

Evaluation of Validity and Reliability of the Eight-Item Body Shape Questionnaire (BSQ-8C) Indonesian Version: Mixture Rasch Model Approach

Feliana Setiawan, Bryan Andhika, Ananta Yudiarso

Master of Psychology, Faculty of Psychology, Universitas Surabaya

Article Information

Submitted date 21-01-2025
Revised date 03-03-2025
Accepted date 05-03-2025

Keywords:

Body Shape Questionnaire (BSQ); mixture Rasch model; validation.

Kata kunci:

Body Shape Questionnaire (BSQ); mixture Rasch model; validasi.

Correspondence concerning this article should be addressed to Ananta Yudiarso, Raya Kalirungkut Street, Surabaya, East Java, Indonesia 60293. Email: ananta@staff.ubaya.ac.id

Abstract

This research aims to evaluate and validate the eight-item Body Shape Questionnaire (BSQ-8C) instrument in Indonesian using the Rasch model and mixture Rasch model approaches with 408 respondents. The analysis showed that the instrument was unidimensional (eigenvalue 12.5; 60.9%), with high reliability (person reliability .85; item reliability .99). The rating scale functioned well, with the Andrich threshold increasing regularly (-1.21 to 1.47). The item fit analysis showed measure values from -.63 to .73 and PT-Measure Corr from .68 to .77. The Wright map indicated an adequate distribution of respondents' abilities, while differential item functioning analysis showed no significant bias based on gender. The mixture Rasch model results identified two latent classes, reflecting diverse response patterns, and supported the two-class model as the best, based on the decrease in AIC, BIC, and CAIC values. Overall, the Indonesian version of the BSQ-8C showed good validity and reliability in measuring body shape perception among the Indonesian population. The results of the Rasch model analysis showed that the scale implications of this instrument tended to measure respondents with moderate levels of agreement, suggesting that improvements were needed on certain items.

Abstrak

Penelitian ini bertujuan untuk mengevaluasi dan memvalidasi instrumen delapan butir Body Shape Questionnaire (BSQ-8C) dalam bahasa Indonesia menggunakan pendekatan *Rasch model* dan *mixture Rasch model* dengan 408 responden. Analisis menunjukkan bahwa instrumen ini bersifat unidimensional (eigenvalue 12,5; 60,9%), dengan reliabilitas tinggi (reliabilitas *person* 0,85; reliabilitas *item* 0,99). Skala penilaian berfungsi dengan baik, dengan *Andrich threshold* yang meningkat secara teratur (-1,21 hingga 1,47). Analisis kecocokan *item* menunjukkan nilai *measure* -0,63 hingga 0,73 dan PT-Measure Corr 0,68 hingga 0,77. *Wright map* menunjukkan distribusi kemampuan responden yang memadai, sementara analisis *differential item functioning* tidak menunjukkan bias signifikan berdasarkan jenis kelamin. Hasil *mixture Rasch model* mengidentifikasi dua kelas laten, yang mencerminkan pola respons yang beragam, dan mendukung model dua kelas sebagai yang terbaik, berdasarkan penurunan nilai AIC, BIC, dan CAIC. Secara keseluruhan, BSQ-8C versi Indonesia menunjukkan validitas dan reliabilitas yang baik dalam mengukur persepsi bentuk tubuh di antara populasi Indonesia. Hasil analisis *Rasch model* menunjukkan bahwa implikasi skala instrumen ini cenderung mengukur responden dengan tingkat persetujuan sedang, yang menunjukkan bahwa diperlukan perbaikan pada item-item tertentu.



INTRODUCTION

Body dissatisfaction is a significant psychological issue that has a wide impact on individual well-being, especially related to low self-esteem, unhealthy behaviors, and the risk of mental health disorders such as anxiety and depression (Cash & Smolak, 2012). In the digital era, beauty standards are often unrealistic and based on ideal body images, which are massively exposed through various online platforms (Holland & Tiggemann, 2016). Factors such as prevailing beauty standards in society, media representation of ideal body shapes, and social influences from family and friends play an important role in influencing an individual's body perception (Tiggemann, 2011). In Indonesia, social and cultural changes have also increased pressure on individuals, especially adolescents and young adults, to meet certain expectations regarding body image.

This dissatisfaction with body shape often not only impacts mental health but also affects physical behavior, such as extreme dieting and excessive exercise (Grogan, 2017), associated with increased body image disturbances and unhealthy eating behaviors, such as extreme dieting, eating disorders, and low self-esteem. Psychological approaches such as the self-discrepancy theory explain that body shape dissatisfaction arises when there is a discrepancy between a person's ideal and actual body image (Higgins, 1987). In addition, the objectification theory argues that individuals often judge their bodies based on external perspectives, resulting in social pressure that worsens body dissatisfaction (Fredrickson & Roberts, 1997).

To understand this phenomenon in depth, a valid and reliable measuring instrument is needed to assess body shape dissatisfaction in Indonesia. The Body Shape Questionnaire (BSQ) is one of the widely used measurement instruments internationally (Cooper et al., 1987). The BSQ is a measuring instrument designed to measure body shape dissatisfaction. This instrument was first developed by Cooper et al. (1987) and is widely used in various countries due to its reliability and validity. The BSQ assesses an individual's emotional experience of negative perceptions of body shape, including cognitive and behavioral dimensions often associated with body image disturbances and unhealthy eating patterns. During its development, a shorter version, such as the BSQ-8C, was created to facilitate its use without compromising measurement quality (Evans & Dolan, 1993; Pook et al., 2008).

Adapting the Indonesian version of the BSQ-8C is crucial to ensure that the questions in the instrument are culturally appropriate and sensitive to local values, thereby enhancing the accuracy of the measurement results. In addition, the use of the BSQ-8C can help identify mental health problems related to body image that the Indonesian population may face. Therefore, the Indonesian version of the BSQ-8C can potentially be a valuable tool in mental health research and intervention in Indonesia (Cooper et al., 1987; Evans & Dolan, 1993).

