



# The Short Depression-Happiness Scale—a Brief Measure of Well-Being—New Evidence of Scale Precision Using Item Response Theory

Elena Stănculescu<sup>1</sup>

Accepted: 10 November 2022 / Published online: 21 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Existing literature on measuring well-being (WB) indicates that The Short Depression-Happiness Scale (SDHS) is the only measure built on the negative–positive affect continuum. However, since the previous studies highlighted the psychometric properties of the SDHS exclusively using classical test theory (CTT), this research aims to fill this gap. By investigating the validity of this scale using both CTT and item response theory (IRT) approaches, this study aims to identify the precision of SDHS measurement. The sample consisted of 1326 Romanian participants (44.26% males, ages ranging from 18 to 64 years,  $M_{age}=34.09$ ,  $SD=9.4$ ). Results confirmed construct, convergent, and concurrent validity of the SDHS Romanian version. CFA replicated the unidimensionality found in the original study. The investigation of the scale precision through IRT showed that the SDHS is a highly reliable and accurate tool for capturing clinically low and moderate levels of WB, namely in the area of  $-3.5 SD$  and  $+1 SD$  from the mean of the latent trait. However, the scale does not perform well for higher levels of WB because there is no reliable information in the range between  $+2 SD$  and  $+3 SD$  of the latent trait continuum. Previous literature emphasized a similar pattern when measuring WB using the Warwick–Edinburgh Mental Well-Being Scale (WEMWBS). Additionally, the current study found that the SDHS has the advantage of measuring clinically low WB while the WEMWBS does not. The present research proposes a preliminary cut-off ( $<7$ ), before clinical assessment confirmation. Future clinical studies are needed. The findings of the current research should be useful for further psychometric developments of WB scales. Adding items of high difficulty would contribute to more measurement precision for higher levels of WB.

**Keywords** Psychometric properties · Measurement precision · Item response theory · Hedonic well-being · Eudaimonic well-being

What does it mean for people to feel good psychologically? Seligman (2011) suggested that pleasant life, good life, and meaningful life underpin the WB concept. This

---

✉ Elena Stănculescu  
elena.stanculescu@fpse.unibuc.ro

<sup>1</sup> Faculty of Psychology and Educational Sciences, University of Bucharest, 90 Panduri Street, Bucharest, Romania

operationalization combines two approaches on WB in positive psychology (Huta & Waterman, 2014; Ryan & Deci, 2001): hedonia and eudaimonia. Hedonia refers to the affective dimension of WB and includes happiness, the balance between positive and negative affect, and life satisfaction (Graham, 2017; Kahneman et al., 1999). Therefore, hedonic WB reflects what Seligman named pleasant life. Eudaimonia refers to striving for positive functioning (Ryff & Keyes, 1995; Ryff & Singer, 2008), or what Seligman (2011) called good life and meaningful life. Positive functioning means how well individuals are doing in managing own environment, achieving long-term purposes, making good relationships, and having high self-acceptance, contributing to self-actualization or personal growth. Additionally, a series of concepts in the literature have tried to capture as best as possible what it means for an individual to feel psychologically well: general WB (Longo et al., 2016, 2017, 2018), subjective WB (Diener, 2000; Lucas et al., 1996), psychological WB (Ryff, 1995, 2018), social WB (Keyes, 1998), positive mental health (Keyes, 2013), mental WB (Tennant & Conaghan, 2007), spiritual WB (Ellison, 1983; Fisher et al., 2000; Gomez & Fisher, 2003), and flourishing (Diener et al., 2010). Longo et al. (2018) highlight in a meta-analysis the existence of 14 main conceptualizations of WB in the literature: happiness, vitality, calmness, optimism, involvement, self-awareness, self-acceptance, self-worth, competence, development, purpose, significance, congruence, and connection. The authors emphasize that these constructs represent dimensions of the superordinate factor that they name “general WB.”

