Brazilian Journal of Psychiatry
bjp
Revista Brasileira de Psiquiatria
Brazilian Psychiatric Association
Fully open access
No APCs

BJP PRE-PROOF

BJP PRE-PROOF
(article published as accepted)

Original Article

From Measurement to Modeling: A Systematic Factorial Framework for Comparing Structural Models of Motivation for Seeking Treatment

Paulo Sergio Panse Silveira, Beatriz Helena Tess, Sandra Regina da Silva, Yuan-Pang Wang, Jose Oliveira Siqueira

http://doi.org/10.47626/1516-4446-2025-4672

Submitted: 27-Oct-2025

Accepted: 17-Oct-2026

This is a preliminary, unedited version of a manuscript that has been accepted for publication in the Brazilian Journal of Psychiatry. As a service to our readers, we are providing this early version of the manuscript. The manuscript will still undergo copyediting, typesetting, and review of the resulting proof before it is published in final form. The final version may present slight differences in relation to the present version.

From measurement to modeling: a systematic factorial framework for comparing structural models of motivation for seeking treatment

Paulo Sergio Panse Silveira[1,*], Beatriz Helena Tess[2], Sandra Regina da Silva[3], Yuan-Pang Wang[4], Jose Oliveira Siqueira[1]

1 Departamento de Patologia, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brazil.
2 Departamento de Medicina Preventiva, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brazil.
3 Programa de Pós-graduação em Saúde Pública, Departamento de Medicina Preventiva,
Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brazil.
4 Instituto & Departamento de Psiquiatria (LIM-23), Hospital das Clinicas (HCFMUSP),
Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brazil.

* Corresponding author:
Paulo Sergio Panse Silveira
Departamento de Patologia
Faculdade de Medicina da USP
Av. Dr. Arnaldo, 455 - sala 1103
01246-903, São Paulo, SP, Brasil
e-mail: silveira@usp.br

BJP PRE-PROOF

Abstract

Objectives: Motivation for seeking treatment is a clinically relevant yet underexplored construct in mental health and behavioral research. This study proposes a systematic methodology to identify, compare, and rank structural models representing the construct of motivation, emphasizing structural plausibility and model comparison.

Methods: Using the European Obesity Academy Questionnaire: Expectations of Surgical Treatment (EOAQ) as a working example, we analyzed data from 161 treatment-seeking patients with severe obesity, exploring their motivations for undergoing bariatric surgery. Seventy-nine candidate structural models were developed, incorporating one general factor (G) and two or three specific factors. Multigroup confirmatory factor analysis was used to evaluate model fit and test configural invariance across menopausal groups.

Results: Bifactor models with one general factor and three specific dimensions were the most plausible under the proposed comparison (RMSEA 90%CI=0.038-0.102, p-close = 0.1214). The dimensions of personal, social, and healthcare represented intrinsic, extrinsic, and healthcare-related aspects of treatment motivation. The selected bifactor model reached the configural level of measurement invariance across groups and McDonald's omega (ω) coefficients ranged from 0.13 to 0.62.

Conclusions: This study presents a comprehensive framework for comparing alternative CFA models of motivation for seeking treatment, including unidimensional, first-order, bifactor, and second-order structures. The approach can be flexibly adapted to other instruments and populations, but the empirical structure identified here is preliminary and requiring replication in larger samples.

Keywords: validation study; psychometrics; motivation; treatment seeking; factor analysis; methodological studies

BJP PRE-PROOF

Introduction

What compels patients with chronic psychiatric or medical conditions (e.g., individuals with severe obesity) to pursue treatment after years of stigma, emotional distress, and failed coping strategies? For many, the decision to seek help is more than about favorable outcomes of underlying conditions; it is also about reclaiming autonomy, hope, and a sense of self. This internal drive, known as motivation to seek treatment, is deeply intertwined with treatment expectation: the beliefs patients hold about what treatment will involve and what it might achieve.

Treatment expectation refers to beliefs about what treatment may involve or achieve (Laferton et al. 2025). Motivation for seeking treatment refers to the reasons that lead a person to seek care. These constructs are related and may influence each other (Rief et al. 2022), but they are not equivalent. In the present study, motivation for seeking bariatric surgery is assessed by Part III of the European Obesity Academy Questionnaire: Expectations about Surgical Treatment (EOAQ). In patients with severe obesity, such motivation may include personal, social, and health-related reasons, which can be examined through psychometric modeling.

The literature on treatment expectations is fragmented, often embedded within broader constructs such as health beliefs, motivation, or placebo responsiveness (Laferton et al. 2017). Motivation is shaped by factors such as perceived need, emotional distress, and readiness for change (Hammer et al. 2024). However, the lack of conceptual clarity surrounding motivation has hindered the development of measurement tools and predictive modeling of treatment efficacy. Existing psychometric assessments typically rely on *ad hoc* or single-item measures, which lack the psychometric rigor needed for deeper analytical insights.

Severe obesity is a major public health concern, associated with a pressing range of physical and mental health comorbidities among bariatric surgery candidates (Bray et al. 2017; World Health Organization 2021; World Obesity Federation 2022). Mental health conditions such as depression and binge eating disorder are common among bariatric surgery candidates (Dawes et al. 2016). Individuals with obesity often face comorbidities like hypertension, cardiovascular disease, type II diabetes, and sleep apnea. Post-surgical complications, weight regain, and psychological distress, including self-harm and substance use, are frequent during follow-up (Morgan et al. 2020). Stress can further exacerbate disordered eating, increasing food intake and loss of control (Hill et al. 2022). In the context of obesity, where behavioral, psychological, and physiological factors converge, understanding how patients anticipate treatment outcomes is critical. Yet, despite its recognized influence on therapeutic engagement and clinical

BJP PRE-PROOF

success, motivation remains conceptually vague and methodologically underdeveloped.

While bariatric surgery offers a powerful intervention for individuals with high BMI (Pipek et al. 2024), its long-term success hinges not only on physical outcomes but also on patients' psychological readiness, postoperative satisfaction, and body image concerns, necessitating a holistic approach to care (Busetto et al. 2017). Comparisons between patients pursuing bariatric surgery and those opting for clinical treatment reveal that the former group tends to have higher dissatisfaction with body image and greater weight-loss expectations, while the latter group expresses more hope for treatment success (Gaudrat et al. 2021). Unrealistic expectations can lead to disappointment, weight regain, and emotional distress. Conversely, hope and readiness for change can fuel adherence and resilience. Recognizing this, recent research has begun to explore how expectations shape treatment trajectories.

One promising tool is the EOAQ, designed to capture patient expectations surrounding bariatric surgery (Fischer et al. 2011, 2014). However, its psychometric validation remains limited. This study addresses that need by applying confirmatory factor analysis (CFA) to a newly translated Brazilian-Portuguese version of the EOAQ in a sample of preoperative patients with severe obesity. By identifying plausible latent structures and proposing a systematic model comparison strategy based on structural equation modeling (SEM), this research aims to demonstrate a procedure for comparing alternative CFA models of motivation for seeking treatment. In obesity research, expectations and motivations are shaped by prior treatment failures, societal stigma, and complex emotional factors.

CFA, implemented through SEM, allows researchers to test hypothesized latent structures and evaluate model fit. The number of possible models is constrained by the number of manifest variables and the type of model selected. Four basic models are commonly used (Jensen 1998): Spearman's unidimensional model, Thurstone's first-order model (with orthogonal or correlated factors),

BJP PRE-PROOF

Holzinger's bifactor model, and the second-order model. Each offers a different approach to explaining how latent variables relate to observed data.

The present study aims to compare candidate latent structures of the EOAQ under explicit model-selection criteria, using CFA as a framework for evaluating alternative representations of motivation for seeking treatment. We adapted a two-step strategy (Huang 2017): first, non-estimable models and models with poor RMSEA values were excluded; then, among the remaining options, the most parsimonious model were selected, that is, the models with the lowest information criterion value. This framework enhances psychometric rigor by replacing ad hoc model selection with a transparent strategy for comparing theoretically constrained structural representations of motivation in obesity care. Ultimately, the study contributes to integrating motivation for seeking treatment as a core construct in behavioral and medical intervention models.

Material and Methods

Study Design

This methodological study describes a strategy to explore different model variants for representing the construct of motivation for seeking treatment using confirmatory factor analysis (CFA). We illustrate this original approach with the recently translated and culturally adapted Brazilian-Portuguese version of the EOAQ, in which data were collected cross-sectionally (Silva et al. 2024). We propose criteria that can be applied to rank and select structural models of motivation for seeking treatment.

It is important to emphasize the sense in which the term "confirmatory" is used here: each model was specified before estimation through explicit rules defining which items were allowed to load on which factors (CFA), as opposed to factors being obtained from the data by exploratory factor analysis (EFA). This distinction matters for interpreting the large number of model structures evaluated in this study. The analysis was an exploration, in the ordinary sense, of possible confirmatory factor models, in the technical CFA sense, of the EOAQ within a theoretically delimited model space.

BJP PRE-PROOF

Participants

The sample included patients who were candidates for metabolic and bariatric surgery at a university-based bariatric center. The inclusion criteria were adults aged 18 years or older at the time of the interview, with sufficient schooling and cognitive ability to understand the informed consent and questionnaire, and with a body mass index greater than 40 kg/m^2, or greater than 35 kg/m^2 with associated clinical conditions. The sample size of 160 participants was determined based on an item-to-subject ratio of 1:10. A research assistant administered the questionnaire in the waiting room of the outpatient clinic to patients recruited from a consecutive list provided by the administrative office.

After approaching 163 eligible candidates in the preoperative outpatient clinic, two individuals were excluded due to the following reasons: one participant presented a cerebral disease and another experienced an emotional breakdown, which prevented both from completing the questionnaire. The final sample included 161 participants, consisting of 137 women and 24 men.

This unbalanced sex distribution in the sample did not allow meaningful comparison between male and female participants, thus male participants were excluded for multi-group (MG) CFA. Since women from two non-overlapping groups are necessary to establish a measurement invariance of the EOAQ, two age groups of 137 women under and over 48 year-old were formed, according to the median age of menopause in Brazilian women (Pompei et al. 2022). Since menopausal status was not collected during clinical procedures, this age-based split was not intended to test whether menopausal status changes treatment motivation. It was used as a clinically interpretable grouping variable to evaluate whether the candidate measurement structure remained stable across younger and older women.

The research protocol was approved by the institutional board of ethics (CAAE: 4.488.229/4.735.640). All participants signed a written consent after being informed about the study. For this study, permission was obtained from the authors (Fischer et al. 2011, 2014) to use the original copyrighted questionnaire.

