JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia)
DOI: http://dx.doi.org/10.15408/jp3i.v15.i1.51044
http://journal.uinjkt.ac.id/index.php/jp3i

Adaptation and Validation of the Indonesian Version of Behavioral Activation for Depression Scale – Short Form (BADS-SF) Using Confirmatory Factor Analysis and Rasch Modeling

Salma Salma[1,2], Restu Tri Handoyo[3], Ajeng J. Puspitasari[4], Rahmat Hidayat[3], Rachel Leonard[5], Jonathan Kanter[6]

Psychology Doctoral Program, Faculty of Psychology, Universitas Gadjah Mada, Indonesia[1]
Faculty of Psychology, Diponegoro University, Indonesia[2]
Faculty of Psychology, Universitas Gadjah Mada, Indonesia[3]
Department of Psychological and Brain Sciences, University of Wisconsin-Milwaukee, USA[4]
Guidelight Health, USA[5]
Department of Psychology, University of Washington, USA[6]
restu.tri.handoyo@ugm.ac.id

Abstract

The Behavioral Activation for Depression Scale – Short Form (BADS-SF) is a brief instrument designed to measure activation levels in individuals undergoing behavioral activation (BA) therapy for depression. With the broad potential for implementing BA therapy, including in Indonesia, the BADS-SF needs to be cross-culturally adapted. While the original validation study supported a two-factor model, subsequent adaptations, particularly in non-Western settings, have yielded inconsistent results. In addition, given the practical application of the BADS-SF total score as an overall indicator of behavioral activation, it is essential to evaluate its psychometric properties as a unidimensional measure. The objectives of this study are: (1) to culturally adapt the BADS-SF into the Indonesian language (Bahasa Indonesia); (2) to examine its underlying factor structure and convergent validity; and (3) to evaluate its psychometric properties using item response theory (IRT), specifically Rasch modeling. The cultural adaptation process demonstrated good content validity based on Aiken's V (.75 to 1 for all items) and confirmed the scale readability. The confrmatory factor analysis (CFA) results revealed a different factor structure than the original, with a 6-item one-factor model providing the best fit and confirmed convergent validity. Rasch analysis further showed that the Indonesian version of BADS-SF had good psychometric properties, including excellent item reliability, acceptable person reliability, and supported unidimensionality following the exclusion of a misfitting item (Item 8). Based on the overall results, the 6-item version of BADS-SF Bahasa Indonesia is recommended to be used for both research and clinical practice. Further implications of the findings are discussed.

Keywords: behavioral activation, depression, measurement, Rasch analysis, reliability, validity.

Abstrak

Behavioral Activation for Depression Scale – Short Form (BADS-SF) merupakan instrumen singkat yang disusun untuk mengukur tingkat aktivasi individu yang sedang menjalani terapi behavioral activation (BA) untuk depresi. Dengan potensi luas untuk mengimplementasikan terapi BA, termasuk di Indonesia, BADS-SF perlu diadaptasi secara lintas budaya. Meski studi awal BADS-SF menghasilkan model dua-faktor, sejumlah adaptasi, khususnya di setting non-Barat, menunjukkan hasil yang tidak konsisten. Selain itu, evaluasi properti psikometri BADS-SF sebagai pengukuran unidimensi penting dilakukan mengingat secara praktis skor total BADS-SF digunakan sebagai indikator keseluruhan dari aktivasi perilaku. Penelitian ini bertujuan: (1) melakukan adaptasi budaya BADS-SF ke dalam bahasa Indonesia; (2) menilai struktur faktor BADS-SF dan validitas konvergennya; dan (3) mengevaluasi properti psikometri BADS-SF menggunakan item response theory (IRT), khususnya pemodelan Rasch. Proses adaptasi budaya menunjukkan validitas isi yang baik berdasarkan Aiken's V (0,75 hingga 1 untuk semua item) dan mengkonfirmasi keterbacaan skala. Analisis faktor konfirmatori (CFA) menghasilkan struktur faktor yang berbeda dari versi asli, dengan model satu-faktor enam-item sebagai model terbaik disertai validitas konfergen yang terkonfirmasi. Analisis Rasch lebih lanjut menunjukkan bahwa BADS-SF versi bahasa Indonesia memiliki properti psikometri yang baik, termasuk reliabilitas butir yang sangat baik, reliabilitas person yang cukup baik, serta terkonfirmasinya unidimensionalitas skala setelah mengeksklusi item misfit (Item 8). Berdasarkan keseluruhan hasil, versi BADS-SF Bahasa Indonesia 6-item direkomendasikan untuk digunakan dalam penelitian maupun praktik klinis di Indonesia. Implikasi teoretis dan praktis lebih lanjut dari temuan penelitian didiskusikan.

Kata kunci: *aktivasi perilaku, analisis Rasch, depresi, pengukuran, reliabilitas, validitas*

Introduction

Depression is among the most prevalent mental health disorders, having been studied since the early development of psychological science (Arroisi & Himaya, 2023), and remains a leading contributor to the global burden of disease (Abbafati et al., 2020). This is particularly true in low- and middle-income countries (LMICs) such as Indonesia, where treatment gaps are more pronounced (Luitel et al., 2017; Mekonen et al., 2021; Sagar et al., 2017; World Health Organization, 2022). Behavioral Activation (BA) is an evidence-based intervention for depression supported by several meta-analyses (Cuijpers et al., 2023; Ekers et al., 2014; Mazzucchelli et al., 2009; Stein et al., 2021; Uphoff et al., 2019, 2020) and recommended by international health organizations, such as the World Health Organization (WHO) and the National Institute for Health and Care Excellence (NICE) (National Institute for Health and Clinical Excellence, 2022; World Health Organization, 2016). Due to its theoretical parsimony, cultural adaptability, and ease of implementation, BA has been recommended for dissemination in low-resource settings and diverse cultural contexts (Anvari et al., 2023; Kanter et al., 2012; Kanter & Puspitasari, 2016; Lehmann & Bördlein, 2020).

Given the broad potential implementation of BA, including in Indonesia, valid and culturally appropriate measures of behavioral activation are required. To address these need, the present study aimed to: (1) conduct a cross-cultural adaptation of the BADS-SF following the International Test Commission Guidelines; (2) examine underlying factor structure of the Indonesian version of the BADS-SF in an adult sample with depressive symptoms using CFA and its convergent validity; and (3) evaluate its psychometric properties using Item Response Theory, specifically Rasch modeling, including its unidimensionality.

BA is grounded in the behavioral model of depression (Ferster, 1973), which conceptualizes depressive symptoms as a consequence of reduced access to positive reinforcement from the environment, often resulting from adverse life events and maladaptive response patterns (Dimidjian et al., 2008). Individuals with depression typically exhibit diminished interest in daily activities and social withdrawal. BA seeks to interrupt this cycle by systematically increasing behavioral engagement, thereby enhancing contact with sources of environmental reward and facilitating problem-solving (Kanter et al., 2012). A range of behavioral strategies are employed in BA, including self-monitoring, activity scheduling and structuring, problem-solving, and social skills training (Dimidjian et al., 2008; Dimidjian et al., 2011).

To monitor therapeutic progress of BA therapy and support the theoretical underpinnings of BA, Kanter et al. (2007) developed the Behavioral Activation for Depression Scale (BADS), a 25-item measure comprising four subscales: Activation, Avoidance/Rumination, Work/School Impairment, and Social Impairment. The BADS has since been adapted into multiple languages and validated across diverse populations (Krings et al., 2021; Mohammadi & Amiri, 2010; Oh et al., 2017; Raes et al., 2010; Teismann et al., 2016). In an effort to improve its brevity and practical utility while maintaining sound psychometric properties, Manos et al. (2011) developed the BADS – Short Form (BADS-SF), which consists of nine items and two subscales: Activation and Avoidance. The BADS-SF has also been translated and validated in various populations (Fuhr et al., 2016; González-Roz et al., 2018; Koşan & Kaya, 2023; Wagener et al., 2015).

Despite its increasing use, empirical gaps remain in the measurement of behavioral activation, particularly in LMICs and non-Western settings. Studies in these contexts are scarce, and findings regarding the optimal factor structure of the BADS-SF remain inconsistent, especially with respect to item-level factor loadings (Fuhr et al., 2016; González-Roz et al., 2018; Koşan & Kaya, 2023; Wagener et al., 2015). Furthermore, although the BADS-SF is often used as a unidimensional measure of activation, to the best of our knowledge, none of the existing studies have rigorously examined its unidimensionality using item response theory approaches such as Rasch modeling (Sumintono & Widhiarso, 2013). Most existing studies rely solely on confirmatory factor analysis (CFA) to test the two-factor structure and assess measurement invariance. Although BADS-SF was originally

conceptualized as comprising two factors (activation and avoidance), both factors collectively represent overall behavioral activation, with avoidance items reverse-scored in the total scale (Manos et al., 2011). Conceptually, avoidance reflects behavioral inactivation, as avoidant behavior functions as a primary impediment to meaningful engagement with reinforcing activities (Jacobson et al., 2001). Therefore, examining the unidimensionality of the BADS-SF is theoretically consistent with the construct of behavioral activation.