## Psychometric Properties of the SDHS

In the realm of clinical psychology, as well as positive psychology, psychometricians have developed tools that measure either depressive symptoms or subjective happiness. Considering the polarity of affectivity and that WB means more than the presence of positive affect and the absence of negative affect, an instrument was needed to measure WB on the depression-happiness continuum. At first glance, it might seem counterintuitive that a WB scale should include items that measure depression. But is not the case, because WB indicates if individuals' lives are getting better or worse. However, scientists have provided empirical support for the fact that coping with negative emotions significantly contributes to the state of WB (Gross & John, 2003; Mayordomo-Rodríguez et al., 2015). When maladaptive or defensive mechanisms appear instead of adaptive coping strategies, it can suggest ill-being and psychological disorders, including affective disorders such as depression. In brief, WB is a complex psychological state that reflects positive emotional experiences, individual's balance of positive over negative states (Lyubomirsky et al., 2005), or ability to cope with negative states.

Clinicians need psychometrically robust tools to measure the progress achieved in therapeutic interventions—progress that is quantified not only in the remission of symptoms but also in the improvement of WB. Since positive psychology appeared only three decades ago, clinicians quantify progress according to the reduction of dysfunctions (), more precisely the decrease of scores on the scales of depression, anxiety, stress, etc. (Holmqvist et al., 2015). Specialists in positive psychology have proposed a series of WB measures, focusing on the hedonic and eudaimonic approaches of this complex construct. From the hedonic perspective, the following scales were developed: Subjective Happiness Scale (Lyubomirsky et al., 2005), Satisfaction With Life Scale (Diener et al., 1985; Pavot & Diener, 2008), and subjective well-being (Diener, 2000; Diener et al., 2010). The

eudaimonic approach includes measures such as: psycho-social well-being (Ryff, 1989, 1995; Ryff & Keyes, 1995), The Meaning in Life Questionnaire (Steger et al., 2006), spiritual well-being (Ellison, 1983; Ellison & Smith, 1991), and The Questionnaire for Eudaimonic Well-Being (Waterman et al., 2010).

Increasingly more, specialists needed to articulate the two perspectives (Guidi et al., 2013; Keyes, 2005) to elaborate measures for assessing the progress obtained following therapeutic interventions. Joseph et al. (2004) put together the depression and happiness constructs and built an instrument that captures WB. The authors argued that the assessment of depressive symptoms could theoretically be extended through the zero point into measure of happiness. Items measuring depression (reverse scored) combined with happiness scores reflect the latent construct or factor measured by the SDHS, namely WB. The findings obtained in the original study proved the scale is psychometrically robust. Joseph et al. (2004) underlined the usefulness of the SDHS in research and progress assessments obtained during therapeutic interventions.

It is important to note that Joseph et al. (2004) have the merit of proposing for the first time an instrument that includes the depression-happiness continuum, combining the clinical model with positive psychology. The items of the SDHS fit into the hedonic approach of WB, because they underlie the affective component of WB. The same researchers proposed a more comprehensive scale (Longo et al., 2017, 2018) to measure WB, more precisely general WB. It is more comprehensive measure of WB, since it combines both hedonic with eudaimonic approaches of WB. However, the authors emphasized that they developed the SDHS to be a useful tool for measuring the levels of negative and positive affective states with rapid application in clinical settings.

## Current Research

As aforementioned, the original study highlighted good psychometric qualities of the SDHS. The literature also mentions several validation studies of adapted versions of the SDHS in Spanish (Martínez et al., 2018) and Arabic (Yildirim & Balahmar, 2020). To the author's knowledge, no study has yet been performed to validate the adapted versions in Romania. Such adaptation studies are thus needed so that researchers can conduct more cross-cultural research. In addition, previous validation studies were carried out using only the CTT approach. None used IRT analysis, which outperforms CTT (specifically CFA) by considering not only raw scores (as CFA does) but also score patterns and item properties, such as discrimination and difficulties parameters (Embretson & Reise, 2013). For this reason, the current study aims to address this gap in the literature, by including the modern psychometric approach of IRT, more precisely graded response model (GRM) to explore the precision of the SDHS. More precisely, this study seeks to evaluate the contribution of each SDHS item to the latent construct as well as the location of the items on the latent construct continuum.

