Pertanika J. Soc. Sci. & Hum. 33 (6): 2459 - 2476 (2025)

PERTANIKA JOURNALS

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Psychometric Evaluation of Subjective Happiness Scale in the Indonesian Context

Clement Eko Prasetio[1,2]*, Mellia Christia[1], Airin Yustikarini Saleh[1], and Agnes Sianipar[1]

[1]Cognition, Affect, and Well-being Lab, Faculty of Psychology, Universitas Indonesia, Jalan Lingkar Kampus Raya, Jalan. Prof. DR. R. Slamet Iman Santoso, Pondok Cina, Kecamatan Beji, Depok City, West Java, 16424, Indonesia

[2]Indonesian Psychological Healthcare Center, Bursa Efek Indonesia Tower 1, Jl. Jenderal Sudirman Kav 52-53 28th Floor, Unit 2805, Senayan, Kebayoran Baru, South Jakarta, Jakarta, 12190, Indonesia

ABSTRACT

The Subjective Happiness Scale (SHS) has been widely used to measure global happiness. However, only a few studies have critically evaluated the psychometric properties of the Indonesian version of the SHS using rigorous techniques. This study aimed to address this gap by conducting a comprehensive assessment of the SHS using complementary Classical Test Theory (CTT) and Item Response Theory (IRT) frameworks to provide a robust and nuanced evaluation of the scale's reliability, validity, and item-level performance. Indonesian-speaking participants completed the Indonesian SHS, the General Health Questionnaire-12 (GHQ-12), and the Satisfaction with Life Scale (SWLS). Results indicated that the Indonesian SHS demonstrated acceptable reliability and validity. Nevertheless, Item 4 consistently underperformed in capturing the happiness construct. These findings highlight the importance of cultural and linguistic adaptation when using the SHS in diverse cultural contexts. The study offers robust psychometric evidence of the scale's internal structure, item functioning, and construct validity of the Indonesian SHS, supporting its use in future well-being research.

Keywords: Indonesian, item response theory, reliability, subjective happiness scale, validity

ARTICLE INFO

Article history:
Received: 01 December 2024
Accepted: 02 September 2025
Published: 30 December 2025

DOI: https://doi.org/10.47836/pjssh.33.6.14

E-mail addresses:
clementeko.work@gmail.com (Clement Eko Prasetio)
mellia-m@ui.ac.id (Mellia Christia)
airinys@ui.ac.id (Airin Yustikarini Saleh)
agnes.nauli@ui.ac.id (Agnes Sianipar)
* Corresponding author

INTRODUCTION

The term "happiness" is quite broad and encompasses many aspects. Generally, "happiness" can be categorised into three domains (Iskandarsyah, 2022): the health domain (Quality of Life) measured by instruments such as WHOQOL (World Health Organisation, 2012) or EQ-

5D-5L (Purba et al., 2018), the socio-economic domain typically measured by socio-economic status/wealth, and the psychological domain. Happiness in the psychological domain can be distinguished into two schools of thought: eudaimonic (e.g., Deci & Ryan, 1985; Ryff & Keyes, 1996) and hedonic (e.g., Diener, 1984) well-being. This study will focus on the perspective of hedonic well-being.

Measurement within the domain of hedonic well-being varies considerably. Some approaches operationalize it as life satisfaction (Diener et al., 1985), while another approach defines it as the frequency of an individual's affective state, encompassing both negative and positive affect (Diener et al., 2009; Watson et al., 1988). However, there was no existing tool that measured global happiness. Lyubomirsky and Lepper (1999) developed the Subjective Happiness Scale (SHS) to address this gap.

The Subjective Happiness Scale (SHS) measures global happiness as individuals experience it, if most individuals can make assessments and judgments regarding their life experiences, including their happiness level. Such happiness assessment is based on a subjective and global assessment of being happy or unhappy across various life experiences. Therefore, someone may assess themselves as 'very happy' even if certain aspects of their life are *somewhat* pleasant (Lyubomirsky & Lepper, 1999). The SHS consists of four items: two items ask respondents to evaluate themselves without involving any social comparison (or absolute ratings), while the other two items ask respondents to evaluate themselves in terms of whether they would fit descriptions of happy and unhappy individuals, relative to their peers.

The SHS has been translated into various languages, such as Malay (Swami, 2008), Tagalog and German (Swami et al., 2009), Chinese (Nan et al., 2014), Greek (Karakasidou et al., 2016), Myanmar and Vietnamese (Powell et al., 2020), and Nigerian (Agbo, 2021). The SHS has demonstrated excellent reliability (Agbo, 2021; Kocjan et al., 2022; Powell et al., 2020) and validity (e.g., construct validity, convergent validity), indicating that the SHS can accurately measure general happiness (Agbo, 2021; Kocjan et al., 2022; Powell et al., 2020). Some studies have also utilised the SHS to measure subjective happiness across various cultural groups via cross-cultural comparisons (Kocjan et al., 2022; Swami et al., 2009).

The SHS has also been used to measure individuals' global happiness in the Indonesian population. For instance, it has been applied in research on happiness among low-income urban communities (Kharisma, 2022) and in a study examining Jakarta residents' perceptions of public open spaces and their well-being (Supriyanto, 2020). Despite its extensive use in the Indonesian population, the psychometric equivalence of Indonesian SHS and the original SHS has remained underexplored. At the outset of the current study, we had identified only one unpublished undergraduate thesis (Lumban Tobing, 2009) and one published research article (Powell et al., 2020), both

with small sample sizes, that specifically examined the psychometric properties of SHS in the Indonesian context. Hence, the primary objective of this study is to address this gap by utilising a larger and more diverse Indonesian sample to conduct a comprehensive psychometric validation of the SHS in Indonesia. These efforts are crucial to ensure that the Indonesian version of the SHS is equally valid and reliable as the original version (International Test Commission [ITC], 2018).

