. Viladrich C, Angulo-Brunet A, Doval E (2017) Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *An Psicol* 33(3):755–782
42. Chen FF, Sousa KH, West SG (2005) Teacher’s corner: testing measurement invariance of second-order factor models. *Struct Equ Model Multidiscip J* 12:471–492. https://doi.org/10.1207/s15328007sem1203_7
43. Smith JM, Smith JE, McLaughlin EA, Belon KE, Serier KN, Simmons JD et al (2018) Body dissatisfaction and disordered eating in Native American, Hispanic, and White college women. *Eat Weight Disord*. <https://doi.org/10.1007/s40519-018-0597-8>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.