Previous validity testing of the eight-item version of the Body Shape Questionnaire (BSQ-8C) instrument has been widely conducted using traditional psychometric approaches such as confirmatory factor analysis (CFA) and internal consistency reliability. For example, Rosen et al. (1996) showed that the BSQ-8C has good construct validity with factor loadings ranging from .65 to .84, indicating that the BSQ-8C items consistently reflect the dimensions of body shape dissatisfaction. Additionally, the reliability test showed a Cronbach's alpha value of .93, indicating a high level of internal consistency (Pook et al., 2008). The use of the BSQ-8C in various populations also demonstrates the instrument's adaptability, as shown by Cooper et al. (1987), who found that the BSQ-8C could distinguish between individuals at high and low risk for eating disorders. These results support the idea that the BSQ-8C is a valid and reliable instrument for measuring body perception. However, more modern analysis methods, such as the Rasch model, are still needed to reveal psychometric qualities in more detail.

However, the BSQ must be re-evaluated to ensure its suitability to the Indonesian cultural context. In the Indonesian cultural context, where media and varying social norms often influence beauty standards, the use of the BSQ-8C may provide deeper insight into body perception among individuals. As suggested by the Standards for Educational and Psychological Testing from the American Educational Research Association (AERA, 2014), instruments used in cross-cultural environments must meet the requirements of validity, reliability, and fairness to produce accurate measurements. Therefore, this research aims to evaluate the validity and reliability of the Indonesian version of the BSQ-8C using modern approaches, such as the Rasch model and the mixture Rasch model.

Calibration or adjustment of instruments, such as the BSQ-8C, using the Rasch model or the mixture Rasch model approach is crucial to ensure that each item in the questionnaire functions consistently and accurately across the spectrum of respondent abilities. In the Rasch model, calibration aims to set item parameters to reflect individual differences in ability. This enables the identification and revision of items that are not functioning properly, thereby enhancing the validity and reliability of the measurement. By conducting proper calibration, the instrument will be more sensitive in capturing variations in body image perceptions that cultural and social factors may influence in Indonesia, providing more representative results to support evidence-based interventions in mental health (Cooper et al., 1987; Evans & Dolan, 1993).

The validity of the Indonesian version of the BSQ-8C can be tested through item-total correlation (ITC) analysis, which evaluates the extent to which each item is correlated with the total score of the instrument as a whole. A high ITC value indicates that the item significantly contributes to measuring the intended construct, with a value greater than .30 generally considered adequate for item validity (Nunnally & Bernstein, 1994). In validating the BSQ-8C in international populations, Cooper et al. (1987) reported ITC ranging from .50 to .80, indicating that each item consistently measures the dimensions of body dissatisfaction. Similar research by Pook et al. (2008) showed that low ITC on certain items can indicate the need for revision of the item or deletion to improve the overall quality of the instrument. Thus, ITC analysis on the Indonesian version of the BSQ-8C is an important step to ensure that each item contributes to the accuracy and consistency of measuring body dissatisfaction in the local cultural context.

This research evaluates the validity and reliability of the Indonesian version of the BSQ-8C using the Rasch model and the mixture Rasch model approaches to answer these needs. The Rasch model approach provides an objective and linear measurement framework by examining item fit, reliability, and item difficulty distribution in relation to respondent ability (Bond et al., 2021). Meanwhile, the mixture Rasch model can identify heterogeneity in response patterns within a population and classify respondents into multiple latent classes, enabling a deeper analysis of data variation (Nylund et al., 2007; von Davier & Carstensen, 2007).

In this research, the relevant validity is the internal structure validity, which shows the unidimensionality of the Indonesian version of the BSQ-8C. It is essential to ensure that all items in the questionnaire consistently measure the same construct: body image perception. This research employed Cronbach's alpha coefficient, which assesses reliability by considering both the reliability of items and persons, as well as the interaction between items and persons. This measurement provides strong evidence of the instrument's consistency and accuracy in the Indonesian cultural context (AERA, 2014).

This research aimed to evaluate the validity and reliability of the Indonesian version of the BSQ-8C and identify heterogeneity of response patterns using the mixture Rasch model. It is expected to make a significant contribution to the development of psychological measurement instruments tai-

lored to the Indonesian cultural context. Furthermore, the results of this research can help psychology practitioners and researchers understand the dynamics of body shape dissatisfaction in young Indonesian adults and develop more effective interventions.

METHODS

Respondents

The research population was young adults aged 18–25 throughout Indonesia. The sample consisted of 408 respondents selected through a convenience sampling technique. Respondents comprised 99 males (24.3%) and 309 females (75.7%). Based on marital status, 128 respondents (31.4%) were dating, 10 respondents (2.4%) were married, and 270 respondents (66.2%) were single. The inclusion criteria included respondents aged 18–25 years, residing in Indonesia, and willing to participate.

This research employed non-random sampling techniques, specifically convenience sampling, which enables data collection from individuals who are readily accessible to researchers. This technique was chosen because of its ease in reaching a large number of respondents in a relatively short time. The determination of the number of samples took into account a margin of error of 4.80%, as calculated using Raosoft software. The number of respondents recommended by Raosoft is approximately 400 people, ensuring that the research results are more representative (“Sample Size Calculator,” 2004). This shows an adequate level of precision for generalizing the results to the target population.

Although the convenience sampling technique was used in this research, data analysis can be optimized with the Rasch model approach to validate the measurement and identify the level of item difficulty based on the logit. The Rasch model converts ordinal data into an interval scale, allowing for a more precise analysis of respondent ability and item characteristics in the study (Bond et al., 2021). By using logit as the unit of measurement, this model provides a more in-depth picture of the distribution of respondent ability compared to the distribution of item difficulty levels, thereby strengthening the validity of the research results. This approach is relevant to minimize bias that may arise from the non-random sampling technique used.