BJP PRE-PROOF

Measurement Tool

The EOAQ was proposed by Fischer and colleagues during an expert meeting (Fischer et al. 2014), with the objective of examining patients' expectations regarding bariatric surgery. The EOAQ has three sections. Part I records demographic and clinical data. Part II explores expectations of body image and desired weight after surgery. Part III identifies motivations for undergoing surgery via a 14-item, 5-point Likert scale. Table 1 shows the original questionnaire in English and its translation into Brazilian-Portuguese (Silva et al. 2024). To our knowledge, this questionnaire has never been translated into any non-English language, nor has it received a validity study, even in its original English version. The only additional study on EOAQ prospectively investigated the proportion of unrealistic expectations of candidate patients and highlighted the different profiles of men and women who sought surgical treatment (Fischer et al. 2011).

The Part III of EOAQ addresses patients' motivations for surgery, emphasizing emotional and social improvements alongside physical and health-related benefits. Data were subjected to reliability analysis, as well as content and construct validity assessments.

Table 1 Translation and cultural adaptation to Brazilian-Portuguese of the European Obesity Academy Questionnaire (EOAQ)

Item	Original	Translation to Portuguese
	Please mark what you think best fits your opinion.	Por favor, marque o que você acha que melhor se encaixa na sua opinião.
	Feel free to use the whole range of answers.	Sinta-se livre para usar toda a gama de respostas.
	How important are the following issues on a scale of 1 (not important for you) to 5 (very important for you)?	Qual o grau de importância desses motivos em uma escala de 1 (sem importância) a 5 (muito importante para você)?
Q1	Improved mental health	Melhorar o bem estar emocional
Q2	Improved physical activity	Melhorar a capacidade de realizar exercícios físicos
Q3	Improved in intimacy and partnership	Melhorar o relacionamento íntimo com o(a) companheiro(a)
Q4	Reduction in clothing size	Reduzir o tamanho das roupas
Q5	Improved fertility	Melhorar a fertilidade
Q6	Improved social life (e.g., culture, meeting friends)	Melhorar a vida social (por ex. encontro com amigos, ir às festas)
Q7	Pain reduction	Diminuir dores

BJP PRE-PROOF

Q8	Improved comorbidity such as diabetes	Melhorar o controle do diabetes, hipertensão, distúrbio do sono, etc.
Q9	Reduced medication need	Diminuir o uso de medicamentos
Q10	Increased employment chance	Melhorar a oportunidade de conseguir um emprego
Q11	Better work performance	Melhorar o desempenho no trabalho
Q12	Weight loss	Perder peso
Q13	Improved self-esteem	Melhorar a autoestima
Q14	Increased life expectancy	Aumentar os anos de vida

Statistical Analysis

We analyzed the 14-item part III of the EOAQ scale of patient's motivations for bariatric surgery to ensure a self-contained scenario with a manageable number of model variations. The other parts consist of descriptive items and body image figures that are not intended as validated scales. The structural models were implemented and estimated using the robust full information maximum likelihood (FIML) method. The significance level was set at 5%. For the purpose of model identification, we did not assume *a priori* whether the structural model is unidimensional or multifactorial, therefore different structures could be explored (Figure 1):

- Unidimensional model: a general factor (G) causes all items.
- First-order model: independent specific factors present possible correlations to explain a subset of items.
- Bifactor model: combines the characteristics of unidimensional and first-order models by proposing a G and specific factors that simultaneously influence scale items (Jensen 1998). No correlations are defined among the specific factors, nor between the specific factors and the general factor G (Holzinger and Swineford 1937).
- Second-order model: this is another intermediate variation which proposes a hierarchy in such a way that a general factor G is mediated by specific factors that influence the items.

BJP PRE-PROOF

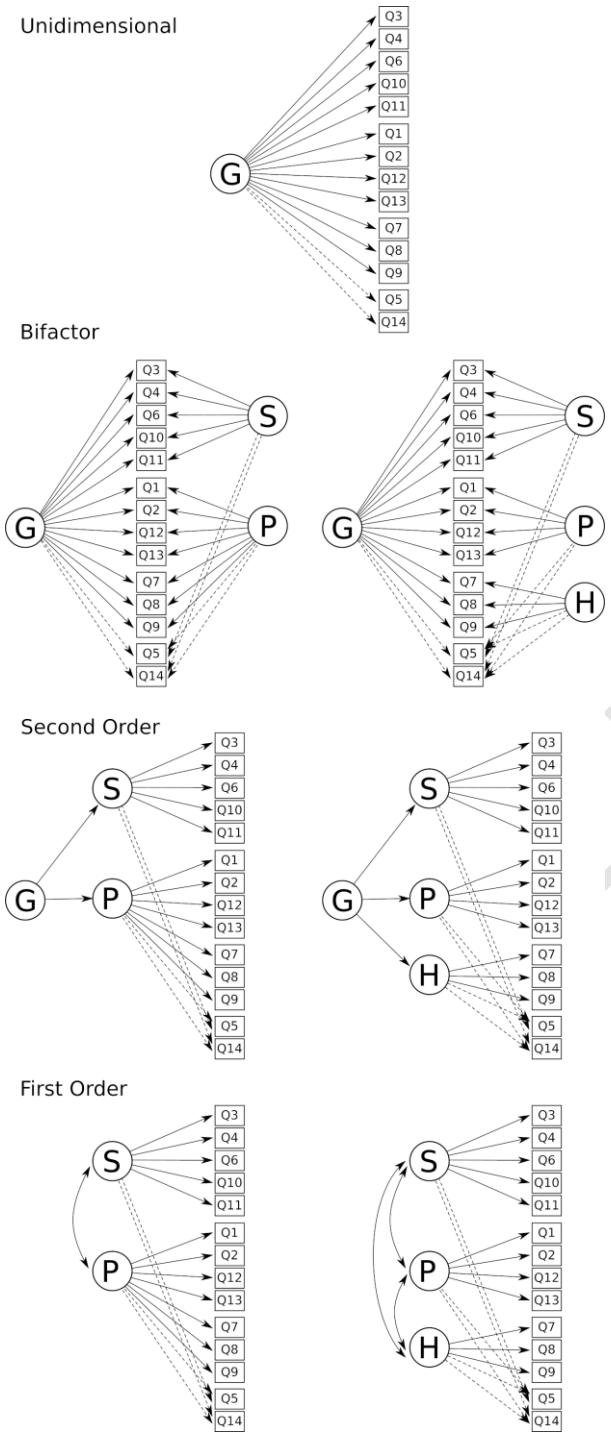


Figure 1 Psychometric models were tested, incorporating a general factor (G) along with either two specific factors (P = personal including Q1, Q2, Q7, Q8, Q9, Q12, Q13; S = social including Q3, Q4, Q6, Q10, Q11) or three specific factors (P = personal including Q1, Q2, Q12, Q13; S = social including Q3, Q4, Q6, Q10, Q11; H = healthcare including Q7, Q8, Q9). Items Q5 and Q14 were explored in relation to multiple proposed factors (represented by dotted lines), resulting in a total of 79 combinations

PROOF

In the context of Fig. 1, factors, dimensions, or domains (the preferred term may vary across applied research areas) refer to the same latent quantities and are represented by ellipses. Items, observed variables, or indicators refer to the responses of participants and are represented by rectangles. The unidimensional and first-order models represent the two extreme situations: a single-factor structure, in which all items are linked to one general factor G, and multifactorial structures, in which items are linked only to two or more specific domains. Bifactor and second-order models represent intermediate structures. Although both include a general factor G, they are not unidimensional because they preserve two or more specific domains. These intermediate structures may allow an overall score without implying that all items measure only a single latent dimension. The distribution of reliable variance across the general and specific factors can be quantified, as described below.

In a preliminary exploratory analysis (not shown), hypothetical models suggested two or three specific factors as candidate models, with items allocated according to all possible structures of Figure 1. For two specific factors, a personal (P) factor included the items Q1, Q2, Q7, Q8, Q9, Q12, and Q13, which represent intrinsic motivations, and a social factor (S) included the items Q3, Q4, Q6, Q10, and Q11, which are related to interpersonal relationships or extrinsic motivations. For the model with three specific factors, some items from the P factor were reallocated to a healthcare-related motivation factor (H), namely Q7, Q8, and Q9.

Item analyses indicated that Q5 and Q14 showed substantial floor and ceiling effects, respectively, and that both items were semantically ambiguous for allocation to specific factors. Q5 (“Improve fertility”) can be interpreted as a motivation related to health, personal or life-project, while Q14 (“Increase life expectancy”) can reflect either health-related or a more general assessment of the benefit of treatment. Their treatment was therefore defined as part of the model-generation rule before ranking the candidate models: both items were allowed either to be suppressed or to load on theoretically interpretable factors, and the same rule was applied across all CFA model families. By suppressing or placing

BJP PRE-PROOF

them in different factors, 79 different models were tested for all combinations (Figure 1).

For comparison and decision of the best structure, there are several indices to evaluate how well the model fits to data. The popular chi-square (χ^2) test is an index of exact fit, has a null tolerance and is overly strict. A non-significant χ^2 ($p \geq 0.05$) suggests an acceptable model fit, denoting that the model could plausibly represent the data. However, this test is sensitive to sample size and almost all models would be rejected ($p < 0.05$), regardless of their plausibility. Also, the χ^2 assumes multivariate normality and does not penalize overly complex models. As such, the test does have limitations that researchers need to be cautious about.

Alternatively, root mean square error of approximation (RMSEA) is an index that assesses how well the theoretical model reproduces the observed covariance matrix, taking model degrees of freedom into account. Unlike the chi-squared test of exact fit, RMSEA allows a numerical tolerance for approximate fit. Psychometric methods often use 0.05 as a reference value for close fit; the corresponding p-close is the p-value for testing whether the population RMSEA is equal to the reference value. Thus, the 90% confidence interval of RMSEA provides information about the precision of the fit estimate. For interpretation, the model was considered “highly plausible” if the upper limit of the 90% confidence interval was less than 0.05, “plausible” if 0.05 was within this interval, and “not plausible” if the lower limit was greater than 0.05 (MacCallum et al. 1996; Curran et al. 2003; Rosseel 2012).

On the other hand, the sample-size-adjusted Bayesian information criterion (SABIC) was calculated to rank the models (Lin et al. 2017). Overall, SABIC is an information criterion that balances model fit and parsimony, with lower values indicating better relative model quality. It is useful for ranking competing non-nested models, especially in smaller samples. We also report Comparative Fit Index (CFI) and the Parsimony-Adjusted Normed Fit Index (PNFI), which are widely reported in the related CFA literature.

McDonald’s omega (ω) coefficients were calculated for the general and specific factors of each structural model (Geldhof et al. 2014). In multidimensional

BJP PRE-PROOF

models, ω helps to describe how reliable variance is distributed across the general and specific factors. In the present study, however, ω was used as an admissibility check for the estimated factor structure rather than as a threshold-based classification of reliability magnitude. Because ω represents a proportion of reliable variance attributable to a factor, its estimates are expected to lie within the [0,1] interval. Values outside this range were interpreted as indicators of numerical instability or inadmissible model solutions.