Moreover, no study to date has evaluated the BADS-SF at the item level using Rasch analysis. Item-level analysis is critical for deepening our understanding of the latent construct of behavioral activation, the primary target of BA therapy. Given that both classical test theory (CTT) and item response theory (IRT) offer complementary perspectives, employing both frameworks provides a more comprehensive psychometric evaluation (Petrillo et al., 2015). Rasch modeling, in particular, is well-suited for assessing unidimensionality and item-level functioning, as it estimates measurement precision based on the interaction between person ability (or endorsement) and item difficulty. Detailed findings from both CTT and IRT approaches are reported and discussed in light of their theoretical and practical implications.

Methods

To evaluate the reliability and validity of the Indonesian version of the BADS-SF, multiple complementary methods were employed to examine key psychometric properties, including reliability, content validity, construct validity, convergent validity, and item-level functioning. The rationale for employing these methods was to provide a comprehensive psychometric profile of the adapted scale, thereby strengthening theoretical interpretation and informing its practical application. The combined use of qualitative and quantitative approaches to validity evaluation, as well as the integration of classical test theory (CTT) and item response theory (IRT; Rasch modeling), is supported by methodological literature (Petrillo et al., 2015). Detailed descriptions of the methodological procedures are provided in the following sections.

Translation/ Adaptation Procedures

The translation/ adaptation procedures in this study followed the International Test Commission (ITC) guidelines for test adaptation (International Test Commission, 2017). Specifically, a systematic five-step process was employed to adapt the BADS-SF scale into Bahasa Indonesia. First, a forward translation was conducted by two independent bilingual psychologists to ensure accuracy and clarity in rendering the original content into the target language. This was followed by a back-translation by a professional translator, ensuring that the translated version reflected the meaning of the original items. Third, the content validity of the translated items was assessed by five subject matter experts (clinical psychologists), who provided judgments on a 5-point Likert scale. The evaluation was based on four key criteria: semantic, idiomatic, experiential, and conceptual equivalence between the original and translated versions. Fourth, a small-scale pilot study involving 30 participants was conducted, incorporating cognitive debriefing to assess the readability and clarity of the items. Finally, the feedback was integrated to finalize the Indonesian version of the BADS-SF scale, ensuring its suitability for use in future research and practice. This final Indonesian version then underwent further reliability and validity analyses.

Participants and Procedures

This study received ethical approval from the Ethics Committee of the Faculty of Psychology, Universitas Gadjah Mada, Indonesia (Approval No. 2798/UN1/FPSi.1.3/SD/PT.01.04/2024). Participants were recruited via online advertisements distributed on social media platforms using convenient sampling. Adults experiencing symptoms of depression or low mood in the past two weeks were invited to participate. Informed consent was obtained electronically before data collection began. Participants were presented with a Google Form containing demographic questions and four study questionnaires, including the BADS-SF and three other questionnaires for convergent validity testing.

As an incentive, participants received electronic payment of 10,000 Indonesian Rupiah (approximately $1 USD) upon completing the survey.

A total of 354 participants completed the survey. The dataset was subjected to outlier detection separately for classical test theory (CTT) and item response theory (IRT) analyses (Anselmi et al., 2019; Field, 2009). For the CTT analysis (i.e., Confirmatory Factor Analysis), boxplot procedures identified nine univariate outliers, resulting in a final analytic sample of 345 participants. For the IRT analysis (i.e., Rasch modeling), person-fit statistics were used to identify misfitting cases whose pattern of responses contradicts the model's expectation. Two rounds of outlier detection were conducted, leading to the exclusion of 159 participants and final sample of 195 participants. The high number ot outliers in Rasch modeling might be due to several reasons, such as random responding, carelessness, and differing interpretation among sub-groups of responders (Sumintono & Widhiarso, 2013; Widhiarso & Sumintono, 2016). While this reduced the sample of Rasch analysis, the exclusion was a deliberate step to ensure the invariance and stability of the item parameter estimates (Mousavi & Cui, 2020). By removing data that does not contribute to the latent trait of behavioral activation, we ensured that the resulting psychometric properties of the Indonesian BADS-SF are based on reliable and logically consistent response behaviors. In addition, it should be noted that the CFA and Rasch analyses were conducted on the same primary dataset (albeit with different outlier exclusion criteria). While this approach allows for a direct comparison of psychometric properties within this specific context, it represents a limitation regarding cross-validation. The lack of an independent sample means that the findings may be partially sensitive to the specific characteristics of this group, and future research should aim to replicate these findings in a separate, independent sample to ensure the generalizability of the Indonesian BADS-SF.

The mean age of participants in the CTT and IRT samples was 24.53 years (SD = 6.24) and 25.54 years (SD = 6.00), respectively. Detailed demographic characteristics for each sample are presented in Table 1.

Instruments

The primary instrument evaluated in this study was the BADS-SF. In addition, participants completed three supplementary questionnaires assessing theoretically related constructs, i.e., depression severity, quality of life, and mindfulness, which were administered to examine convergent validity.

Behavioral Activation for Depression Scale–Short Form (BADS-SF)

The Behavioral Activation for Depression Scale–Short Form (BADS-SF; Manos et al., 2011) is a brief version of the original BADS (Kanter et al., 2007), consisting of nine items designed to assess behavioral activation in individuals experiencing depressive symptoms. Participants rate how well each statement described their condition over the past week, including the day of completion, using a 7-point Likert-type scale ranging from 0 (Not at all) to 6 (Completely true). Items 1, 6, 7, and 8 are reverse-scored when calculating the total score. Higher total and subscale scores indicate greater behavioral activation or avoidance.

Table 1. The Demographic Characteristics of the Study Respondents

Demographic variables	Data set 1 (for CFA)		Data set 2 (for Rasch modeling)	
	Frequency	Percentage	Frequency	Percentage
Gender				
Male	56	16.23%	33	16.92%
Female	289	83.77%	162	83.08%
Religion				
Moslems	313	90.99%	178	91.75%
Christians (Protestant)	16	4.65%	7	3.61%
Christians (Catholic)	6	1.74%	2	1.03%
Buddhists	4	1.16%	3	1.55%
Hindu	3	.87%	2	1.03%
Others	2	.58%	2	1.03%
Area of residence				
Java	304	88.12%	172	88.20%
Sumatra	22	6.38%	11	5.64%
Kalimantan	8	2.32%	6	3.08%
Sulawesi	4	1.16%	2	1.03%
Bali and Nusa Tenggara	5	1.45%	3	1.54%
Papua	1	.29%	0	0%
Outside Indonesia	1	.29%	1	0.51%
Education level				
Junior high school	1	.29%	1	.513%
Senior high school	142	41.16%	83	42.564%
Diploma/Bachelor degree	172	49.85%	93	47.692%
Master degree	30	8.70%	18	9.231%
Marital status				
Unmarried	298	86.38%	167	85.64%
Married	45	13.04%	26	13.33%
Divorced/ widowed	2	.58%	2	1.03%
Employment status				
Unemployed	219	63.48%	124	63.59%
Employed	126	36.52%	71	36.41%
Perceived characteristic of living area				
Urban	228	66.09%	129	66.15%
Rural	117	33.91%	66	33.85%

Note: All variables on the first and second data set have N1=345 and N2=195 except for religion (N1=344, N2=194); Source: Personal data (2024).

Patient Health Questionnaire–9 (PHQ-9)

The Patient Health Questionnaire–9 (PHQ-9; Kroenke et al., 2001) is a self-report measure used to assess the presence and severity of depressive symptoms. It consists of nine items rated on a 4-point Likert-type scale ranging from 0 (Not at all) to 3 (Nearly every day), based on how frequently

http://journal.uinjkt.ac.id/index.php/jp3i

symptoms were experienced over the past two weeks. In the present study, the PHQ-9 demonstrated good internal consistency (α = .88, ω = .88).

World Health Organization Quality of Life–Brief Version (WHOQOL-BREF)

The WHOQOL-BREF (The WHOQOL Group, 1998; WHO, 1996) is a 26-item short version of the WHOQOL-100, developed to assess quality of life across four domains: physical health, psychological well-being, social relationships, and environmental conditions. Each item is rated on a 5-point Likert-type scale assessing frequency or intensity. Domain scores range from 4 to 20, with higher scores indicating better perceived quality of life. In the current study, Cronbach's alpha and McDonald's omega for the four domains were .50/.58, .55/.70, .56/.62, and .74/.74, respectively.

Mindfulness Attention Awareness Scale (MAAS)

The Mindfulness Attention Awareness Scale (MAAS; Brown & Ryan, 2003) consists of 15 items measuring an individual's general tendency to be attentive to and aware of present-moment experiences in daily life. The unidimensional scale yields a single total score. Items are rated on a 6-point Likert-type scale ranging from 1 (Almost always) to 6 (Almost never). In this study, the MAAS demonstrated high internal consistency (α = .89, ω = .89).