Instrument

This research employed a quantitative research design, utilizing psychometric analysis methods based on the Rasch model and the mixture Rasch model. This design was used to evaluate the validity, reliability, and item performance of the short version of the BSQ-8C, which consists of eight items. The Rasch model approach enables the objective evaluation of item and respondent consistency, while the mixture Rasch model is employed to identify patterns of heterogeneity within the population (Bond et al., 2021).

The Rasch model is often referred to as the Rasch unidimensional model because it operates under the assumption of unidimensionality, that is, that each set of measured items represents the same psychological construct or dimension (Wright & Stone, 1979). In this context, unidimensionality refers to the uniformity of the relationship between the items and the measured attributes, thus ensuring that the scores obtained reflect only variability on one specific dimension. This assumption enables a more transparent and in-depth interpretation of the data, particularly in psychometric measurements, as the results are free from the bias introduced by other dimensions.

The advantage of the Rasch unidimensional model lies in its ability to produce accurate interval measures from ordinal data while ensuring that items and respondents are on the same scale. This is done by modeling the probabilistic relationship between individual ability and item difficulty, expressed in logits. Thus, the Rasch model not only allows for the evaluation of consistency between items but also detects potential misfits that could compromise the validity of the measurement instru-

ment. This approach is crucial in psychometric research, particularly in the development and validation of psychological scales (Bond et al., 2021).

Other terms, such as the finite mixture Rasch model or latent class Rasch model in the scientific literature, often refer to the mixture Rasch model. This highlights the nature of the model, which combines Rasch analysis with a finite mixture approach or latent classification. This model is designed to identify heterogeneity in the population that cannot be explained by the simple unidimensionality assumption of the traditional Rasch model (von Davier & Carstensen, 2007). This approach is useful for capturing the presence of latent subgroups in the data, which may have unique response patterns or characteristics. By modeling data into multiple latent classes, the mixture Rasch model offers flexibility in uncovering complex population dynamics, particularly when additional factors, such as cultural differences, age groups, or specific psychological conditions, influence individual response patterns (Pook et al., 2008).

Table 1.
Blueprint of the Indonesian Version of the BSQ-8C Measurement Tool

Code	Questions	
	Bahasa Indonesia	English
4	Apakah Anda pernah merasa khawatir bahwa Anda dapat menjadi gemuk (atau bahkan menjadi lebih gemuk)?	Have you ever worried that you might gain weight (or even put on more)?
6	Apakah Anda pernah merasa gemuk setelah merasa kenyang (seperti setelah makan dalam jumlah banyak)?	Have you ever felt fat after feeling full (such as after eating a large meal)?
13	Apakah memikirkan bentuk tubuh Anda pernah mengganggu fokus Anda dalam melakukan aktivitas (contoh: menonton televisi, membaca, atau mendengarkan percakapan)?	Has thinking about your body shape ever distracted you from focusing on activities (e.g., watching television, reading, or listening to conversations)?
16	Pernahkah Anda membayangkan untuk mengoperasi bagian tubuh Anda yang berlemak?	Have you ever imagined having surgery on a fatty part of your body?
19	Apakah Anda pernah merasa sangat amat besar dan bulat?	Have you ever felt very big and round?
23	Apakah Anda pernah berpikir bahwa (tubuh) Anda dalam bentuk sekarang ini terjadi karena Anda kurang (memiliki) pengendalian diri?	Have you ever thought that your current shape may be due to a lack of self-control?
29	Apakah melihat bayangan Anda (contoh: di depan kaca, atau jendela toko) pernah membuat Anda merasa tidak bahagia dengan bentuk (tubuh) Anda?	Has looking at your reflection (e.g., in a mirror or a shop window) ever made you feel unhappy with your shape?
33	Apakah Anda lebih merasa terganggu akan bentuk tubuh Anda saat Anda bersama orang lain?	Are you more self-conscious about your body shape when you are with others?

Data were collected using the BSQ-8C questionnaire, which was distributed online via Google Forms. This instrument uses a 6-point Likert scale, ranging from *never* (tidak pernah) to *always* (selalu), to measure body shape dissatisfaction based on an individual's emotional perception (Pook et al., 2008). Before being distributed, this instrument was tested on a small scale to ensure respondents understood it.

The use of a 6-point Likert scale in the BSQ-8C aims to capture more detailed response variations related to the level of individual dissatisfaction with body shape. This scale avoids middle options, thus encouraging respondents to provide a more assertive assessment of their emotional experiences (Chyung et al., 2017). Without a neutral option, the data obtained can provide a clearer picture of the level of extremity of individual perceptions of their body shape. This is important in the context of research related to body image disorders, where sensitivity to emotional and perceptual nuances can influence the results of the analysis (Joshi et al., 2015).

Data Analysis

Data analysis techniques were employed with the help of Winsteps 5.0 for Rasch model analysis and Jamovi 2.0 from The Jamovi Project (2021) for mixture Rasch model analysis. The analysis includes unidimensionality, item and respondent reliability, item fit analysis through infit and outfit mean square (MNSQ) values, and respondent distribution mapping using Wright maps. Additionally, the mixture Rasch model analysis was employed to identify heterogeneous response groups (Linacre, 2020; von Davier & Carstensen, 2007). Winsteps 5.0 enables in-depth evaluation of item and scale performance, while Jamovi 2.0 offers an intuitive interface for identifying heterogeneity patterns in respondent data.

This approach ensures a comprehensive evaluation of the validity and reliability of the BSQ-8C instrument. The Rasch model provides an objective framework for measuring item characteristics, while the mixture Rasch model supports the exploration of hidden patterns that may exist in heterogeneous populations (Bond et al., 2021).

RESULTS

Unidimensionality

Unidimensional analysis aims to ensure that the Indonesian version of the BSQ-8C instrument measures a single main construct: dissatisfaction with body shape. In the Rasch model approach, unidimensionality is a basic assumption that must be met, where all items in the instrument are expected to contribute to one particular dimension or construct (Linacre, 2020). An analysis of the raw variance explained by measures and unexplained variance in the first contrast was conducted to test this assumption. Raw variance explained by measures represents the proportion of total variance explained by the main construct, while the unexplained variance in the first contrast reflects the remaining variance that the main model cannot explain (Bond et al., 2021).

