Functionality Appreciation Scale (FAS): Item Response Theory Examination

Joshua Marmara (✉ joshua.marmara@live.vu.edu.au)

Victoria University

Daniel Zarate

Victoria University

Research Article

Keywords: Functionality Appreciation, Positive Body Image, Item Response Theory, Psychometric Properties, Positive Psychology, Gender

Posted Date: February 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1148688/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The present study considers a measure of positive body image, the Functionality Appreciation Scale (FAS), which assesses one's perception of appreciating, respecting, and honouring the body for what it can do [3]. Differential functioning of the scale across groups (i.e., gender) is yet to be investigated. The present study contributes to this area of knowledge via the employment of Item Response Theory (IRT) analyses.

Methods: A sample of 386 adults from Canada, Australia, New Zealand, Ireland, the United Kingdom (UK), and United States of America (USA) were assessed online (N = 394, 54.8% men, 43.1% women, $M_{age} = 27.48$, SD = 5.57).

Results: The two-parameter logistic model employed to observe IRT properties indicated that all items demonstrated, although variable, strong discrimination capacity. Considering the DIF across men and women, all items demonstrated psychometric invariance across groups indicating that FAS measures FA equally in men and women.

Conclusions: The items showed increased reliability for latent levels of ± 2 SD from the mean level of Functionality Appreciation. The implications and interpretations of the findings for clinical practice are discussed.

Introduction

Body image is a multifaceted construct that represents one's cognitions, perceptions, and behaviours relating to one's body [10; 27; 36]. Contemporary research has focused on a unidimensional component of body image by emphasizing a negative connotation, primarily related to mental health treatment seekers [5; 12; 35]. Alternatively, positive psychology, an outlook entrenched in hygiology (the promotion of health; [35]), provides a conceptual structure to guide the study of positive body image as distinct from negative body image.

One area of positive body image that has yet to be explored in depth, is *functionality appreciation* (FA). Alleva et al. [3] characterised FA as appreciating, respecting, and honouring the body for what it is capable of doing" (p. 29). Given its importance to positive body image [30], FA has been employed in interventions aimed at enhancing positive feelings and attitudes towards one's body (i.e., [1; 2; 11]). Given the lack of sufficient measures for assessing FA, Alleva et al. [3] pioneered the Functionality Appreciation Scale (FAS). An exploratory factor analysis adhered to a unidimensional solution with seven items, and a confirmatory factor analysis maintained the construct's unidimensionality and yielded invariance across gender [3]. The FAS has shown adequate convergent, criterion-related, and divergent validity, adequate test-retest reliability, and good internal consistency [3; 23; 28]. Moreover, Alleva et al. [3] outlined that FAS scores gauged good divergent, adequate convergent, and criterion-related validity, indicated by positive and significant associations with measures pertaining to body image (i.e., body appreciation), positive self-care (i.e., self-compassion), and psychological well-being (i.e., life satisfaction).

Considering that functionality appreciation is suggested to vary across men and women, previous research has employed measurement invariance (MI) to investigate the stability of FAS psychometric properties across these groups [3; 31; 32; 33; 34]. Results indicate that FAS is invariant across males and females in adult samples from Italy, Malaysia, Romania, the United States of America (USA), and the United Kingdom (UK; [3; 31; 32; 33; 34]). However, item intercept values (i.e., scalar invariance) were different for men and women in samples from the UK and Malaysia [34]. Thus, specific items (i.e., “*I acknowledge and appreciate when my body feels good and/or relaxed*”) are suggested to function differently for men and women [34].

Overall, the FAS presents seemingly sound psychometric properties using Classical Test Theory (i.e., confirmatory factor analysis, measurement invariance, etc.). However, FAS psychometric properties have yet to be investigated using newer methodologies such as Item Response Theory (IRT).

Item Response Theory (IRT).

Previous literature has indicated that IRT outperforms CTT [13] psychometric estimation for two main reasons [16]. Firstly, whilst CTT explains associations among items and a construct, IRT facilitates examining associations between the construct and individuals with different levels of the latent trait (i.e., item-participant relationships; [6; 13; 14; 19]). Secondly, unlike CTT, IRT can estimate reliability coefficients at the test and item level [16]. Analysing reliability coefficients at the item level can provide greater insights into measurement reliability, enabling a robust evaluation of internal construct and item validity [13; 24].