## Methods

### Participants and Procedure

The current study sampled 1326 participants (44.26% male) from Romania, age ranging from 18 to 64 years old ( $M_{age}=34.09$ ,  $SD=9.4$ ). Participants' level of education showed

that 31.9% participants completed high school, 38.9% completed a bachelor's degree, and 29.2% completed a master's degree. A series of campaigns carried out on WhatsApp, SNSs, and e-mail between January and April 2022, using non-probabilistic sample based on the snowball technique. Five hundred eighty participants enrolled in various postgraduate training courses at the University of Bucharest were invited by e-mail to get involved in this study and 514 accepted the invitation. They were asked to share the research link in their WhatsApp contact list and Facebook network. Thus, another 841 questionnaires were collected. Twenty-one questionnaires were removed because the participants answered the two filler questions incorrectly. The eligibility criteria were (i) to be at least 18 years old and (ii) to be native speakers of the Romanian language. Eight participants were ineligible because they were under 18 years old. Thus, in the end, 1326 questionnaires were collected. Participants expressed their informed consent to complete the battery of questionnaires before engaging in research. Various details related to the main aim of the study, voluntary and anonymous participation, and possibility to withdraw during the research without any explanation being necessary were mentioned in the informed consent. To conduct test–retest reliability analysis at 2-week interval, we renewed the invitation for all 514 participants who were initially invited by e-mail and 125 participants agreed to continue their involvement in this research.

## Ethics

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation. The study adhered to the tenets of the Helsinki Declaration 1975 as revised in 2000.

## Measures

**Sociodemographic variables** Sociodemographic data were collected on age, gender, and educational level.

*The Short Depression-Happiness Scale* (SDHS; Joseph et al., 2004) is a 6-item scale that assesses well-being on a continuum of depression and happiness. Three items are negatively worded (i.e., “I felt cheerless”), and the other three were positively worded (i.e., “I felt happy”), which evaluates the frequency of some mood states in the past week. Each response is scored on a 4-point Likert scale, ranging from 0 (never) to 3 (often). The total score of items assessing depression ranges from 0 to 9, and a higher score indicates a higher level of depression. The total score of items assessing happiness ranges from 0 to 9, and a higher score indicates a higher level of happiness. The reverse scoring of negatively worded items supposes that higher scores imply lower levels of WB. The overall score measuring WB is computed by adding the reverse scores on depression items to scores on happiness. It ranges between 0 and 18. A higher score indicates a higher level of well-being. In the original study (Joseph et al., 2004), Cronbach's alpha of the SHDH ranged from 0.77 to 0.92. The translation of the SDHS into Romanian was achieved following a forward–backward translation procedure, in accordance with the recommended methodological approach described by Sousa and Rojjanasrirat (2011).

*The Brief Resilience Scale* (BRS; Smith et al., 2008) is a 6-item scale which measures the ability to bounce back or recover from stress. Items include both positively and negatively worded sentences, such as “I usually come through difficult times with little trouble” and “It is hard for me to snap back when something bad happens.” Each item is scored from 1 (disagree) to 5 (strongly disagree). The total score of the BRS ranges from 6 to 30, and a higher score indicates a higher level of resilience. In the current study, the BRS proved very good psychometric properties in terms of (i) internal consistency ( $\omega=0.83$ , 95% CI [0.81, 0.85];  $CR=0.86$ ); (ii) convergent validity ( $AVE=0.51$ ); and (iii) construct validity ( $CFI=0.98$ ,  $TLI=0.98$ ,  $RMSEA=0.05$ , CI [0.03, 0.08],  $SRMSEA=0.05$ ,  $\lambda_s$  ranges from 0.58 to 0.89).