The first adaptation and psychometric testing of the Indonesian version of the SHS were conducted by Lumban Tobing (2009) using a general population sample (n = 583) from urban areas (Jakarta, Bogor, Tangerang, Depok, Bekasi), with 54% of participants being female and 54.1% of participants in young adulthood. The study found that the SHS had good reliability (Cronbach's alpha = 0.791) and demonstrated good divergent validity, with moderate but significant correlations with constructs such as depression (r = -0.318, p = 0.002) and self- actualisation characteristics (r = 0.310, p = 0.003). In another study, Powell et al. (2020) conducted psychometric testing of the SHS in four Southeast Asian countries, including Indonesia (n = 231), using CFA and IRT analyses to assess construct validity. Powell et al. (2020) found that while the SHS demonstrated good overall reliability and construct validity, the fourth item showed deficient psychometric scores across several countries, including Myanmar, Vietnam, Indonesia, and Malaysia. However, given the small sample size and the selective characteristics of the Indonesian participants (all of whom were university students from a limited number of universities), some selection bias may have affected the findings (Powell et al., 2020).

This study aims to conduct a comprehensive psychometric evaluation of the Indonesian version of the Subjective Happiness Scale (SHS) to strengthen the evidence for its validity and reliability. Although earlier work has laid the groundwork for such efforts, including a preliminary study by Lumban Tobing (2009), that manuscript remains unpublished, has not been peer-reviewed, and focusses on the urban populations. A recent study (Powell et al., 2020) used a modest sample size of Indonesians to collect the psychometric evidence. Such sampling constraints may compromise the robustness and generalisability of psychometric findings (Hair et al., 2019). To address the limitation of generalizability, the present study will use a larger sample and include participants from various regions in Indonesia.

During the adaptation process, certain biases may still arise. First, there is the potential for construct bias, as happiness may be understood differently across cultures. To minimise this bias, we conducted expert judgments to ensure the construct remains relevant and widely recognised by the Indonesian population. Statistically, we also employed two approaches to gather evidence of validity and reliability: Classical Test Theory or CTT (i.e., reliability, CFA, convergent validity, and test criterion) and Item Response Theory or IRT.

The use of CTT and IRT, two complementary approaches, contributes theoretically by capturing different aspects of measurement quality. CTT offers accessible indices, such as reliability, convergent validity, and factor structure, through confirmatory factor analysis, while IRT provides item-level precision and parameter invariance that align more closely with latent trait theory (De Ayala, 2022). The integration of both methods enhances the robustness of the validation process and advances theoretical understanding of cross-cultural measurement (Kaplan & Sacuzzo, 2018).

Second, method bias is a concern, particularly regarding the range of the response scale (from 1 to 7) and the test administration process. To address this, we conducted pilot tests with a small group of individuals to ensure that the instructions, items, and response options were well understood and that the test could be administered correctly. Finally, item bias may occur during translation, particularly if sentences are too lengthy or complex or if specific changes alter the intended meaning. To mitigate this, expert evaluations and readability tests were performed. Additionally, item discrimination analysis assessed each item's ability to differentiate individual happiness levels.

METHODS

Research Design

This study aimed to collect psychometric evidence of SHS and employed a cross-sectional design. Various analytical techniques were utilised, including reliability analysis, validity analysis, and the application of Item Response Theory (see the Analysis section).

Participants and Procedures

This study employed convenience sampling, targeting Indonesian citizens in urban areas with Internet access. The participants were classified as young adults (18–39 years old), proficient in the Indonesian language and had completed at least high school education. Information about participating in the study was disseminated through the researchers' social networks and social media platforms, such as LinkedIn. The total number of participants was 1,464, aged 18-39 ($M = 22.53$, $SD = 3.57$). Most participants were female (80.3%), undergraduate students (65.8%), and residents from West Java (35.8%), Jakarta (26.8%), Banten (10.4%), East Java (10%), and Central Java (6.5%), and other areas in Indonesia (e.g., Yogyakarta, Aceh, Maluku; 10.5%). Each participant provided informed consent before participation. Those who agreed to participate in the study completed online questionnaires, including the Subjective Happiness Scale (Lyubomirsky & Lepper, 1999), the Satisfaction with Life Scale (Diener et al., 1985), and the General Health Questionnaire-12 (Goldberg & Blackwell, 1970). The SWLS (Muttaqin, 2022) and GHQ-12 (Primasari & Hidayat, 2016) questionnaires have been adapted into Bahasa Indonesia and have demonstrated good psychometric properties. The Ethics Committee of the Faculty of Psychology, Universitas Indonesia (163/FPsi.Komite

Etik/PDP.04.00/2022) approved all procedures and steps in this research.

Instruments

Subjective Happiness Scale (SHS)

The Subjective Happiness Scale is a psychological measurement of individuals' subjective happiness (Lyubomirsky & Lepper, 1999). This scale consists of four items (the fourth is reverse scored), which relate to the subjective judgment about their happiness. Each item has Likert-scale responses ranging from 1 to 7. More specifically, two items assess self- perceptions based on absolute ratings of well-being and ratings relative to peers. The other two present brief descriptions of happy and unhappy individuals and ask respondents to evaluate the extent to which the descriptions fit themselves. A single final score is calculated by averaging the scores of the four items. Thus, the possible final score will range from 1 to 7, with a higher score representing greater subjective happiness.

When translating the Subjective Happiness Scale, we used a forward-backward translation procedure. Two experts performed the forward translation process; one was an author of the current study (AN), and the other was a professional translator with expertise in happiness studies. For the back-translation, two professional translators with backgrounds in psychology and a familiarity with the topic of happiness were employed. Each translator worked independently at every stage of the translation process. The authors' team (MC, CEP, AYS, and AN) discussed the results of each translator's work. Expert judgment was conducted by two experts in the field of happiness and well-being. The results of the expert judgment were used to enhance the translational quality and to ensure cultural relevance, construct equivalence, and linguistic fluency of the Indonesian version of the SHS.

Satisfaction with Life Scale (SWLS)

The Satisfaction with Life Scale (SWLS, henceforth) (Diener et al., 1985; Muttaqin, 2022) is one of the most widely used instruments for measuring subjective well-being and happiness in psychology. The SWLS measures the cognitive aspect of well-being using five items. Each item has a 7-point Likert response format. The scale's total score is determined by averaging the scores of the items. The higher score reflects higher well-being.