Data Analysis

Descriptive statistics and psychometric evaluations were conducted using both classical test theory (CTT) and item response theory (IRT) approaches. The rationale for this dual-approach lies in their synergistic strengths: while CTT provides traditional indices of reliability and factor structure suitable for comparison with previous BADS-SF studies, IRT (specifically Rasch modeling) offers a more granular, item-level analysis that is independent of the sample distribution. This ensures a comprehensive evaluation of the scale's performance, from broad construct validity to the specific functioning of the rating scale categories (Cappelleri et al., 2014; Jabrayilov et al., 2016; Petrillo et al., 2015).

Classical Test Theory (CTT)

Analyses based on CTT were performed using JASP version 0.19.3.0. CTT assumes that observed scores are composed of true scores and measurement error. It focuses on test-level properties and treats all items as equally contributing to the measured construct. In addition, CTT is sample-dependent (Cappelleri et al., 2014).

Reliability was evaluated using Cronbach's alpha and McDonald's omega (Hayes & Coutts, 2020). Item-total correlations were also examined (Anselmi et al., 2019). Construct validity was assessed using confirmatory factor analysis (CFA). Convergent validity was tested by correlating BADS-SF scores with WHOQOL-BREF and MAAS scores using Pearson or Spearman correlation coefficients, depending on the distribution of the data.

Due to non-normal distribution in several items, robust maximum likelihood (MLR) estimation was used in CFA (Brown, 2015). Model fit was evaluated using Chi-square (χ^2), root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker–Lewis index (TLI). Given the sensitivity of the chi-square statistic to sample size, RMSEA was used as the primary fit index. Model fit was considered acceptable if at least two of the following criteria were met: (a) RMSEA < .05, (b) RMSEA 90% confidence interval includes .05, or (c) $p > .05$ (Umar & Nisa, 2020). Additionally, CFI and TLI values ≥ .90 were considered acceptable, and values ≥ .95 indicated a good fit (Hu & Bentler, 1999; Sun, 2005).

Item Response Theory (IRT) – Rasch Modeling

Rasch modeling, an application of IRT, was conducted using Winsteps software (Linacre, 2012). Rasch modeling transforms ordinal data into interval-level measurements in logit units (Bond & Fox, 2015). Unlike CTT, Rasch modeling produces sample-independent item and person estimates and provides a detailed evaluation of individual item characteristics (Boone et al., 2014).

The Rasch model assumes: (a) each individual has a latent ability level, (b) each item has a specific level of difficulty, (c) both can be placed along a common latent continuum, and (d) the probability of a correct or endorsed response is a function of the difference between the individual's ability and item difficulty (Bond & Fox, 2015). The Rasch analyses reported in this study included: descriptive statistics (e.g., item logit values, Wright map); reliability indices, including Cronbach's alpha, item reliability, person reliability, and the person separation index; and construct validity assessments using principal component analysis (PCA) of residuals and item fit statistics. Item fit was evaluated based on three criteria: (a) infit mean-square (MnSq) values less than the sum of its mean and standard deviation (Sumintono & Widhiarso, 2013), (b) outfit MnSq values between 0.50 and 1.60 (Linacre, 2020), and (c) point-measure correlation values between .40 and .85 (Sumintono & Widhiarso, 2013). Additional analyses included differential item functioning (DIF) to assess item bias and evaluation of response scale functioning to ensure appropriate category usage.

Results and Discussion

Results

Translation/ Adaptation Result

Forward translation by two independent bilingual psychologists resulted in T1 and T2 versions of the BADS-SF with minimum discrepancy. After deciding on the synthesized version (T12), back-translation was conducted by a professional translator (BT version). The original, T12, and BT versions of the BADS-SF were then evaluated by five experts in psychology, including experts in clinical psychology research and clinical practice. Based on the evaluation of the equivalence between the original and translated version, the content validity (Aiken's V) coefficient was calculated with results ranging from .75 to 1. Item 1 was the only item with Aiken's V .75. After evaluating the translated version, the wording was preserved because it had clearly represented the meaning of the original version.

The following step was pilot testing involving 30 participants with diverse ages and levels of education attainment. The result of the cognitive debriefing process showed that Item 1 was the most difficult item to understand among participants with four participants (13.33%) who reported difficulty. Most of other items had one participant (3.33%) reporting difficulty in understanding the statement, i.e., items 2, 3, 4, 5, 6, and 8. Participants with difficulty in understanding item statements had diverse educational attainment, ranging from elementary school, junior high school, senior high school, and diploma/ bachelor degree. The age of participants was also diverse, ranging from early adulthood to late adulthood. Three participants with the highest reported difficulties were within the age range of 50-59 years old and reported education lower than a bachelor degree. For items with reported difficulty to understand, participants needed more explanation to help them understand the items correctly. Additionally, it was found that some participants interpreted the items in slightly different ways. As a highlight, items 6 and 8 seemed to be interpreted positively by some participants, not as problematic avoidant behaviors as they were intended. "Escaping from or avoiding something unpleasant" (Item 6) was interpreted as avoiding a bad thing and a way to keep oneself happy or in a positive emotional state. "Engaging in activities that would distract oneself from feeling bad" (Item 8) was interpreted positively as a form of emotional regulation. Although several notes were taken based on the pilot testing step, the wording of translated version was still preserved before stronger evidence from field testing and empirical validation were found.

http://journal.uinjkt.ac.id/index.php/jp3i

Reliability Analysis, Confirmatory Factor Analysis, and Convergent Validity

The internal consistency of the Indonesian version of the BADS-SF was acceptable, with Cronbach's α = .68 and McDonald's ω = .72. Examination of individual items indicated that removing Items 6 and 8 improved the overall reliability. Both items showed negative item-total correlations, suggesting poor alignment with the overall construct. Table 2 presents the reliability coefficients if each item were deleted, as well as the item-total correlations for all BADS-SF items.

Table 2. Individual Item Reliability Statistics

	If item dropped				
Item	**McDonald's ω**	**Cronbach's α**	**Item-rest correlation**	**Mean**	**SD**
bads1	0.695	0.643	0.427	2.032	1.585
bads2	0.649	0.612	0.572	2.617	1.536
bads3	0.678	0.643	0.423	2.664	1.821
bads4	0.632	0.602	0.614	2.928	1.536
bads5	0.651	0.614	0.557	2.962	1.559
bads6	0.744	0.705	0.090	1.365	1.355
bads7	0.700	0.651	0.395	1.849	1.487
bads8	0.762	0.743	-0.166	1.409	1.311
bads9	0.697	0.667	0.315	3.751	1.516

Source: Personal data (2024).

Confirmatory Factor Analysis

Prior to conducting factor analysis, assumptions were examined. Normality was not fully supported. Although the sample size in this study was considered large (N > 200), the Shapiro–Wilk test was significant for all items, likely due to its sensitivity in large samples and subsequent over-rejection of the null hypothesis. Visual inspection of item histograms revealed non-normal distributions for Items 6 and 8. Further data screening showed z-scores for skewness ranging from –1.04 to 7.89 and z score for kurtosis values between –3.21 and 2.04. Using approximate cut off of 3 for both z score of skewness and kurtosis, the data were considered to violate normal distribution assumption (Field, 2009).

Due to the limited availability of data, it was not feasible to split the sample into separate datasets for exploratory and confirmatory factor analyses. Therefore, item-level reliability statistics were used to inform the confirmatory factor analysis (CFA). As shown in Table 2, Items 6 and 8 demonstrated near-zero or negative correlations with the other items. Based on these findings, Items 6 and 8 were excluded from the CFA.

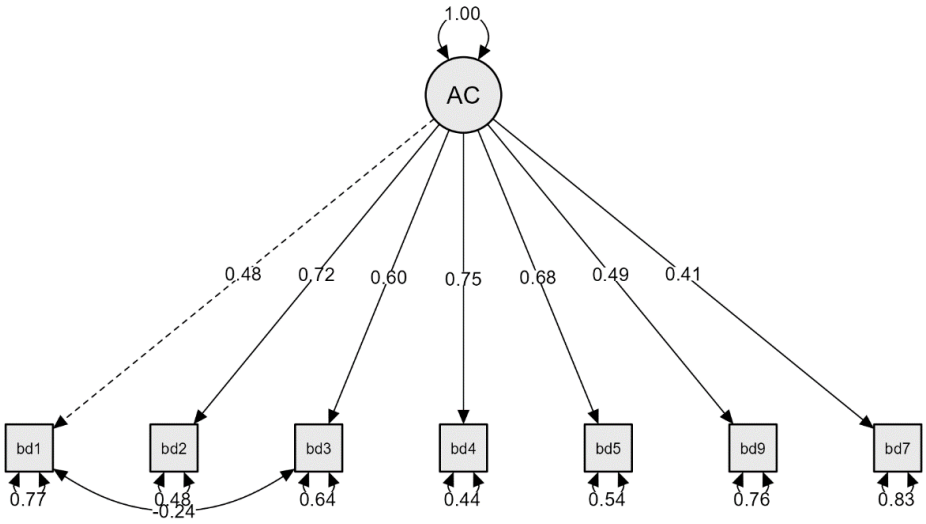
Based on the remaining items, two one-factor confirmatory factor analysis (CFA) models were tested. The first model included Item 7 within the activation factor, consistent with the original study, resulting in a 7-item, one-factor model. The second model included only the six items originally assigned to the activation factor, forming a 6-item model.