In IRT contexts, the item-participant relationship is represented by the probability that participants with a certain level of the latent trait (in this case FA) will endorse a particular item [16]. This is graphically represented by the item response function (IRF; [16]) through a nonlinear (logit) regression line. The probability that a participant will respond to a particular item is contingent on several item parameters including difficulty (β) and discrimination (α) [16]. Difficulty (β) specifies the level of the latent trait required where a participant will endorse a specific item or criterion [18; 24]. For example, ‘easier’ items have lower β values and their IRF is displayed closer to the horizontal axis [16]. Discrimination (α) outlines how steeply the rate of a positive response from an individual differs in accordance with their level of the latent trait [18]. Thus, items that are more robustly associated with the latent variable show steeper IRF functions, in turn accurately discriminating different levels of the latent trait (i.e., FA; [24; 36]).

IRT models vary according to the estimated number of parameter logistic (PL; [13]). While Rasch models assume equal α across different items, Graded Response Model (GRM) or Generalised Partial Credit Model (GPCM) assume free estimation of β and α across items [16; 24]. Both GRM and GPCM accommodate ordered polytomous data (i.e., Likert scale); however, GRM has been the preferred model as it enables comparisons across non-adjacent categories of responses (i.e., “*strongly disagree*” and “*strongly agree*”; [18; 24; 36]).

Additionally, differential item functioning (DIF) can be employed to assess whether different groups (i.e., men and women) respond differently to certain items within a scale (i.e., FAS; [24; 26]). According to Camilli and Shepard [9] there are three reasons why IRT methods are more suitable than CTT methods to detect DIF [15]. Firstly, a graphic illustration of the item characteristic curve (ICC) for each group (i.e., men and women) is exhibited, which in turn, increases the comprehensibility of items displaying DIF. Secondly, statistical properties of items are enhanced through IRT (as opposed to CTT), as this method is able to locate where the item functions differently (i.e., discrimination or difficulty). And finally, item parameter estimates derived from IRT are less confounded and influenced with sample specific characteristics [15; 24].

Method

Participants

Upon receiving approval from the Victoria University Ethics Committee, participants were recruited online via a crowd sourcing platform (Prolific.co) and were awarded \$2.50 each for their time. As part of a larger study, 394 participants completed an online survey including the FAS. Omission of items was not allowed by the Qualtrics-setting parameters. These included 216 men and 170 women, whilst eight participants identified as non-binary. These eight participants were excluded in the present analyses targeting gender differences. The remaining participants' age ranged from 18 to 39 years ($M = 27.54$, $SD = 5.58$). Only the 386 full responses were utilised for statistical analyses resulting to a maximum random sampling error of .089 for a 95% confidence interval and .117 for a 99% confidence interval. Most participants undergraduate degree (40.4%), reported Caucasian ethnicity (57.8%), lived in the USA (54.9%), were heterosexual (80.5%), and worked full-time (44.3%).

Measures

The FAS [3] contains seven items rated on a 5-point Likert scale ranging from 1 (*“strongly disagree”*) to 5 (*“strongly agree”*). Scores from each item (i.e., *“I appreciate that my body allows me to communicate and interact with others”*) are averaged, which produces an aggregate score, where higher scores reflect greater functionality appreciation. Table 1 presents a description of the items and descriptive statistics for the current sample. In the present study, the internal consistency of the FAS was acceptable (Cronbach's $\alpha = .93$, McDonald's $\omega = .95$) and consistent with previous research (Cronbach's $\alpha = .86$ [3]).

Table 1.