General Health Questionnaire (GHQ)

The General Health Questionnaire is a widely used and valid instrument for screening psychological distress. The scale was developed by Goldberg and Blackwell (1970) and adapted for use in Bahasa Indonesia by Primasari and Hidayat (2016). There are 12 items, and each item uses a 4-point Likert scale. The total score of GHQ-12 is computed by summing all the items. GHQ-12 also has a cut-point score to discriminate between people who have psychological distress and those who do not have psychological distress. The cut-point is 11, with a score above/equal

to the cut-off point representing having psychological distress.

Analysis

Reliability Analysis

The reliability analysis method used in this research was internal consistency. Internal consistency aimed to see the consistency between items and the measured construct. Cronbach's alpha (α) and Construct reliability (Hair et al., 2019) were used to estimate the internal consistency. The reliability was higher than or equal to 0.70, indicating that the instrument has good reliability (Hair et al., 2019). We also checked the item discriminant with a cut-off score higher than 0.30 (Hair et al., 2019). Jamovi was used to analyse reliability.

Validity Analysis

We conducted a validity analysis by collecting two sources of validity evidence: evidence based on the internal test structure (or construct validity) and evidence based on relations to other variables (or convergent validity and test criterion) (American Educational Research Association et al., 2014; Urbina, 2014). All the evidence was analysed using Jamovi Software.

Evidence of the Internal Test Structure (Construct Validity). We used confirmatory factor analysis (CFA) to collect evidence based on internal structure. Confirmatory factor analysis is a statistical analysis technique that uses a hypothetical model to estimate the population covariance matrix in comparison to the observed covariance matrix (Hair et al., 2019). There are some statistics criteria to decide the fitness of the SHS model according to observed data (Hair et al., 2019; Kline, 1998, 2005), including p-value (>0.05), $\chi 2/df$ (≤ 3), CFI (≥ 0.90), GFI (≥ 0.90), RMSEA (≤ 0.07), and SRMR (< 0.08). Factor loadings higher than 0.30 were acceptable (Hair et al., 2019). The maximum likelihood method was used for estimation.

Evidence based on Relations to Other Variables (Convergent Validity). We used convergent validity and the test-criterion relationship (contrasted-group method) to collect evidence based on relationships with other variables. The Subjective Happiness Scale was correlated with the Satisfaction with Life Scale (SWLS) for convergent validity. Convergent validity is applied when a measurement correlates with other tests that measure the same construct as the intended measures (Kaplan & Saccuzzo, 2018). In this research, SHS and SWLS have the same or similar construct (i.e., happiness/subjective well- being). Pearson correlations were used to analyse the relationship between variables.

For the test-criterion relationship, we used GHQ-12 as a criterion for mental health conditions. Individuals with a GHQ-12 score below 11, as indicated, are considered to have good mental health (Primasari & Hidayat, 2016). A GHQ-12 score higher than 11 indicates poor mental health. Based on the GHQ-12 cut-off score, we compared the scores of SHS between the two groups. The t-test was used to determine whether there were significant differences in SHS's scores.

IRT Analysis

IRT analysis was conducted using Jamovi software. Before performing the IRT analysis, we tested the assumptions of unidimensionality and local independence. Unidimensionality refers to the concept that all items in a measurement tool assess a single, underlying construct. This assumption was tested using Confirmatory Factor Analysis (CFA), and the results will be presented alongside the evidence collected based on the internal structure. Local independence assumes that each item is strongly associated with the primary construct and weakly associated with other items. This assumption was tested using the Q3 correlation matrix (Christensen et al., 2017).

The IRT analysis employed the Polytomous Rasch model (rating scale). Model fit indicators, such as the infit mean square and outfit mean square, were examined to determine whether the items aligned with the hypothesised construct. The infit mean square was generally more sensitive to the contribution of each measurement item within the expected model and was more commonly used. In contrast, the outfit mean square was more sensitive to errors or outliers in the model (Bond et al., 2021). According to Bond et al. (2021), reasonable infit and outfit mean square values ranged from 0.6 to 1.4 for rating scales. We also presented Item Characteristic Curves (ICC), rating scale deltas/thresholds, and item difficulty (beta) for each item. Rating scale deltas (thresholds) were used to assess the functioning of the Likert scale, determining whether participants' responses aligned with the scale's expectations (ranging from 1 = "not at all" or "less happy person" to 7 = "a great deal" or "happier person") (Bond et al., 2021). The interpretation of thresholds is that the lower (or more negative) the delta values, the lower the individual's ability (i.e., happiness); whereas the higher (or more positive) the deltas, the higher the individual's ability (i.e., happiness). Item difficulty (beta) indicates how difficult it is for individuals to respond to a particular item. The lower (or more negative) the beta, the easier the item is to endorse; likewise, the higher (or more positive) the beta, the more difficult the item is to endorse.

RESULTS

Reliability

The reliability analysis indicated that the SHS is reliable ($\alpha = 0.856$, construct reliability = 0.795). Regarding item analysis, specifically item discrimination, three items of the SHS (items 1-3) effectively distinguished between individuals with high and low subjective happiness. However, item 4 showed a low discrimination index. For more detailed information on the discrimination indices for each item, refer to Table 1.

We also checked the inter-item correlations between the items in the SHS (Table 2). Table 2 shows that all items had significant correlations, but item 4 had a weak correlation with the other items, ranging from $r = 0.163$ to $r = 0.167$. In other words, item 4 did not provide as much information about 'happiness' as the other three items and indicated that this specific item contributed the least to the construct.