The analysis results indicate that the raw variance explained by the measures on the Indonesian version of the BSQ-8C has an eigenvalue of 12.5, accounting for 60.9% of the total variance. This value indicates that most of the variance in the data can be explained by the main construct, namely body dissatisfaction. This percentage of variance exceeds the threshold of 50%, indicating that this instrument has good conceptual clarity to measure one construct (Reckase, 2009). Thus, these results support the internal validity of the BSQ-8C as a measuring instrument that focuses on the dimensions of body dissatisfaction. Additionally, the analysis of unexplained variance in the first contrast reveals an eigenvalue of 1.8, accounting for 9.0% of the total variance. This value is below the threshold of 15%, indicating that there are no significant additional dimensions in the data (Smith, 2002). The low unex-

plained variance confirms that the Indonesian version of the BSQ-8C is unidimensional, following the standards set out in Rasch’s theory (Linacre, 2020).

Unidimensionality is an important factor in ensuring measurement consistency in a psychometric context. Unidimensional instruments can provide more accurate and relevant results because they are not influenced by external factors unrelated to the main construct being measured (Boone et al., 2014). Therefore, these results support the use of the Indonesian version of the BSQ-8C as a valid measuring instrument for evaluating body shape dissatisfaction in research and clinical assessments. Overall, the Indonesian version of the BSQ-8C has met the unidimensionality criteria required for psychometric instruments. The analysis results indicate that this instrument has a strong and consistent structure in measuring one main dimension. Thus, the BSQ-8C can be used as a reliable measuring instrument to assess body shape dissatisfaction in the Indonesian population.

Rating Scale

A rating scale analysis was conducted to evaluate the performance of the scale categories in the Indonesian version of the BSQ-8C. This evaluation aims to ensure that each scale category used can clearly and consistently distinguish respondents’ responses. The results of the analysis include the frequency distribution of observations, the average response value (observed average), the fit of the category to the model (infit and outfit MNSQ), and the category threshold (Andrich threshold). These indicators provide important information about how respondents use the scale and whether the scale is functioning optimally (Linacre, 2020).

Good scale category performance reflects the respondent’s ability to understand the differences between categories and use the scale according to the intensity of their body dissatisfaction. If the scale functions well, the observed average value should increase gradually with higher categories, and the infit and outfit values should be within acceptable limits (.60–1.40). Additionally, regular category thresholds suggest that each category has sufficient space to accurately distinguish individual responses (Bond et al., 2021). These results serve as a basis for evaluating the reliability of the response scale in the BSQ-8C instrument.

Table 2.
Rating Scale Results

Level	<i>f</i>	%	Category Measure	Infit MNSQ	Outfit MNSQ	Andrich Threshold
1 <i>Never (Tidak Pernah)</i>	623	19	-2.58	1.22	1.24	NONE
2 <i>Infrequently (Jarang)</i>	582	18	-1.18	1.00	1.07	-1.21
3 <i>Sometimes (Kadang-Kadang)</i>	498	15	-.39	.89	1.01	-.44
4 <i>Frequently (Sering)</i>	566	17	.27	.87	.80	-.23
5 <i>Very Frequently (Sangat Sering)</i>	573	18	.40	.88	.88	.40
6 <i>Always (Selalu)</i>	422	13	1.47	1.08	1.09	1.47

The results of the rating scale analysis on the Indonesian version of the BSQ-8C showed adequate scale category performance in differentiating respondents’ responses. This instrument employs a six-

point Likert scale, with scores ranging from 1 to 6. The distribution of scores across categories varies from 13% to 19% in each category. This balanced distribution proportion indicates that respondents use all categories without any categories being ignored. This condition is important to ensure the scale's validity, where each category provides relevant information about respondents' responses (Linacre, 2020).

Category measures showed a consistent upward trend with increasing category scores. This trend reflects a logical progression, with higher categories accurately indicating greater levels of body dissatisfaction. This confirms that the responses obtained reflect the intensity of respondents' experiences in a structured manner, consistent with the basic principles of psychometric scale construction (Wright & Masters, 1982). This progressive trend also supports the theoretical fit between the dimensions measured and how respondents responded to the scale.

Statistically, the infit MNSQ values range from .88 to 1.22, while the outfit MNSQ values range from .88 to 1.24. This range is within the acceptable range of .60–1.40, indicating that the scale categories function as expected without any significant anomalies (Bond et al., 2021). These values indicate that no items or categories cause a misfit to the model, indicating a good fit between the empirical data and the theoretical model used in this analysis (Smith, 2002).

The optimal function of the scale categories also ensures that respondents' responses can be interpreted validly and reliably. With no indication of problematic categories, these results suggest that the BSQ-8C scale can be used to measure body dissatisfaction in the Indonesian population accurately. In addition, these results provide empirical evidence that the Likert scale with six categories is adequate to capture the diversity of emotional responses measured by the BSQ-8C (Boone et al., 2014).

Reliability

Reliability is one of the leading indicators in psychometrics used to evaluate the consistency and reliability of an instrument in measuring the intended construct. High reliability indicates that the measurement results are reliable and have a minimal level of random error, thereby providing confidence that the instrument can be used consistently across various contexts (Linacre, 2020). Reliability is a crucial aspect of measuring body shape dissatisfaction, ensuring that instruments such as the BSQ-8C yield stable and valid results in the target population. Reliability evaluation also provides a strong basis for further analysis regarding the instrument's validity (Bond et al., 2021).

Reliability analysis in this research includes measuring person and item reliability, as well as person separation and item separation values. Person reliability measures the extent to which an instrument can differentiate between respondents in terms of ability or body dissatisfaction levels, while item reliability assesses the quality of items in supporting measurement (Wright & Masters, 1982). The separation value indicates the instrument's ability to group respondents or items based on their level of ability or difficulty in logit measurement. The results of the reliability analysis not only reflect the technical quality of the instrument but also provide insight into the extent to which the instrument can function optimally in various population groups (Linacre, 1994).