Descriptive Statistics for FAS 7 items (N = 386)

	Overall				Men	Women
	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	<i>M</i>	<i>M</i>
1. I appreciate my body for what it is capable of doing.	3.47	0.98	-0.46	-0.15	3.52	3.45
2. I am grateful for the health of my body, even if it isn't always as healthy as I would like it to be.	3.68	1.09	-0.67	-0.14	3.60	3.79
3. I appreciate that my body allows me to communicate and interact with others.	3.80	1.12	-0.90	0.10	3.74	3.91
4. I acknowledge and appreciate when my body feels good and/or relaxed.	3.78	1.06	-0.79	0.12	3.78	3.79
5. I am grateful that my body enables me to engage in activities that I enjoy or find important.	3.85	1.08	-0.79	0.01	3.79	3.96
6. I feel that my body does so much for me.	3.59	1.15	-0.61	-0.35	3.63	3.60
7. I respect my body for the functions that it performs.	3.75	1.07	-0.57	-0.43	3.74	3.82

Note. *M* = Mean; *SD* = Standard Deviation

Statistical Analysis

To address the outlined aims, FAS psychometric properties were examined using IRT and DIF analyses via IRTPRO 5.0. Local independence and unidimensionality assumptions were assessed prior to the analysis. Local independence assumes that item scores do not correlate when holding the latent trait constant. This is determined by residual correlations on items with values < .1 as sufficient proof of having met this assumption [13]. Nonetheless, following suggestions outlined in previous research [24; 36], we employed a Confirmatory Factor Analysis (CFA; [6]) to observe relevant fit indices and thus validate the unidimensional structure of the FAS.

Additionally, to assess IRT-GRM goodness of fit we employed marginal likelihood information statistics (M_2 ; [7; 8]) and RMSEA (given that M_2 is sensitive to samples > 200; 25). We also considered (1) the loglikelihood index of fit [13]; (2) the Bayesian Information Criterion (BIC); and (3) the Akaike Information Criterion (AIC), with lower values indicating improved fit [13; 20]. Visual examination was then conducted by the item information function (IIF; [8] and Item Characteristic Curves (ICC; α , β). Test Information Function (TIF) and the Test Characteristic Curve (TCC; [8]) was used to assess reliability and functionality at the construct level. Additionally, we used FAS raw scores of ∓ 2 SD from the mean as suggested cut-off points to identify significant deviations from expected scores [13].

Results

Confirmatory Factor Analysis

First, CFA was performed to verify the unidimensionality of the scale. To address considerations in relation to CFA performance for ordinal data, we employed two estimators (Maximum likelihood with robust errors, MLR; and Weighted Least Square Mean Variance, WLSMV; [22; 29]). As investigated by a CFA, the WEMWBS demonstrated acceptable fit with MLR ($\chi^2 = 38.178$, $df = 14$, $p < .001$, CFI = .979, TLI = .969, RMSEA = .088, SRMR = .023) and WLSMV ($\chi^2 = 31.816$, $df = 14$, $p = .004$, CFI = .972, TLI = .958, RMSEA = .057, SRMR = .023). Figure 1 below illustrates FAS' unidimensionality including item loadings.

Psychometric IRT Properties

Following past recommendations [7; 8], we employed marginal likelihood information statistics with one and two-way marginal table to assess goodness of fit ($M_2[658] = 1478.42$, $p < .001$; $\chi^2_{\text{Loglikelihood}} = 5558.74$; RMSEA = 0.04; BIC = 5975.65; AIC = 5698.74). Given that M_2 is sensitive to sample size, RMSEA was sufficient to determine that our GRM represented good fit to data [25]. The acceptable fit to data provided by the GRM validates that a unidimensional factorial structure appropriately represents the FAS.

Considering item parameters, discrimination power for all items were above the very high range (0 = non discriminative; 0.01-0.34 = very low; 0.35-0.64 = low; 0.65-1.34 = moderate; 1.35-1.69 = high; > 1.70 = very high; 4) between 2.79 (α item 1) and 4.08 (α item 7). The descending sequence of the items' discrimination power is 7, 5, 6, 3, 4, 2, and 1 (see Table 2). Regarding the item difficulty parameters (β), there were fluctuations between the different thresholds across the 7 items. Indicatively, for the first threshold the ascending item sequence of difficulty was 3, 6, 2, 7, 4, 5, and 1, with item 1 as the most difficult item. Considering the fourth threshold, this alternated to 1, 6, 2, 4, 7, 3, and 5. Nonetheless, the threshold difficulty parameters gradually increased between the first and the last threshold across all items (see Table 2 and Figure 2). In sum, IRT showed that: (i) increasing item scores correctly described increasing levels of FA behaviours across all items, (ii) the rate of these increases differs from item to item, and (iii) different thresholds perform differently from item to item considering their level of difficulty.