The results indicated that neither model demonstrated an adequate fit to the observed data. For the 7-item model, the fit indices were: $\chi^2(14) = 38.39$, $p < .001$; CFI = .957; TLI = .935; RMSEA = .071

(90% CI [.045, .098]). For the 6-item model, the fit indices were: $\chi^2(9) = 23.64$, $p = .005$; CFI = .971; TLI = .952; RMSEA = .069 (90% CI [.035, .103]).

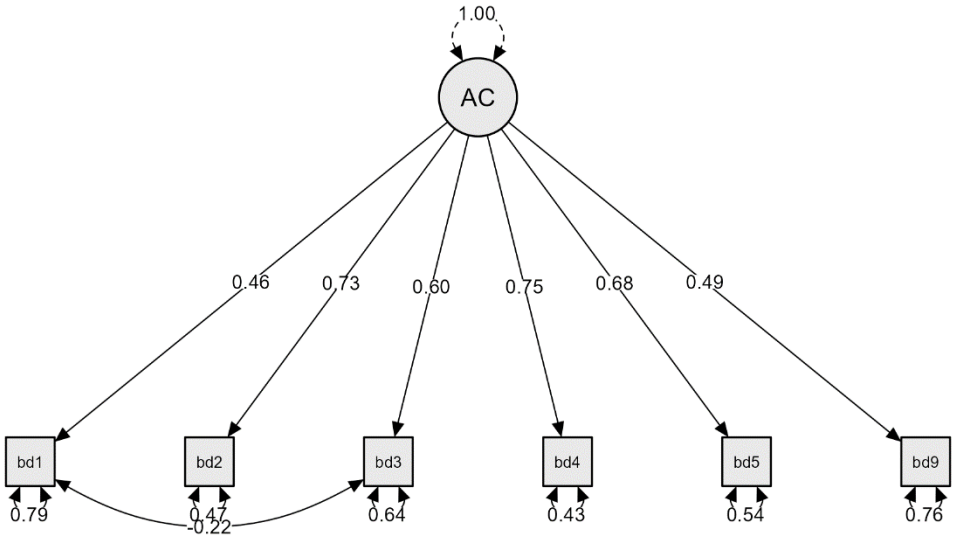
Modification indices suggested a residual covariance between Item 1 and Item 3. Including this covariance improved model fit for both models. The modified 7-item model showed acceptable to excellent fit: $\chi^2(13) = 23.02$, $p = .041$; CFI = .982; TLI = .971; RMSEA = .047 (90% CI [.009, .078]). The modified 6-item model also demonstrated excellent fit: $\chi^2(8) = 10.77$, $p = .215$; CFI = .995; TLI = .990; RMSEA = .032 (90% CI [.000, .075]).

Figures 1 and 2 display the path diagrams for the final 7-item and 6-item models, respectively. The standardized factor loadings for the modified 7-item model ranged from .407 to .724, with Item 2 having the highest and Item 7 the lowest loading. In the other hand, the loadings for the modified 6-item model ranged from .459 to .755; Item 4 had the highest and Item 1 the lowest loading.



Source: Personal data (2024)

Figure 1. Diagram of the One-factor 7-item Model



Source: Personal data (2024)

Figure 2. Diagram of the One-factor 6-item Model

Convergent Validity

The next analysis examined the convergent validity of the BADS-SF by assessing its correlations with depression and quality of life, as measured by the PHQ-9 and WHOQOL-BREF, respectively. Based on the CFA results, which supported both the 7-item and 6-item versions of the BADS-SF, total scores were recalculated accordingly. Correlation analyses indicated that the activation subscale was negatively correlated with depression and positively correlated with both quality of life and mindfulness, supported the theoretical relationship between constructs (see Table 3).

Table 3. Convergent Validity of the Indonesian Version of BADS-SF

Variables		1	2	3	4	5	6	7	8
1. 7-item BADS-SF	Pearson's r	—							
	p-value	—							
2. 6-item BADS-SF	Pearson's r	0.982 ***	—						
	p-value	< .001	—						
3. PHQ-9	Pearson's r	-0.490***	-0.444 ***	—					
	p-value	< .001	< .001	—					
4. QoL Domain 1	Pearson's r	0.636 ***	0.609 ***	-0.542***	—				
	p-value	< .001	< .001	< .001	—				
5. QoL Domain 2	Pearson's r	0.696 ***	0.647 ***	-0.644***	0.675 ***	—			
	p-value	< .001	< .001	< .001	< .001	—			
6. QoL Domain 3	Pearson's r	0.512 ***	0.481 ***	-0.311***	0.513 ***	0.562 ***	—		
	p-value	< .001	< .001	< .001	< .001	< .001	—		
7. QoL Domain 4	Pearson's r	0.456 ***	0.434 ***	-0.337***	0.527 ***	0.477 ***	0.450***	—	
	p-value	< .001	< .001	< .001	< .001	< .001	< .001	—	
8. MAAS	Pearson's r	0.335 ***	0.274 ***	-0.456***	0.456 ***	0.468 ***	0.240***	0.239***	—
	p-value	< .001	< .001	< .001	< .001	< .001	< .001	< .001	—

*p<.05 **p<.01 ***p<.001

Note: BADS-SF = Behavioral Activation for Depression – Short Form; PHQ-9 = Nine-item Patient Health Questionnaire; QoL = Quality of Life measured by WHO Quality of Life – Brief Version (WHOQOL-BREF), Domain 1 = physical health, Domain 2 = psychological, Domain 3 = Social relationship, Domain 4 = environment; MAAS = Mindfulness Attention Awareness Scale.

Source: Personal data (2024).

Summary Statistics of Rasch Modeling, Item Logits, and Person-Item Map

The results of the Rasch model analysis showed that the mean person logit was −0.45 (SD = 0.80), indicating that, on average, participants tended to slightly disagree with the item statements. The mean INFIT MNSQ and OUTFIT MNSQ values for persons were 0.98 (SD = 0.45) and 0.99 (SD = 0.44), respectively, both close to the ideal value of 1.0, suggesting good model fit. Similarly, the mean INFIT ZSTD and OUTFIT ZSTD were 0.0 (SD = 1.0), reflecting near-perfect fit to the Rasch model expectations.

Regarding the items, the mean item logit was 0.0 (SD = 0.78), suggesting that items had equal probability of being endorsed or not by respondents. The mean INFIT MNSQ and OUTFIT MNSQ values for items were 1.00 (SD = 0.22) and 0.99 (SD = 0.22), respectively, also indicating ideal model fit. The mean INFIT ZSTD and OUTFIT ZSTD values were −0.2 (SD = 2.3) and −0.3 (SD = 2.3), respectively, consistent with good item-level fit. Overall, these summary statistics indicate that the items in the Indonesian version of the BADS-SF demonstrated excellent psychometric quality under the Rasch model.

Further analysis of item difficulty is presented in Table 4, which displays item logits from highest to lowest. Item 9 had the lowest logit value, indicating that it was the easiest for participants to endorse. Items 4, 5, 2, and 3 followed, with similar logit values, suggesting moderate agreement among respondents. More difficult items included Items 1, 7, and 6, while Item 8 had the highest logit value, making it the most difficult or least likely to be endorsed.

Table 4. Item Logit of the Indonesian Version of BADS-SF Items

Item Number	Item Statement	Logit Value
8	I engaged in activities that would distract me from feeling bad.*	1.19
6	Most of what I did was to escape from or avoid something unpleasant.*	1.1
7	I spent a long time thinking over and over about my problems.*	.44
1	There were certain things I needed to do that I didn't do.*	.31
3	I engaged in many different activities.	-.24
2	I am content with the amount and type of things I did.	-.33
5	I was an active person and accomplished the goals I set out to do.	-.57
4	I made good decisions about what type of activities and/or situations I put myself in.	-.61
9	I did things that were enjoyable.	-1.27

*Reverse scored. Source: Personal data (2024).

