HEALTH PSYCHOLOGY REPORT · 2026
ORIGINAL ARTICLE

Patrycja Uram 1 · A,B,C,D,E,F,G
Robert Balas 1 · E,F
Anna Kwiatkowska 1 · C,D
Joanna Wąsowicz 1 · A,F
Sebastian B. Skalski-Bednarz 2,3 · C,D

The Polish version of the Body, Eating, and Exercise Comparison Orientation Measure: adaptation, validation, and psychometric evaluation

BACKGROUND
The study aimed to adapt the Body, Eating, and Exercise Comparison Orientation Measure (BEECOM) for the Polish population and assess its psychometric properties, including factor structure, reliability, validity, and temporal stability. Social comparison processes, particularly in the domains of body image, eating, and exercise, are increasingly recognized as important factors influencing psychological well-being and the development of disordered eating behaviors.

PARTICIPANTS AND PROCEDURE
Study 1 involved 408 participants (M_{age} = 26.26, *SD* = 9.60; 49.9% female) and focused on scale adaptation through exploratory and confirmatory factor analyses. Study 2 (*N* = 70; M_{age} = 41.23, *SD* = 13.21) and Study 3 (*N* = 129; M_{age} = 20.68, *SD* = 2.77) assessed the temporal stability of the 18-, 12-, and 9-item BEECOM versions across three one-month intervals.

RESULTS
A revised factor structure led to the Polish BEECOM-R, which demonstrated good internal consistency, convergent validity (with PACS and SATAQ-3), and measurement invariance across gender and age groups. Temporal stability analyses showed satisfactory to strong consistency (*r* = .48-.87) and minimal mean-level change, though some variability appeared in the Body and Exercise subscales. The 18- and 9-item versions outperformed the 12-item version in terms of stability.

CONCLUSIONS
The Polish BEECOM-R is a valid and reliable tool for measuring appearance-related social comparisons and can be effectively used in both cross-sectional and longitudinal research.

KEY WORDS
transcultural adaptation; psychometrics; body comparison; eating comparison; exercise comparison

ORGANIZATION – 1: Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland · 2. Faculty of Philosophy and Education, Catholic University of Eichstätt-Ingolstadt, Eichstätt, Germany · 3. Institute of Psychology, Ignatianum University in Cracow, Cracow, Poland

AUTHORS' CONTRIBUTIONS – A: Study design · B: Data collection · C: Statistical analysis · D: Data interpretation · E: Manuscript preparation · F: Literature search · G: Funds collection

CORRESPONDING AUTHOR – Patrycja Uram, Institute of Psychology, Polish Academy of Sciences, 1 Jaracza Str., 00-378 Warsaw, Poland, e-mail: patrycja.uram@sd.psych.pan.pl

TO CITE THIS ARTICLE – Uram, P., Balas, R., Kwiatkowska, A., Wąsowicz, J., & Skalski-Bednarz, S. B. (2026). The Polish version of the Body, Eating, and Exercise Comparison Orientation Measure: adaptation, validation, and psychometric evaluation. *Health Psychology Report*. https://doi.org/10.5114/hpr/214692

RECEIVED 31.07.2025 · REVIEWED 15.11.2025 · ACCEPTED 25.11.2025 · ONLINE PUBLICATION 11.02.2026

BACKGROUND

Festinger's (1954) theory of social comparison suggests that individuals are inclined to gauge themselves against others, motivated by a need to evaluate their value and assess their thoughts, skills, or abilities. Initially, these comparisons are typically made using personally established benchmarks. However, when these benchmarks are lacking or underdeveloped, individuals seek patterns within their immediate surroundings. The theory proposes that individuals prefer role models within their environment whose qualities are akin to their skills and are subjectively similar in other aspects rather than those markedly different (Lin & Kulik, 2002). The social comparison theory posits that we typically compare ourselves to individuals perceived as similar. However, in the domain of physical appearance, it is also posited that individuals compare themselves to those considered significantly more attractive. This distinction gives rise to upward and downward comparisons. Upward comparisons of appearance are considered consistent predictors of eating disorders and body dissatisfaction (compared to someone perceived to be doing better) and downward comparisons (compared to someone perceived to be doing worse). For example, the study by Rancourt et al. (2016) found that upward comparisons were associated with higher levels of eating disorders and lower body satisfaction for women of all racial/ethnic groups. Downward comparisons, on the other hand, were found to be harmful, primarily among Hispanic and Latino women, but protective of Asian and white women. Although social comparison theory has proven helpful in comparing opinions and abilities, it is now gaining importance because of comparisons of personal characteristics, including physical appearance (Schaefer & Thompson, 2014).

Historical research trends indicate that women use social comparisons more frequently than men (Fatt et al., 2019). Moreover, the most common focus of comparison lies in the evaluation of one's own body and appearance (Fardouly et al., 2015). Research on social comparison among men is more limited, and the results are less consistent than for women. Previous reports, however, indicate associations between men's propensity to compare appearance and self-esteem, muscular aspirations, or mandatory exercise (Cash & Smolak, 2011; McCreary & Saucier, 2009). Studies emphasize that the social comparison process, spontaneously and subconsciously, is triggered by exposure to body images in various media formats, impacting both men and women (Fatt et al., 2019; Peng et al., 2019). Previously, this process mainly involved body images displayed in newspapers or on television (Tiggemann, 2003); however, contemporary research focuses primarily on body images displayed on social media platforms, often digitally altered using graphic design software (McLean et al., 2015). The growing interest in this area has prompted researchers to create tools to accurately and reliably measure the propensity to make comparisons (see Appendix 1 in Supplementary materials).

BEECOM SCALE

Currently, researchers are constructing scales that focus on more specific aspects of social comparison. This is dictated, among other things, by including comparison studies of the impact of social media and the content presented there. Another tool that allows for quantitative measurement in terms of social comparison, additionally allowing for the assessment of associations of social comparison with a bias towards eating disorders, body satisfaction, and exercise-related comparisons, is the Body, Eating, and Exercise Comparison Orientation Measure (BEECOM; Fitzsimmons-Craft et al., 2012). The scale is constructed from 18 items that fall into three subscales. Each subscale is made up of six items. The first subscale deals with aspects of the body, the second with food, and the third with exercise.

The tool shows very good psychometric properties. Internal consistency is between .93 and .96, relevance with eating disorders was .96, and body dissatisfaction was between .61 and .75. Test-retest reliability among US female students was between .85 and .89 (Fitzsimmons-Craft et al., 2012).

Scale scores were further characterized by temporal stability and annual predictive validity for body dissatisfaction and eating disorders (Fitzsimmons-Craft & Bardone-Cone, 2014). Thus, the tool is extremely interesting from the point of view of research development and is characterized in its original version by good indicators. At the same time, the division into three comparison categories seems very important for yet another reason. There is a growing trend in various social networks to publish photos of one's meals and photos from gyms showing people before, during, or after physical activity. The authors intend to publish photos of eating and exercising to motivate others to engage in similar behavior (Vaterlaus et al., 2015). However, it can also strongly influence the trend toward comparisons, precisely in areas measured with BEECOM.

The comprehensiveness of BEECOM has meant that the tool has already seen several local adaptations: Iranian (Sahlan et al., 2020), Brazilian (Hudson et al., 2023), and Spanish (Paterna et al., 2023). BEECOM studies have been completed on clinical samples – people with eating disorders (Saunders et al., 2019) – and non-clinical samples (Sahlan et al., 2020; Paterna et al., 2023). Moreover, there are also differences in terms of sample selection. There are validation studies in which the samples consist only of women (Saunders et al., 2019; Hudson et al., 2023),

while there are studies in which the sample consists of both women and men (Sahlan et al., 2020; Paterna et al., 2023). The version proposed by Sahlan et al. (2020) carried out on a non-clinical sample of Iranian university students – women (Body: 2, 4, 9, 12, 13; Eating: 1, 7, 8, 11; Exercise: 5, 6, 10, 14, 15, 16, 18) and men (Body: 2, 4, 9, 12, 13, 17; Eating: 1, 3; Exercise: 5, 6, 10, 14, 15, 18) – consists of 15 items. In the Spanish version, the authors (Paterna et al., 2023) compared two versions of the scale: an extended version built with 18 items (BEECOM-L) and a short version built with nine items (BEECOM-S), which was a proposal of BEECOM-R by Saunders et al. (2019). Their analysis showed that the nine-item version is more adequate (Body: 4, 9, 12; Eating: 3, 7, 11; Exercise: 6, 14, 15). To date, validation studies suggest some inconsistency in the factor structure of BEECOM, which, as indicated by studies and the authors of local papers themselves (Sahlan et al., 2020; Paterna et al., 2023), may be related to socio-cultural differences.