Evaluation of the reliability of the Indonesian version of the BSQ-8C is a crucial step in ensuring that this instrument consistently measures the body dissatisfaction construct in the target population. With high person and item reliability values, this instrument is expected to separate respondents based on their level of body dissatisfaction accurately. In addition, a good separation value indicates the ability of the BSQ-8C to identify group differences based on respondents' abilities or item difficulty levels. These findings not only support the use of the BSQ-8C in psychological research but also

contribute to the validation of culturally relevant instruments in the Indonesian context (Bond et al., 2021).

Table 3.
Reliability Results

Pearson Measure (<i>N</i> = 408)		Item Measure (<i>N</i> = 8)	
Reliability	Separation	Reliability	Separation
.85	2.38	.99	8.21

The reliability analysis results of the Indonesian version of the BSQ-8C show that this instrument performs very well in measuring dissatisfaction with body shape. The respondents' reliability value obtained was .85, indicating the consistency of responses between respondents when answering the items on this instrument. This value indicates that the data generated from respondents can be relied upon to measure the level of body dissatisfaction consistently. According to Linacre (2020), a reliability value above .80 is considered good, indicating that the BSQ-8C has a high level of reliability as a measuring instrument.

The person separation value of 2.38 indicates that the Indonesian version of the BSQ-8C instrument can differentiate groups of respondents based on their level of body dissatisfaction. However, it has not reached the optimal level. According to Bond et al. (2021), the ideal separation value should be above 3 to ensure the instrument can consistently separate respondents into more than two clear levels. A separation value below this optimal threshold may indicate that the research sample tends to be homogeneous, suggesting that the variability in respondents' ability to measure body dissatisfaction is limited. Another factor that may influence this is the quality of the items in the instrument, which are less capable of capturing more diverse levels of ability (Linacre, 2020).

The high homogeneity of the sample, as evident in the dominance of female respondents (75.7%) and young adult age (18–25 years), may be one of the reasons for the low person separation value. A population with more heterogeneous characteristics, such as variations in age, gender, or cultural background, can help increase variability in respondents' abilities, thereby improving the separation value (Boone et al., 2014). Therefore, further research is recommended to recruit more diverse samples, ensuring the instrument can be applied more widely and provide more valid results for various populations.

Regarding item quality, the item reliability value is very high, which is .99. This value indicates that the items in this instrument have very good quality in supporting measurement consistency. This value also shows that the measurement model is very stable, even when applied to different samples. Support for this result is also obtained from the item separation value of 8.21, which demonstrates the instrument's ability to differentiate the difficulty levels between items optimally. According to Bond et al. (2021) and Boone et al. (2014), an item separation value above 2.0 is sufficient to indicate that the instrument has good measurement quality. Therefore, the value of 8.21 is far above the minimum standard.

This analysis also shows that each item in the BSQ-8C provides a significant contribution to measuring the construct of body dissatisfaction. With high reliability and separation values, no problematic or irrelevant items were found with the main construct. These results strengthen the instrument's validity as a comprehensive and efficient measure of a specific psychological dimension, namely body dissatisfaction (Wright & Masters, 1982).

Item Misfit

Item fit is one of the important aspects of psychometric analysis, which aims to evaluate the extent to which empirical data from each item aligns with the measurement model used. In the context of the Rasch model, item fit analysis ensures that each item in the instrument can measure the intended construct consistently and validly (Linacre, 2020). This evaluation provides a solid foundation for assessing the individual performance of each item on the scale, enabling the instrument to be utilized optimally in research or practice. Therefore, item fit analysis is an essential step in the validation process of psychometric instruments, including the Indonesian version of the BSQ-8C (Bond et al., 2021).

The main parameters in item fit analysis include infit and outfit MNSQ, which indicate the consistency between respondents' responses and the theoretical model. The ideal MNSQ value ranges from .60 to 1.40, where values outside this range indicate that an item may be too noisy or irrelevant in measuring the intended construct (Wright & Masters, 1982). Additionally, the evaluation of point-measure correlation (PT-Measure Corr) is used to assess the relationship between item scores and the total score of the construct. A low PT-Measure Corr value may indicate that an item does not provide an optimal contribution to the overall measurement (Bond et al., 2021).

The results of the item fit analysis provide important insights to assess the reliability and validity of each item in the Indonesian version of the BSQ-8C instrument. Items with inappropriate MNSQ or PT-Measure Corr values can serve as the basis for revising or deleting items to enhance the overall quality of the instrument. In addition to ensuring measurement quality, this analysis also helps maintain the fit between empirical data and the Rasch model, ensuring that measurement results are reliable and relevant within the context of the target population (Linacre, 2020). This approach provides a strong foundation for improving the accuracy of measuring body dissatisfaction in research in Indonesia.

Table 4.
Item Misfit Rasch Model

Item Coding	Measure	Infit		Outfit		PT-Measure Corr
		MNSQ	ZSTD	MNSQ	ZSTD	
4	-.63	1.18	2.4	1.15	1.9	.72
6	-.58	.88	-1.7	.85	-2.1	.76
13	.31	1.11	1.6	1.14	1.8	.71
16	.73	1.39	4.7	1.34	3.6	.68
19	.08	.90	-1.4	.84	-2.2	.77
23	-.11	.77	-3.5	.77	-3.3	.77
29	-.02	.93	-1.0	1.16	2.1	.72
33	.21	.97	-.4	.94	-.7	.74

The results of the item fit analysis using the Rasch model show the item performance of the Indonesian version of the BSQ-8C. The measure values range from -.63 to .73, indicating that the level of item difficulty varies evenly. Item coding 16 has the highest measure value (.73), indicating that this item is more difficult than the other items, while item coding 4 has the lowest measure value (-.63), indicating that it is the easiest item. This distribution ensures that the items cover a sufficient range to measure body dissatisfaction at various levels of respondent ability (Linacre, 2020).