Table 1
Subjective happiness scale's item discriminant score (English and Bahasa Indonesia)

Item (in English)	Item (in Bahasa Indonesia)	Item discriminant score
SHS 1. In general, I consider myself… (options: *(1) not a very happy person… (7) a very happy person*)	SHS 1. Secara umum, saya menganggap diri saya sebagai orang yang.. (opsi: *(1) sangat tidak bahagia… (7) sangat bahagia*)	0.685
SHS 2. Compared with most of my peers, I consider myself… (options: *(1) less happy… (7) more happy*)	SHS 2. Dibandingkan dengan teman-teman saya, saya menganggap diri saya sebagai orang yang… (opsi: *(1) lebih tidak bahagia… (7) lebih bahagia*)	0.665
SHS 3. Some people are generally very happy. They enjoy life regardless of what is going on, getting the most out of everything. To what extent does this characterization describe you? (options: *(1) not at all… (7) a great deal*)	SHS 3. Ada orang-orang yang secara umum sangat bahagia, mereka menikmati hidup mereka tanpa terpengaruh oleh apa yang sedang terjadi, dan sebisa mungkin menikmati segalanya secara penuh. Sejauh manakah pernyataan ini menggambarkan diri Anda? (opsi: *(1) tidak sama sekali menggambarkan… (7) sangat menggambarkan*)	0.626
SHS 4. Some people are generally not very happy. Although they are not depressed, they never seem as happy as they might be. To what extent does this characterization describe you? (options: *(1) not at all… (7) a great deal*)	SHS 4. Ada orang-orang yang secara umum tidak bahagia, meskipun mereka tidak tertekan/ depresi, mereka tidak terlihat bahagia seperti yang seharusnya. Sejauh manakah pernyataan ini menggambarkan diri Anda? (opsi: *(1) tidak sama sekali menggambarkan… (7) sangat menggambarkan*)	0.186

Note. *Options = in the SHS, each item has a Likert-scale model response ranging from 1 to 7

Table 2
Inter-item correlations (Pearson)

Items	SHS 1	SHS 2	SHS 3	SHS 4_R
SHS 1	-	-	-	-
SHS 2	0.743**	-	-	-
SHS 3	0.671**	0.658**	-	-
SHS 4_R	0.166**	0.163**	0.167**	-

Note. **p<0.01

Validity

Validity Evidence Based on Internal Test Structure (Construct Validity)

Based on the CFA results, it was evident that the structural model of the SHS demonstrates a good fit. This was supported by all fitness indicators of the SHS structural model meeting the predefined criteria (as outlined in the methods section): p-value = 0.574; $\chi 2/df = 0.556$; CFI = 1.00; GFI = 1.00; RMSEA = 0.000; SRMR = 0.005. All items in the SHS showed significant factor loadings

($p < 0.01$), except for item SHS 4, which had factor loadings less than 0.3 (Figure 1 and Table 3). Psychometrically, this indicated that item 4 had a weak relationship with the happiness constructed by the SHS. Consequently, a revision for item 4 is recommended or may be considered for exclusion in the Indonesian version of the SHS.

Validity Evidence Based on Relations to Other Variables (Convergent Validity and Test-criterion Relationships)

We found a significant positive correlation between the SHS and the SWLS ($r(1462) = 0.680$, $p < 0.01$). This indicated that the construct measured by the SHS (i.e., subjective happiness) aligns well with another well-established and widely used happiness measurement tool, the SWLS.

This study also found that the Subjective Happiness Scale (SHS) could distinguish between happiness levels in individuals who indicated experiencing psychological distress (GHQ-12) and those who did not (test-criterion relationship). In the group showing high psychological distress, there was a lower level of happiness ($N = 910$, $M = 16.59$, $SD = 4.00$) compared to the group not indicated for

Table 3
Factor loadings of the subjective happiness scale

Item	Unstandardized loadings (Std. Error) Standardised Factor	Standardised Factor Loading	Z-score (p-value)
SHS1	1.04 (0.027)	0.87	38.46 (< 0.01)
SHS2	1.17 (0.030)	0.85	39.17 (< 0.01)
SHS3	1.18 (0.032)	0.77	36.92 (< 0.01)
SHS4	0.34 (0.055)	0.20	6.08 (< 0.01)

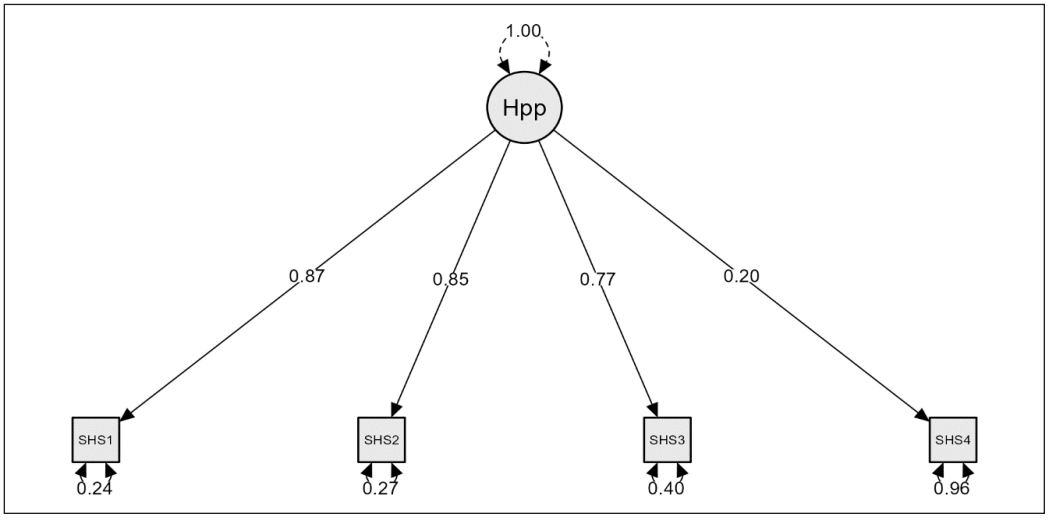


Figure 1. Model of SHS in confirmatory factor analysis (CFA)

psychological distress (N = 554, M = 20.92, SD = 3.34). The results of the t-test analysis revealed a significant difference between the group with psychological distress and the group without psychological distress ($t(1325.6) = 22.27$, $p < 0.001$).