BEECOM TEMPORAL STABILITY

The temporal stability of the original BEECOM was assessed by Fitzsimmons-Craft and Bardone-Cone (2014), who administered it twice over one year to female college students. They reported test-retest correlations and used regression analyses to predict Time 2 scores from Time 1 scores, finding high temporal stability for the overall BEECOM score and its subscales. This suggests that comparison tendencies are relatively stable traits. However, most subsequent BEECOM adaptation studies (e.g., Paterna et al., 2023; Sahlan et al., 2020; Saunders et al., 2019) have focused on other psychometric properties such as factor structure and internal consistency, rather than temporal stability. This omission limits a comprehensive understanding of the BEECOM's utility in longitudinal and clinical research. Since the initial assessment of the BEECOM in the Polish context did not include an analysis of temporal stability, focusing primarily on validity and reliability, the question remains open as to whether this version of the tool demonstrates equally satisfactory temporal stability, highlighting the need for further research.

EXISTING SOCIAL COMPARISON MEASURES

The increasing interest in comparisons and one's appearance has resulted in researchers developing measurement scales to measure the relationships obtained in research. Indeed, one of the most common scales used to measure general social comparisons is the Iowa-Netherlands Comparison Orientation Measure (INCOM), validated by Gibbons and Buunk (1999). However, considering this research and the specific tendencies observed in social comparisons focusing on appearance and assessing satisfaction with one's body, researchers have also started constructing scales dedicated to the specific phenomena being measured. One of these, for example, is the Physical Appearance Comparison Scale (PACS), a very brief (5-item) measure that assesses the degree to which individuals tend to compare their appearance with others. The PACS scale is used to assess the extent to which an individual compares their appearance with others (Thompson et al., 1991). It is worth mentioning that the PACS has seen several updates over the past decade or so, on the one hand resulting from numerous studies on body image and comparison bias and the other related to the psychometric difficulties of the original version of the PACS (Schaefer & Thompson, 2014, 2018).

Although Thompson et al. (1991) reported adequate scale reliability for the original version of the PACS, some reports have indicated marginal or poor reliability (Vander Wal & Thelen, 2000). Moreover, authors of PACS studies indicate that a single item with an inverted score in a 5-item measure often undermines the scale's reliability and should be removed to achieve adequate internal consistency (Davison & McCabe, 2005; Keery et al., 2004). Because the tendency to compare can be up/down, another tool used is the Upward and Downward Physical Appearance Comparisons Scale, created by O'Brien et al. (2009). The scale is often used to assess how an individual perceives their body appearance regarding being overweight or underweight. We can also distinguish between the Body Image Comparison Scale (BICS), which is concerned with comparing body image in social situations (Faith et al., 1997). A similar scale is the Body Comparison Scale (BCS). The scale evaluates how a person assesses different parts of their body against the same parts in other people (Thompson et al., 1999).

PRESENT STUDY

There is a lack of comprehensive, validated tools in Poland to measure social comparisons, aside from the PACS-PL (Dzielska et al., 2017). Existing tools generally assess only the overall tendency to compare, not specific domains. This article presents three studies. Study 1 aimed to adapt the BEECOM scale to Polish and assess its psychometric properties (exploratory factor analysis, confirmatory factor analysis, reliability, validity, invariance). Studies 2 and 3 examined the temporal stability of the 18-, 12-, and 9-item Polish BEECOM versions over three months. Study 2 used Pearson correlations and paired-samples *t*-tests to assess relative and absolute stability. Study 3 applied latent growth curve modeling (LGCM) to analyze average and individual changes over time, accounting for measurement error.

PARTICIPANTS AND PROCEDURE

PARTICIPANTS

Study 1

In the study, there were N = 408 participants aged 18 to 70 (M = 26.26, SD = 9.60), with 49.9% being female (n = 203). Details are shown in Appendix 2 in Supplementary materials.

Study 2

A total of 70 Polish adults were recruited for Study 2 (M_{age} = 20.34, SD = 2.74, age range = 18-37) and completed all three waves of data collection (March, April, and May 2024). Among the retained participants, 61% were female and 39% were male. Details are shown in Appendix 2 in Supplementary materials.

Study 3

A total of 129 Polish adults (M_{age} = 20.68, SD = 2.77, age range = 18-65) participated in a three-wave longitudinal second study (March, April, and May 2025). Among the retained participants, 86.8% were female and 10.1% were male. Details are shown in Appendix 2 in Supplementary materials.

PROCEDURES

The first study began with the procedure for translating the original BEECOM version (see Appendices 3 and 4 in Supplementary materials). In the first step, two independent translators translated the method into Polish (the name of the questionnaire, the 18 items of the questionnaire, the names of the individual subscales, and the response scale). In the second step, psychologists fluent in English also completed the translation. In the third step, the translated versions were compared, establishing a common version. In the fourth step, the common version was back-translated and compared with the original BEECOM version (Hornowska & Paluchowski, 2004). The translation turned out to be very similar to the content of the original version. The final translated version was included in a Google Form designed for this study, which also contained the measurement scales and demographic questionnaire described below.

Participants for Studies 1, 2, and 3 were recruited through social media. The invitation to participate in the survey was sent out via social media using the snowball method. This recruitment strategy allowed for the inclusion of participants from different age groups, educational backgrounds, and regions of Poland, although participation was based on self-selection. The first study was conducted from June 2021 to July 2022. The second study was conducted from March 2024 to May 2024. The third study was conducted from March 2025 to May 2025. The studies were anonymous and voluntary, did not involve any financial gratification, and participants were informed of their right to withdraw at any time. The research procedure was positively reviewed by the Ethics Committee of the Polish Academy of Sciences, Institute of Psychology.

MATERIALS

Study 1

The following measurement scales were used in the study:

Body, Eating, and Exercise Comparison Orientation Measure (BEECOM; Fitzsimmons-Craft et al., 2012), consisting of 18 items measuring social comparison tendencies across three dimensions: Body (e.g., "I pay attention to whether or not I am as thin as, or thinner than, my peers"), Eating (e.g., "I look at the amount of food my peers leave on their plate in comparison to me"), and Exercise (e.g., "I pay attention to the length of time others exercise"). Responses are rated on a 7-point scale from 1 (*never*) to 7 (*always*).

Physical Appearance Comparison Scale (PACS-PL; Thompson et al., 1991; Polish adaptation by Dzielska et al., 2017), assessing general appearance-based comparison tendencies through 5 items (e.g., "At social events, I compare how I am dressed to how others are dressed"), using a 5-point scale from 0 (*never*) to 4 (*always*). PACS-PL reliability: α = .62; PACS-R-PL (recommended version): α = .90.

Sociocultural Attitudes Toward Appearance Questionnaire-3 (SATAQ-3; Thompson et al., 2004; Polish adaptation by Izydorczyk & Lizińczyk, 2020), measuring the influence of media on appearance-related attitudes and behaviors. The 28-item scale includes four subscales: Internalization Pressure (α = .94), Internalization-Information Seeking (α = .76), Internalization-Athlete (α = .84), and Information (α = .89).

The demographic information was gathered using a questionnaire. It contained questions about age, gender, place of residency, education, marital status, and occupation.

Studies 2 and 3

The Polish version of the BEECOM from the first study was used. Participants completed the full 18-item BEECOM-PL. For subsequent analyses, scores were also calculated for the 12-item (BEECOM-R-12) and 9-item (BEECOM-R-9) versions as defined in previously presented study.

The demographic information was gathered using a questionnaire. It contained questions about age, gender, education, marital status, and occupation.

STATISTICAL ANALYSES

All analyses were conducted using IBM SPSS Statistics (version 28.0), AMOS (version 28.0), and JASP (version 0.19.1). Statistical significance was set at $\alpha = .05$. Before the first study began, exploratory factor analysis (EFA) was conducted as a pilot study to verify the original structure of the scale. For details, see Supplementary materials (Appendices 3 and 4).