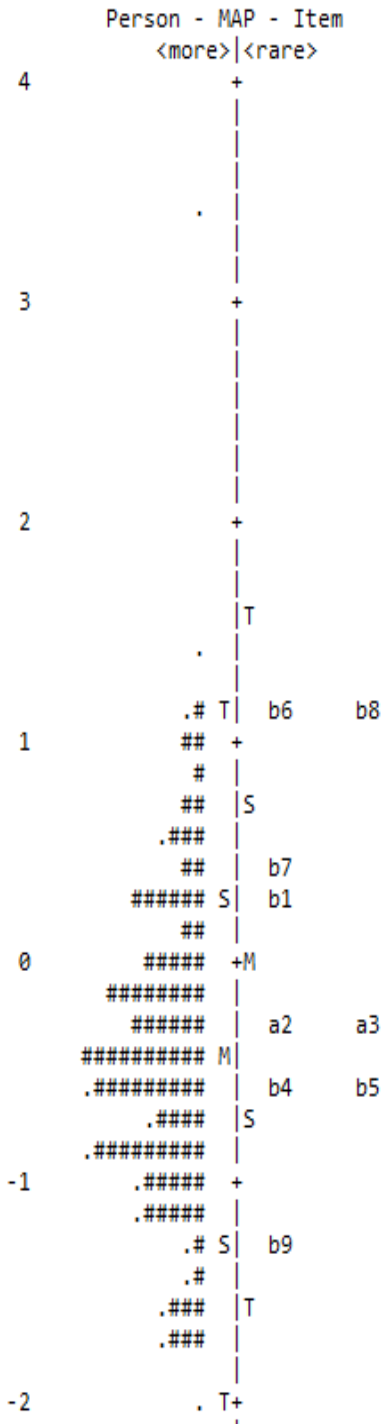
Furthermore, the interaction between item difficulty (or agreeability) and person ability (or agreeableness) is illustrated in the Wright map (see Figure 3). As shown in the figure, the items of the BADS-SF were distributed within a logit range of −2 to 2. Although the BADS-SF appeared to lack items at both the higher end (around 2 logits) and the lower end (around −2 logits) of the difficulty spectrum, this distribution is acceptable given that the BADS-SF is a shortened version of the full scale. Despite the limited number of items, the scale adequately captures individuals with varying levels of behavioral activation.

Rasch-Based Reliability: Scale, Item, and Person Reliability

The reliability of the Indonesian version of the BADS-SF was evaluated using three primary indices: item reliability, person reliability, and the person separation index. Additionally, Cronbach's alpha was

reported as a supplementary reliability measure. The results indicated excellent item reliability, with a value of .99. Item reliability reflects the consistency of item difficulty estimates; a high value suggests that the item difficulty hierarchy is stable across different samples.

Person reliability was .79, which is considered moderate to acceptable, indicating that the scale can reliably distinguish between individuals with different levels of the latent trait. The person separation index was 1.95, which is regarded as good. This value suggests that the scale can differentiate approximately three distinct groups based on their latent behavioral activation levels. Lastly, the Cronbach's alpha coefficient for the scale was .80, indicating good internal consistency. Cronbach's alpha reflects the overall reliability derived from the interaction between item and person parameters.



Source: Personal data (2024)

Figure 3. Item-Person Interaction Map (Wright Map)

Item Fit

Table 8 presents the item fit indices for all BADS-SF items. Item fit statistics are used to determine whether each item conforms to the expectations of the Rasch model. In this study, three criteria were used to assess item fit: (1) the item was classified as misfitting if the infit mean square (MnSq) exceeded the mean plus one standard deviation (in this case, greater than 1.22); (2) the outfit MnSq should fall within the acceptable range of 0.5 to 1.5; and (3) the point-measure correlation should fall between .40

http://journal.uinjkt.ac.id/index.php/jp3i

and .85. Based on these criteria, Item 8 was identified as misfitting, particularly due to exceeding the cutoff for the infit MnSq value.

Table 8. Item Fitness Indices of BADS-SF Bahasa Indonesia

Item Number	INFIT		OUFIT		Point Measure Correlation	Item Fitness
	MNSQ	SZTD	MNSQ	SZTD		
8	**1.31**	2.8	1.32	2.9	.35	Misfit
1	1.20	2.0	1.19	1.8	.64	Fit
3	1.17	1.7	1.17	1.6	.65	Fit
6	1.09	.9	1.06	.6	.50	Fit
9	1.06	.6	1.05	.5	.52	Fit
7	1.03	.3	1.02	.2	.67	Fit
2	.76	-2.6	.76	-2.7	.71	Fit
5	.68	-3.6	.69	-3.6	.71	Fit
4	.68	-3.7	.67	-3.8	.71	Fit
Mean	1.00	-.2	.99	-.3		
SD	.22	2.3	.22	2.3		

Note: Bold number indicate misfit index. Items were ordered according to item fitness. Source: Personal data (2024).

Unidimensionality of the Scale

The BADS-SF comprises two subscales—activation and avoidance. However, all items may be summed to represent overall activation levels, with reverse scoring applied to avoidance items. To evaluate the dimensionality of the Indonesian version of the BADS-SF, principal component analysis (PCA) of the residuals was conducted within the Rasch model framework. As shown in Table 9, the total variance explained by the Rasch measures was 56.1%. The unexplained variances in the first, second, and third contrasts were 22.3%, 15.1%, and 12.7%, respectively. The total explained variance exceeded 20%, supporting the assumption of unidimensionality.

However, the unexplained variance in the first contrast exceeded the commonly accepted threshold of 15% for an ideal unidimensional model, suggesting the potential presence of a secondary dimension within the scale. To explore whether the misfitting item (Item 8) contributed to this secondary dimension, PCA was repeated using the 8-item version of the BADS-SF (excluding Item 8). The results indicated improved unidimensionality, with the unexplained variance in the first contrast falling below 15%. These findings support the unidimensionality of the revised 8-item version of the BADS-SF. Detailed unidimensionality test results are presented in Table 9.

Table 9. The Result of Principal Component Analysis of the Indonesian Version of BADS-SF

Description	9-items BADS-SF	8-items BADS-SF
Total raw variance in observations	100%	100%
Raw variance explained by measures	56.1%	55.3%
Raw variance explained by persons	14.8%	17.0%
Raw variance explained by items	41.2%	38.3%

Raw unexplained variance (total)	43.9%	44.7%
Unexplained variance in first contrast	22.3%	9.7%
Unexplained variance in second contrast	15.1%	7.3%
Unexplained variance in third contrast	12.7%	6.7%
Unexplained variance in fourth contrast	12.0%	6.3%
Unexplained variance in fifth contrast	11.5%	5.4%

Source: Personal data (2024).

Differential Item Functioning

An important aspect of scale adaptation is the evaluation of differential item functioning (DIF), which identifies items that may exhibit bias or favor a particular group, thereby threatening measurement equivalence. In this study, DIF was assessed by gender. Table 10 presents the results of the gender-based DIF analysis.

As shown in Table 10, Item 7 was the only item that demonstrated significant DIF by gender. This item was more favorable for female respondents than for male respondents, indicating that females were more likely to endorse the item, regardless of their overall level of activation. Figure 4 provides a visual representation of the DIF patterns across all nine items of the BADS-SF. Specifically, male respondents tended to score lower on Items 1, 2, and 7; higher on Items 3, 4, 6, and 8; and relatively similarly on Items 5 and 9. Despite these trends, Item 7 was the only item that showed a statistically significant gender-related DIF.

Table 10. Differential Item Functioning (DIF) Analysis on Gender

Item	DIF		DIF Contrast	t	Prob.	DIF Category
	Male	Female				
1	.17	.34	.17	.91	.365	No DIF
2	-.46	-.31	.15	.79	.430	No DIF
3	-.09	-.27	.18	-.97	.335	No DIF
4	-.37	-.66	.29	-1.60	.115	No DIF
5	-.57	-.57	.00	.00	1.00	No DIF
6	1.30	1.06	.23	-1.11	.273	No DIF
7	.02	.53	.50	2.70	**.009**	Moderate DIF
8	1.37	1.16	.21	-.99	.327	No DIF
9	-1.31	-1.27	.04	.19	.849	No DIF

Note: positive t-value indicated that item was more favorable among female group, and vice versa. Bold value indicated the DIF. Source: Personal data (2024).

Response Scale Functioning

The BADS-SF uses a 7-point Likert-type rating scale, ranging from 0 to 6. In the present study, the adapted Indonesian version retained the original response format. To evaluate the appropriateness of the rating scale, the Rasch model was applied.

Results from the Rasch analysis indicated that both the observed averages and the Andrich thresholds for response categories 0 through 6 were ordered as expected, progressing logically from the lowest (0) to the highest (6) response option. This finding suggests that the response scale functioned appropriately and that each category was used consistently and meaningfully by respondents. Table 11 presents the detailed values of the observed averages and Andrich thresholds for each category across all items.

Table 11. Observed Average and Andrich Threshold of the 7-point Likert Scale of the Indonesian Version of BADS-SF

http://journal.uinjkt.ac.id/index.php/jp3i

Rating Scale	Observed Average	Andrich Threshold	Category Measure
0	-1.82	NONE	(-3.57)
1	-1.29	-2.22	-1.98
2	-.75	-1.41	-.85
3	-.14	-.17	.05
4	.35	.34	.90
5	.88	1.35	1.93
6	1.49	2.11	(3.47)

Source: Personal data (2024).