Based on these findings, the SHS in the young adult population of Indonesia exhibits an internal structure akin to the hypothesized model. SHS and other measurement instruments assessing happiness-related constructs are also significantly correlated. Further details will be explained in the subsequent section. Moreover, SHS can effectively discriminate between individuals who have experienced psychological distress and those who have not.

Item Response Theory (IRT)

Testing for Unidimensionality and Local Independence

The results of the unidimensionality assumption test using CFA indicate that the SHS measures only one construct (p-value = 0.574; $\chi^2/df = 0.556$; CFI = 1.00; GFI = 1.00; RMSEA = 0.000; SRMR = 0.005). For more details, refer to the results on validity based on the internal test structure. The local independence assumption test results using the Q3 Correlation Matrix show high correlations between items, exceeding 0.2 or 0.3 (Table 4). This indicates the presence of local dependence. Therefore, caution is needed when interpreting the results from the IRT. Previous research has found that local dependence can lead to bias in certain indicators, such as item parameter estimates (Reese, 1995), theta scores (Zenisky et al., 2002), and an overestimation of overall test reliability (Sireci et al., 1991).

IRT Analysis

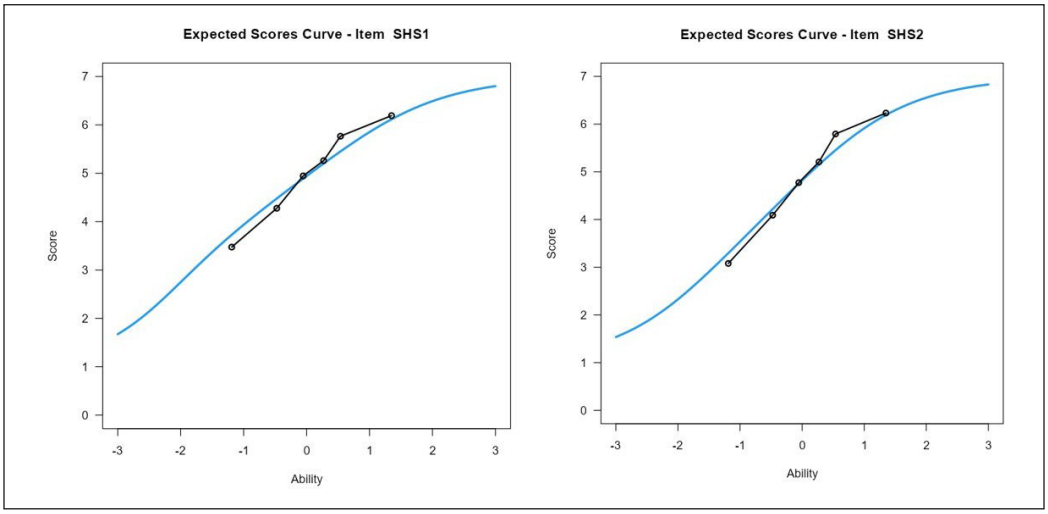
Based on the model fit analysis results (i.e., infit and outfit mean squares), it was found that items 1-3 of the SHS fall within the recommended score range (Bond et al., 2021), which is between 0.6 and 1.4. However, for item 4, the infit and outfit mean squares obtained were 1.6. This indicates that item 4 is slightly less productive in measuring the construct (Table 5). However, it does not significantly diminish the overall psychometric quality of the SHS. Thus, we must examine the ICC of each item to better understand model fit (Figure 2). Based on the ICC, it was found that items 1-3 have response patterns (jagged lines) that closely approximate the expected or hypothesised patterns. There are only minor deviations, which are insignificant

Table 4
Q3 correlation matrix for unidimensionality

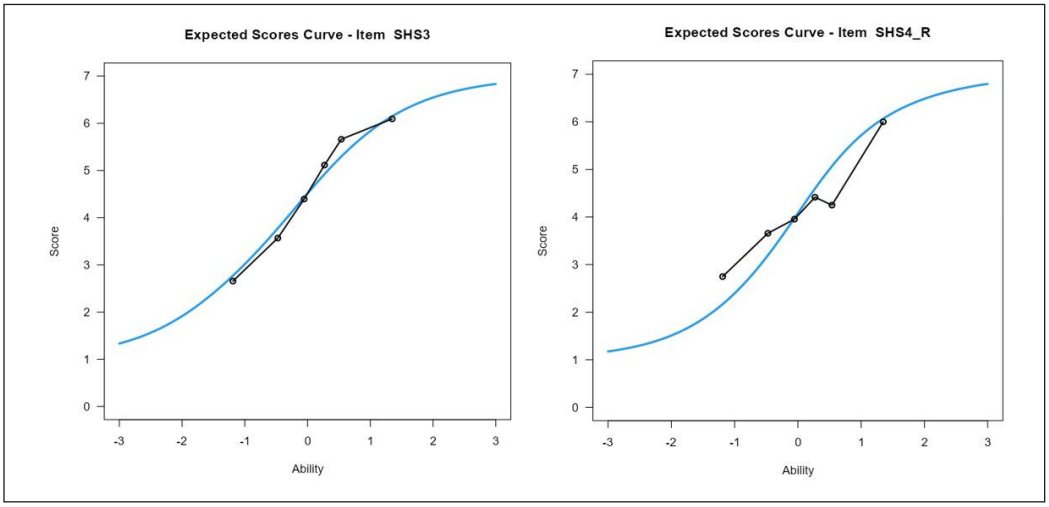
Items	SHS 1	SHS 2	SHS 3	SHS 4_R
SHS 1	-	-	-	-
SHS 2	0.320	-	-	-
SHS 3	0.158	0.161	-	-
SHS 4_R	-0.528	-0.515	-0.486	-

Table 5
Item statistics of the rating scale model

	Measure	S.E.Measure	Infit	Outfit
SHS1	-1.615	0.0235	0.591	0.621
SHS2	-1.502	0.0232	0.744	0.771
SHS3	-1.270	0.0226	0.923	0.923
SHS4_R	-0.977	0.0222	1.677	1.687



(a)



(b)

Figure 2. Item Characteristic Curve (ICC): (a) ICC for item 1 (SHS1) and 2 (SHS2); and (b) ICC for item 3 (SHS3) and 4 (SHS4_R)
Source: Authors' work

in shifting the items' infit/outfit scores. These results suggest that the response patterns for items 1-3 are consistent with the hypothesised/expected model. However, item 4 shows a noticeable deviation between the actual response pattern (jagged line) and the expected pattern, which aligns with the item's unacceptable infit/outfit scores. Despite this result, the actual response pattern for item 4 appears to be slightly aligned with the expected pattern.