Study 1 (N = 408)

A confirmatory factor analysis (CFA) was performed using maximum likelihood estimation. The model (three factors, 18 items) was tested iteratively. Items with cross-loadings were removed to preserve discriminant validity (Byrne, 2016). Model fit was evaluated using the following indices: χ^2/df (< 2), root mean square error of approximation (RMSEA; < .08), root mean square residual (RMR; < .08), goodness-of-fit index (GFI) and adjusted GFI (AGFI; > .90), parsimony goodness-of-fit index (PGFI; > .50), incremental fit index (IFI), Tucker-Lewis index (TLI), and confirmatory fit index (CFI), all with acceptable values > .90 (Brown, 2015; Hu & Bentler, 1999; Jöreskog & Sörbom, 1993; Schermelleh-Engel et al., 2003). Additionally, model fit was assessed using the Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Vandenberg & Grelle, 2009).Convergent validity was examined via correlations with the Perceived Appearance Comparison Scale (PACS) and the Sociocultural Attitudes Toward Appearance Questionnaire-3 (SATAQ-3). Reliability was assessed using McDonald's omega (ω). Pearson's correlation coefficients were used to examine associations among variables and with demographic data.

Study 2 (N = 70)

Relative stability was assessed using Pearson's correlations between scores at Times 1 (T1), 2 (T2), and 3 (T3) for the total score and subscales (18-, 12-, and 9-item versions).

Mean-level stability was evaluated with paired-samples *t* tests for T1–T2, T2–T3, and T1–T3 comparisons. If the assumption of normality was violated (based on Shapiro-Wilk tests), Wilcoxon signed-rank tests were used. Means and standard deviations were reported (Lodder et al., 2022).

Study 3 (N = 129)

Latent growth curve modeling (LGCM) was used to examine change trajectories in BEECOM scores (total and subscales) across three time points. Models included latent intercept and slope factors (Hertzog et al., 2006). Model fit was evaluated using χ^2, CFI (acceptable ≥ .90, good ≥ .95), RMSEA (acceptable ≤ .08, good ≤ .05), and standardized root mean square residual (SRMR; ≤ .08) (Hu & Bentler, 1999; Byrne & Crombie, 2003). A nonsignificant mean slope indicated overall stability; a nonsignificant slope variance suggested homogeneity of change (Hertzog et al., 2008).

RESULTS

PILOT STUDY

The Polish version of the BEECOM

Exploratory factor analysis

The verification of the EFA results conducted allowed for the evaluation of items according to the established inclusion and exclusion criteria, thus including items with low commonality, items cross-loading more than one factor, and items with loadings below .40. Details are shown in Appendix 3 in Supplementary materials.

STUDY 1

Confirmatory factor analysis

Hypothesized model – Three factors/18 items
CFA was used to test the original 18-item, three-factor structure of the BEECOM scale. Although the chi-square test was significant ($\chi^2(131, n = 408) = 383.12$, $p < .001$), this result is common with larger samples and not a sole indicator of poor model fit. Other fit indices (RMSEA, CFI, TLI) also fell below acceptable thresholds, suggesting that the model did not fit well. Based on these results, previous research, and factor loadings, a shortened version of the scale was developed by removing items that weakened the model, aiming to improve its psychometric properties.

Respecified Model 2 – Three factors/12 items
The first step was to create a model that would allow, after verification, its acceptance. Re-analyzing the individual statements allowed the following statements to be removed from the initial model (BEECOM-18): for the subscale Body: 2, 17, for the subscale Eating: 1, 11, and for the subscale Exercise: 5, 10. The resulting model consisted of 12 assertions. Its fit indices are shown in Table 1. In this case, the χ^2 was significant, but the χ^2/df ratio was quite good (below 3). The RMSEA values were more acceptable (below .08). The remaining fit indices were also satisfactory, providing evidence for model acceptance (Table 1).

Respecified Model 3 – Three factors/9 items
However, using the proposal of Saunders et al. (2019) and Paterna et al. (2023), who created a shortened version of the scale (BEECOM-R-9) in analysis, it was

Table 1

Model adequacy and goodness of fit indices of BEECOM-18, -12, and -9 item models (N = 204), and results of multigroup invariance analysis across age and sex groups (N = 408)

Model	χ^2	*df*	*p*	χ^2/df	RMSEA	LO	HI	PCLOSE	GFI	AGFI	CFI	PGFI	RMR	IFI	TLI	AIC	BIC
BEECOM-18 3 factor	383.12	131	< .001	2.93	.097	.086	.109	.000	.827	.774	.924	.633	.211	.924	.911	465.12	642.16
BEECOM-12 3 factor	99.29	51	< .001	1.95	.068	.048	.088	.067	.928	.890	.978	.607	.167	.978	.971	159.29	258.83
BEECOM-9 3 factor	26.45	24	.331	1.10	.022	.000	.063	.841	.973	.949	.998	.519	.081	.998	.997	74.45	154.09
Results of multigroup invariance analysis across sex groups																	
Model	CMIN/*df*	RMSEA [90% CI]	Δ RMSEA	CFI	Δ CFI												
Model 1. Unconstrained	2.38	.059 [.053, .066]		.944													
Model 2. Configural invariance	2.33	.058 [.052, .064]	.001	.943	.001												
Model 3. Metric invariance	2.31	.058 [.052, .064]	.000	.940	.003												
Model 4. Scalar invariance	2.30	.057 [.052, .063]	.001	.939	.001												
Results of multigroup invariance analysis across age groups																	
Model 1. Unconstrained	2.67	.065 [.059, .071]		.935													
Model 2. Configural invariance	2.59	.064 [.057, .070]	.001	.935	.000												
Model 3. Metric invariance	2.52	.062 [.056, .068]	.002	.933	.002												
Model 4. Scalar invariance	2.48	.061 [.056, .067]	.001	.933	.000												

Note. BEECOM – Body, Eating, and Exercise Comparison Orientation Measure; RMSEA – root mean square error of approximation; LO – low RMSEA; HI – high RMSEA; PCLOSE – *p* of close fit; GFI – goodness of fit index; AGFI – adjusted goodness of fit index; CFI – comparative fit index; PGFI – parsimony goodness of fit index; RMR – root mean square residual; TLI – Tucker-Lewis index; IFI – incremental fit index; AIC – Akaike information criterion; BIC – Bayesian information criterion. The re-specifications of models were based on error covariance modification indices.

decided to test such a model in this case as well (Figure 1). In this case, items 2, 13, and 17 were removed for the Body subscale, and items 3, 11, and 8 were removed for the Eating subscale. Items 5, 10, and 18 were removed for the Exercise subscale. BEECOM-R-9 allowed full reproduction of the proposed shortened version of the original scale for two subscales: Body (4, 9, 12) and Exercise (6, 14, 15). For the subscale Eating, the following statements were left in place: 1, 7, 16. In this case, the χ^2 was not significant. RMSEA and other fit indices were acceptable, providing evidence for model acceptance (Table 1).

Based on previous reports, all versions of BEECOM-18 (18 items), BEECOM-R-12 (12 items), and BEECOM-R-9 (9 items) were included for further analysis of the psychometric properties of BEECOM.

Invariance testing for sex and age

To test for the equivalence of item measurements and theoretical factorial structure of the BEECOM across sex and age groups, we performed multiple-group analysis with AMOS v.28. We transformed a continuous age variable into a categorical age variable with two values (1 – young, 2 – old), by choosing age = 30 years as the cutpoint. Thus, we established two age groups: the young group: $N = 321$; $M = 22.60$, $SD = 2.88$; and the old group: $N = 74$; $M = 42.27$, $SD = 11.67$. The results of the tested models' comparisons are presented in Table 1. In both analyses, the differences between RMSEA values met the cut-off criterion of .015, as well as differences between CFI values (Δ CFI < .01). Thus, equivalence of the BEECOM across sex and age groups was confirmed.

Internal consistency reliability and convergent evidence of validity

Next, the reliability of the scale (Table 2) and the convergent validity were evaluated, similarly to the previous evaluations for the 18-item, 12-item, and 9-item versions. Convergent validity was estimated by assessing the correlation coefficients between the BEECOM and PACS scales. The results showed that BEECOM-18, BEECOM-R-12, and BEECOM-R-9 positively correlated equally with PACS, PACS-R, and SATAQ-3 subscales. For the demographic variables analyzed, such as place of residence, education, and marital status or occupation, correlation analysis showed that only occupation was positively related to the exercise subscale. Negative correlations were observed between gender and age and the tendency to make comparisons. The body subscale was related to age, while gender was related to the body subscale and eating subscale. The correlation coefficients are shown in Table 3.