Discussion

This study aimed to conduct cross-cultural validation of the BADS-SF into Bahasa Indonesia, investigate its underlying structure using CTT, and evaluate its psychometric properties using IRT. To the best of our knowledge, this is the first study to adapt the BADS-SF for the Indonesian adult population and to examine its factor structure and psychometric properties using both CTT and IRT. Despite being recommended by the scale developers for its practicality in clinical settings (Manos et al., 2011), the BADS-SF has still received limited empirical attention, especially in non-Western populations. Previous studies on the BADS-SF have either focused on cross-cultural adaptation (e.g., González-Roz et al., 2018; Koşan & Kaya, 2023; Wagener et al., 2015) or explored its longitudinal and mediating relationships with depression and quality of life (Shudo et al., 2017; Shudo & Yamamoto, 2017). This study addresses the gap by contributing to the cross-cultural validation of the BADS-SF in a collectivist, non-Western context and using a more comprehensive approach (i.e., CFA and Rasch modeling) to examine the factor structure, item-level psychometric properties and scale unidimensionality. In general, results from this study demonstrate that the Indonesian BADS-SF is valid and reliable for behavioral activation measurement among Indonesian adults.

In terms of classical test theory, the Indonesian version of the 9-item BADS-SF demonstrated acceptable internal consistency (α = .68, ω = .72). This reliability level was comparable with results from Turkey (Koşan & Kaya, 2023) and Germany (Fuhr et al., 2016), although it is lower than those reported in the original (Manos et al., 2011) and French versions (Wagener et al., 2015), and studies involving adolescent populations (Petts et al., 2016). Notably, item-total correlation analysis revealed that Item 6 and Item 8 had near-zero or negative correlations with the overall scale, indicating possible conceptual or linguistic misalignment. Deleting Item 8 improved reliability (α = .74, ω = .76), supporting its exclusion.

Confirmatory factor analysis (CFA) did not support the original two-factor structure proposed by Manos et al. (2011), likely due to non-positive latent covariances. Modified one-factor models (7- and 6-item versions) with a residual covariance between Items 1 and 3 provided better model fit. These findings contrast with those from previous CFA studies in various cultural contexts, such as Spain (González-Roz et al., 2018), Belgium (Wagener et al., 2015), and Turkey (Koşan & Kaya, 2023), which supported two-factor structures but with different item compositions. Wagener et al.'s (2015) study in French supported a two-factor model with the same items on the avoidance factor, but excluding item 1 and 9 on the activation factor due to low factor loading (<.7). Fuhr et al.'s (2016) study found the French model as having better fit than original model and model obtained from EFA in their study. Last, Koşan & Kaya's (2023) study also supported two-factor model, but with different item composition and exclusion of item 8. It can be concluded that this study supported the activation factor from the original study, but not the avoidance factor. Cultural factors and linguistic nuances likely contributed to the variability in factor structures across these studies (van de Vijver & Tanzer, 2004).

The exclusion of Items 6 and 8 may be attributed to wording complexity and cultural interpretations. Cognitive debriefing during the adaptation process revealed that individuals with lower education levels struggled to comprehend these items. Moreover, behaviors described in these items,

such as avoiding unpleasant experiences or distracting oneself from negative emotions, may be culturally interpreted as socially acceptable or effective coping strategies rather than maladaptive avoidance. For example, a previous study demonstrated that both approach and avoidant coping styles were used by Indonesians during the COVID-19 pandemic (Lubis et al., 2024). The tendency of Indonesians to seek for harmony amidst diversity or conflict also evident (Peristianto et al., 2025; Setiyawati et al., 2025; Sugiarto & Suparno, 2025). These interpretations reflect cultural norms around emotional regulation and collectivist values (Heine, 2020; Ryder et al., 2002).

Convergent validity analyses supported the expected associations between behavioral activation, as measured by the BADS-SF, and related constructs, including depression severity, quality of life, and mindfulness. Consistent with prior research, behavioral activation was negatively associated with depression severity (Li et al., 2014; Raes et al., 2010) and positively correlated with quality of life (Shudo & Yamamoto, 2017) and mindfulness (Crego et al., 2021).

Rasch analysis further provided item-level psychometric insights. The scale demonstrated excellent item reliability (.99), acceptable person reliability (.79), and a good person separation index (1.95), indicating its capability to differentiate individuals across activation levels. These findings are consistent with other studies that emphasize the advantages of IRT for detailed item diagnostics (Bond & Fox, 2015). Rasch model indices also indicated that Item 8 was misfitting, further supporting its exclusion. Unidimensionality was confirmed after removing Item 8, with principal component analysis of residuals meeting the recommended criteria (Linacre, 2020).

Differential item functioning (DIF) analysis revealed that Item 7 favored female respondents, while other items demonstrated negligible gender bias. The response scale functioning analysis showed that the 7-point Likert scale operated effectively, with ordered thresholds and expected step calibrations, supporting its continued use.

Several limitations warrant consideration. First, item rewording was not implemented in this study as a potential solution for items demonstrating poor reliability or misfit. Because avoidance is a theoretically central component of behavioral activation and a primary target of behavioral activation therapy, revising the wording of avoidance items may improve content validity while preserving their conceptual relevance. Second, the same sample was used for both confirmatory factor analysis and Rasch analyses, and exploratory factor analysis could not be conducted due to the limited sample size. This design precluded cross-validation of the factor structure and limits confidence in the stability and replicability of the measurement model across independent samples. Third, the sample was not representative of the broader Indonesian population, as participants were predominantly female and primarily resided on Java Island. Consequently, the generalizability of the findings to more diverse demographic and national populations remains limited. Finally, participants were not recruited from clinical settings or behavioral activation trials, which limits the ability to evaluate predictive validity and sensitivity to therapeutic change.

Future research should employ systematic design procedures to revise or replace culturally sensitive or misfitting items (e.g., Items 6 and 8) and evaluate these modifications through iterative testing. The incorporation of cognitive interviewing and other qualitative approaches is recommended to clarify how respondents interpret avoidance-related behaviors within the Indonesian cultural context. Studies using larger and more representative samples, as well as alternative research designs such as longitudinal approaches, would further strengthen the evidence base and extend the generalizability of the findings. Finally, validation of the Indonesian BADS-SF in clinical populations undergoing behavioral activation therapy is needed to establish its clinical utility and treatment sensitivity.

Conclusion

The Indonesian version of the BADS-SF demonstrated satisfactory psychometric performance across classical test theory and Rasch frameworks. The results supported a unidimensional construct of behavioral activation, with a 6-item one-factor model providing the most parsimonious and best-fitting

solution. Evidence of validity included strong content validity, acceptable confirmatory factor analysis fit, theoretically coherent convergent validity, and Rasch-based support for item functioning and unidimensionality after removal of a misfitting item. Reliability evidence included acceptable internal consistency, excellent item reliability, and adequate person reliability. This study provides the first comprehensive psychometric evaluation of the Indonesian BADS-SF using complementary measurement frameworks and item-level modeling. The findings support the use of the 6-item Indonesian BADS-SF as a brief and psychometrically supported measure of behavioral activation for research and clinical assessment in Indonesia, thereby facilitating the implementation and evaluation of behavioral activation interventions. Further evaluation of avoidance items and validation in clinical populations, longitudinal designs, and more representative samples is warranted.

Acknowledgment

This study is part of dissertation project of the first author and fully funded by the Indonesia Endowment Fund for Education Agency (Lembaga Pengelola Dana Pendidikan), Ministry of Finance, Republic of Indonesia, through doctoral program scholarship awarded to the first author.

Conflict of Interest

JK is the developer of the original version of BADS and together with RL developed the original brief version of the measurement (BADS-SF). JK and RL do not received any financial incentive upon the BADS-SF adaptation study. All authors declare that they have no conflict of interest.

Authors Contribution

SS, RTH, AP, and RH designed the study. JK provided the original BADS-SF and guidance on how to use it. AP translated the original version of BADS-SF into the Bahasa Indonesia version. SS was responsible for data collection and analysis. SS drafted the article, which was reviewed and approved by all other authors.

References

Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abegaz, K. H., Abolhassani, H., Aboyans, V., Abreu, L. G., Abrigo, M. R. M., Abualhasan, A., Abu-Raddad, L. J., Abushouk, A. I., Adabi, M., Adekanmbi, V., Adeoye, A. M., Adetokunboh, O. O., … Murray, C. J. L. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet, 396(10258), 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2

Anselmi, P., Colledani, D., & Robusto, E. (2019). A Comparison of Classical and Modern Measures of Internal Consistency. Frontiers in Psychology, 10. https://doi.org/10.3389/fpsyg.2019.02714

Anvari, M. S., Hampton, T., Tong, M. P., Kahn, G., Triemstra, J. D., Magidson, J. F., & Felton, J. W. (2023). Behavioral Activation Disseminated by Non–Mental Health Professionals, Paraprofessionals, and Peers: A Systematic Review. Behavior Therapy, 54(3), 524–538. https://doi.org/10.1016/j.beth.2022.12.007

Arroisi, J., & Himaya, N. N. (2023). Abu Zayd Al-Balkhi's Perspective on Depression: Countering Sadness with Cognitive Theory in the Book of Mashalih al Abdan wa al Anfus. TAZKIYA Jounal of Psychology, 11(1), 11–12. https://doi.org/https://doi.org/10.15408/tazkiya.v11i1.29913

Bond, T. G., & Fox, C. M. (2015). Applying the Rasch model: Fundamental measurement in the human sciences (Third edition). Routledge.