Table 2.

Item Discrimination, Difficulty, and Loadings of the FAS (N = 386)

Item	Label	a	s.e.	b_1	s.e.	b_2	s.e.	b_3	s.e.	b_4	s.e.
1	FAS_1	2.79	0.37	-2.14	0.26	-1.07	0.14	0.06	0.10	1.30	0.16
2	FAS_2	3.27	0.44	-1.84	0.21	-1.07	0.13	-0.45	0.09	0.70	0.12
3	FAS_3	3.30	0.46	-1.73	0.19	-1.10	0.13	-0.55	0.10	0.49	0.11
4	FAS_4	3.29	0.44	-2.08	0.25	-1.12	0.13	-0.36	0.09	0.68	0.11
5	FAS_5	3.85	0.54	-2.09	0.24	-1.23	0.13	-0.50	0.09	0.46	0.10
6	FAS_6	3.52	0.48	-1.75	0.19	-0.89	0.11	-0.12	0.09	0.81	0.12
7	FAS_7	4.08	0.58	-1.95	0.21	-1.06	0.12	-0.30	0.09	0.52	0.10

Note: a defines the capacity of an item to discriminate between varying levels of FA (θ). β defines the level of behaviour intensity, where subsequent response rates are more probable than their previous rate. S.e. is the standard error.

Considering the items' reliability across the different levels of the latent trait, controlling concurrently for the different levels of items' difficulty, meaningful variations were confirmed. Indicatively, the IIF of items 5 and 7 provided the highest levels of information/reliability, although with some variability (within 1 SD), in the range between 2 SD above and below the mean. The IIFs of items 2, 4, and 6 showed better performance in the range between 2 SD above and below the mean with significant drops in the areas of 3 SD above and below the mean. Items 1 and 3 showed the lowest (yet acceptable) level of reliability in the area between 3 SD below the mean and 2 SD above the mean, with a significant drop for behaviours exceeding 1 SD above the mean (see Figure 3). However, a common feature across *all* items was a reduction in the reliability index after 1 SD.

Considering the performance of the scale as whole, this is visualized by the Test Characteristic Curve (TCC) and the Test Information Function (TIF) figures following. The TCC graph illustrates that the trait of FA inclined steeply, as the total score reported increased (see Figure 3). Considering the information provided by the scale, improved information (TIF) scores were around 1.5 SD below the mean, up to about 1 SD above the mean (see Figure 3). Interestingly, participants with high levels of FAS seemed to endorse the most difficult category (i.e., "*strongly agree*") on all seven items. This can be visualised in the right side of the TCC curve (Figure 4 left panel) suggesting that this test could not accurately assess participants with 2 SD above the mean. Indeed, reliability indices drop significantly after 1.5 SD above the mean (TIF, Figure 4 right panel) as items seem 'too easy' for participants with high levels of the latent trait.

These findings suggest that the FAS provides a sufficient and reliable psychometric measure for assessing individuals in the range between 2 SDs below and 1 SD above the mean. Nevertheless, it may not be an ideal measure for individuals with extremely low (i.e., below 2 SD), or relatively high FA behaviours (i.e., 1 SD above the mean). FA at the levels of 2 SD below and above the mean trait level correspond with mean scores of 13 and 35 respectively, and based on these, it could be suggested as

conditional (before clinical assessment confirmation) diagnostic cut-off points [18]. Accordingly, 0% of the participants scored below 2 SD and 30.5% scored above 2 SD of extreme range and over-confidence in FA. However, this consideration needs to be applied with caution as FAS cannot seem to accurately detect significant differences for individuals with high FA levels.

Considering DIF of FAS across men and women, sources of non-invariance at the item level were not detected. DIF statistics were observed (see Table 3) for all items, with no significant discrepancies across groups (total $\chi^2 p < 0.05$). This indicates that items are invariant across gender groups.

Table 3.