Figure 1

Structure of the Polish version of BEECOM-R-9

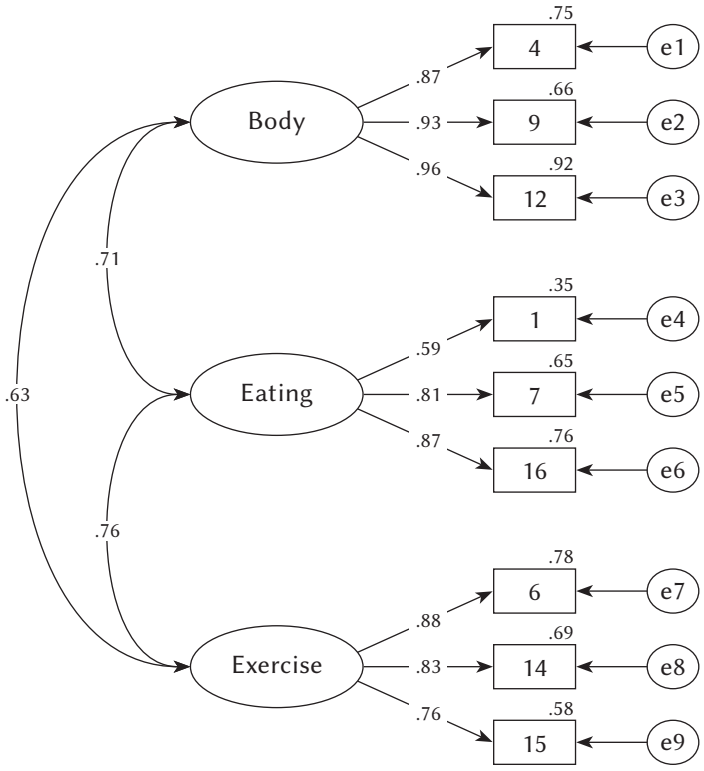


Table 2

Reliability coefficients and descriptive statistics of results for scales and individual subscales of BEECOM-Total, BEECOM-R-12 and BEECOM-R-9 (N = 408)

Factors	Number of questions	Cronbach's α	McDonald ω	Guttman	*M*	*SD*	Min	Max
BEECOM-Total-18	18	.96	.96	.96	57.76	25.44	18	126
BEECOM-Body-6	6	.92	.95	.95	23.50	10.69	6	42
BEECOM-Eating-6	6	.95	.90	.90	19.44	9.17	6	42
BEECOM-Exercise-6	6	.90	.92	.92	14.82	8.73	6	42
BEECOM-R-12	12	.94	.94	.94	33.32	17.24	12	84
Body-R-4	4	.94	.94	.94	16.11	7.38	4	28
Eating-R-4	4	.88	.88	.88	13.50	6.39	4	28
Exercise-R-4	4	.91	.91	.91	9.71	6.01	4	28
BEECOM-R-9	9	.91	.90	.91	29.39	12.60	9	63
Body-R-3	3	.93	.94	.94	12.44	5.54	3	21
Eating-R-3	3	.77	.78	.78	9.66	4.67	3	21
Exercise-R-3	3	.87	.87	.87	7.29	4.49	3	21

Note. BEECOM – Body, Eating, and Exercise Comparison Orientation Measure.

STUDY 2 (*N* = 70)

Test-retest correlations (Pearson's r) for BEECOM-18, BEECOM-12, and BEECOM-9 total scores and subscales across T1, T2, and T3

To assess the temporal stability of the BEECOM instrument, test-retest correlations were calculated (Lodder et al., 2022) for three versions of the scale (BEECOM-18, BEECOM-12, and BEECOM-9) across three time points: T1, T2, and T3. All versions of the scale demonstrated high stability, with total score correlations ranging from $r = .77$ to $r = .87$.

For the BEECOM-18, the highest correlation was observed between T2 and T3 ($r = .87$, $p < .001$). Similar values were obtained for BEECOM-12 ($r = .87$) and BEECOM-9 ($r = .87$), indicating good measurement consistency over time.

The analysis of subscales revealed varying levels of stability. The highest correlations were found for the Body subscale (e.g., $r = .86$ between T1 and T2 in both the 18- and 12-item versions), while the lowest correlation was noted for the Eating subscale in the 9-item version ($r = .48$ between T1 and T3), suggesting greater variability in this domain of behavior (Table 4).

Paired t-test for BEECOM-18, BEECOM-12, and BEECOM-9 total scores and subscales across T1, T2, and T3

Paired-samples *t*-tests and Wilcoxon signed-rank tests indicated overall high stability of scores across the BEECOM-18, BEECOM-12, and BEECOM-9 scales over three time points (T1, T2, T3). No significant differences were observed for total scores or for the majority of subscales (Eating, Exercise, Body) across assessments, supporting the reliability and temporal consistency of the instrument.

Notable exceptions were found in a few comparisons: significant decreases were observed for the Exercise subscale of BEECOM-18 between T2 and T3 ($p = .018$), for the Body subscale of BEECOM-12 between T1 and T3 ($p = .042$), and for the Body subscale of BEECOM-9 between T1 and T2 ($p = .036$), as well as between T1 and T3 ($p = .014$). These findings suggest that certain aspects related to body perception or physical activity may be more susceptible to change over time (Table 5).

STUDY 3 (*N* = 129)

Latent growth curve model

Table 6 presents the fit indices for each latent growth curve (LGC) model, as well as the results of the Wald tests (Hertzog et al., 2008), which indicate whether the mean and variance of the latent intercepts and slopes significantly differed from zero. In the LGC analysis, the mean slope was non-significant across all models, including the BEECOM-18 ($M = -1.98$, $p > .05$), BEECOM-12 ($M = -1.22$, $p > .05$), and BEECOM-9 ($M = -0.72$, $p > .05$) total scores. This indicates that, on average, participants' comparison tendencies did

Table 3

Correlation coefficients between BEECOM-18 scale, BEECOM-R-12, BEECOM-R-9 scale and subscale (Body-R, Eating-R, Exercise-R), PACS-PL, PACS-PL-R and SATAQ-3 (N = 408)

	M (*SD*)	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.
1.	57.76 (25.44)	1																	
2.	33.32 (17.24)	.988***	1																
3.	29.39 (12.60)	.986***	.988***	1															
4.	12.44 (5.54)	.856***	.865***	.867***	1														
5.	9.66 (4.67)	.848**	.838***	.860***	.601***	1													
6.	7.29 (4.49)	.829***	.741***	.842***	.573***	.632***	1												
7.	10.29 (3.93)	.744***	.741***	.738***	.757***	.587***	.525***	1											
8.	6.59 (3.61)	.742***	.742***	.740***	.789***	.572***	.507***	.937***	1										
9.	29.75 (13.33)	.635***	622***	.621***	.661***	.461***	.446***	.668***	.668***	1									
10.	15.43 (6.05)	.359***	.350***	.340***	.355***	.258***	.248***	.361***	.384***	.553***	1								
11.	11.47 (4.50)	.505**	.526***	.522***	.504***	.352***	.477***	.439***	.467***	.502***	.222***	1							
12.	11.71 (5.85)	.491***	.478***	.470***	.437***	.384***	.380***	.539***	.524***	.721***	.497***	.371***	1						
13.	3.70 (1.94)	.011	.014	.010	−.011	.050	−.010	−.016	.006	−.022	−.032	−.007	−.116*	1					
14.	4.36 (0.73)	−.007	−.002	−.013	−.069	−.013	.064	−.080	−.074	−.039	−.030	.011	−.021	.126*	1				
15.	1.65 (0.71)	.080	.071	.070	.057	.030	.096	.082	.053	.011	.025	−.004	.036	.045	.247***	1			
16.	2.30 (0.89)	.098*	.097*	.100*	.079	.052	.129**	.060	.049	.084	.096	.056	.072	.057	−.006	.046	1		
17.	26.26 (9.60)	−.135**	−.137**	−.136**	−.154**	−.095	−.092	−.108*	−.158**	−.102*	−.039	−.213***	−.044	−.052	.320***	.393***	.045	1	
18.		−.145**	−.137**	−.130**	−.127*	−.130**	−.074	−.226***	−.218***	−.238***	−.183***	.013	−.206***	.070	−.066	−.018	.030	−.011	1

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. BEECOM – Body, Eating, and Exercise Comparison Orientation Measure; PACS – Physical Appearance Comparison Scale; SATAQ-3 – Sociocultural Attitudes Toward Appearance Questionnaire. 1 – BEECOM-Total; 2 – BEECOM-R-12; 3 – BEECOM-R-9; 4 – Body-R; 5 – Eating-R; 6 – Exercise-R; 7 – PACS; 8 – PACS-R; 9 – SATAQ-3 Internalization Pressure; 10 – SATAQ-3 Internalization Information Seeking; 11 – SATAQ-3 Internalization Athlete; 12 – SATAQ-3 Information; 13 – place of residency; 14 – education; 15 – marital status; 16 – occupation; 17 – age; 18 – gender (0 – male, 1 – female).