One can examine the rating scale deltas/thresholds to understand the response pattern in options. Based on the rating scale deltas, it was found that there is an increase in delta values from category 1 (i.e., response 1 = not at all/less happy) to category 7 (i.e., response 7 = a great deal/happier). This indicates that the higher the response category, the higher the ability (level of happiness) individuals possess to respond to that option. Thus, the response options of the SHS items function effectively (see Table 6, the tauCat columns). Furthermore, the item difficulty of the SHS items tends to be low, although item 4 (or SHS4_R) appeared to be the most difficult to answer among the four SHS items.

DISCUSSION

This study provides empirical evidence that the Indonesian version of the SHS is reliable, and its scores adequately represent the construct of happiness. The SHS demonstrated good reliability. The gathered validity evidence, based on internal structure and relationships with other variables, supports the accurate interpretation of the SHS scores in the Indonesian version. Regarding internal structure, the results of the CFA analysis indicate that the relationships between items and test components are consistent with the theoretically proposed SHS construct model. Additionally, the SHS showed associations with other measurement tools, such as the Satisfaction with Life Scale (SWLS) and psychological distress indicators, as measured by the GHQ-12.

Furthermore, the SHS scores were able to differentiate individuals with high psychological distress from those with relatively low distress. In the IRT analysis, the response patterns for each SHS item generally aligned with the expected SHS construct. We also found that the overall

Table 6
Rating scale deltas/thresholds and item difficulty

Item	Beta*	tauCat 1	tauCat 2	tauCat 3	tauCat 4	tauCat 5	tauCat 6	tauCat 7
SHS1	-1.615	-5.900	-0.303	-0.192	0.683	1.103	1.917	2.692
SHS2	-1.502	-5.900	-0.303	-0.192	0.683	1.103	1.917	2.692
SHS3	-1.270	-5.900	-0.303	-0.192	0.683	1.103	1.917	2.692
SHS4_R	-0.977	-5.900	-0.303	-0.192	0.683	1.103	1.917	2.692

Note. *Beta = Item difficulty
Source: Authors' work

difficulty level of the SHS items was low, and the response options functioned as expected, showing an increase in ability with higher response choices.

Overall, the SHS remains a suitable measure of happiness. The findings of this study are consistent with various previous studies in multiple countries, such as Malaysia (Swami, 2008), Greece (Karakasidou et al., 2016), and Nigeria (Agbo, 2021), which also found high reliability of the SHS and adequate CFA validity evidence (Agbo, 2021; Karakasidou et al., 2016). The results of the CFA analysis also suggested that the SHS is a unidimensional construct, namely Happiness. This is consistent with other studies using CFA (Powell et al., 2020). Furthermore, the SHS and SWLS were moderately positively correlated, suggesting that both measures assess related constructs of happiness, though they focus on different aspects. The SWLS measures the cognitive dimension of happiness, while the SHS measures a broader, subjective evaluation that encompasses both mental and affective happiness. The correlation between the SHS and GHQ-12 was significant but negative, indicating that higher happiness levels were associated with lower levels of psychological distress. The Indonesian version of the SHS also effectively differentiated between individuals experiencing psychological distress and those who were not. Individuals with a GHQ score of 11 or higher tended to report lower levels of happiness compared to those without psychological distress. The IRT analysis further showed that SHS items aligned well with individual happiness levels—individuals with higher happiness were more likely to select higher response options on the SHS. This finding is consistent with the research by Powell et al. (2020). Additionally, the local dependence observed in this study and in Powell et al.'s (2020) research indicates a strong relationship between certain SHS items, particularly between items 1 and 2. However, further research is needed to confirm potential overlap with these items.

Upon closer examination of the SHS items, item 4 exhibited poor item discrimination ($r = 0.186$), weak inter-item correlations ($r = 0.163–0.167$), low factor loading (0.20), and misfit conditions (infit = 1.68; outfit = 1.69). This finding is consistent with several previous studies (O'Connor et al., 2015; Powell et al., 2020), which also reported that item 4 performed poorly statistically. This suggests that item 4 may not assess happiness as effectively as the other three items. The item's poor psychometric properties could be attributed to its linguistic structure, specifically its use of long sentences with multiple negative phrases, such as "not happy" or "not very happy." Lengthy and negatively phrased items can make it more difficult for individuals to understand the question's intent (O'Connor et al., 2015; van Sonderen et al., 2013).

The IRT analysis also indicated that item 4 was the most challenging in the SHS. We hypothesise that Indonesians may find it difficult to imagine a general state of "unhappiness," particularly when

this state is dissociated from psychological distress (e.g., depression). In the Indonesian mindset, if an individual feels happy, others around them are expected to feel happy too. This expectation is linked to the strong social bonds between individuals, which maintain social harmony (Uchida & Ogihara, 2012; Miyamoto et al., 2017). However, when presented with the idea that "there are people who are generally unhappy even though they are not experiencing depression or pressure," Indonesians tend to feel somewhat confused. This is because, in their perception, good social harmony should correlate with the absence of unhappiness. This concept is tied to the strong connection between the self and others in collectivist cultures (Miyamoto et al., 2017). Moreover, expressions of happiness or positive emotions in collectivist cultures are often more subdued than in individualistic cultures, as extreme displays of emotion may disrupt social harmony (Miyamoto et al., 2017). Nevertheless, further research is needed to explore and clarify these findings and hypotheses.