Item fit analysis based on ZSTD values in the table shows that several items have ZSTD values outside the ideal range of -2 to 2, such as item 4 (infit ZSTD = 2.4; outfit ZSTD = 1.9), item 16 (infit ZSTD = 4.7; outfit ZSTD = 3.6), and item 29 (outfit ZSTD = 2.1). ZSTD values that exceed these

limits indicate a potential mismatch between the respondents' empirical response and the Rasch model, which can occur because the item is too sensitive to certain characteristics or contains bias (Linacre, 2020). Conversely, ZSTD values that are too low, such as in item 23 (infit ZSTD = -3.5; outfit ZSTD = -3.3), may indicate that the item lacks variation in its responses, thereby limiting its contribution to measurement (Bond et al., 2021).

Most of the infit and outfit MNSQ values were within the acceptable range (.60–1.40), with an average infit and outfit of 1.02. However, coding item 16 showed infit (1.39) and outfit (1.34) values that were close to the upper limit, indicating potential inconsistency in this item and requiring revision to improve clarity or relevance (Bond et al., 2021), while coding item 23 had infit (.77) and outfit (.77) values, indicating optimal and stable item performance.

The PT-Measure Corr values ranged from .68 to .77, indicating that all items positively correlate with the measured construct. Item coding 16 has the lowest correlation (.68), further supporting the need to revise this item to improve its consistency with the entire instrument. Thus, overall, the Indonesian version of the BSQ-8C exhibits good item performance. However, certain items, such as coding 16, require improvement to provide more optimal and accurate measurements.

Wright Map

Wright map analysis is used to evaluate the distribution of respondent ability levels (person ability) and item difficulty levels (item difficulty) in the Indonesian version of the BSQ-8C. Wright Map visualizes these two aspects on a single logit scale, allowing for the evaluation of how well items can cover various levels of respondent ability. The map's left side displays respondent ability distribution, while the right side displays the item difficulty.

This analysis aims to ensure that the item distribution encompasses the full range of respondent abilities, enabling the instrument to measure body dissatisfaction in individuals with varying abilities, including those with low, medium, and high abilities. If there are gaps in the item distribution or an imbalance between respondent ability and item difficulty, this may indicate the need for item revision or addition (Bond et al., 2021). The Wright map also provides information about items that may be too easy or too difficult, which could potentially affect the instrument's sensitivity. The Wright map results offer valuable insights into how the BSQ-8C accurately captures respondents' overall abilities, providing a basis for enhancing measurement quality by ensuring a balanced and representative item distribution (Linacre, 2020).

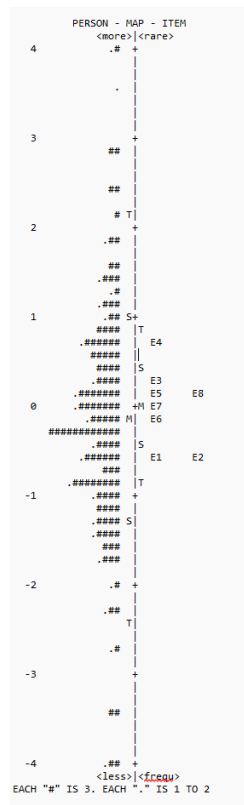


Figure 1.
Wright Map

The Wright map results of the Indonesian version of the BSQ-8C provide an overview of the distribution of respondent abilities and item difficulty levels on the same logit scale. On this map, the vertical axis shows the distribution of respondent abilities (person) on the left and item difficulty levels (items) on the right. The majority of respondents are concentrated around the zero logit value, indicating that the average respondent's ability to answer the questionnaire is at a level that is balanced with the item difficulty level. This indicates that the BSQ-8C aligns with the target population and the level of difficulty of the items being measured (Linacre, 2020).

Item coding 16 is at a higher logit than the other items, indicating that it is more difficult or requires a higher level of body dissatisfaction for respondents to answer affirmatively. In contrast, items coding 4 and 16 are at a lower logit, indicating that these two items are easier and tend to be answered affirmatively by respondents with lower levels of dissatisfaction. This distribution indicates that the BSQ-8C covers a relatively good range of item difficulty levels, enabling it to measure body dissatisfaction among respondents with varying abilities, from low to high.

The Wright map on the Indonesian version of the BSQ-8C shows a fairly adequate distribution of respondents' ability (person ability) and item difficulty levels (item difficulty), with most items being around the respondent's average ability. However, this map also shows that the existing items do not fully cover some individuals with very high or very low ability, as no items are located at the upper or lower extremes on the logit scale. This suggests that the instrument has limitations in measuring body dissatisfaction in individuals who are either very satisfied or very dissatisfied with their body shape (Bond & Fox, 2015). To increase the instrument's reach, additional items that are more challenging or easier are suggested, allowing the instrument to cover the entire spectrum of respondents' abilities more comprehensively (Linacre, 2020). This solution not only expands the scope of measurement but also increases the instrument's validity in a more heterogeneous population.

Differential Item Function (DIF)

A differential item functioning (DIF) analysis was conducted to evaluate whether the items in the Indonesian version of the BSQ-8C function equally well among subgroups of respondents, such as those based on gender or other characteristics. DIF is used to identify potential bias in items where respondents from different subgroups may give different responses at the same ability level. The parameters analyzed in DIF include DIF contrast values, the Mantel-Haenszel chi-square statistic, and Welch’s test.

The DIF contrast value indicates the difference in item difficulty levels between two groups of respondents, with a value greater than .50 logits indicating significant potential bias (Linacre, 2020). The Mantel-Haenszel test provides statistical information to evaluate whether this difference is statistically significant. In contrast, Welch’s test examines the suitability of the data in the context of the respondent group. The results of the DIF analysis are crucial in ensuring that the BSQ-8C can be used reliably across groups in the Indonesian population. If items with potential bias are identified, further revision or adaptation is necessary to enhance the fairness and validity of the instrument (Bond et al., 2021).