Table 4

Test-retest correlations (Pearson's r) for BEECOM-18, BEECOM-12, and BEECOM-9 total scores and subscales across T1, T2, and T3

Scale/Subscale	T1–T2	T1–T3	T2–T3
BEECOM-18 Total	.85***	.80***	.87***
Eating	.81***	.68***	.80***
Exercise	.71***	.77***	.83***
Body	.86***	.76***	.81***
BEECOM-12 Total	.83***	.78***	.87***
Eating	.77***	.65***	.77***
Exercise	.66***	.75***	.84***
Body	.86***	.77***	.81***
BEECOM-9 Total	.83***	.77***	.87***
Eating	.61***	.48***	.74***
Exercise	.63***	.73***	.82***
Body	.84***	.76***	.82***

Note. BEECOM – Body, Eating, and Exercise Comparison Orientation Measure; T1 – Time 1; T2 – Time 2; T3 – Time 3; ***$p < .001$.

Table 5

Paired t-test for BEECOM-18, BEECOM-12, and BEECOM-9 total scores and subscales across T1, T2, and T3

Scale	Subscale	Comparison	*M*1	*SD*1	*M*2	*SD*2	Test	Statistic	*p*
BEECOM-18	Total	T1-T2	58.69	27.11	58.59	26.79	*t*-test	$t(69) = 0.06$	.955
BEECOM-18	Total	T1-T3	58.69	27.11	55.59	25.00	*t*-test	$t(69) = 1.56$	.123
BEECOM-18	Total	T2-T3	58.59	26.79	55.59	25.00	*t*-test	$t(69) = 1.88$	.065
BEECOM-6	Eat	T1-T2	20.01	8.90	20.13	9.10	*t*-test	$t(69) = -0.17$	.863
BEECOM-6	Eat	T1-T3	20.01	8.90)	19.26	8.83	*t*-test	$t(69) = 0.89$	.378
BEECOM-6	Eat	T2-T3	20.13	9.10	19.26	8.83	Wilcoxon	$W = 1198.50$ $Z = 1.06$	.289
BEECOM-6	Exercise	T1-T2	15.69	9.56	16.37	9.14	*t*-test	$t(69) = -0.80$	.426
BEECOM-6	Exercise	T1-T3	15.69	9.56	14.87	8.51	*t*-test	$t(69) = 1.10$	.276
BEECOM-6	Exercise	T2-T3	16.37	9.14	14.87	8.51	*t*-test	$t(69) = 2.43$	.018*
BEECOM-6	Body	T1-T2	21.27	9.96	22.09	10.47	*t*-test	$t(69) = -1.26$	.210
BEECOM-6	Body	T1-T3	21.27	9.96	21.44	8.33	*t*-test	$t(69) = -0.22$	.828
BEECOM-6	Body	T2-T3	22.09	10.47	21.44	8.33	*t*-test	$t(69) = 0.89$	.379
BEECOM-12	Total	T1-T2	39.40	18.24	39.17	17.95	*t*-test	$t(69) = 0.18$	.855
BEECOM-12	Total	T1-T3	39.40	18.24	37.30	16.58	*t*-test	$t(69) = 1.52$	.134
BEECOM-12	Total	T2-T3	39.17	17.95	37.30	16.58	*t*-test	$t(69) = 1.73$	.088
BEECOM-4	Eat	T1-T2	13.63	6.26	13.57	6.23	*t*-test	$t(69) = 0.11$	.911
BEECOM-4	Eat	T1-T3	13.63	6.26	12.99	5.89	*t*-test	$t(69) = 1.06$	.293
BEECOM-4	Eat	T2-T3	13.57	6.23	12.99	5.89	Wilcoxon	$W = 1052$ $Z = 1.01$	.313

Table 5 continues

Table 5

Table 5 continued

Scale	Subscale	Comparison	*M*1	*SD*1	*M*2	*SD*2	Test	Statistic	*p*
BEECOM-4	Exercise	T1-T2	9.97	6.47	10.64	6.14	*t*-test	$t(69) = -1.08$	.284
BEECOM-4	Exercise	T1-T3	9.97	6.47	9.73	5.78	Wilcoxon	$W = 795.50$ $Z = 0.46$	.645
BEECOM-4	Exercise	T2-T3	10.64	6.14	9.73	5.78	Wilcoxon	$W = 726$ $Z = 2.03$	.042*
BEECOM-4	Body	T1-T2	15.80	7.16	14.96	7.10	Wilcoxon	$W = 924$ $Z = 1.85$	.065
BEECOM-4	Body	T1-T3	15.80	7.16	14.59	6.67	*t*-test	$t(69) = 2.15$	.035*
BEECOM-4	Body	T2-T3	14.96	7.10	14.59	6.67	Wilcoxon	$W = 1140.50$ $Z = 0.67$	.500
BEECOM-9	Total	T1-T2	29.29	13.01	29.40	13.25	Wilcoxon	$W = 1245.50$ $Z = 0.44$	.660
BEECOM-9	Total	T1-T3	29.29	13.01	28.10	12.60	Wilcoxon	$t(69) = 1.15$	.255
BEECOM-9	Total	T2-T3	29.40	13.25	28.10	12.60	Wilcoxon	$W = 1361.50$ $Z = 1.39$	.165
BEECOM-3	Eat	T1-T2	9.57	3.52	9.97	4.61	*t*-test	$t(69) = -0.90$	.373
BEECOM-3	Eat	T1-T3	9.57	3.52	9.60	4.56	*t*-test	$t(69) = -0.06$	.955
BEECOM-3	Eat	T2-T3	9.97	4.61	9.60	4.56	*t*-test	$t(69) = 0.94$	.351
BEECOM-3	Exercise	T1-T2	7.47	4.82	7.97	4.69	*t*-test	$t(69) = -1.02$	.312
BEECOM-3	Exercise	T1-T3	7.47	4.82	7.41	4.48	*t*-test	$t(69) = 0.14$	.890
BEECOM-3	Exercise	T2-T3	7.97	4.69	7.41	4.48	*t*-test	$t(69) = 1.68$	.098
BEECOM-3	Body	T1-T2	12.23	5.43	11.46	5.35	*t*-test	$t(69) = 2.14$	.036*
BEECOM-3	Body	T1-T3	12.23	5.43	11.09	5.03	Wilcoxon	$W = 859.50$ $Z = 2.46$	.014*
BEECOM-3	Body	T2-T3	11.46	5.35	11.09	5.03	Wilcoxon	$W = 1024.50$ $Z = 0.81$	.419

Note. BEECOM – Body, Eating, and Exercise Comparison Orientation Measure; *M*1 – mean for the first measurement point; *SD*1 – standard deviation for the first measurement point; *M*2 – mean for the second measurement point; *SD*2 – standard deviation for the second measurement point; *W* – Wilcoxon test statistic; *Z* – standardized test value; *p* – significance level (*p*-value). Measurement points: T1 – Time 1, T2 – Time 2, T3 – Time 3; *$p < .05$.

not change significantly over the three-month period, supporting the stability of the BEECOM across time at the group level.

In contrast, for the BEECOM Eating 3 subscale, although the mean slope was also non-significant ($M = -0.11$, $p > .05$), the variance of the slope was significant (Var = 3.82, $p < .001$). This result suggests that while no overall group-level change occurred, individual participants showed significant variation in their trajectories over time: some demonstrated increased comparison behaviors related to eating, others showed decreases, and some remained stable (Hertzog et al., 2006).

An exception to this overall pattern was observed in the BEECOM Exercise 3 subscale, for which the LGC model produced a negative slope variance (Var = −0.13, $p > .05$), indicating an inadmissible solution commonly referred to as a Heywood case. This suggests that the model was misspecified or that there was insufficient individual variability in change over time to estimate the trajectory reliably (see Figure 2) (Shen & Li, 2023).