In terms of the conceptualisation of the construct, the SHS contains items that can measure global happiness. Compared to several previous measurement tools, such as Cantril's Ladder (Cantril, 1965), the SWLS (Diener et al., 1985), and the SPANE (Diener et al., 2009), most of the SHS items tend to be more straightforward, fewer in number, and address the global nature of happiness. Cantril's Ladder consists of only one item, which raises concerns about its psychometric properties. The items of the Indonesian version of the SHS appear to be easily understood by the Indonesian urban population. In other words, these items represent the operationalisation of global happiness as individuals experience it.

The Indonesian version of the SHS is well-suited for use in the general population, particularly among non-clinical groups. This is evidenced by the diverse backgrounds of the research sample of the current study, none of whom had specific diagnoses. However, based on the findings of this study, it appears that the SHS also has the potential to be used in clinical populations, such as individuals with depression or anxiety disorders. This assumption is supported by the finding that the SHS can significantly differentiate between groups of individuals with high and low levels of psychological distress. Furthermore, using SHS in other age groups, such as older adults or adolescents, may need further consideration through additional research. In adolescent or childhood age groups, it may be slightly more challenging to understand the existing items (especially item 4), and the concept of "happiness" may not be fully understood yet. On the other hand, in the older adult age group, this version's items may need to be presented more concisely and with fewer negative words to facilitate comprehension.

This study has some significant contributions to developing and adapting the Subjective Happiness Scale (SHS). First, based on a large and diverse sample, it provides a comprehensive psychometric evaluation of the Indonesian SHS using both

Classical Test Theory and Item Response Theory. Second, it reveals cultural nuances in how happiness and unhappiness are perceived, particularly the consistent misfit of item 4. Lastly, these results highlight the importance of culturally sensitive adaptations of global well-being measures and offer a theoretical basis for future cross-cultural research on subjective happiness.

CONCLUSION

In conclusion, various tools are available to assess happiness levels, including the Subjective Happiness Scale (SHS), which measures an individual's overall subjective happiness. Psychometric analysis based on Classical Test Theory (CTT) indicates that the SHS demonstrates good reliability, with all three items showing solid discriminatory power. Moreover, several pieces of valid evidence support the interpretation of SHS scores as a measure of happiness. To complement CTT, Item Response Theory (IRT) was also employed to analyse the strength of each item in representing the construct of happiness. Both analytical approaches suggest that the items on the SHS and the scale adequately capture the construct of happiness. However, item 4 exhibited weaker psychometric properties, implying its contribution to the happiness construct is relatively low. Therefore, revising item 4 in future applications of the Indonesian version of the SHS may be advisable. It is recommended that experts in well-being and linguistics be involved in this revision process. However, Lyubomirsky and Lepper (1999) and Lyubomirsky (2025) state that item 4 may be omitted when measuring happiness. Overall, the SHS is suitable for research and for measuring global subjective happiness in various areas. We also recommend revising the item 4 and conducting follow-up studies to evaluate its quality in a cross-cultural context.

Limitations and Recommendations

Several limitations were identified in this study. Firstly, regarding the sample. Although the sample size was sufficient, most respondents were from Jakarta and West Java. Obtaining data from more regions (predominantly rural areas) could broaden the generalizability of SHS usage and provide a more comprehensive understanding of happiness and the age range of respondents. Most respondents were young adults. Therefore, the representation of responses from other age groups (predominantly middle and late adulthood) must be improved. Representation from middle and late adulthood age groups is crucial as changes in happiness levels occur during these ages (Blanchflower, 2021). For future studies, research should involve a broader range of respondents, particularly age and residential location. Secondly, the validity of the evidence should be considered, specifically the relationship between the SHS and the measurement tools used to assess other variables. This study only used the SWLS to determine a variable like the SHS (convergent validity). The SWLS only measures happiness based on cognitive aspects, while the SHS measures global happiness (both cognitive and affective).

Therefore, further research should also employ measures of the affective component of happiness, such as the PANAS and SPANE.

ACKNOWLEDGEMENT

The authors sincerely thank Luthfi Setya Hilmawan, Vanessa Maudy Putri, and Nurul Salsabila for their valuable assistance in data collection. This study received no external funding, and the authors report no conflicts of interest. CEP conducted data analysis and drafted the manuscript; MC conducted data analysis and reviewed the manuscript; AYS reviewed the manuscript; and AS collected the data and contributed to writing and reviewing the manuscript. All authors read and approved the final version of the manuscript.

REFERENCES

Agbo, A. A. (2021). Assessment of the psychometric properties of the subjective happiness scale among emerging, young, and middle-aged adult populations in Nigeria. *Current Psychology*, *42*(7), 5765-5773. https://doi.org/10.1007/s12144-021-01884-4

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Barrington-Leigh, C. P. (2022). Trends in conceptions of progress and well-being. In. J.F. Helliwell, R. Layard, J.D. Sachs, J-E. De Neve, L.B. Aknin, S.Wang (Eds.), *World happiness report 2022* (pp. 53-74).

Blanchflower, D. G. (2021). Is happiness U-shaped everywhere? Age and subjective well-being in 145 countries. *Journal of Population Economics*, *34*(2), 575-624. https://doi.org/10.1007/s00148-020-00797-z

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th Ed.). Taylor and Francis. https://doi.org/10.4324/9780429030499

Cantril, H. (1965). *The pattern of human concerns*. Rutgers University Press.

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement*, *41*(3), 178-194. https://doi.org/10.1177/0146621616677520

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour*. Plenum Publishers.

De Ayala, R. J. (2022). *The Theory and Practice of Item Response Theory*. (2nd ed). The Guilford Press.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3).

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71-75.

Diener, E., Wirtz, D., Biswas-Diener, R., Tov, W., Kim-Prieto, C., Choi, D., & Oishi, S. (2009). New measures of wellbeing. In E. Diener (Ed.), *Assessing well-being: The collected works of Ed Diener* (Vol. 39). Springer Netherlands. https://doi.org/10.1007/978-90-481-2354-4

Fugl-Meyer, A. R., Melin, R., & Fugl-Meyer, K. S. (2002). Life satisfaction in 18- to 64-year-old Swedes: In relation to gender, age, partner, and immigrant status. *Journal of Rehabilitation Medicine*, *34*, 239-246.