Multi-group Confirmatory Factor Analysis (MG-CFA) is a statistical technique used to test measurement invariance, which assesses whether a psychological construct is measured in the same way across different groups. Measurement invariance was developed in layers to test between-group equality (Brown 2015; Cheung and Rensvold 2002; Dimitrov 2010; Wicherts and Dolan 2010). We adopted the interval (quantitative) level of measurement for testing scale items (Stevens 1946). The following seven levels were investigated:

- preliminary, global: the model design is plausible for the whole sample,
- preliminary, group ≤ 48 years-old: the model design is plausible for group 1,
- preliminary, group > 48 years-old: the model design is plausible for group 2,
- configural: equality of model design,
- metric: equality of loading factors,
- scalar: equality of item means, and
- strict: equality of variance of item error terms.

When a layer is tested and considered plausible, the next layer can be meaningfully tested. The underlying construct is interpreted only up to the highest level of measurement equivalence supported by the data (Cizek et al. 2011; Franzen 2000; Markus and Borsboom 2013). Measurement invariance analysis assesses how far the same measurement rule is supported across groups. Thus, measurement invariance is a prerequisite for interpreting between-group comparisons, because it evaluates whether group differences reflect the construct rather than artifacts of the measurement rule.

BJP PRE-PROOF

Data Availability

The analyses were performed with R 4.2.2 (R Core Team 2022) using the packages *lavaan* (Rosseel 2012) and *semTools* (Jorgensen et al. 2022). Data and R scripts are available at Harvard Dataverse (https://doi.org/10.7910/DVN/FCBOGB). This supplementary material provides our results and presents a systematic approach to model selection and validation using confirmatory factor analysis. It includes detailed statistical procedures, application guidelines, access to data sets and computer code, and a step-by-step method to select and rank competing models using psychometric indices.

Results

The EOAQ was completed by 161 adult patients with obesity and candidates for bariatric surgery. From these, only females (n=137) were selected for construct validity testing and arranged in two groups, with age less or equal to 48 or above 48 years old. The clinical and demographic characteristics of these two groups are summarized in Table 2. It was observed that, apart from the age grouping required for the multi-group analysis, patients did not differ in their core variables, including body composition, educational level, occupation, or commonly prevalent comorbidities.

Table 2 Characteristics of female participants (n = 137), stratified by age.

Variable	**< 48 years old (n=82)**	**≥ 48 years old (n=55)**	**p**
Age (years)			
Range	23 - 47	48 - 68	
Mean (sd)	39.57 (5.55)	57.11 (5.18)	
Body mass index			0.091
Range	33.79 - 80.47	32.46 - 71.92	
Mean (sd)	48.73 (8.53)	46.09 (7.53)	
Schooling (years)			0.320
< 14	59 (72.0%)	46 (83.6%)	
≥ 14	23 (28.0%)	9 (16.4%)	
Working status			0.053
Working	53 (64.6%)	19 (34.6%)	
Not working	29 (35.4%)	36 (65.5%)	
Diagnoses of hypertension			0.579
yes	50 (61.0%)	45 (81.8%)	
No	32 (39.0%)	10 (18.2%)	

BJP PRE-PROOF

Diagnoses of diabetes			0.999
Yes	24 (29.3%)	31 (56.4%)	
No	58 (70.7%)	24 (43.6%)	

Because items Q5 (“Improve fertility” and Q14 (“Increase life expectancy”) of the EOAQ had preliminarily shown floor and ceiling effects, all model combinations were considered by suppressing or including one or both items in each factor. Accounting for seven possible basic model structures (Figure 1), a total of 79 different models included the unidimensional models (with only the general factor, G), and three specific factors, named personal (P), social (S), and healthcare (H) motivations. The H factor was formed by removing some items from the P factor (Q7: “Pain reduction”, Q8: “Improve comorbidity such as diabetes”, and Q9: “Reduced medication need”).

Table 3 displays the indices of construct validity. Models are presented into four blocks of plausibility: (1) plausible models at the configural level of construct validity according to the criterion of RMSEA confidence interval, (2) plausible models only at the preliminary level (when all patients were in a single group, regardless of menopause age), (3) not plausible models, and (4) models with convergence problems. Each block is sorted by SABIC. The statistical decision of non-rejection is inferential, based on p-close ≥ 0.05, which usually coincides with model plausibility given by the value 0.05 within the 90% confidence interval of RMSEA. The maximum level of measurement invariance achieved in the current EOAQ example was configural invariance, meaning that both groups share the same model structure but differ in factor loadings (metric invariance), and further levels (scalar and strict invariance) could not be meaningfully tested.

BJP PRE-PROOF

Table 3 Construct validation: model plausibility ordered by SABIC

Model	X^2	df	p	RMSEA	CI 90% lower	CI 90% upper	pclose	CFI	PNFI	SABIC	acceptance	up to
1. Bifactor G+PS (no Q5 and Q14)	153.0	84	<0.0001	0.061	0.000	0.095	0.3049	0.937	0.431	3777.5	plausible	configural
2. Bifactor G+PSH (no Q5 and Q14)	166.8	84	<0.0001	0.076	0.038	0.108	0.1099	0.902	0.412	3792.0	plausible	configural
3. Bifactor G+PSH (G:Q5 H:Q5)	226.5	104	<0.0001	0.073	0.039	0.101	0.1153	0.907	0.415	3896.7	plausible	configural
4. Bifactor G+PS (G:Q14 P:Q14)	185.1	104	<0.0001	0.055	0.000	0.087	0.3955	0.939	0.426	3914.0	plausible	configural
5. Bifactor G+PS (G:Q14 S:Q14)	176.4	104	<0.0001	0.066	0.024	0.096	0.2200	0.914	0.437	3924.8	plausible	configural
6. Bifactor G+PSH (G:Q14 P:Q14)	207.1	104	<0.0001	0.068	0.029	0.098	0.1849	0.908	0.397	3926.4	plausible	configural
7. Bifactor G+PSH (G:Q14 H:Q14)	211.1	104	<0.0001	0.069	0.031	0.098	0.1745	0.906	0.392	3926.8	plausible	configural
8. Bifactor G+PSH (G:Q14 S:Q14)	208.2	104	<0.0001	0.073	0.038	0.102	0.1214	0.893	0.396	3933.3	plausible	configural
9. Bifactor G+PSH (G:Q5+Q14 P:Q14 H:Q5)	282.5	126	<0.0001	0.066	0.032	0.092	0.1888	0.911	0.387	4032.0	plausible	configural
10. Bifactor G+PSH (G:Q5+Q14 H:Q5+Q14)	272.4	126	<0.0001	0.069	0.038	0.095	0.1368	0.901	0.398	4036.0	plausible	configural
11. Bifactor G+PSH (G:Q5+Q14 S:Q14 H:Q5)	261.6	126	<0.0001	0.070	0.039	0.096	0.1290	0.899	0.410	4038.0	plausible	configural
12. 1st Order PSH (no Q5 and Q14)	084.2	51	0.0024	0.060	0.018	0.092	0.2893	0.901	0.538	3986.9	plausible	preliminary
13. 2nd Order G->PSH (no Q5 and Q14)	083.5	51	0.0027	0.060	0.018	0.092	0.2893	0.901	0.540	3986.9	plausible	preliminary
14. 1st Order PS (no Q5 and Q14)	092.8	53	0.0006	0.071	0.038	0.101	0.1298	0.856	0.534	3999.3	plausible	preliminary
15. 1st Order PSH (P:Q14)	106.7	62	0.0004	0.060	0.026	0.088	0.2703	0.883	0.517	4173.2	plausible	preliminary
16. 2nd Order G->PSH (G:Q14 P:Q14)	106.0	62	0.0004	0.060	0.026	0.088	0.2703	0.883	0.519	4173.2	plausible	preliminary
17. 1st Order PSH (H:Q14)	106.2	62	0.0004	0.060	0.024	0.087	0.2855	0.886	0.518	4173.9	plausible	preliminary
18. 2nd Order G->PSH (G:Q14 H:Q14)	103.1	62	0.0008	0.060	0.024	0.087	0.2854	0.886	0.527	4173.9	plausible	preliminary
19. 2nd Order G->PSH (G:Q14 S:Q14)	106.2	62	0.0004	0.061	0.026	0.089	0.2650	0.881	0.519	4175.9	plausible	preliminary
20. 1st Order PSH (S:Q14)	106.0	62	0.0004	0.061	0.026	0.089	0.2650	0.881	0.519	4175.9	plausible	preliminary
21. 1st Order PS (P:Q14)	116.0	64	<0.0001	0.069	0.040	0.095	0.1230	0.840	0.509	4185.5	plausible	preliminary
22. 1st Order PS (S:Q14)	114.9	64	<0.0001	0.070	0.040	0.096	0.1218	0.838	0.512	4188.3	plausible	preliminary
23. Bifactor G+PS (G:Q5 S:Q5)	071.3	52	0.0392	0.031	0.000	0.068	0.7638	0.973	0.511	4298.9	plausible	preliminary
24. Bifactor G+PS (G:Q5 P:Q5)	074.6	52	0.0216	0.037	0.000	0.070	0.7063	0.963	0.504	4299.0	plausible	preliminary
25. Bifactor G+PSH (G:Q5 S:Q5)	075.6	52	0.0180	0.045	0.000	0.078	0.5725	0.946	0.502	4309.9	plausible	preliminary
26. Bifactor G+PSH (G:Q5 P:Q5)	078.6	52	0.0100	0.048	0.000	0.080	0.5161	0.938	0.495	4311.4	plausible	preliminary
27. 1st Order PSH (S:Q5)	103.6	62	0.0007	0.061	0.027	0.089	0.2601	0.880	0.525	4323.4	plausible	preliminary
28. 2nd Order G->PSH (G:Q5 S:Q5)	101.5	62	0.0012	0.061	0.027	0.089	0.2600	0.880	0.531	4323.4	plausible	preliminary