DISCUSSION

STUDY 1

The attempt to adapt the BEECOM scale to local cultural conditions (Hornowska & Paluchowski, 2004)

Table 6

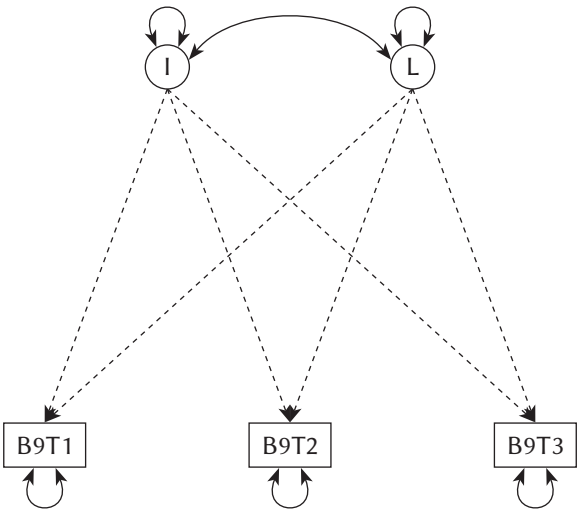
Fit indices and individual changes for BEECOM-18, BEECOM-12, BEECOM-9, and the BEECOM-9 subscales in relation to the mean and variance of the latent intercept and slope

	BEECOM-18	BEECOM-12	BEECOM-9	BEECOM EAT 3	BEECOM EXERCISE 3	BEECOM BODY 3
Model fit	*N* = 129	*N* = 129	*N* = 129	*N* = 129	*N* = 129	*N* = 129
χ^2	308.90	292.80	285.18	164.17	215.66	254.21
RMSEA (95% CI)	0.09 [0.00, 0.27]	0.09 [0.00, 0.27]	0.09 [0.00, 0.27]	0.14 [0.00, 0.31]	0.17 [0.04, 0.34]	0.04 [0.00, 0.24]
SRMR	0.02	0.02	0.02	0.03	0.04	0.01
CFI	0.99	0.99	0.99	0.98	0.98	0.99
Latent growth parameters						
Mean intercept	59.09***	42.93***	29.35***	9.51***	7.72***	12.06***
Variance intercept	621.16***	274.53***	143.67***	13.46***	12.65***	21.70***
Mean slope	−1.98	−1.22*	−0.72	−0.11	−0.12	−0.51
Variance slope	43.19	20.84	11.85	3.82***	−0.13	1.15

Note. BEECOM – Body, Eating, and Exercise Comparison Orientation Measure; *$p < .05$, ***$p < .001$.

Figure 2

Structure of the testing model



Note. B9T1 – BEECOM 9 Time 1; B9T2 – BEECOM 9 Time 2; B9T3 – BEECOM 9 Time 3; I – intercept; L – linear slope.

and evaluate its psychometric properties, to the best of our knowledge, is the first of its kind in Poland. The confirmatory analyses conducted showed that BEECOM-18 did not achieve satisfactory indicators for acceptance (RMSEA = .09; GFI = .83; AGFI = .77). The modifications made for the initial model resulted in a 12-item acceptable version (RMSEA = .07; GFI = .93; AGFI = .89). At the same time, similarly, the BEECOM-R-9 version of the scale, on the other hand, proved to be a good fit in general population studies focusing on the analysis of relationships and correlations in terms of the tendency to make comparisons, focusing on the body, eating, and exercise (RMSEA = .02; GFI = .97; AGFI = .95). Both BEECOM-R-12 and BEECOM-R-9 have very good reliability indices; in addition, BEECOM-R-9 indices are consistent with previous reports using this scale (Fitzsimmons-Craft et al., 2012; Paterna et al., 2023; Sahlan et al., 2020; Saunders et al., 2019). The removal of six (BEECOM-R-12) and nine items (BEECOM-R-9) from the original 18-item version did not affect the reliability of the scale, as evidenced by the analyses conducted, as well as the retained three-factor structure of the scale. It is also worth noting that the elimination of statements, with a high degree of certainty, was dictated primarily by their limited relevance to the local culture, resulting in a final version that represents a culturally valid instrument (Hudson et al., 2023; Paterna et al., 2023).

It is noteworthy that BEECOM-R-9 allowed full replication of the structure for two of the three subscales. Thus, it can be concluded that the Exercise and Body subscales show universality, regardless of culture and individual participant differences. For the Eating subscale, on the other hand, the differences in deleted and retained items compared to other

versions may be due to cultural and social behaviors related to eating. For the Eating subscale, deleted items (items 3 and 11) focus on comparing what I eat to what others eat. Previous literature suggests that local eating habits, especially in group social situations (family celebrations or formal occasions), focus more on comparing what others eat to what I eat (Polivy, 2017).

Conversely, the opposite situation occurs when the food choices are dictated, for example, by the motivation to lose weight (item 8). At the same time, it is worth mentioning that the original version of the scale was created more than ten years ago, before the trend of developing conscious eating habits began in Poland. Moreover, the original version of the scale was developed in the context of dietary habits using different measurement units (e.g., gallons) than those used in the population of the adapted version of the scale (e.g. grams). It is also worth noting that the sheer size of foods and meals consumed can differ between Europe and America, which can also affect how an individual makes self–other eating comparisons (Polivy, 2017).

Attempts to adapt BEECOM to local conditions to date have mainly focused, as recommended, on analyses using tools to measure eating disorders. Other attempts to adapt the scale also involved a reduction in the number of statements, which was one of the guidelines when developing this version of the scale (Paterna et al., 2023; Saunders et al., 2019). Moreover, BEECOM was used more often in samples consisting of women (Fitzsimmons-Craft et al., 2012), including clinical groups (Saunders et al., 2019). The version of BEECOM-R-9 presented in this study was developed for a population of male and female adult participants, consistent with previous studies of BEECOM (Hudson et al., 2023; Paterna et al., 2023).

Despite the poor values of the indices obtained in the confirmatory analysis for BEECOM-18, it is worth noting that the scale demonstrated high reliability, with McDonald's omega ω = .96 for the total scale and strong values for the subscales: Body: ω = .95; Eating: ω = .90; Exercise: ω = .92. Similar values were obtained for BEECOM-R-12 and BEECOM-R-9, which showed very good internal reliability (Body-R ω = .94; Eating-R ω = .78; Exercise-R ω = .87), and factor relevance. Moreover, the BEECOM-18 scale, BEECOM-R-12, and subscales of BEECOM-R-9 were correlated with higher intensity of all SATAQ-3 subscales: Internalization Pressure, Internalization Information Seeking, Internalization Athlete, and Information. Evidence of BEECOM-R-9 convergent validity was clearly demonstrated by the strong positive correlation with PACS (r = .74, p < .001). In conclusion, the results indicate that the developed version of the BEECOM scale is effective, and the results allow it to be used in research. The proposed BEECOM-R-9 version for the Body (4, 9, 12) and Exercise (6, 14, 15) scales is consistent with the studies by Saunders et al. (2019) and Paterna et al. (2023). Differences are shown by the Eating scale. The study by Saunders et al. (2019) and Paterna et al. (2023) contained statements 3, 7, and 11, while the Polish version of the Eating scale contains statements 1, 7, and 16.

Consistent with prior studies (e.g., Hudson et al., 2023; Paterna et al., 2023), the invariance analysis confirmed that BEECOM yields comparable results across gender and age groups. Although most body comparison research focuses on women and younger individuals, the tool is also valid for men and older adults. Given that comparison tendencies decline with age, BEECOM can be effectively used to explore age-related differences and potential protective factors.

Previous BEECOM adaptations have mainly used the EDE-Q and focused on populations at risk for eating disorders (e.g., Sahlan et al., 2020; Saunders et al., 2019). While the current study supports the BEECOM-R-9's use in research and clinical settings, it did not assess absolute stability. Future research should examine the scale's temporal stability, especially in clinical groups previously involved in other validations. Although the sample size met general research standards, it may still be considered small, suggesting a need for further, more comprehensive validation – both cross-sectional and longitudinal. Additionally, since the current study was conducted on a general population, future work should replicate these analyses in clinical samples to better assess the scale's relevance for individuals with eating disorders or addictions (Harriger et al., 2022; Polivy, 2017).