Differential Item Functioning (DIF) Statistics for Graded Items (N = 386)

Item numbers in:											
Group 1 (Men)	Group 2 (Women)	Total χ^2	<i>df</i>	<i>p</i>	χ^2_a	<i>df</i>	<i>p</i>	χ^2_{cla}	<i>df</i>	<i>p</i>	
1	1	4.8	5	0.4451	0.5	1	0.4973	4.3	4	0.3665	
2	2	6	5	0.3037	0	1	0.9624	6	4	0.1959	
3	3	3.7	5	0.5969	2.8	1	0.0975	0.9	4	0.9219	
4	4	2.5	5	0.7702	0	1	0.8876	2.5	4	0.6409	
5	5	1.1	5	0.9558	0.2	1	0.6414	0.9	4	0.9297	
6	6	3.6	5	0.6120	0.4	1	0.5090	3.1	4	0.5349	
7	7	0.7	5	0.9802	0	1	0.9935	0.7	4	0.9452	

Discussion

The present study is the first of this type to employ item response theory (IRT) procedures to assess the psychometric properties of the FAS at both the scale and the item level for an English-speaking sample. Although all items presented with high discrimination capacity, this fluctuated according to the following descending sequence of items 7, 5, 6, 3, 4, 2, and 1. Similarly, items' difficulty parameters differed across the different item thresholds. Considering the FAS as a whole scale, it seems to perform sufficiently and reliably for examining FA levels between 2 SDs below and above the mean. However, this measure of FA may not be ideal for individuals experiencing extremely high FA (scores that lie 2 SD above the mean). Finally, considering the DIF across men and women, all items demonstrated psychometric invariance across groups indicating that FAS measures FA equally in men and women.

Scale and Item Discrimination, Difficulty, and Reliability

The findings from the IRT analysis supported the unidimensionality of the FAS. Considering that IRT principles relate to the identification of most appropriate items for the evaluation of a specific level of a latent trait, items were evaluated and ranked in relation to their discrimination, difficulty, and reliability [16]. We considered various aspects of IRT including discrimination, difficulty, and information functions across thresholds of the latent trait and considering different levels. Specifically, most items yielded very high discriminative power apart from four items. The items that yielded high discrimination were, *“I respect my body for the functions that it performs”* and *“I am grateful that my body enables me to engage in activities that I enjoy or find important”*. This shows that these two items were most distinguishable between high FA and low FA among groups. Specifically, clinicians should be more inclined to focus on items pertaining to respectfulness and gratefulness regarding what one’s body can do with activities that are enjoyable and important between those experiencing high and low levels of FA among gender.

Further, while the level of difficulty of endorsing an item increased between the first (*“strongly disagree”*) and last options (*“strongly agree”*) of the Likert scale, the sequence of item difficulty varied across thresholds. Specifically, the ascending order of endorsed items between the first (*“strongly disagree”*) and second (*“disagree”*) options of the Likert scale was 3, 6, 2, 7, 4, 5, and 1. However, the ascending order of endorsed items between the fourth (*“agree”*) and last (*“strongly agree”*) options of the Likert scale was 1, 6, 2, 4, 7, 3, and 5. This suggests that participants felt more inclined to endorse *“strongly disagree”* or *“disagree”* being appreciative of what the body is capable of doing and feeling that the body does so much. Alternatively, participants felt more inclined to endorse *“agree”* or *“strongly agree”* being appreciative that the body allows them to communicate and interact with others and gratefulness regarding what one’s body can do with activities that are enjoyable and important. Therefore, it is proposed that items should be interpreted differently when conducting clinical assessment of FA. Specifically, more difficult items (i.e., 3, 6, and 2) should be prioritised with those who have high levels of FA, and easier items (i.e., 7, 4, 5, and 1) should be prioritised when focusing on low levels of FA. Lastly, a common feature across all items was a reduction in the reliability index after 1 SD. Thus, when conducting clinical assessment of FA, the FAS may not be reliable with scores beyond 1 SD—this is because the items within the FAS may not have been challenging enough for the subjects (i.e., ceiling effect; [17]).