BJP PRE-PROOF

Model	X^2	df	p	RMSEA	CI 90% lower	CI 90% upper	pclose	CFI	PNFI	SABIC	acceptance	up to
29. 1st Order PSH (P:Q5)	103.0	62	0.0008	0.061	0.026	0.089	0.2614	0.881	0.527	4323.6	plausible	preliminary
30. 2nd Order G->PSH (G:Q5 P:Q5)	102.2	62	0.0010	0.061	0.026	0.089	0.2614	0.881	0.529	4323.6	plausible	preliminary
31. 1st Order PSH (H:Q5)	104.0	62	0.0007	0.061	0.027	0.089	0.2539	0.879	0.525	4324.1	plausible	preliminary
32. 2nd Order G->PSH (G:Q5 H:Q5)	103.2	62	0.0008	0.061	0.027	0.089	0.2539	0.879	0.527	4324.1	plausible	preliminary
33. 1st Order PS (S:Q5)	112.2	64	0.0002	0.070	0.040	0.096	0.1199	0.837	0.519	4335.8	plausible	preliminary
34. 1st Order PS (P:Q5)	111.6	64	0.0002	0.070	0.040	0.097	0.1205	0.837	0.521	4336.0	plausible	preliminary
35. Bifactor G+PS (G:Q5+Q14 P:Q14 S:Q5)	094.3	63	0.0065	0.036	0.000	0.067	0.7421	0.959	0.498	4486.5	plausible	preliminary
36. Bifactor G+PS (G:Q5+Q14 P:Q5+Q14)	097.5	63	0.0035	0.040	0.000	0.069	0.6871	0.949	0.491	4486.5	plausible	preliminary
37. Bifactor G+PS (G:Q5+Q14 S:Q5+Q14)	094.4	63	0.0064	0.037	0.000	0.067	0.7279	0.957	0.498	4487.1	plausible	preliminary
38. Bifactor G+PS (G:Q5+Q14 P:Q5 S:Q14)	097.8	63	0.0032	0.040	0.000	0.069	0.6778	0.948	0.491	4487.2	plausible	preliminary
39. Bifactor G+PSH (G:Q5+Q14 P:Q14 S:Q5)	092.8	63	0.0087	0.041	0.000	0.072	0.6438	0.945	0.501	4495.0	plausible	preliminary
40. Bifactor G+PSH (G:Q5+Q14 S:Q5 H:Q14)	099.5	63	0.0023	0.047	0.000	0.076	0.5341	0.928	0.487	4498.5	plausible	preliminary
41. Bifactor G+PSH (G:Q5+Q14 S:Q5+Q14)	102.7	63	0.0012	0.049	0.000	0.077	0.4991	0.923	0.481	4499.3	plausible	preliminary
42. Bifactor G+PSH (G:Q5+Q14 P:Q5 H:Q14)	100.6	63	0.0018	0.049	0.000	0.077	0.5020	0.923	0.485	4499.6	plausible	preliminary
43. Bifactor G+PSH (G:Q5+Q14 P:Q5+Q14)	104.0	63	0.0009	0.048	0.000	0.077	0.5115	0.925	0.478	4499.7	plausible	preliminary
44. Bifactor G+PSH (G:Q5+Q14 P:Q5 S:Q14)	101.4	63	0.0016	0.050	0.000	0.078	0.4829	0.921	0.483	4500.3	plausible	preliminary
45. 1st Order PSH (P:Q14 S:Q5)	127.9	74	0.0001	0.060	0.031	0.085	0.2507	0.864	0.504	4509.8	plausible	preliminary
46. 2nd Order G->PSH (G:Q5+Q14 P:Q14 S:Q5)	125.4	74	0.0002	0.060	0.031	0.085	0.2506	0.864	0.509	4509.8	plausible	preliminary
47. 1st Order PSH (P:Q5+Q14)	127.2	74	0.0001	0.060	0.031	0.085	0.2519	0.864	0.505	4510.0	plausible	preliminary
48. 2nd Order G->PSH (G:Q5+Q14 P:Q5+Q14)	126.3	74	0.0001	0.060	0.031	0.085	0.2519	0.864	0.507	4510.0	plausible	preliminary
49. 1st Order PSH (S:Q5 H:Q14)	127.3	74	0.0001	0.060	0.030	0.084	0.2644	0.866	0.505	4510.4	plausible	preliminary
50. 2nd Order G->PSH (G:Q5+Q14 S:Q5 H:Q14)	123.9	74	0.0002	0.060	0.030	0.084	0.2641	0.866	0.513	4510.4	plausible	preliminary
51. 1st Order PSH (P:Q14 H:Q5)	128.2	74	<0.0001	0.060	0.031	0.085	0.2449	0.862	0.503	4510.5	plausible	preliminary
52. 2nd Order G->PSH (G:Q5+Q14 P:Q14 H:Q5)	127.9	74	0.0001	0.060	0.031	0.085	0.2449	0.862	0.503	4510.5	plausible	preliminary
53. 1st Order PSH (P:Q5 H:Q14)	126.6	74	0.0001	0.060	0.030	0.084	0.2654	0.866	0.506	4510.6	plausible	preliminary
54. 2nd Order G->PSH (G:Q5+Q14 P:Q5 H:Q14)	123.4	74	0.0003	0.060	0.030	0.085	0.2653	0.866	0.514	4510.6	plausible	preliminary
55. 1st Order PSH (H:Q5+Q14)	127.7	74	0.0001	0.060	0.030	0.085	0.2578	0.865	0.504	4511.1	plausible	preliminary
56. 2nd Order G->PSH (G:Q5+Q14 H:Q5+Q14)	124.3	74	0.0002	0.060	0.030	0.085	0.2577	0.865	0.512	4511.1	plausible	preliminary
57. 1st Order PSH (S:Q5+Q14)	127.0	74	0.0001	0.061	0.031	0.085	0.2464	0.862	0.506	4512.4	plausible	preliminary

PRE-PROOF

Model	X^2	df	p	RMSEA	CI 90% lower	CI 90% upper	pclose	CFI	PNFI	SABIC	acceptance	up to
58. 2nd Order G->PSH (G:Q5+Q14 S:Q5+Q14)	125.8	74	0.0002	0.061	0.031	0.085	0.2464	0.862	0.508	4512.4	plausible	preliminary
59. 2nd Order G->PSH (G:Q5+Q14 P:Q5 S:Q14)	126.5	74	0.0001	0.061	0.031	0.085	0.2470	0.862	0.507	4512.6	plausible	preliminary
60. 1st Order PSH (P:Q5 S:Q14)	126.4	74	0.0001	0.061	0.031	0.085	0.2470	0.862	0.507	4512.6	plausible	preliminary
61. 2nd Order G->PSH (G:Q5+Q14 S:Q14 H:Q5)	127.4	74	0.0001	0.061	0.031	0.086	0.2400	0.860	0.505	4513.1	plausible	preliminary
62. 1st Order PSH (S:Q14 H:Q5)	127.4	74	0.0001	0.061	0.031	0.086	0.2400	0.860	0.505	4513.1	plausible	preliminary
63. 1st Order PS (P:Q14 S:Q5)	137.2	76	<0.0001	0.068	0.042	0.091	0.1174	0.822	0.494	4522.1	plausible	preliminary
64. 1st Order PS (P:Q5+Q14)	136.5	76	<0.0001	0.068	0.042	0.092	0.1179	0.822	0.496	4522.3	plausible	preliminary
65. 1st Order PS (S:Q5+Q14)	135.9	76	<0.0001	0.068	0.042	0.092	0.1164	0.820	0.497	4524.8	plausible	preliminary
66. 1st Order PS (P:Q5 S:Q14)	135.3	76	<0.0001	0.068	0.042	0.092	0.1165	0.819	0.499	4525.0	plausible	preliminary
67. Unidimensional (no Q5 and Q14)	104.3	54	<0.0001	0.084	0.055	0.112	0.0291	0.796	0.510	4013.9	not plausible	none
68. Unidimensional (G:Q14)	128.8	65	<0.0001	0.080	0.054	0.105	0.0295	0.781	0.482	4200.9	not plausible	none
69. Unidimensional (G:Q5)	123.7	65	<0.0001	0.081	0.055	0.106	0.0288	0.779	0.496	4350.6	not plausible	none
70. Unidimensional (G:Q5+Q14)	149.8	77	<0.0001	0.078	0.054	0.100	0.0300	0.764	0.469	4537.7	not plausible	none
71. 2nd Order G->PS (no Q5 and Q14)	091.0	52	0.0007	-	-	-	-	-	0.529	4001.0	no RMSEA	not available
72. 2nd Order G->PS (G:Q14 P:Q14)	114.2	63	<0.0001	-	-	-	-	-	0.506	4187.3	no RMSEA	not available
73. 2nd Order G->PS (G:Q14 S:Q14)	113.1	63	0.0001	-	-	-	-	-	0.509	4190.0	no RMSEA	not available
74. 2nd Order G->PS (G:Q5 S:Q5)	110.5	63	0.0002	-	-	-	-	-	0.516	4337.6	no RMSEA	not available
75. 2nd Order G->PS (G:Q5 P:Q5)	109.9	63	0.0002	-	-	-	-	-	0.517	4337.8	no RMSEA	not available
76. 2nd Order G->PS (G:Q5+Q14 P:Q14 S:Q5)	135.4	75	<0.0001	-	-	-	-	-	0.492	4523.8	no RMSEA	not available
77. 2nd Order G->PS (G:Q5+Q14 P:Q5+Q14)	134.7	75	<0.0001	-	-	-	-	-	0.494	4524.0	no RMSEA	not available
78. 2nd Order G->PS (G:Q5+Q14 S:Q5+Q14)	134.1	75	<0.0001	-	-	-	-	-	0.495	4526.6	no RMSEA	not available
79. 2nd Order G->PS (G:Q5+Q14 P:Q5 S:Q14)	133.5	75	<0.0001	-	-	-	-	-	0.496	4526.8	no RMSEA	not available

Model and factor structure: general factor (G); specific factors P (personal), S (social), and H (health). See source document for full definitions and explanatory notes.

BJP PRE-PROOF

Only bifactor models achieved configural invariance. According to SABIC, the two bifactor models without Q5 and Q14 were the best ones (models 1 and 2, Table 3). Models 3 to 8 included only one of them. It seems that Q14 can be placed in different factors, but Q5 was only admissible in factor H (model 3). Models 9 to 11 used all the questionnaire items and were formed with three specific factors.

Table 4 shows the reliability analysis based on McDonald's omegas (ω), presenting the models in the same sequence as Table 3. For bifactor models, ω was reported separately for the general factor and for each specific factor, and it was computed by group when configural or higher levels of invariance were reached. These coefficients were used to check whether the estimated factor structure produced admissible proportions of reliable variance. Many models showed numerical instability, resulting in ω estimates outside the expected [0,1] interval. Considering both global fit and admissible ω estimates, the valid models are 7, 8, and 10.