In conclusion, comparing with each other, the baseline model (BEECOM-18) and the alternative models (BEECOM-R-12 and BEECOM-R-9), it should be noted that the results obtained allow the use of both the 12-item and 9-item versions (Appendix 4 in Supplementary materials). However, it is worth noting that BEECOM-R-12 has only been validated in the local population, whereas BEECOM-R-9 has been examined in multiple other studies using this tool (Paterna et al., 2023; Saunders et al., 2019).

STUDIES 2 AND 3

This study provides the first comprehensive examination of the temporal stability of the Polish BEECOM using multiple methodological approaches across two independent samples. Our findings offer nuanced support for the scale's stability, with important variations across different versions and subscales that have implications for longitudinal research applications (Lodder et al., 2022).

The assessment of relative temporal stability in Study 1, using Pearson correlation coefficients, revealed statistically significant associations across

time points for all BEECOM versions. Correlation values for the total BEECOM scores and their subscales ranged from r = .48 (moderate) to r = .87 (strong), generally indicating good rank-order stability over the three months, consistent with benchmarks for trait-like constructs (e.g., Moreau, 2025).

To evaluate mean-level (absolute) temporal stability, paired-sample t-tests and Wilcoxon signed-rank tests were conducted. The results showed no significant changes for most comparisons, indicating stable average levels over time (Lodder et al., 2022). However, some statistically significant differences emerged at the subscale level. For example, in the BEECOM-18 model, a significant decrease was observed in the Exercise subscale scores between T2 and T3 ($t(69) = 2.43$, p = .018). In the BEECOM-R-12 model, the Exercise subscale also showed a significant decrease from T2 (median = 10.64) to T3 (median = 9.73), p = .042. These findings suggest that certain subscales – particularly Exercise and Body – may be more sensitive to time-related or contextual fluctuations, even if global mean-level stability is preserved.

To further assess absolute temporal stability, latent growth curve modeling (LGCM) was applied (Byrne & Crombie, 2003; Hertzog et al., 2008; Shen & Li, 2023). Model fit was evaluated using χ^2, CFI, SRMR, and RMSEA. For several models (e.g., total scores for BEECOM-18, BEECOM-R-12, and BEECOM-R-9), CFI and SRMR values suggested acceptable to good fit (see Table 1). However, RMSEA values were problematic for some models. Notably, the BEECOM-Eating-3 (RMSEA = .14) and BEECOM-Exercise-3 (RMSEA = .17) subscale models exhibited poor fit according to RMSEA benchmarks (MacCallum et al., 1996), although their CFI and SRMR values were acceptable. The wide confidence intervals for RMSEA in some models also suggest some imprecision. These models with poorer fit should be interpreted with caution. Decisions to retain models were based on a holistic view of fit indices and theoretical considerations, but limitations due to fit are acknowledged (Shen & Li, 2023).

The BEECOM-18, BEECOM-R-9, and the Exercise and Body subscales of BEECOM-9 showed evidence of absolute temporal stability, as indicated by non-significant slopes and minimal interindividual variability. In contrast, the BEECOM-R-12 model demonstrated a significant average slope and variance, indicating a lack of mean-level stability over time. Additionally, for the BEECOM Eating-3 subscale (within the BEECOM-R-9 model), while the mean slope was not significant ($M_{slope} = -0.11$, $p > .05$), indicating stable group-level means, the variance of the slope was significant ($Var_{slope} = 3.82$, $p < .001$). This suggests that although the average tendency to compare eating did not change, there were significant individual differences in how participants' eating comparisons changed over time (Lodder et al., 2022; Shen & Li, 2023).

LIMITATIONS AND FUTURE DIRECTIONS

The present research has several limitations that should be acknowledged. First, the recruitment of participants through social media and the use of the snowball sampling method may have introduced self-selection bias, as individuals with specific characteristics or interests might have been more likely to take part. Moreover, the relatively small sample sizes, particularly in the longitudinal studies, limit the generalizability of the findings and may reduce the statistical power of some analyses. Future studies should therefore aim to include larger and more diverse samples, using recruitment procedures that minimize potential sampling bias and allow for broader representativeness of the population.

METHODOLOGICAL CONTRIBUTIONS AND IMPLICATIONS

The use of two distinct Polish samples (Study 1: broader adult sample for correlational and mean-comparison analyses; Study 2: younger adult/student sample for LGCM) and three assessment time points allowed for a multifaceted evaluation of the BEECOM-PL's temporal stability. The application of rank-order correlations, paired comparisons of means, and latent growth modeling enabled a triangulated assessment, addressing different facets of stability and mitigating limitations of any single method (Lodder et al., 2022).

To the best of our knowledge, this is the first study to systematically evaluate the temporal stability of any BEECOM version using LGCM to assess absolute change trajectories, and the first to provide such a comprehensive temporal stability assessment for its Polish adaptations. Overall, the findings provide preliminary support for the Polish BEECOM, particularly the 9-item version (BEECOM-R-9 total score and its Body subscale), as a relatively stable instrument over a three-month period. However, issues with model fit (e.g., high RMSEA for some subscale LGCMs) and an inadmissible solution for the Exercise-3 LGCM – specifically, a negative variance estimate for the latent slope (i.e., a Heywood case) – highlight areas needing further investigation before unqualified use in longitudinal research. The 12-item version showed a slight mean decrease over time.

Our findings support conceptualizing comparison orientation as a relatively stable individual difference variable, consistent with theoretical frameworks positioning social comparison tendencies as trait-like constructs (Gibbons & Buunk, 1999). However, the observed subscale-specific variations suggest that comparison tendencies may comprise both stable trait components and domain-specific state variations.

For researchers and clinicians, these results indicate that the Polish BEECOM can reliably assess comparison tendencies over time, making it suitable for intervention studies and longitudinal research. However, investigators should carefully consider which version best suits their research questions, with the BEECOM-18 offering optimal stability for detecting individual differences over time, while shorter versions may be more sensitive to intervention effects.

ACKNOWLEDGEMENTS

We would like to thank Ellen E. Fitzsimmons-Craft for her kind permission to validate the Body, Eating, and Exercise Comparison Orientation Measure Polish version.

Supplementary materials are available on the journal's website.

Disclosures

This research received no external funding.
The study was approved by the Ethics Committee of the Polish Academy of Sciences, Institute of Psychology (Approval No. 32/XII/2020).
The authors declare no conflict of interest.

References

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

Byrne, B. M. (2016). *Structural equation modelling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.

Byrne, B. M., & Crombie, G. (2003). Modeling and testing change: an introduction to the latent growth curve model. *Understanding Statistics, 2*, 177–203. https://doi.org/10.1207/S15328031US0203_02

Cash, T. F., & Smolak, L. (2011). Understanding body images: Historical and contemporary perspectives. In T. F. Cash & L. Smolak (Eds.), *Body image: a handbook of science, practice, and prevention* (pp. 3–11). Guilford Press.

Davison, T. E., & McCabe, M. P. (2005). Relationships between men's and women's body image and their psychological, social, and sexual functioning. *Sex Roles, 52*, 463–475. https://doi.org/10.1007/s11199-005-3712-z

Dzielska, A., Mazur, J., Nałęcz, H., Oblacińska, A., & Strucińska, M. (2017). Polish adaptation and validation of the Physical Appearance Comparison Scale (PACS) – an analysis among young people in late adolescence in the context of obesity. *Developmental Period Medicine, 21*, 213–223. https://doi.org/10.34763/devperiodmed.20172103.213223

Faith, M. S., Leone, M. A., & Allison, D. B. (1997). The effects of self-generated comparison targets, BMI, and social comparison tendencies on body image appraisal. *Eating Disorders, 5*, 128–140. https://doi.org/10.1080/10640269708249216

Fardouly, J., Diedrichs, P. C., Vartanian, L. R., & Halliwell, E. (2015). Social comparisons on social media: The impact of Facebook on young women's body image concerns and mood. *Body Image, 13*, 38–45. https://doi.org/10.1016/j.bodyim.2014.12.002

Fatt, S. J., Fardouly, J., & Rapee, R. M. (2019). #malefitspo: Links between viewing fitspiration posts, muscular-ideal internalisation, appearance comparisons, body satisfaction, and exercise motivation in men. *New Media & Society, 21*, 1311–1325. https://doi.org/10.1177/1461444818821064

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140. https://doi.org/10.1177/001872675400700202