As illustrated by the TIF scale, enhanced information performance was evident among ± 2 SDs of the mean. Though, appreciable disparities were observed regarding the amount of information precision displayed by each criterion. Particularly, findings outlined that item 7 (*“I respect my body for the functions that it performs”*) yielded the strongest information/reliability between ± 2 SD and ± 1.5 SD of the mean. Items 5 (*“I am grateful that my body enables me to engage in activities that I enjoy or find important”*), and 6 (*“I feel that my body does so much for me”*) yielded a considerable amount of information/reliability between ± 2 SDs of the mean. Last, items 2 (*“I am grateful for the health of my body, even if it isn’t always as healthy as I would like it to be”*) and 1 (*“I appreciate my body for what it is”*)

capable of doing") consistently provided a low amount of information/reliability among ± 3 SDs of the mean. However, the aforementioned items in company with item 3 ("I appreciate that my body allows me to communicate and interact with others") and 4 ("*I acknowledge and appreciate when my body feels good and/or relaxed*") yielded the most information among 2 and 3 SDs below the mean. This demonstrates that items pertaining to: (i) "*I appreciate my body for what it is capable of doing*", (ii) "*I am grateful for the health of my body, even if it isn't always as healthy as I would like it to be*", (iii) and "*I acknowledge and appreciate when my body feels good and/or relaxed*", should take precedence when identifying participants with significantly low levels of FA.

Additionally, the Test Characteristic Curve (TCC) demonstrated an appropriate steepness indicating that FAS clearly identifies increments in FA as the overall score increases. This favours FAS as a sufficient psychometric measure for the assessment of individuals with high and low levels of FA. Nonetheless, the instruments performance significantly decreases to differentiate very low (-3 SD) and very high ($+3$ SD) FA levels.

Moreover, considering the DIF analysis, results revealed that all items did not differ between men and women. This supports the MI analysis from previous literature, where non-invariance as the intercept level between men and women did not differ amongst all items [3; 31; 32; 33; 34].

Overall, the FAS presents as a useful questionnaire for clinicians interested in gaining a fine-grained understanding of body image and appreciation of one's body functionality. Specifically, considering the high discrimination capacity that FAS items demonstrate, clinicians can adequately assess and identify participants with different levels of body functionality. However, whilst the FAS is a psychometrically sound scale, its use should be complemented with formal assessment (i.e., clinical interviews) or other tests concurrently (i.e., employment of another scale) given the ceiling effect observed within the study. Therefore, clinicians interested in identifying/assessing individuals with significantly high levels of FA may wish to explore this in further depth with the use of other tools.

Conclusion, Limitations And Further Research

Firstly, IRT analysis, using a GRM determined that the scale meets the assumptions fit to IRT analysis for discrimination and difficulty assessment. Following this, we found differing discriminative power across items with "*I respect my body for the functions that it performs*" and "*I am grateful that my body enables me to engage in activities that I enjoy or find important*" as having the strongest degree of discrimination. These items should be considered to differentially assess high and low levels of FA than other items on the FAS. Item difficulty also indicated that the scale is most reliable at assessing FA in non-clinical populations, but its reliability decreases as scores deviate from the normative levels, particularly at clinically low levels. Future research utilising the FAS should also consider psychological disorder diagnostics and exclude those meeting clinically significant criteria for psychological disorders relating to FA. Alternatively, more discriminative items should be used to assess individuals with an extremely high

or low state of FA as outlined in this study. Results reported from this study provide information for clinicians and researchers to determine the appropriate use of the FAS for their population of interest [24].

This analysis compliments and extends upon existing research [3; 31; 32; 33; 34] and is a worthwhile tool regarding increasing the quality of psychological questionnaires and psychological examination. Notwithstanding the unique innovative influence this study makes to the appraisal of FAS psychometric properties, numerous limitations should be highlighted. The employed sample included adult English speakers from developed countries and may lack a wide generalisability of application to samples involving non-English speakers, youth, and older adults. Additionally, IRT properties may not accurately reflect those experiencing pathological mental illness as a community sample of healthy adults was employed. Future studies may wish to address the shortcomings of the present study to improve and expand upon assessment practices typified by the FAS.