BJP PRE-PROOF

Table 4 Reliability analysis: effect size estimated by McDonald's omega

Model	≤48 ωG	≤48 ωP	≤48 ωS	≤48 ωH	>48 ωG	>48 ωP	>48 ωS	>48 ωH	plausibility up to	ω range	offending count	estimates sum
1. Bifactor G+PS (no Q5 and Q14)	0.61	5.59	0.24		0.56	0.15	0.42		configural	invalid	1	14.77
2. Bifactor G+PSH (no Q5 and Q14)	0.61	0.16	0.23	0.42	0.31	0.43	0.58	3.40	configural	invalid	2	5.71
3. Bifactor G+PSH (G:Q5 H:Q5)	0.58	0.17	0.23	0.26	0.31	0.42	0.58	1.18	configural	invalid	2	3.26
4. Bifactor G+PS (G:Q14 P:Q14)	0.61	3.39	0.24		0.58	0.19	0.39		configural	invalid	1	11.52
5. Bifactor G+PS (G:Q14 S:Q14)	0.63	4.25	0.21		0.56	0.04	0.42		configural	invalid	1	12.78
6. Bifactor G+PSH (G:Q14 P:Q14)	0.60	2.39	0.24	0.43	0.36	0.20	0.58	1.40	configural	invalid	3	8.27
7. Bifactor G+PSH (G:Q14 H:Q14)	0.60	0.16	0.24	0.44	0.33	0.43	0.59	0.01	configural	valid	1	0.12
8. Bifactor G+PSH (G:Q14 S:Q14)	0.62	0.17	0.20	0.42	0.34	0.33	0.55	0.13	configural	valid	1	0.04
9. Bifactor G+PSH (G:Q5+Q14 P:Q14 H:Q5)	0.58	1.79	0.24	0.26	0.35	0.29	0.58	1.45	configural	invalid	2	8.61
10. Bifactor G+PSH (G:Q5+Q14 H:Q5+Q14)	0.58	0.16	0.24	0.60	0.35	0.38	0.58	0.27	configural	valid	3	2.74
11. Bifactor G+PSH (G:Q5+Q14 S:Q14 H:Q5)	0.61	0.16	0.19	0.26	0.35	0.39	0.54	1.28	configural	invalid	1	3.14
12. 1st Order PSH (no Q5 and Q14)	-	-	-	-	-	0.76	0.59	0.55	preliminary	valid	-	-
13. 2nd Order G->PSH (no Q5 and Q14)	-	-	-	-	-	0.76	0.59	0.55	preliminary	valid	-	-
14. 1st Order PS (no Q5 and Q14)	-	-	-	-	-	0.68	0.60	-	preliminary	valid	-	-
15. 1st Order PSH (P:Q14)	-	-	-	-	-	0.74	0.60	0.55	preliminary	valid	-	-
16. 2nd Order G->PSH (G:Q14 P:Q14)	-	-	-	-	-	0.74	0.60	0.55	preliminary	valid	-	-
17. 1st Order PSH (H:Q14)	-	-	-	-	-	0.76	0.59	0.53	preliminary	valid	-	-
18. 2nd Order G->PSH (G:Q14 H:Q14)	-	-	-	-	-	0.76	0.59	0.53	preliminary	valid	-	-
19. 2nd Order G->PSH (G:Q14 S:Q14)	-	-	-	-	-	0.76	0.59	0.55	preliminary	valid	-	-
20. 1st Order PSH (S:Q14)	-	-	-	-	-	0.76	0.59	0.55	preliminary	valid	-	-
21. 1st Order PS (P:Q14)	-	-	-	-	-	0.68	0.60	-	preliminary	valid	-	-
22. 1st Order PS (S:Q14)	-	-	-	-	-	0.68	0.60	-	preliminary	valid	-	-
23. Bifactor G+PS (G:Q5 S:Q5)	-	-	-	-	0.54	0.05	0.45	-	preliminary	valid	1	0.05
24. Bifactor G+PS (G:Q5 P:Q5)	-	-	-	-	0.54	0.04	0.43	-	preliminary	valid	1	0.11
25. Bifactor G+PSH (G:Q5 S:Q5)	-	-	-	-	0.56	0.09	0.40	0.42	preliminary	valid	-	-
26. Bifactor G+PSH (G:Q5 P:Q5)	-	-	-	-	0.51	0.09	0.43	0.42	preliminary	valid	-	-
27. 1st Order PSH (S:Q5)	-	-	-	-	-	0.76	0.58	0.55	preliminary	valid	-	-
28. 2nd Order G->PSH (G:Q5 S:Q5)	-	-	-	-	-	0.76	0.58	0.55	preliminary	valid	-	-

BJP PRE-PROOF

Model	≤48 ωG	≤48 ωP	≤48 ωS	≤48 ωH	>48 ωG	>48 ωP	>48 ωS	>48 ωH	plausibility up to	ω range	offending count	estimates sum
29. 1st Order PSH (P:Q5)	-	-	-	-	-	0.68	0.60	0.55	preliminary	valid	-	-
30. 2nd Order G->PSH (G:Q5 P:Q5)	-	-	-	-	-	0.68	0.60	0.55	preliminary	valid	-	-
31. 1st Order PSH (H:Q5)	-	-	-	-	-	0.76	0.59	0.46	preliminary	valid	-	-
32. 2nd Order G->PSH (G:Q5 H:Q5)	-	-	-	-	-	0.76	0.59	0.46	preliminary	valid	-	-
33. 1st Order PS (S:Q5)	-	-	-	-	-	0.68	0.58	-	preliminary	valid	-	-
34. 1st Order PS (P:Q5)	-	-	-	-	-	0.66	0.60	-	preliminary	valid	-	-
35. Bifactor G+PS (G:Q5+Q14 P:Q14 S:Q5)	-	-	-	-	0.55	0.05	0.45	-	preliminary	valid	-	-
36. Bifactor G+PS (G:Q5+Q14 P:Q5+Q14)	-	-	-	-	0.55	0.03	0.43	-	preliminary	valid	1	0.05
37. Bifactor G+PS (G:Q5+Q14 S:Q5+Q14)	-	-	-	-	0.54	0.05	0.45	-	preliminary	valid	1	0.05
38. Bifactor G+PS (G:Q5+Q14 P:Q5 S:Q14)	-	-	-	-	0.55	0.03	0.43	-	preliminary	valid	1	0.09
39. Bifactor G+PSH (G:Q5+Q14 P:Q14 S:Q5)	-	-	-	-	0.55	0.11	0.39	0.43	preliminary	valid	-	-
40. Bifactor G+PSH (G:Q5+Q14 S:Q5 H:Q14)	-	-	-	-	0.56	0.06	0.40	0.39	preliminary	valid	-	-
41. Bifactor G+PSH (G:Q5+Q14 S:Q5+Q14)	-	-	-	-	0.55	0.06	0.40	0.42	preliminary	valid	-	-
42. Bifactor G+PSH (G:Q5+Q14 P:Q5 H:Q14)	-	-	-	-	0.52	0.09	0.43	0.40	preliminary	valid	-	-
43. Bifactor G+PSH (G:Q5+Q14 P:Q5+Q14)	-	-	-	-	0.54	0.11	0.40	0.43	preliminary	valid	-	-
44. Bifactor G+PSH (G:Q5+Q14 P:Q5 S:Q14)	-	-	-	-	0.52	0.09	0.42	0.42	preliminary	valid	-	-
45. 1st Order PSH (P:Q14 S:Q5)	-	-	-	-	-	0.74	0.58	0.55	preliminary	valid	-	-
46. 2nd Order G->PSH (G:Q5+Q14 P:Q14 S:Q5)	-	-	-	-	-	0.74	0.58	0.55	preliminary	valid	-	-
47. 1st Order PSH (P:Q5+Q14)	-	-	-	-	-	0.67	0.60	0.55	preliminary	valid	-	-
48. 2nd Order G->PSH (G:Q5+Q14 P:Q5+Q14)	-	-	-	-	-	0.67	0.60	0.55	preliminary	valid	-	-
49. 1st Order PSH (S:Q5 H:Q14)	-	-	-	-	-	0.76	0.58	0.53	preliminary	valid	-	-
50. 2nd Order G->PSH (G:Q5+Q14 S:Q5 H:Q14)	-	-	-	-	-	0.76	0.58	0.53	preliminary	valid	-	-
51. 1st Order PSH (P:Q14 H:Q5)	-	-	-	-	-	0.74	0.60	0.46	preliminary	valid	-	-
52. 2nd Order G->PSH (G:Q5+Q14 P:Q14 H:Q5)	-	-	-	-	-	0.74	0.60	0.46	preliminary	valid	-	-
53. 1st Order PSH (P:Q5 H:Q14)	-	-	-	-	-	0.68	0.60	0.53	preliminary	valid	-	-
54. 2nd Order G->PSH (G:Q5+Q14 P:Q5 H:Q14)	-	-	-	-	-	0.68	0.60	0.53	preliminary	valid	-	-
55. 1st Order PSH (H:Q5+Q14)	-	-	-	-	-	0.76	0.59	0.45	preliminary	valid	-	-
56. 2nd Order G->PSH (G:Q5+Q14 H:Q5+Q14)	-	-	-	-	-	0.76	0.59	0.46	preliminary	valid	-	-
57. 1st Order PSH (S:Q5+Q14)	-	-	-	-	-	0.76	0.58	0.55	preliminary	valid	-	-

BJP PRE-PROOF

Model	≤48 ωG	≤48 ωP	≤48 ωS	≤48 ωH	>48 ωG	>48 ωP	>48 ωS	>48 ωH	plausibility up to	ω range	offending count	estimates sum
58. 2nd Order G->PSH (G:Q5+Q14 S:Q5+Q14)	-	-	-	-	-	0.76	0.58	0.55	preliminary	valid	-	-
59. 2nd Order G->PSH (G:Q5+Q14 P:Q5 S:Q14)	-	-	-	-	-	0.68	0.60	0.55	preliminary	valid	-	-
60. 1st Order PSH (P:Q5 S:Q14)	-	-	-	-	-	0.68	0.60	0.55	preliminary	valid	-	-
61. 2nd Order G->PSH (G:Q5+Q14 S:Q14 H:Q5)	-	-	-	-	-	0.76	0.59	0.46	preliminary	valid	-	-
62. 1st Order PSH (S:Q14 H:Q5)	-	-	-	-	-	0.76	0.59	0.46	preliminary	valid	-	-
63. 1st Order PS (P:Q14 S:Q5)	-	-	-	-	-	0.68	0.58	-	preliminary	valid	-	-
64. 1st Order PS (P:Q5+Q14)	-	-	-	-	-	0.66	0.60	-	preliminary	valid	-	-
65. 1st Order PS (S:Q5+Q14)	-	-	-	-	-	0.68	0.58	-	preliminary	valid	-	-
66. 1st Order PS (P:Q5 S:Q14)	-	-	-	-	-	0.66	0.60	-	preliminary	valid	-	-
67. Unidimensional (no Q5 and Q14)	-	-	-	-	0.65	-	-	-	none	valid	-	-
68. Unidimensional (G:Q14)	-	-	-	-	0.66	-	-	-	none	valid	-	-
69. Unidimensional (G:Q5)	-	-	-	-	0.64	-	-	-	none	valid	-	-
70. Unidimensional (G:Q5+Q14)	-	-	-	-	0.65	-	-	-	none	valid	-	-
71. 2nd Order G->PS (no Q5 and Q14)	-	-	-	-	-	0.68	0.60	-	not available	undefined	-	-
72. 2nd Order G->PS (G:Q14 P:Q14)	-	-	-	-	-	0.68	0.60	-	not available	undefined	-	-
73. 2nd Order G->PS (G:Q14 S:Q14)	-	-	-	-	-	0.68	0.60	-	not available	undefined	-	-
74. 2nd Order G->PS (G:Q5 S:Q5)	-	-	-	-	-	0.68	0.58	-	not available	undefined	-	-
75. 2nd Order G->PS (G:Q5 P:Q5)	-	-	-	-	-	0.66	0.60	-	not available	undefined	-	-
76. 2nd Order G->PS (G:Q5+Q14 P:Q14 S:Q5)	-	-	-	-	-	0.68	0.58	-	not available	undefined	-	-
77. 2nd Order G->PS (G:Q5+Q14 P:Q5+Q14)	-	-	-	-	-	0.66	0.60	-	not available	undefined	-	-
78. 2nd Order G->PS (G:Q5+Q14 S:Q5+Q14)	-	-	-	-	-	0.68	0.58	-	not available	undefined	-	-
79. 2nd Order G->PS (G:Q5+Q14 P:Q5 S:Q14)	-	-	-	-	-	0.66	0.60	-	not available	undefined	-	-

McDonald's omega calculated for factors G, P, S, and H. Offending estimates: factor loadings outside the [−1, 1] range.