Fitzsimmons-Craft, E. E., & Bardone-Cone, A. M. (2014). One-year temporal stability and predictive and incremental validity of the Body, Eating, and Exercise Comparison Orientation Measure (BEECOM) among college women. *Body Image, 11*, 27–35. https://doi.org/10.1016/j.bodyim.2013.09.003

Fitzsimmons-Craft, E. E., Bardone-Cone, A. M., & Harney, M. B. (2012). Development and validation of the Body, Eating, and Exercise Comparison Orientation Measure (BEECOM) among college women. *Body Image, 9*, 476–487. https://doi.org/10.1016/j.bodyim.2012.07.007

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology, 76*, 129–142. https://doi.org/10.1037/0022-3514.76.1.129

Harriger, J. A., Evans, J. A., Thompson, J. K., & Tylka, T. L. (2022). The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms. *Body Image, 41*, 292–297. https://doi.org/10.1016/j.bodyim.2022.03.007

Hertzog, C., Lindenberger, U., Ghisletta, P., & von Oertzen, T. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods, 11*, 244–252. https://doi.org/10.1037/1082-989X.11.3.244

Hertzog, C., von Oertzen, T., Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling, 15*, 541–563. https://doi.org/10.1080/10705510802338983

Hornowska, E., & Paluchowski, W. J. (2004). Kulturowa adaptacja testów psychologicznych [Cultural adaptation of psychological tests]. In J. Brzeziński

(Ed.), *Metodologia badań psychologicznych. Wybór tekstów* [Psychological research methodology. Selected texts] (pp. 151–191). Wydawnictwo Naukowe PWN.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria to fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. https://doi.org/10.1080/10705519909540118

Hudson, T. A., Ferreira, L. K., Amaral, A. C. S., & Ferreira, M. E. C. (2023). Body, Eating, and Exercise Comparison Orientation Measure (BEECOM): Cross-cultural adaptation and psychometric properties among Brazilian female university students. *Body Image, 45*, 323–330. https://doi.org/10.1016/j.bodyim.2023.03.013

Izydorczyk, B., & Lizińczyk, S. (2020). The Polish adaptation of the socio-cultural attitudes towards appearance SATAQ 3 questionnaire. *Health Psychology Report, 8*, 68–82. https://doi.org/10.5114/hpr.2020.91443

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Scientific Software International.

Keery, H., Van den Berg, P., & Thompson, J. K. (2004). An evaluation of the Tripartite Influence Model of body dissatisfaction and eating disturbance with adolescent girls. *Body Image, 1*, 237–251. https://doi.org/10.1016/j.bodyim.2004.03.001

Lin, L. F., & Kulik, J. A. (2002). Social comparison and women's body satisfaction. *Basic and Applied Social Psychology, 24*, 115–123. https://doi.org/10.1207/S15324834BASP2402_4

Lodder, P., Kupper, N., Mols, F., Emons, W. H., & Wicherts, J. M. (2022). Assessing the temporal stability of psychological constructs: an illustration of type D personality, anxiety and depression. *Journal of Research in Personality, 101*, 104299. https://doi.org/10.1016/j.jrp.2022.104299

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149. https://doi.org/10.1037/1082-989X.1.2.130

McCreary, D. R., & Saucier, D. M. (2009). Drive for muscularity, body comparison, and social physique anxiety in men and women. *Body Image, 6*, 24–30. https://doi.org/10.1016/j.bodyim.2008.09.002

McLean, S. A., Paxton, S. J., Wertheim, E. H., & Masters, J. (2015). Photoshopping the selfie: Self photo editing and photo investment are associated with body dissatisfaction in adolescent girls. *International Journal of Eating Disorders, 48*, 1132–1140. https://doi.org/10.1002/eat.22449

Moreau, D. (2025). Temporal stability in measurements matters only for stable constructs. *Nature Reviews Psychology, 4*, 361. https://doi.org/10.1038/s44159-025-00435-z

O'Brien, K. S., Caputi, P., Minto, R., Peoples, G., Hooper, C., Kell, S., & Sawley, E. (2009). Upward and downward physical appearance comparisons: Development of scales and examination of predictive qualities. *Body Image, 6*, 201–206. https://doi.org/10.1016/j.bodyim.2009.03.003

Paterna, A., Alcaraz-Ibáñez, M., & Sicilia, A. (2023). Psychometric examination of the Body, Eating, and Exercise Comparison Orientation Measure (BEECOM) among Spanish adolescents and young adults. *Nutrients, 15*, 626. https://doi.org/10.3390/nu15030626

Peng, C. T., Wu, T. Y., Chen, Y., & Atkin, D. J. (2019). Comparing and modeling via social media: The social influences of fitspiration on male Instagram users' work out intention. *Computers in Human Behavior, 99*, 156–167. https://doi.org/10.1016/j.chb.2019.05.011

Polivy, J. (2017). What's that you're eating? Social comparison and eating behavior. *Journal of Eating Disorders, 5*, 18. https://doi.org/10.1186/s40337-017-0148-0

Rancourt, D., Schaefer, L. M., Bosson, J. K., & Thompson, J. K. (2016). Differential impact of upward and downward comparisons on diverse women's disordered eating behaviors and body image. *International Journal of Eating Disorders, 49*, 519–523. https://doi.org/10.1002/eat.22470

Sahlan, R. N., Saunders, J. F., Fitzsimmons-Craft, E. E., & Taravatrooy, F. (2020). The validity and reliability of the Farsi version of the Body, Eating, and Exercise Comparison Orientation Measure (F-BEECOM) among Iranian male and female students. *Body Image, 34*, 72–84. https://doi.org/10.1016/j.bodyim.2020.05.006

Saunders, J. F., Eaton, A. A., & Fitzsimmons-Craft, E. E. (2019). Body-, eating-, and exercise-related comparisons during eating disorder recovery and validation of the BEECOM-R. *Psychology of Women Quarterly, 43*, 494–508. https://doi.org/10.1177/0361684319851718

Schaefer, L. M., & Thompson, J. K. (2014). The development and validation of the Physical Appearance Comparison Scale-Revised (PACS-R). *Eating Behaviors, 15*, 209–217. https://doi.org/10.1016/j.eatbeh.2014.01.001

Schaefer, L. M., & Thompson, J. K. (2018). The development and validation of the Physical Appearance Comparison Scale-3 (PACS-3). Psychological Assessment, 30, 1330–1341. https://doi.org/10.1037/pas0000576

Shen, L., & Li, S. S. (2023). Quantifying emotional flow: Testing the emotional flow hypothesis from a longitudinal latent growth curve (LGC) modeling approach. *Media Psychology, 26*, 436–459. https://doi.org/10.1080/15213269.2022.2156886

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive good-

ness-of-fit measures. *Methods of Psychological Research Online, 8*, 23–74.
Thompson, J. K., Heinberg, L. J., Altabe, M., & Tantleff-Dunn, S. (1999). *Exacting beauty: Theory, assessment, and treatment of body image disturbance.* American Psychological Association. https://doi.org/10.1037/10312-000
Thompson, J. K., Van Den Berg, P., Roehrig, M., Guarda, A. S., & Heinberg, L. J. (2004). The sociocultural attitudes towards appearance scale-3 (SATAQ-3): Development and validation. *International Journal of Eating Disorders, 35*, 293–304. https://doi.org/10.1002/eat.10257
Thompson, J. K., Fabian, L. J., Moulton, D. O., Dunn, M. E., & Altabe, M. N. (1991). Development and validation of the Physical Appearance Related Teasing Scale. *Journal of Personality Assessment, 56*, 513–521. https://doi.org/10.1207/s15327752jpa5603_12
Tiggemann, M. (2003). Media exposure, body dissatisfaction and disordered eating: Television and magazines are not the same. *European Eating Disorders Review, 11*, 418–430. https://doi.org/10.1002/erv.502
Vander Wal, J. S., & Thelen, M. H. (2000). Eating and body image concerns among obese and average-weight children. *Addictive Behaviors, 25*, 775–778. https://doi.org/10.1016/S0306-4603(00)00061-7
Vandenberg, R. J., & Grelle, D. M. (2009). Alternative model specifications in structural equation modeling: Facts, fictions and truths. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences* (pp. 165–191). Routledge.
Vaterlaus, J. M., Patten, E. V., Roche, C., & Young, J. A. (2015). # Gettinghealthy: The perceived influence of social media on young adult health behaviors. *Computers in Human Behavior, 45*, 151–157. https://doi.org/10.1016/j.chb.2014.12.013