Table 5.
Differential Item Function Results by Gender

Item Coding	DIF Contrast	Mantel-Haenszel (prob)	Welch (Prob)
4	.19	1.5302 (.2161)	1.65 (.1001)
6	.04	.1261 (.7225)	.31 (.7580)
13	.00	.0680 (.7943)	.00 (1.000)
16	.02	.0001 (.9941)	.19 (.8514)
19	-.06	.5143 (.4733)	-.49 (.6263)
23	.00	.1135 (.7362)	.00 (1.000)
29	-.03	.0518 (.8200)	-.30 (.7651)
33	-.17	1.9862 (.1587)	-1.47 (.1425)

The results of the DIF analysis from the table show differences in the difficulty level between respondent groups on the Indonesian version of the BSQ-8C. The DIF contrast values for all items range from -.17 to .19, with item 4 having the highest DIF contrast value of .19. However, this value remains within reasonable limits, below the significance threshold of .50 logit (Linacre, 2020), indicating no significant item bias between groups.

The Mantel-Haenszel test yields the highest chi-square value for items 33 (1.9862) and 4 (1.5302), but the probability remains insignificant, indicating that this difference can be disregarded. This value indicates that the performance of items 4 and 33 exhibits slight variations, but these are not large enough to be considered significant bias. Other items, such as items 6, 13, and 16, show very small Mantel-Haenszel values, indicating consistency in performance between respondent groups.

Welch’s test further supports these results, with insignificant t-values for all items. Item 4 has the highest value of 1.65, while item 33 has the lowest value of -1.47. These values indicate that the variations detected through DIF contrast and Mantel-Haenszel do not have a statistically significant impact.

Overall, the results of the DIF analysis indicate that all items in the Indonesian version of the BSQ-8C have fair performance and are free from significant bias between groups. Although items 4

and 33 show small variations, these values do not exceed the significant limit, so this instrument can be said to be consistent in measuring body shape dissatisfaction in various groups of respondents.

Mixture Rasch Model

The mixture Rasch model is a statistical approach that combines the basic principles of the Rasch model with latent group analysis to identify heterogeneity in respondents' response patterns to an instrument. Unlike the classical Rasch model, which assumes a population of respondents as a homogeneous group, the mixture Rasch model enables the identification of latent subgroups with distinct characteristics within the same data (von Davier & Carstensen, 2007). In psychometrics, this model is very useful for uncovering hidden variations, such as differences in response patterns based on certain demographic or psychological factors. Jamovi, a user-friendly statistical software interface, offers a dedicated module for running mixture Rasch model analysis, enabling researchers to easily and efficiently identify latent groups. Using the mixture Rasch model through Jamovi provides additional flexibility in evaluating the validity and reliability of instruments across populations.

Table 6.
Model Fit

Class	AIC	AIC3	BIC	CAIC	Log-likelihood
1	11399	11439	11559	11599	-5659
2	10271	10352	10596	10677	-5055

The results of the model fit analysis in Table 6 indicate that the two-class model has a BIC value of 10.596, which is lower than that of the one-class model (BIC = 11.559). The lower BIC value in the two-class model indicates that this model is better at explaining the respondents' response patterns than the one-class model. However, it is essential to note that the BSQ-8C is not conceptually designed to have subclasses, as it measures a single construct: body dissatisfaction (Rosen et al., 1996). Therefore, the existence of two classes in this mixture Rasch model is more likely to reflect variations in respondent characteristics, such as differences in respondent ability levels or experiences, rather than the existence of subclasses in the instrument structure itself. This suggests that external factors outside the instrument design may influence heterogeneity in respondent response patterns.

The presence of two classes may indicate heterogeneity in the population that a single model does not fully accommodate. Factors such as cultural background, gender, or age can influence response patterns on the BSQ-8C, despite the instrument measuring a single construct (Bond et al., 2021; Linacre, 2020). However, the absence of subclasses suggests that these differences are more related to respondent characteristics than the instrument's structure. The relative homogeneity of the sample, such as the predominance of female respondents and the young adult age group, may also influence the formation of the two classes. Thus, this model indicates the instrument's sensitivity to inter-individual variation but does not suggest the presence of subclasses in measuring body dissatisfaction.

DISCUSSION

Evaluation of the Indonesian version of the BSQ-8C revealed that the responses to the item may be less consistent or not in accordance with the Rasch model, making this item less relevant for accurately measuring body dissatisfaction (Linacre, 2020). Additionally, this evaluation also revealed a low contribution to the overall measurement. This indicates that the item may not be sensitive enough to capture variations in respondent abilities optimally. This mismatch can potentially reduce the instrument's validity in detecting more diverse levels of body dissatisfaction in the target population (Bond & Fox, 2015).

Wright map shows that BSQ-8C items tend to be centered on average ability, with an uneven distribution for respondents with very high or very low levels of body dissatisfaction. The absence of items occupying the upper or lower extremes of the logit scale suggests that this instrument may not be suitable for respondents with extreme abilities. This suggests that the scope of the instrument remains limited, particularly for populations with more heterogeneous levels of body dissatisfaction (Wright & Masters, 1982). This problem can reduce the instrument's sensitivity in studies with broader or heterogeneous populations. Therefore, it is necessary to add items designed to measure extreme abilities to expand the scope of measurement and improve data accuracy.

To overcome this problem, one solution that can be applied is to revise items with inappropriate ZSTD and PT-Measure Corr values. For example, item 16 can be revised or retested on a more diverse population to ensure its relevance and sensitivity in the Indonesian cultural context (Linacre, 2020). Additionally, adding new items with varying levels of difficulty can enhance the instrument's ability to accommodate the entire spectrum of respondent abilities. This approach should also be supported using more heterogeneous samples to identify more varied response patterns. Thus, the Indonesian version of the BSQ-8C can be optimized to provide more comprehensive and valid measurements.