BJP PRE-PROOF

Among these three plausible models at the configural level, it is still necessary to verify the loading factors, which are Pearson's correlations ranging from -1 to 1. Model 10 had three problematic estimates at the configural level: 1.93 for H to Q7 and 1.26 for P to Q12 for participants aged 48 years or younger, and 2.55 for H to Q7 for those older than 48 (not shown). In contrast, Model 8 exhibited the smallest problematic estimate with 1.04 for G to Q12 for participants older than 48 (Figure 2, upper panel). This was closely followed by Model 7, with an estimate of 1.12 for P to Q12 for participants aged 48 years or younger (Figure 2, lower panel). Considering the difficulties caused by a small sample, these are the only two models with almost all loading factors in the correct range, both bifactor models with three specific factors, without Q5 and with Q14 placed in the S factor (model 8) or H factor (model 7). Although apparently successful, all remaining models presented offending loading factors.

BJP PRE-PROOF

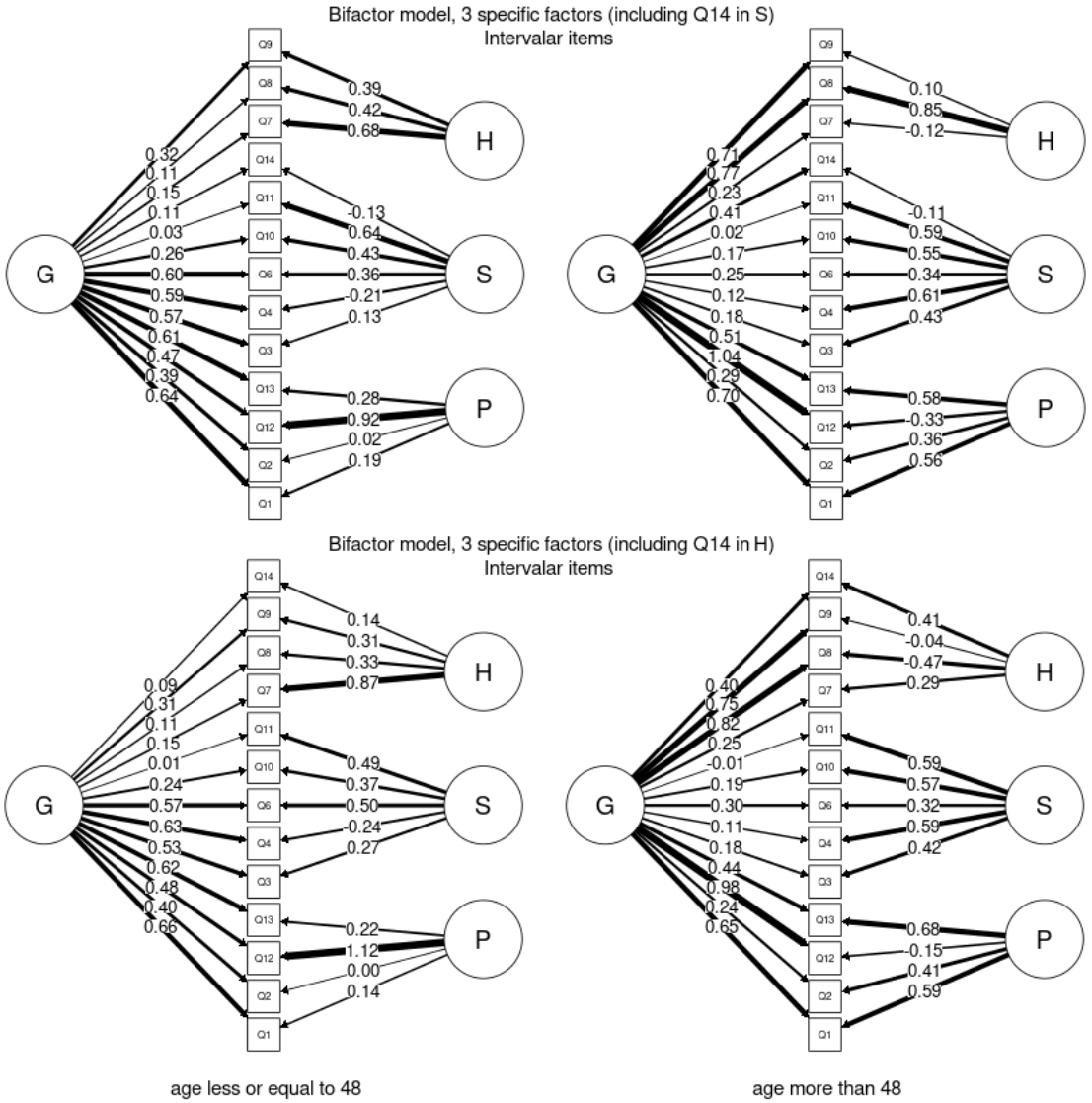


Figure 2 Graphical representation of model 8 (upper panel) and model 7 (lower panel). Bifactor models with three specific factors, personal (P), social (S), and healthcare (H), including Q14 in social and healthcare specific factors, respectively

Models 12 through 66 are plausible only at the preliminary level, and their ω ranges are valid. At this preliminary level, these structures can be interpreted as plausible representations of the EOAQ motivation items in this sample of women with obesity, but they provide only limited structural support and were not retained for subsequent multi-group invariance testing. Numerical instability likely did not occur at this level because these models are less demanding. More complex models, such as those at the configural level, may require larger samples to achieve stability.

According to SABIC rankings, the best models should exclude items Q5 and Q14, followed by those including only Q14 or only Q5. Models containing both Q5 and Q14 tended to rank lower. All unidimensional models (Models 67 to 70) were rejected. Among the remaining, all are second-order models with two specific factors presenting convergence problems, whose SABIC-based ranking is as follows: without Q14 and Q5, including only Q14, including only Q5, and finally, the full models with all fourteen items, which performed the worst.

In summary, models 8 and 7 emerged as the best ones after testing 79 models, each verified across 7 levels (preliminarily with the entire sample and with each subgroup, as well as configural, metric, scalar, and strict invariances). They indicated acceptable fit measures and achieved admissible ω values. These models are almost complete, with only item Q5 excluded.

Discussion

For the first time, a systematic confirmatory factor analysis (CFA) framework is proposed to identify and refine candidate structural equation models, using fit indices and reliability metrics to compare competing structures. This process was demonstrated through its application to Part III of the EOAQ in treatment-seeking individuals with severe obesity. Two bifactor solutions emerged as the most plausible candidates under the proposed decision rules after evaluating 79 competing models that included unidimensional, bifactor, first-order, and second-order structures. Within this dataset, the EOAQ motivation items were better represented by a general motivation factor, alongside personal, social, and healthcare-related dimensions. Both bifactor models retained nearly all items, excluding one malfunctioning item (Q5: “improve fertility”), and achieved acceptable RMSEA p-close values and favorable SABIC rankings. McDonald’s omega coefficients and factor loadings were used as additional criteria to identify unstable or inadmissible models.

Measurement invariance helps to locate where group equivalence is supported and where it stops (Meredith 1993; Putnick and Bornstein 2016). It also contributes to evaluating the extent to which score properties and interpretations

BJP PRE-PROOF

can be generalized across populations, settings, tasks, or experimental conditions (Wu et al. 2007; Dimitrov 2010). In the present study, the selected bifactor models reached the configural level, which may appear incomplete if strict invariance is treated as the expected endpoint. However, the purpose of invariance testing is to identify the level of measurement equivalence supported by the data, rather than to force the strongest possible level (Robitzsch and Lüdtke 2023). In the present example, the same item–factor architecture (configural) was not rejected across the two age-defined groups, but equality constraints on factor loadings (metric), item intercepts (scalar), and residual variances (strict) were not supported. This differs from conventional statistical comparisons between groups, which usually begin after measurement has already produced the numbers to be compared. These procedures may compare means, variances, or regression coefficients, but they usually leave implicit the assumption that the same underlying condition would generate the same observed value in each group. Psychometric invariance analysis brings this assumption into the model by examining the ruler itself (Dimitrov 2010). In this study, the ruler had the same architecture across groups, but the data did not support the same units, origin, or residual structure. The comparison did not fail after reaching the observed scores; it stopped earlier, at the measurement ruler. One virtue of psychometric analysis is that it locates where group nonequivalence appears in the available data. An analogy with serum creatinine may clarify this distinction.