Conclusively, the present findings indicate that FA evaluations and associations within gender based on FAS should be interpreted with caution because of response pattern differences, which affect the metric and the scale properties of the instrument. Moreover, the instrument may not perform well for clinically low and high (+1 SD) FA levels and therefore, its use should be complemented with formal assessment (i.e., clinical interviews). Accordingly, as approximately one third of participants scored above 2 SD and were at risk for presenting FA in the extreme range, further assessment should investigate these underlying causes or traits (i.e., narcissism; [21]) to provide more clarity on excessive levels of heightened FA. Last, items differ considering their suitability to discriminate participants with different levels of the latent trait with certain items.

Declarations

Ethical approval and consent to participate:

Ethics approval granted by the Victoria University Ethics Committee (VUHREC). The current study only involved adult subjects (+18 years old) and informed consent was obtained in all cases. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication:

Not applicable.

Availability of data and materials:

All data generated or analysed during this study are included in this published article [and its supplementary information files]. Data cannot be disseminated/distributed – it must be used to reproduce results only.

Competing Interests:

Nil

Funding:

Nil

Authors' contributions:

JM contributed to the article's conceptualization, project administration, writing of the original draft, methodology, formal analysis, data curation, and reviewing and editing the original and final draft. DZ contributed to project administration, article's conceptualization, data curation, writing of the original draft, methodology, formal analysis, and reviewing and editing the original and final draft.

Acknowledgments:

The authors would like to thank Dr Vasileios Stavropoulos for his unconditional support and guidance.

References

1. Alleva, J. M., Diedrichs, P. C., Halliwell, E., Martijn, C., Stuijtzand, B. G., Treneman-Evans, G., & Rumsey, N. (2018). A randomised-controlled trial investigating potential underlying mechanisms of a functionality-based approach to improving women's body image. *Body Image*, 25, 85–96.
2. Alleva, J. M., Diedrichs, P. C., Halliwell, E., Peters, M. L., Dures, E., Stuijtzand, B. G., & Rumsey, N. (2018). More than my RA: A randomized trial investigating body image improvement among women with rheumatoid arthritis using a functionality-focused intervention program. *Journal of Consulting and Clinical Psychology*, 86(8), 666.
3. Alleva, J. M., Tylka, T. L., & Van Diest, A. M. K. (2017). The Functionality Appreciation Scale (FAS): Development and psychometric evaluation in US community women and men. *Body image*, 23, 28–44.
4. Baker, F. B. (2001). *The basics of item response theory*. For full text: <http://ericae.net/irt/baker..>
5. Barnes, M., Abhyankar, P., Dimova, E., & Best, C. (2020). Associations between body dissatisfaction and self-reported anxiety and depression in otherwise healthy men: A systematic review and meta-analysis. *PLoS ONE*, 15(2), e0229268.
6. Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
7. Cai, L., & Monroe, S. (2014). A New Statistic for Evaluating Item Response Theory Models for Ordinal Data. CRESST Report 839. *National Center for Research on Evaluation, Standards, and Student*

Testing (CRESST).

8. Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological methods*, 16(3), 221.
9. Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Sage.
10. Cash, T. F. (2004). Body image: Past, present, and future. 1-5.
11. Cerea, S., Todd, J., Ghisi, M., Mancin, P., & Swami, V. (2021). Psychometric properties of an Italian translation of the Functionality Appreciation Scale (FAS). *Body Image*, 38, 210–218.
12. Chan, C. Y., Lee, A. M., Koh, Y. W., Lam, S. K., Lee, C. P., Leung, K. Y., & Tang, C. S. K. (2020). Associations of body dissatisfaction with anxiety and depression in the pregnancy and postpartum periods: A longitudinal study. *Journal of Affective Disorders*, 263, 582–592.
13. De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.
14. DeMars, C. E. (2010). A comparison of limited-information and full-information methods in Mplus for estimating IRT parameters for non-normal populations.
15. Diaz, E., Brooks, G., & Johanson, G. (2021). Detecting Differential Item Functioning: Item Response Theory Methods Versus the Mantel-Haenszel Procedure. *International Journal of Assessment Tools in Education*, 8(2), 376–393.
16. Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.
17. Garin O. (2014) Ceiling Effect. In: Michalos A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_296
18. Gomez, R., Vance, A., & Stavropoulos, V. (2018). Test-retest measurement invariance of clinic referred children's ADHD symptoms. *Journal of Psychopathology and Behavioral Assessment*, 40(2), 194–205.
19. Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.
20. Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1–55.
21. Jackson, L. A., Ervin, K. S., & Hodge, C. N. (1992). Narcissism and body image. *Journal of Research in Personality*, 26(4), 357–370.
22. Li, C (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res* 48, 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
23. Linardon, J., Messer, M., Lisboa, J., Newton, A., & Fuller-Tyszkiewicz, M. (2020). Examining the factor structure, sex invariance, and psychometric properties of the Body Image Acceptance and Action Questionnaire and the Functionality Appreciation Scale. *Body Image*, 34, 1–9.
24. Marmara, J., Zarate, D., Vassallo, J., Patten, R., & Stavropoulos, V. (2021). Warwick Edinburgh Mental Well-Being Scale (WEMWBS): measurement invariance across genders and item response theory