Goldberg, D. P., & Blackwell, B. (1970). Psychiatric illness in general practice. A detailed study using a new method of case identification. *British Medical Journal*, *2*, 439-443.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.

International Test Commission (ITC). (2018). ITC guidelines for translating and adapting tests (Second Edition). *International Journal of Testing*, *18*(2), 101-134. https://doi.org/10.1080/15305058.2017.1398166

Iskandarsyah, A. (2022). Wellbeing concept: relevance, challenge, and boundaries in Indonesian context. In Z. Abidin, F. A. Abidin, J. R. Siregar, P. Joefiani, & R. U. Purwono (Eds.), *Wellbeing: concepts, research, and implementation in Indonesia* (1st ed., pp. 18-30). PT. Remaja Rosdakarya.

Kaplan, R. M., & Saccuzzo, D. P. (2018). *Psychological testing: principles, applications, & issues*. Cengage Learning.

Karakasidou, E., Pezirkianidis, C., Stalikas, A., & Galanakis, M. (2016). Standardization of the subjective happiness scale (SHS) in a Greek sample. *Psychology*, *07*(14), 1753-1765. https://doi.org/10.4236/psych.2016.714164

Kharisma, V. (2022). Experience-based food insecurity and subjective happiness: A case study of socioeconomically disadvantaged people in three urban areas of Indonesia. *International Journal of Community Well-being*, *5*(4), 777-791. https://doi.org/10.1007/s42413-022-00181- 7

Kocjan, G. Z., Jose, P. E., Sočan, G., & Avsec, A. (2022). Measurement invariance of the subjective happiness scale across countries, gender, age, and time. *Assessment*, *29*(4), 826-841. https://doi.org/10.1177/1073191121993558

Lumban Tobing, L. P. (2009). *Cultural adaptation of subjective happiness scale, satisfaction with life scale, and schedule for evaluation of individual quality of life-direct weighing* [Unpublished undergraduate thesis]. Universitas Indonesia.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, *46*, 137-1999.

Lyubomirsky, S. (2025, March 19). *Home: Subjective Happiness Scale*. https://sonjalyubomirsky.com/

Miyamoto, Y., Ma, X., & Wilken, B. (2017). Cultural variation in pro-positive versus balanced systems of emotions. *Current Opinion in Behavioral Sciences*, *15.* 27-32. https://doi.org/10.1016/j.cobeha.2017.05.014

Muttaqin, D. (2022). Measurement invariance of satisfaction with life-scale Indonesian Version.

Jurnal Psikologi Ulayat, 9(2). https://doi.org/10.24854/jpu438

Nan, H., Ni, M. Y., Lee, P. H., Tam, W. W. S., Lam, T. H., Leung, G. M., & McDowell, I. (2014). Psychometric evaluation of the Chinese version of the subjective happiness scale: Evidence from the Hong Kong family cohort. *International Journal of Behavioral Medicine*, *21*(4), 646-652. https://doi.org/10.1007/s12529-014-9389-3

O'Connor, B. P., Crawford, M. R., & Holder, M. D. (2015). An item response theory analysis of the subjective happiness scale. *Indicators Research*, *124*(1), 249-258. https://doi.org/10.1007/sl

Powell, M., Bateman, C. J., Gerasimova, D., Bateman, A., & Peltzer, K. (2020). Psychometric properties of the subjective happiness scale in four Asian countries. *Journal of Wellbeing Assessment*, *4*(3), 495-509. https://doi.org/10.1007/s41543-021-00045-5

Primasari, I., & Hidayat, R. (2016). General Health Questionnaire-12 (GHQ-12) as adjustment disorder's screening instrument. *Jurnal Psikologi*, *43*(2), 121-134.

Purba, F. D., Hunfeld, J. A. M., Iskandarsyah, A., Fitriana, T. S., Sadarjoen, S. S., Passchier, J., & Busschbach, J. J. V. (2018). Quality of life of the Indonesian general population: Test-retest reliability and population norms of the EQ-5D-5L and WHOQOL-BREF. *PLoS ONE*, *13*(5). https://doi.org/10.1371/journal.pone.0197098

Reese, L. M. (1995). *The impact of local dependencies on some LSAT outcomes*. Statistical Report. LSAC Research Report Series. Retrieved June 30, 2024, from https://eric.ed.gov/?id=ED469244

Ryff, C. D., & Keyes, C. L. M. (1996). The structure of psychological well-being. *Journal of Personality and Social Psychology*, *35*(5), 719-727. https://doi.org/10.2307/2093340

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the Reliability of Testlet-Based Tests. *Journal of Educational Measurement Fall*, *28*(3), 237-247.

Supriyanto. (2020). Sense of community and subjective happiness in Jakarta people who use RPTRA facilities. *Psikologi Positif Menuju Mental Wellness*.

Swami, V. (2008). Translation and validation of the Malay subjective happiness scale. *Social Indicators Research*, *88*(2), 347-353. https://doi.org/10.1007/s11205-007-9195-2

Swami, V., Stieger, S., Voracek, M., Dressler, S. G., Eisma, L., & Furnham, A. (2009). Psychometric evaluation of the Tagalog and German subjective happiness scales and a cross-cultural comparison. *Social Indicators Research*, *93*(2), 393-406. https://doi.org/10.1007/s11205-008-9331-7

Uchida, Y., & Ogihara, Y. (2012). Personal or interpersonal construal of happiness: A cultural psychological perspective. *International Journal of Wellbeing*, 354-369. https://doi.org/10.5502/ijw.v2.i4.5

Urbina, S. (2014). *Essentials of Psychological Testing* (2nd ed.). Wiley.

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068967

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070.

World Health Organisation. (2012). *WHOQOL User Manual*.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2002). Identification and evaluation of local item dependencies in the medical college admissions test. *Journal of Educational Measurement Winter*, *39*(4), 291-309.