When a conventional *t* test compares serum creatinine as a proxy for renal function, the analysis begins after the observed values have already been produced. Although the laboratory procedure may correctly measure the circulating concentration of creatinine, its relation to renal function may not be the same across groups because it is affected by muscle mass, age, sex, sarcopenia, amputation, or diet. For example, consider a comparison between young adult men and women, who differ in muscle mass. The same reduction in renal function may not produce the same increase in serum creatinine in the two groups; conversely, the absence of a difference between group means does not necessarily mean that renal function is equivalent. A *t* test may reject or fail to reject the null hypothesis of

BJP PRE-PROOF

equality between means, but it does not evaluate whether the inference from creatinine to renal function uses an equally calibrated ruler across the conditions represented by the groups. In this analogy, the configural level corresponds to adopting serum creatinine as an indicator of renal function in both groups when the same basic measurement architecture is assumed: renal function is linked to serum creatinine, while group-related conditions such as muscle mass, age, sex, and diet may affect how this link operates (of course, this is a didactic analogy: in psychometric models a latent variable is not well identified with only one indicator). The metric level then tests whether the same change in the latent factor corresponds to similar changes in the observed indicator across groups (i.e., similarity of loadings). In the creatinine analogy, this would ask whether the same change in renal function produces the same expected change in serum creatinine across groups. Assuming equality of factor loadings, that is, comparable slopes in the relation between the latent and the observed variables, the scalar level concerns the origin of the scale, represented by the intercepts. In the analogy, even if the relation between renal function and serum creatinine had the same slope in both groups, the expected creatinine value at the same level of renal function could still be shifted upward or downward for reasons not directly attributable to renal function. This shift corresponds to a difference in the baseline expected value of the observed values. In psychometric terms, equivalent changes in motivation may produce similar changes in a given item (metric equivalence), but the same latent level of motivation may correspond to different expected item responses across groups (scalar non-equivalence). The strict level adds equivalence of residuals. In the analogy, this would require that, in addition to the same structure (configuration), relationship (slope), and baseline (intercept), the unexplained part of the observed value be comparable across groups. If renal function were treated as a latent variable in a psychometric model, measured creatinine values could be compared directly as indicators of renal function only if the data supported an equally calibrated measurement ruler across groups. Otherwise, renal function should be compared between groups only to the level at which its observed values showed measurement equivalence. This is a central

BJP PRE-PROOF

contribution of psychometric modeling: while conventional group comparisons, such as a *t* test, assume that the observed scores were generated by the same measurement ruler, measurement invariance analysis makes this assumption explicit and tests whether it is supported by the data.

The same reasoning applies to the EOAQ: in the current data, measurement equivalence stopped at the configural level; therefore, observed or latent mean comparisons were not pursued. Larger samples would clarify whether the configural-only result reflects limited power and estimation instability, an incomplete specification of the candidate models, or a true difference in the measurement rule across groups. With more observations per group, the same model-comparison procedure could provide more stable estimates of factor loadings, intercepts, residual variances, and omega coefficients. It could also determine whether the bifactor solutions retained here remain the most plausible candidates, whether models currently classified as preliminary become admissible at higher invariance levels, or whether specific items, such as Q5 or Q14, show stable evidence of differential functioning.

Although the baseline variables summarized in Table 4 did not show statistically detected differences between the two age-defined groups apart from age itself, possible group heterogeneity is not, in itself, a problem for CFA or MG-CFA. Invariance analysis exists precisely because groups may differ. What MG-CFA tests is whether the same ruler can be used across groups, that is, whether the measurement instrument remains sufficiently similar despite demographic or clinical differences. In the present study, the age-defined groups were not used to estimate the effect of age, menopausal status, or baseline clinical variables on motivation for bariatric surgery. Because the selected bifactor models reached only the configural level, group comparisons of observed or latent motivation means were not pursued. If larger samples do not support stronger invariance levels, an alternative would be to reconsider the organization of EOAQ items and test whether another theoretically interpretable factor configuration provides a more stable measurement rule, at least at the metric invariance level.

BJP PRE-PROOF

This study is not intended to dictate the definitive structure of the EOAQ. Its purpose is to demonstrate a systematic procedure that combines theoretical restrictions with exhaustive testing within a predefined model space. The candidate models were constrained by item content, interpretable domains of motivation, and standard CFA families. Within this theoretically delimited space, all admissible combinations were tested under explicit decision rules. The same procedure can be reused in larger EOAQ samples or adapted to other psychometric instruments in which the latent structure is uncertain.

Our core findings echo international research on obesity-related motivations (Sarwer et al. 2018; Zomer et al. 2016), highlighting personal, social, and healthcare factors. After exploring multiple models, the best fitting models reflected the same structure: bifactor models with secondary factors (P, S, and H) plus a general factor influencing the entire item set. In Brazil, body image holds significant cultural importance, as evidenced by the high prevalence of cosmetic surgery and body dissatisfaction among Brazilian individuals (Rodrigues et al. 2025). Additionally, the psychosocial burden of obesity, including societal pressure and employment discrimination, further reinforces patients' motivations. This sociocultural emphasis on appearance may amplify the social motivation for bariatric surgery, influencing patients' motivations for seeking surgery. Additionally, healthcare accessibility and public health policies impact decision-making, as Brazil's Unified Health System (SUS) provides free bariatric surgery, albeit with long waiting lists (Zomer et al. 2016). Patients often undergo extensive bureaucratic processes before receiving surgical approval, which may help explain why healthcare-related motivations emerged as a key factor in our model. The present study contributes by showing how universal psychological factors may interact with Brazilian-specific cultural and structural influences. Given the limited research on this topic in Brazil, our study has addressed this gap and encourages future investigations into long-term psychosocial outcomes, as well as comparisons with other Latin American populations (Kelles et al. 2015).

Our proposed approach relies on its emphasis on thoroughly assessing a broad range of models, in contrast with traditional frameworks that often overlook

BJP PRE-PROOF

or fail to test alternative structures. Limiting model exploration in covariance structure modeling carries risks, and specification searches should be approached with caution. Evaluating too few models may cause researchers to miss viable configurations. Scholars have argued that rigid adherence to cutoff values for fit indices can be misleading, and that model evaluation should instead involve a contextual examination of multiple models and fit criteria (Marsh et al. 2004).

Rather than testing multiple competing models, some researchers rely on modification indices to adjust a limited number of underperforming models. This practice may result in overfitting and statistically acceptable but theoretically unsupported solutions (MacCallum 1986; MacCallum et al. 1992). Simulation studies have warned against the uncritical use of specification searches without theoretical justification and have shown that systematically evaluating a broader range of models, especially when guided by an explicit analytical strategy, can increase the likelihood of identifying well-fitting structures (Silvia and MacCallum 1988). Therefore, comprehensive assessment of diverse alternative models should become standard practice in validating any new construct or measurement tool.

Researchers face a combinatorial challenge when searching for the best-fitting model, especially when there is uncertainty about which items belong to which factors. Manually exploring dozens of models is time-consuming, which helps explain why only a limited number of models appear in the literature; the sheer volume of results can complicate interpretation. However, as demonstrated here with the EOAQ, using fit indices to rank multiple models and automating the calculation of variations makes it feasible to evaluate numerous factor–item combinations. In this context, model selection depends heavily on the choice of evaluation indices.

We assessed both statistical and practical significance. Statistical plausibility was determined using RMSEA and its associated p-close value. Practical significance was examined through McDonald's ω, which can be interpreted as an effect-size estimator because it represents the proportion of reliable variance attributable to a factor. In the present study, however, ω was used primarily as an admissibility criterion for comparing candidate models: estimates outside the [0,1]

BJP PRE-PROOF

interval were taken as evidence of numerical instability or inadmissible model solutions. Factor loadings were also inspected, with values outside the [–1,1] interval also treated as offending estimates. Because the present study aimed to compare candidate structures constrained by the original two- and three-factor organization of the EOAQ, rather than to dictate a definitive structure for the instrument, ω was not used to classify the retained factors as weak or strong effects.

Potential limitations should be kept in mind when interpreting the results. First, the sample of 137 female patients could limit the generalizability of the findings to male patients or demographically and culturally diverse groups (Dimitrov 2010). Future studies could benefit from including a more diverse population with an expanded sample size. Statistically speaking (Kyriazos 2018), small sample sizes may lead to estimation instability, with recommendation of at least 200 for a single group or 100 per group in multi-group CFA, but it could be solved with Bayesian methods. Although our sample size falls below this general guideline, many CFA models we tested for single- or multi-group analyses successfully converged with acceptable fit indices. Second, there is no “gold standard” instrument of a similar motivation construct available for its original English or translated versions of the questionnaire to compare. Third, testing a combinatorial set of structural models requires implementation in a computational environment; however, the R implementation of the proposed approach is described step by step and is available in the Supplementary Material.

In conclusion, this study presents a flexible framework that emphasizes the value of a systematic approach to structural model comparison in CFA. In this illustrative dataset, two bifactor solutions were identified as plausible candidates for representing both general and domain-specific dimensions of motivation for seeking treatment. The proposed CFA-based framework offers a replicable strategy for comparing complex structural models when the dimensionality of a construct is uncertain. This methodological approach allows the evaluation of multiple competing models and may be adapted to other instruments and research

BJP PRE-PROOF

contexts. The empirical findings regarding the EOAQ should be interpreted as preliminary and require replication in larger and independent samples.

Acknowledgment

No specific funding was associated with this study. The National Council for Scientific and Technological Development (CNPq) sponsors Y.-P.W.

Author contributions (CRediT)

P.S.P.S.: was responsible for conceptualization, methodology, formal analysis, and writing – original draft and review & editing.
B.H.T.: was responsible for conceptualization, methodology, investigation, and writing – review & editing.
S.R.S.: participated in investigation and writing – review & editing.
Y.-P.W.: was responsible for methodology, formal analysis, and writing – original draft and review & editing.
J.O.S.: was responsible for conceptualization, methodology, formal analysis, and writing – original draft and review & editing.
All authors have read and approved the final version submitted and take public responsibility for all aspects of the work.

Disclosure: There are no conflicts of interest concerning the publication of this article.

Edited by: Dr. Marco Scanavino

References:

Bray, G. A., K. K. Kim, J. P. H. Wilding, and World Obesity Federation. 2017. “Obesity: a chronic relapsing progressive disease process. A position statement of the World Obesity Federation.” *Obes Rev* 18 (7): 715–23. https://doi.org/10.1111/obr.12551.

BJP PRE-PROOF

Brown, T. A. 2015. *Confirmatory factor analysis for applied research*. Guilford Press. https://doi.org/10.1521/978.1462515363.

Busetto, L., D. Dicker, C. Azran, et al. 2017. “Practical Recommendations of the Obesity Management Task Force of the European Association for the Study of Obesity for the Post-Bariatric Surgery Medical Management.” *Obes Facts* 10 (6): 597–632. https://doi.org/10.1159/000481825.

Cheung, G. W., and R. B. Rensvold. 2002. “Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance.” *Struct Equ Modeling* 9 (2): 233–55. https://doi.org/10.1207/S15328007SEM0902_5.

Cizek, G. J., H. K. Koons, and S. L. Rosenberg. 2011. “Finding validity evidence: An analysis using the Mental Measurements Yearbook.” In *High-stakes testing in education: Science and practice in K–12 settings*. American Psychological Association. https://doi.org/10.1037/12330-008.

Curran, P. J., K. A. Bollen, F. Chen, P. Paxton, and J. B. Kirby. 2003. “Finite Sampling Properties of the Point Estimates and Confidence Intervals of the RMSEA.” *Sociol Methods Res* 32 (2): 208–52. https://doi.org/10.1177/0049124103256130.