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch analysis in the human sciences. In Rasch Analysis in the Human Sciences. https://doi.org/10.1007/978-94-007-6857-4

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present : Mindfulness and its role in psychological well-being. Journal of Personality and Social Psychology, 84(4), 822–848. https://doi.org/10.1037/0022-3514.84.4.822

Brown, T. A. (2015). Confirmatory factor analysis for applied research (Second edition) (Second edi). The Guilford Press.

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of Classical Test Theory and Item Response Theory for quantitative assessment of items in developing patient-reported outcome measures. Clinical Therapeutics, 36(5), 648–662. https://doi.org/10.1016/j.clinthera.2014.04.006.Overview

Crego, A., Yela, J. R., Gómez-Martínez, M. Á., Riesco-Matías, P., & Petisco-Rodríguez, C. (2021). Relationships between mindfulness, purpose in life, happiness, anxiety, and depression: Testing a mediation model in a sample of women. International Journal of Environmental Research and Public Health, 18(3), 1–16. https://doi.org/10.3390/ijerph18030925

Cuijpers, P., Karyotaki, E., Harrer, M., & Stikkelbroek, Y. (2023). Individual behavioral activation in the treatment of depression: A meta analysis. Psychotherapy Research, 1–12. https://doi.org/10.1080/10503307.2023.2197630

Dimidjian, S., Martell, C. R., Addis, M. E., & Herman-Dunn, R. (2008). Behavioral activation for depression. In D. H. Barlow (Ed.), Clinical handbook of psychological disorders: A step-by-step treatment manual (4th edition) (pp. 328–364). The Guilford Press.

Dimidjian, Sona, Barrera, M., Martell, C., Muñoz, R. F., & Lewinsohn, P. M. (2011). The origins and current status of behavioral activation treatments for depression. Annual Review of Clinical Psychology, 7, 1–38. https://doi.org/10.1146/annurev-clinpsy-032210-104535

Ekers, D., Webster, L., Van Straten, A., Cuijpers, P., Richards, D., & Gilbody, S. (2014). Behavioural activation for depression; An update of meta-analysis of effectiveness and sub group analysis. PLoS ONE, 9(6). https://doi.org/10.1371/journal.pone.0100100

Ferster, C. B. (1973). A functional analysis of depression. American Psychologis, 28, 857–870.

Field, A. (2009). Discovering statistics using SPSS (3rd Editio). Sage Publication.

Fuhr, K., Hautzinger, M., Krisch, K., Berking, M., & Ebert, D. D. (2016). Validation of the Behavioral Activation for Depression Scale (BADS) - Psychometric properties of the long and short form. Comprehensive Psychiatry, 66, 209–218. https://doi.org/10.1016/j.comppsych.2016.02.004

González-Roz, A., Secades-Villa, R., & Muñiz, J. (2018). Validity evidence of the Behavioral Activation for Depression Scale-Short Form among depressed smokers. International Journal of Clinical and Health Psychology, 18(2), 162–169. https://doi.org/10.1016/j.ijchp.2018.03.002

Hayes, A. F., & Coutts, J. J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But…. Communication Methods and Measures, 14(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Heine, S. J. (2020). Cultural psychology (4th edition). W. W. Norton & Company.

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

International Test Commission. (2017). The ITC Guidelines for Translating and Adapting Tests (Second edition). [www.InTestCom.org]. https://doi.org/10.1111/j.1464-0597.1975.tb00322.x

Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. Applied Psychological Measurement, 40(8), 559–572. https://doi.org/10.1177/0146621616664046

Jacobson, N. S., Martell, C. R., & Dimidjian, S. (2001). Behavioral Activation Treatment for Depression: Returning to Contextual Roots. Clinical Psychology: Science and Practice, 8(3), 255–270. https://doi.org/10.1093/clipsy.8.3.255

Kanter, J. W., Mulick, P. S., Busch, A. M., Berlin, K. S., & Martell, C. R. (2007). The Behavioral Activation for Depression Scale (BADS): Psychometric properties and factor structure. Journal of Psychopathology and Behavioral Assessment, 29(3), 191–202. https://doi.org/10.1007/s10862-006-9038-5

Kanter, J. W., & Puspitasari, A. J. (2016). Global dissemination and implementation of behavioural activation. The Lancet, 388(10047), 843–844. https://doi.org/10.1016/S0140-6736(16)31131-X

Kanter, J. W., Puspitasari, A. J., Santos, M. M., & Nagy, G. A. (2012). Behavioural activation: History, evidence and promise. British Journal of Psychiatry, 200(5), 361–363. https://doi.org/10.1192/bjp.bp.111.103390

Koşan, Y., & Kaya, S. (2023). Psychometric Properties and Factor Structure of the Turkish Version of the Short Form of Behavioral Activation for Depression Scale (BADS-SF) in Non-Clinical Adults. E-International Journal of Educational Research, 14(3), 198–211. https://doi.org/https://doi.org/10.19160/e-ijer.1227934

Krings, A., Bortolon, C., Yazbek, H., & Blairy, S. (2021). Psychometric properties and factor structure of the french version of the behavioral activation for depression scale (bads) in non-clinical adults. Psychologica Belgica, 61(1), 20–32. https://doi.org/10.5334/pb.542

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 16(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lehmann, D. C., & Bördlein, C. (2020). A Systematic Review of Culturally Adapted Behavioral Activation Treatments for Depression. Research on Social Work Practice, 30(6), 688–702. https://doi.org/10.1177/1049731520915635

Li, M., Ding, C., Kanter, J. W., Zeng, Q., & Yang, D. (2014). Further evaluation of the psychometric properties of the behavioral activation for depression scale. International Journal of Psychology and Psychological Therapy, 14(1), 45–57.

Linacre, J. M. (2012). Winsteps [Software]. http://www.winsteps. com/index.html.

Linacre, J. M. (2020). A user's guide to Winsteps Ministep Rasch-model computer programs. winsteps.com.

Lubis, F. Y., Abidin, F. A., Qodariah, L., Anindhita, V., & Purba, F. D. (2024). Coping strategies and psychological distress among mothers during COVID-19 pandemic: The moderating role of social support. PLoS ONE, 19(4 April), 1–12. https://doi.org/10.1371/journal.pone.0300365

Luitel, N. P., Jordans, M. J. D., Kohrt, B. A., Rathod, S. D., & Komproe, I. H. (2017). Treatment gap and barriers for mental health care: A cross-sectional community survey in Nepal. PLoS ONE, 12(8), 1–15. https://doi.org/10.1371/journal.pone.0183223

Manos, R. C., Kanter, J. W., & Luo, W. (2011). The Behavioral Activation for Depression Scale-Short Form: Development and Validation. Behavior Therapy, 42(4), 726–739. https://doi.org/10.1016/j.beth.2011.04.004

Mazzucchelli, T., Kane, R., & Rees, C. (2009). Behavioral activation treatments for depression in adults: A meta-analysis and review. Clinical Psychology: Science and Practice, 16(4), 383–411.

https://doi.org/10.1111/j.1468-2850.2009.01178.x

Mekonen, T., Chan, G. C. K., Connor, J. P., Hides, L., & Leung, J. (2021). Estimating the global treatment rates for depression: A systematic review and meta-analysis. Journal of Affective Disorders, 295(September), 1234–1242. https://doi.org/10.1016/j.jad.2021.09.038

Mohammadi, A., & Amiri, M. (2010). Behavioral Activation for Depression Scale: Psychometric properties and confirmatory factor analysis for Persian version. Iranian Journal of Psychiatry and Clinical Psychology, 16(1), 65–73.