- examination. BMC Psychology.
25. Maydeu-Olivares, A. (2014). Evaluating the fit of IRT models. In *Handbook of item response theory modelling*, 129–145, Routledge.
 26. Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., de Graeff, A., Groenvold, M., ... Sprangers, M. A. (2010). Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health and quality of life outcomes*, 8(1), 1–9.
 27. Smolak, L., & Cash, T. F. (2011). *Future challenges for body image science, practice, and prevention*. Guilford Press.
 28. Soulliard, Z. A., & Vander Wal, J. S. (2021). Confirmatory factor analyses of the Body Image-Acceptance and Action Questionnaire and Functionality Appreciation Scale among LGBTQ adults. *Current Psychology*, 40(9), 4278–4286.
 29. Suh, Y. (2015). The performance of maximum likelihood and weighted least square mean and variance adjusted estimators in testing differential item functioning with nonnormal trait distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(4), 568–580.
 30. Swami, V., Furnham, A., Horne, G., & Stieger, S. (2020). Taking it apart and putting it back together again: Using Item Pool Visualisation to summarise complex data patterns in (positive) body image research. *Body image*, 34, 155–166.
 31. Swami, V., Todd, J., & Barron, D. (2021). Translation and validation of body image instruments: An addendum to Swami and Barron (2019) in the form of frequently asked questions. *Body image*, 37, 214–224.
 32. Swami, V., Todd, J., Aspell, J. E., Khatib, N. A. M., Toh, E. K. L., Zahari, H. S., & Barron, D. (2019). Translation and validation of a Bahasa Malaysia (Malay) version of the Functionality Appreciation Scale. *Body image*, 30, 114–120.
 33. Swami, V., Todd, J., Goian, C., Tudorel, O., Barron, D., & Vintilă, M. (2021). Psychometric properties of a Romanian translation of the Functionality Appreciation Scale (FAS). *Body Image*, 37, 138–147.
 34. Todd, J., & Swami, V. (2020). Assessing the measurement invariance of two positive body image instruments in adults from Malaysia and the United Kingdom. *Body image*, 34, 112–116.
 35. Tylka, T. L. (2011). Positive psychology perspectives on body image. In *Body image: A handbook of science, practice, and prevention*, 2nd ed. (pp. 56–64). Guilford Press.
 36. Zarate, D., Marmara J., Potoczny, C., Hosking, W., & Stavropoulos, V. (2021). Body Appreciation Scale (BAS-2): measurement invariance across genders and item response theory examination. BMC Psychology, 9(1), 1–15.

Figures

Figure 1

FAS factorial structure. All items load significantly and saliently on one latent factor (namely FAS).

Figure 2

FAS Item Characteristic Curves ($N = 386$). These plots demonstrate how the probability of endorsing a category of FAS items (i.e., “strongly disagree” to “strongly agree”) change as levels of the latent trait change.

Figure 3

FAS Item Information Function (IIF). These plots demonstrate how reliability indices vary with changes in the latent trait.

Figure 4

FAS Test Characteristic Curve (TCC) and Test Information Function (TIF). The TCC provides a visual representation of expected FAS scores as a function of latent trait levels (i.e., as FAS scores increase, levels of the latent trait increase). The TIF demonstrates the relationship between standard errors and reliability indices (i.e., smaller standard errors result in more information)

Note: Raw score divided by 7.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FunctionalityAppreciationScaleRawDataSupplementaryInformationFile.xlsx](#)