Dawes, A. J., M. Maggard-Gibbons, A. R. Maher, et al. 2016. “Mental Health Conditions Among Patients Seeking and Undergoing Bariatric Surgery: A Meta-analysis.” *JAMA* 315 (2): 150–63. https://doi.org/10.1001/jama.2015.18118.

Dimitrov, D. M. 2010. “Testing for factorial invariance in the context of construct Validation.” *Meas Eval Couns Dev* 43 (2): 121–49. https://doi.org/10.1177/0748175610373459.

BJP PRE-PROOF

Fischer, L., A. Deckert, M. K. Diener, J. B. Zimmermann, M. W. Büchler, and C. M. Seiler. 2011. “Ranking of patient and surgeons’ perspectives for endpoints in randomized controlled trials–lessons learned from the POVATI trial [ISRCTN 60734227].” *Langenbecks Arch Surg* 396 (7): 1061–66. https://doi.org/10.1007/s00423-011-0798-3.

Fischer, L., F. Nickel, J. Sander, et al. 2014. “Patient expectations of bariatric surgery are gender specific–a prospective, multicenter cohort study.” *Surg Obes Relat Dis* 10 (3): 516–23. https://doi.org/10.1016/j.soard.2014.02.040.

Franzen, M. D. 2000. *Reliability and validity in neuropsychological assessment*. Springer. https://doi.org/10.1007/978-1-4757-3224-5.

Gaudrat, B., S. Andrieux, V. Florent, and A. Rousseau. 2021. “Psychological characteristics of patients seeking bariatric treatment versus those seeking medical treatment for obesity: is bariatric surgery a last best hope?” *Eat Weight Disord* 26 (3): 949–61. https://doi.org/10.1007/s40519-020-00934-3.

Geldhof, G. J., K. J. Preacher, and M. J. Zyphur. 2014. “Reliability estimation in a multilevel confirmatory factor analysis framework.” *Psychol Methods* 19 (1): 72–91. https://doi.org/10.1037/a0032138.

Hammer, J. H., D. L. Vogel, P. R. Grzanka, et al. 2024. “The integrated behavioral model of mental health help seeking (IBM-HS): A health services utilization theory of planned behavior for accessing care.” *J Couns Psychol* 71 (5): 315–27. https://doi.org/10.1037/cou0000754.

Hill, D., M. Conner, F. Clancy, et al. 2022. “Stress and eating behaviours in healthy adults: a systematic review and meta-analysis.” *Health Psychol Rev* 16 (2): 280–304. https://doi.org/10.1080/17437199.2021.1923406.

BJP PRE-PROOF

Holzinger, K. J., and F. Swineford. 1937. “The Bi-factor method.” *Psychometrika* 2 (1): 41–54. https://doi.org/10.1007/BF02287965.

Huang, P. H. 2017. “Asymptotics of AIC, BIC, and RMSEA for Model Selection in Structural Equation Modeling.” *Psychometrika* 82 (2): 407–26. https://doi.org/10.1007/s11336-017-9572-y.

Jensen, A. R. 1998. *The g factor: The science of mental ability*. Praeger. https://doi.org/10.5860/choice.36-2443.

Jorgensen, T. D., S. Pornprasertmanit, A. M. Schoemann, and Y. Rosseel. 2022. *semTools: Useful tools for structural equation modeling*. https://CRAN.R-project.org/package=semTools.

Kelles, S. M., F. Diniz Mde, C. J. Machado, and S. M. Barreto. 2015. “Perfil de pacientes submetidos à cirurgia bariátrica, assistidos pelo Sistema Único de Saúde do Brasil: revisão sistemática [The profile of patients undergoing bariatric surgery in the Brazilian Unified National Health System: a systematic review].” *Cad Saude Publica* 31 (8): 1587–601. https://doi.org/10.1590/0102-311X00022714.

Kyriazos, T. A. 2018. “Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General.” *Psychology* 09 (08): 2207–30. https://doi.org/10.4236/psych.2018.98126.

Laferton, J. A. C., T. Kube, S. Salzmann, C. J. Auer, and M. C. Shedden-Mora. 2017. “Patients’ expectations regarding medical treatment: A critical review of concepts and their assessment.” *Front Psychol* 8: 233. https://doi.org/10.3389/fpsyg.2017.00233.

BJP PRE-PROOF

Laferton, J. A. C., W. Rief, and M. Shedden-Mora. 2025. “Improving Patients’ Treatment Expectations.” *JAMA* 334 (2): 171–72. https://doi.org/10.1001/jama.2025.6261.

Lin, L. C., P. H. Huang, and L. J. Weng. 2017. “Selecting path models in SEM: A comparison of model selection criteria.” *Struct Equ Modeling* 24 (6): 855–69. https://doi.org/10.1080/10705511.2017.1363652.

MacCallum, R. C. 1986. “Specification searches in covariance structure modeling.” *Psychol Bull* 100 (1): 107–20. https://doi.org/10.1037/0033-2909.100.1.107.

MacCallum, R. C., M. W. Browne, and H. M. Sugawara. 1996. “Power analysis and determination of sample size for covariance structure modeling.” *Psychol Methods* 1 (2): 130–49. https://doi.org/10.1037/1082-989X.1.2.130.

MacCallum, Robert C., Mary Roznowski, and Lawrence B. Necowitz. 1992. “Model Modifications in Covariance Structure Analysis: The Problem of Capitalization on Chance.” *Psychological Bulletin* 111 (3): 490–504. https://doi.org/10.1037/0033-2909.111.3.490.

Markus, K. A., and D. Borsboom. 2013. *Frontiers of test validity theory: Measurement, causation, and meaning*. Routledge. https://doi.org/10.4324/9780203501207.

Marsh, H. W., K. T. Hau, and Z. Wen. 2004. “In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler’s (1999) findings.” *Struct Equ Modeling* 11 (3): 320–41. https://doi.org/10.1207/s15328007sem1103_2.

Meredith, William. 1993. “Measurement invariance, factor analysis and factorial invariance.” *Psychometrika* 58 (4): 525–43. https://doi.org/10.1007/BF02294825.

BJP PRE-PROOF

Morgan, D. J. R., K. M. Ho, and C. Platell. 2020. “Incidence and Determinants of Mental Health Service Use After Bariatric Surgery.” *JAMA Psychiatry* 77 (1): 60–67. https://doi.org/10.1001/jamapsychiatry.2019.2741.

Pipek, L. Z., W. A. F. Moraes, R. M. Nobetani, et al. 2024. “Surgery is associated with better long-term outcomes than pharmacological treatment for obesity: a systematic review and meta-analysis.” *Sci Rep* 14 (1): 9521. https://doi.org/10.1038/s41598-024-57724-5.

Pompei, L. M., R. Bonassi-Machado, M. L. Steiner, et al. 2022. “Profile of Brazilian climacteric women: results from the Brazilian Menopause Study.” *Climacteric* 25 (5): 523–29. https://doi.org/10.1080/13697137.2022.2088276.

Putnick, Diane L., and Marc H. Bornstein. 2016. “Measurement invariance conventions and reporting: The state of the art and future directions for psychological research.” *Developmental Review* 41: 71–90. https://doi.org/10.1016/j.dr.2016.06.004.

R Core Team. 2022. *R: A language and environment for statistical computing*. https://www.R-project.org/.

Rief, W., M. F. J. Sperl, K. Braun-Koch, et al. 2022. “Using expectation violation models to improve the outcome of psychological treatments.” *Clin Psychol Rev* 98: 102212. https://doi.org/10.1016/j.cpr.2022.102212.

Robitzsch, Alexander, and Oliver Lüdtke. 2023. “Why full, partial, or approximate measurement invariance are not a prerequisite for meaningful and valid group comparisons.” *Structural Equation Modeling: A Multidisciplinary Journal* 30 (6): 859–70. https://doi.org/10.1080/10705511.2023.2191292.

BJP PRE-PROOF

Rodrigues, R. C. B., I. S. O. Pereira, G. R. Bahia, et al. 2025. “Association Between the Predominant Domain of Eating Behavior and Perception of Distortion and Satisfaction with Body Image in People Who Underwent Bariatric Surgery in Brazil.” *Nutrients* 17 (5): 850. https://doi.org/10.3390/nu17050850.

Rosseel, Y. 2012. “lavaan: An R Package for Structural Equation Modeling.” *J Stat Software* 48 (2): 1–36. https://doi.org/10.18637/jss.v048.i02.

Sarwer, D. B., T. A. Wadden, J. C. Spitzer, et al. 2018. “4-Year Changes in Sex Hormones, Sexual Functioning, and Psychosocial Status in Women Who Underwent Bariatric Surgery.” *Obes Surg* 28 (4): 892–99. https://doi.org/10.1007/s11695-017-3025-7.

Silva, S. R. da, Y. P. W, A. C. B. Dantas, et al. 2024. “Translation and Validation of the Brazilian Version of the European Obesity Academy Questionnaire on Patients’ Motivations for Seeking Metabolic and Bariatric Surgery.” *Obes Surg* 34 (10): 3840–47. https://doi.org/10.1007/s11695-024-07460-6.

Silvia, E. S., and R. C. MacCallum. 1988. “Some Factors Affecting the Success of Specification Searches in Covariance Structure Modeling.” *Multivariate Behav Res* 23 (3): 297–326. https://doi.org/10.1207/s15327906mbr2303_2.

Stevens, S. S. 1946. “On the theory of scales of measurement.” *Science* 103 (2684): 677–80. https://doi.org/10.1126/science.103.2684.677.

Wicherts, J. M., and C. V. Dolan. 2010. “Measurement Invariance in Confirmatory Factor Analysis: An Illustration Using IQ Test Performance of Minorities.” *Educ Meas* 29 (3): 39–47. https://doi.org/10.1111/j.1745-3992.2010.00182.x.

World Health Organization. 2021. *Obesity and overweight*. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

BJP PRE-PROOF

World Obesity Federation. 2022. *Prevalence of obesity*. https://www.worldobesity.org/about/about-obesity/prevalence-of-obesity.

Wu, Amery D., Zhen Li, and Bruno D. Zumbo. 2007. “Decoding the Meaning of Factorial Invariance and Updating the Practice of Multi-Group Confirmatory Factor Analysis: A Demonstration with TIMSS Data.” *Practical Assessment, Research & Evaluation* 12 (3): 1–26. http://pareonline.net/getvn.asp?v=12&n=3.

Zomer, E., K. Gurusamy, R. Leach, et al. 2016. “Interventions that cause weight loss and the impact on cardiovascular risk factors: a systematic review and meta-analysis.” *Obes Rev* 17 (10): 1001–11. https://doi.org/10.1111/obr.12433.

BJP PRE-PROOF