Another step that can be taken is to conduct additional analyses, such as CFA, to verify the instrument's unidimensionality. CFA can help ensure that these issues are not caused by additional dimensions or subclasses in the structure of the BSQ-8C (Rosen et al., 1996). Using additional statistical tools, such as the mixture Rasch model in a more heterogeneous population, can also help better understand the identified response patterns. This will not only strengthen the instrument's validity but also provide stronger empirical evidence regarding the effectiveness of the BSQ-8C in measuring body dissatisfaction. These improvements will ensure that the instrument remains relevant and reliable in various research and clinical applications.

CONCLUSION

Evaluation of the Indonesian version of the BSQ-8C revealed that this instrument possesses adequate validity and reliability in measuring body dissatisfaction among young adults, with unidimensionality analysis supporting its consistency. Although it has a person reliability value of .85 and an item reliability of .99, limitations such as a person separation value of 2.38 and an item distribution that does not reach individuals with extreme ability levels reduce the measurement's sensitivity. Several items, such as item 16, have fit statistics outside the ideal range, indicating a need for revision to suit the local cultural context better. Mixture Rasch model analysis identified two latent classes, reflecting response heterogeneity that needs further exploration. Therefore, improvements such as adding and revising items, as well as conducting additional analyses like CFA, are needed to optimize the instrument. With these improvements, the BSQ-8C can provide more accurate results in research and clinical applications, support a deeper understanding of body image in Indonesia, and serve as a basis for more effective mental health policies and interventions.

REFERENCES

- American Educational Research Association. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3rd ed.). New York: Routledge.
- Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (4th ed.). New York: Routledge.

- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Dordrecht: Springer.
- Cash, T. F., & Smolak, L. (Eds.). (2012). *Body Image: A Handbook of Science, Practice, and Prevention* (2nd ed.). New York: Guilford Press.
- Chyung, S. Y. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement*, 56(10), 15–23. <https://doi.org/10.1002/pfi.21727>
- Cooper, P. J., Taylor, M. J., Cooper, Z., & Fairbum, C. G. (1987). The Development and Validation of the Body Shape Questionnaire. *International Journal of Eating Disorders*, 6(4), 485–494. [https://doi.org/10.1002/1098-108X\(198707\)6:4<485::AID-EAT2260060405>3.0.CO;2-O](https://doi.org/10.1002/1098-108X(198707)6:4<485::AID-EAT2260060405>3.0.CO;2-O)
- Evans, C., & Dolan, B. (1993). Body Shape Questionnaire: Derivation of Shortened “Alternate Forms.” *International Journal of Eating Disorders*, 13(3), 315–321. [https://doi.org/10.1002/1098-108X\(199304\)13:3<315::AID-EAT2260130310>3.0.CO;2-3](https://doi.org/10.1002/1098-108X(199304)13:3<315::AID-EAT2260130310>3.0.CO;2-3)
- Fredrickson, B. L., & Roberts, T.-A. (1997). Objectification Theory: Toward Understanding Women’s Lived Experiences and Mental Health Risks. *Psychology of Women Quarterly*, 21(2), 173–206. <https://doi.org/10.1111/j.1471-6402.1997.tb00108.x>
- Grogan, S. (2017). *Body Image: Understanding Body Dissatisfaction in Men, Women and Children* (3rd ed.). New York: Routledge.
- Higgins, E. T. (1987). Self-Discrepancy: A Theory Relating Self and Affect. *Psychological Review*, 94(3), 319–340. <https://doi.org/10.1037/0033-295X.94.3.319>
- Holland, G., & Tiggemann, M. (2016). A Systematic Review of the Impact of the Use of Social Networking Sites on Body Image and Disordered Eating Outcomes. *Body Image*, 17, 100–110. <https://doi.org/10.1016/j.bodyim.2016.02.008>
- Joshi, A., Kale, S., Chandel, S., & Pal, D. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology*, 7(4), 396–403. <https://doi.org/10.9734/BJAST/2015/14975>
- Linacre, J. M. (1994). Sample Size and Item Calibration (or Person Measure) Stability. *Rasch Measurement Transactions*, 7(4), 328.
- Linacre, J. M. (2020). *A User’s Guide to WINSTEPS® MINISTEP Rasch-Model Computer Programs*. Retrieved from <https://www.winsteps.com/ARM/Winsteps-Manual.pdf>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535–569. <https://doi.org/10.1080/10705510701575396>
- Pook, M., Tuschen-Caffier, B., & Brähler, E. (2008). Evaluation and Comparison of Different Versions of the Body Shape Questionnaire. *Psychiatry Research*, 158(1), 67–73. <https://doi.org/10.1016/j.psychres.2006.08.002>
- Reckase, M. (2009). *Multidimensional Item Response Theory*. New York: Springer.

- Rosen, J. C., Jones, A., Ramirez, E., & Waxman, S. (1996). Body Shape Questionnaire: Studies of Validity and Reliability. *International Journal of Eating Disorders*, 20(3), 315–319. [https://doi.org/10.1002/\(SICI\)1098-108X\(199611\)20:3<315::AID-EAT11>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1098-108X(199611)20:3<315::AID-EAT11>3.0.CO;2-Z)
- Sample Size Calculator. (2004). *Raosoft, Inc.* Retrieved from <http://www.raosoft.com/sampleize.html>
- Smith, E. V. (2002). Detecting and Evaluating the Impact of Multidimensionality Using Item Fit Statistics and Principal Component Analysis of Residuals. *Journal of Applied Measurement*, 3(2), 205–231.
- The Jamovi Project. (2021). *Jamovi (Version 2.0)* [Computer software]. Retrieved from <https://www.jamovi.org/>
- Tiggemann, M. (2011). Sociocultural Perspectives on Human Appearance and Body Image. In *Body Image: A Handbook of Science, Practice, and Prevention* (2nd ed., pp. 12–19). New York: Guilford Press.
- von Davier, M. & Carstensen, C. H. (Eds.). (2007). *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*. New York: Springer.
- Wright, B. D., & Stone, M. H. (1979). *Best Test Design*. Chicago: Mesa Press.
- Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis*. Chicago: Mesa Press.