Mousavi, A., & Cui, Y. (2020). The effect of person misfit on item parameter estimation and classification accuracy: A simulation study. Education Sciences, 10(11). https://doi.org/10.3390/educsci10110324

National Institute for Health and Clinical Excellence. (2022). Depression in adults: treatment and management NICE guideline. In NICE Guideline (Issue June). www.nice.org.uk/guidance/ng222

Oh, J.-H., Hwang, N.-R., Cha, Y.-J., Lee, E.-B., Choi, K.-H., & Seo, H.-J. (2017). The Reliability and Validity of the Korean Version of Behavioral Activation for Depression Scale. Journal of Korean Neuropsychiatric Association, 56(2), 89. https://doi.org/10.4306/jknpa.2017.56.2.89

Peristianto, S. V., Subandi, M. A., & Utami, M. S. (2025). Empathy and self-awareness (rasa rumangsa) in the Javanese cultural context. Culture and Psychology, 31(2), 713–733. https://doi.org/10.1177/1354067X251315737

Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. Value in Health, 18(1), 25–34. https://doi.org/10.1016/j.jval.2014.10.005

Petts, R. A., Foster, C. S., Douleh, T. N., & Gaynor, S. T. (2016). Measuring activation in adolescent depression: Preliminary psychometric data on the Behavioral Activation for Depression Scale-Short Form. Behavior Analysis: Research and Practice, 16(2), 65–80. https://doi.org/10.1037/bar0000036

Raes, F., Hoes, D., Van Gucht, D., Kanter, J. W., & Hermans, D. (2010). The Dutch version of the behavioral activation for depression scale (BADS): Psychometric properties and factor structure. Journal of Behavior Therapy and Experimental Psychiatry, 41(3), 246–250. https://doi.org/10.1016/j.jbtep.2010.02.001

Ryder, A. G., Yang, J., & Heine, S. J. (2002). Somatization vs. Psychologization of Emotional Distress: A Paradigmatic Example for Cultural Psychopathology. Online Readings in Psychology and Culture, 10(2), 1–22. https://doi.org/10.9707/2307-0919.1080

Sagar, R., Pattanayak, R. D., Chandrasekaran, R., Chaudhury, P. K., Deswal, B. S., Lenin Singh, R. K., Malhotra, S., Nizamie, S. H., Panchal, B. N., Sudhakar, T. P., Trivedi, J. K., Varghese, M., Prasad, J., & Chatterji, S. (2017). Twelve-month prevalence and treatment gap for common mental disorders: Findings from a large-scale epidemiological survey in India. Indian Journal of Psychiatry, 59(1), 46–55. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_333_16

Setiyawati, Diana, Jatmika, Wulan Nur, Read, Ursula M, & Colucci, Erminia. (2025). "They think we are trying to combine knowledges, whereas we are trying to harmonize them": A visual exploration of what promotes successful collaboration between mental health workers and healers in Indonesia. Transcultural Psychiatry, 13634615251338906. https://doi.org/10.1177/13634615251338906

Shudo, Y., & Yamamoto, T. (2017). Assessing the relationship between quality of life and behavioral activation using the Japanese Behavioral Activation for Depression Scale-Short Form. PLoS

ONE, 12(9), 1–9. https://doi.org/10.1371/journal.pone.0185221

Shudo, Y., Yamamoto, T., & Sakai, M. (2017). Longitudinal Predictions of Depression Symptoms Using the Activation and Avoidance Subscales of the Japanese Behavioral Activation for Depression Scale-Short Form. Psychological Reports, 120(1), 130–140. https://doi.org/10.1177/0033294116680794

Stein, A. T., Carl, E., Cuijpers, P., Karyotaki, E., & Smits, J. A. J. (2021). Looking beyond depression: A meta-analysis of the effect of behavioral activation on depression, anxiety, and activation. Psychological Medicine, 51(9), 1491–1504. https://doi.org/10.1017/S0033291720000239

Sugiarto, R., & Suparno, K. (2025). The Javanese Rasa: Sense, psychology, and spirituality. Culture and Psychology, 31(2), 734–765. https://doi.org/10.1177/1354067X241300978

Sumintono, B., & Widhiarso, W. (2013). Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial. Tim Komunikata Publishing House.

Sun, J. (2005). Assessing goodness of fit in confirmatory factor analysis. Measurement and Evaluation in Counseling and Development, 37(4), 240–256. https://doi.org/10.1080/07481756.2005.11909764

Teismann, T., Ertle, A., Furka, N., Willutzki, U., & Hoyer, J. (2016). The German Version of the Behavioral Activation for Depression Scale (BADS): A Psychometric and Clinical Investigation. Clinical Psychology & Psychotherapy, 23(3), 217–225. https://doi.org/10.1002/cpp.1948

The WHOQOL Group. (1998). Development of the World Health Organization WHOQOL-BREF Quality of Life Assessment. Psychological Medicine, 28, 551–558. https://doi.org/10.1017/S0033291798006667

Umar, J., & Nisa, Y. F. (2020). Uji Validitas Konstruk dengan CFA dan Pelaporannya. Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I), 9(2), 1–11. https://doi.org/10.15408/jp3i.v9i2.16964

Uphoff, E., Ekers, D., Robertson, L., Dawson, S., Sanger, E., South, E., Samaan, Z., Richards, D., Meader, N., & Churchill, R. (2020). Behavioural activation therapy for depression in adults. Cochrane Database of Systematic Reviews, 2020(7). https://doi.org/10.1002/14651858.CD013305.pub2

Uphoff, E., Pires, M., Barbui, C., Barua, D., Churchill, R., Ekers, D., Fottrell, E., Mazumdar, P., Purgato, M., Rana, R., Wright, J., & Siddiqi, N. (2019). Behavioural activation therapies for depression in adults with non-communicable diseases. Cochrane Database of Systematic Reviews, 2019(10). https://doi.org/10.1002/14651858.CD013461

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. Revue Europeenne de Psychologie Appliquee, 54(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Wagener, A., Van Der Linden, M., & Blairy, S. (2015). Psychometric properties of the French translation of the Behavioral Activation for Depression Scale-Short Form (BADS-SF) in non-clinical adults. Comprehensive Psychiatry, 56, 252–257. https://doi.org/10.1016/j.comppsych.2014.10.008

WHO. (1996). WHOQOL-BREF: Introduction, administration, scoring and generic version of the assessment, field trial version, December 1996. In World Health Organization (pp. 1–16). https://apps.who.int/iris/bitstream/handle/10665/63529/WHOQOL-BREF.pdf?sequence=1&isAllowed=y

Widhiarso, W., & Sumintono, B. (2016). Examining response aberrance as a cause of outliers in statistical analysis. Personality and Individual Differences, 98, 11–15.

https://doi.org/10.1016/j.paid.2016.03.099

World Health Organization. (2016). mhGAP Intervention Guide for mental, neurological and substance use disorders in non-specialized health settings: Version 2.0. In World Health Organization. World Health Organization.

World Health Organization. (2022). World mental health report: Transforming mental health for all. World Health Organization.

Appendix

Behavioral Activation for Depression Scale – Short Form (BADS-SF) (Original Version)

Please read each statement carefully and then circle the number which best describes how much the statement was true for you DURING THE PAST WEEK, INCLUDING TODAY.

Responses:

0 = Not at all
1
2 = A little
3
4 = A lot
5
6 = Completely

Items:

1. There were certain things I needed to do that I didn't do.* (AC)
2. I am content with the amount and types of things I did. (AC)
3. I engaged in many different activities. (AC)
4. I made good decisions about what type of activities and/or situations I put myself in. (AC)
5. I was an active person and accomplished the goals I set out to do. (AC)
6. Most of what I did was to escape from or avoid something unpleasant.* (AV)
7. I spent a long time thinking over and over about my problems.* (AV)
8. I engaged in activities that would distract me from feeling bad.* (AV)
9. I did things that were enjoyable. (AC)

*reverse item: score on this item was reversed to compute total score.

Notes: AC = Activation subscale; AV = Avoidance subscale.

Behavioral Activation for Depression Scale – Short Form (BADS-SF) (Indonesian Version)

Silakan baca setiap pernyataan dengan cermat, lalu tandai nomor yang paling menggambarkan sejauh mana pernyataan itu benar bagi anda selama seminggu terakhir, termasuk hari ini.

Pilihan respon:

0 = Sama sekali tidak
1
2 = Sedikit
3
4 = Banyak
5
6 = Sepenuhnya benar

Item:

1. Ada hal-hal tertentu yang perlu saya lakukan, tetapi tidak saya lakukan.* (AC)
2. Saya puas dengan banyak dan jenis aktivitas yang saya lakukan. (AC)
3. Saya terlibat dalam banyak kegiatan yang berbeda. (AC)
4. Saya membuat keputusan yang tepat tentang jenis aktivitas yang saya lakukan dan/atau situasi di mana saya berada di dalamnya. (AC)
5. Saya adalah orang yang aktif dan bisa mencapai tujuan yang saya tetapkan. (AC)
6. Sebagian besar dari apa yang saya lakukan adalah untuk menghindari sesuatu yang tidak menyenangkan.* (AV)
7. Saya menghabiskan banyak waktu untuk terus-menerus memikirkan masalah-masalah saya.* (AV)

8. Saya melakukan kegiatan yang akan mengalihkan saya dari perasaan buruk.* (AV)
9. Saya melakukan hal-hal yang menyenangkan. (AC)

*skor pada item ini dibalik untuk perhitungan skor total.

Catatan: AC = Subskala Aktivasi (Activation); AV = Subskala Penghindaran (Avoidance)

Scoring

BADS-SF is a self-report measure consisting of 9 items and being developed to capture changes during behavioral activation therapy for depression. Each item is rated from 0 (not at all) to 6 (completely) based on the extent of the statement of each item is true for patient during the past week.

Total score of the BADS-SF ranges from 0-54. To compute total score, item 1, 6, 7, and 8 should be reverse scored. Higher score of the BADS-SF represents higher activation or lower avoidance level.

http://journal.uinjkt.ac.id/index.php/jp3i