*The Perceived Stress Scale-10* (PSS-10; Cohen & Williamson, 1991) is a 10-item scale which assesses the global level of perceived stress. It includes both negatively worded questions about how often respondents felt a certain way over the past month (i.e., “In the last month, how often have you felt nervous and stressed”) and positively worded questions (i.e., “In the last month, how often have you felt that things were going your way”). Each item is scored on a 5-point Likert scale from 1 (never) to 5 (always). Scores for positively worded items must be reversed to obtain the total score. The total score of the PSS-10 ranges from 10 to 50, and a higher score indicates a higher level of perceived stress. In the current study, the PSS proved very good psychometric properties in terms of (i) internal consistency ( $\omega=0.80$ , 95% CI [0.78, 0.85];  $CR=0.91$ ); (ii) convergent validity ( $AVE=0.52$ ); and (iii) construct validity ( $CFI=0.96$ ,  $TLI=0.95$ ,  $RMSEA=0.06$ , CI [0.05, 0.08],  $SRMSEA=0.05$ ,  $\lambda_s$  range from 0.61 to 0.80).

## Data Analysis

Data analysis was conducted using Mplus 8.8 (2022), Stata 16.1 (2020), and SPSS 28 (2021). Descriptive statistics of the sample characteristics and analysis of the psychometric properties of the SDHS using CTT and IRT analysis were included. Firstly, the internal consistency of the SDHS based on a congeneric model and McDonald's  $\omega$  (1999) as well as test–retest reliability at 2-week interval were computed. Second, the construct validity by means of a confirmatory factor analysis (CFA) using the robust maximum likelihood (RML) estimation method was tested. Various goodness-of-fit indices were calculated to determine the acceptability of the model: the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMS), the comparative fit index (CFI), and the Tucker-Lewis index (TLI). CFI and TLI values  $>0.90$  are acceptable and  $>0.95$  are good (L. Hu & Bentler, 1999). The RMSEA and SRMR preferably should be less than or equal to 0.08. Third, the convergent validity was computed by calculating the average variance extracted (AVE) with a cut-off criterion of  $>0.50$  (Hair et al., 2018) and composite reliability (CR) with a cut-off criterion of  $>0.70$  (Chin et al., 2003). Fourthly, the concurrent validity based on Pearson correlations between WB, resilience, and perceived stress was tested. Fifthly, IRT analysis was conducted to measure the information given by each item of the scale. The assumptions of local independence were verified. Based on the Fisher information provided by each item's characteristics, that is, the slope/discrimination ( $\alpha$ ) and threshold/difficulty ( $\beta$ ) parameters were estimated whether the SDHS lays out more or less psychometric information about latent construct, i.e., WB. Items with  $\alpha > 1.7$  are considered providing very high information;  $1.35 < \alpha < 1.69$ , high information;  $0.65 < \alpha < 1.34$ , moderate information;  $0.35 < \alpha < 0.64$ , low information (Baker, 2001). In addition to the  $\alpha$  and  $\beta$  parameters, graphs of trace lines, curves of each

item, and all items together were considered, including boundary characteristic curves (BCCs), item information function (IIF), the test information function (TIF), and the test characteristic curve (TCC).

## Results

### Descriptive Statistics

The mean, *SD*, skewness and kurtosis indicators, and standard errors of all items of the SDHS, as well for other research variable, are shown in Table 1. The findings, namely kurtosis values  $< 2$  and kurtosis  $< 7$ , provide evidence for the univariate normality of the data for WB, resilience, and stress.

### Internal Reliability

The results confirmed that SDHS has very good internal consistency, taking into account the McDonald's  $\omega = 0.84$ ; 95% CI [0.82, 0.85]. As can be seen in Table 2, internal reliability could not be improved by removing none of the six items. To compare current results on internal reliability with those obtained in the validation study of the original scale (Joseph et al., 2004), Cronbach's  $\alpha$  coefficient was also computed. A very similar value was obtained: Cronbach's  $\alpha = 0.84$ ; 95% CI [0.83, 0.85]. All corrected item-rest correlations were positive and ranged between 0.50 and 0.71 (see Table 2). In addition, the inter-item correlation matrix proved that all values were higher than the cut-off criterion ( $> 0.30$ ) recommended in the literature (Cohen, 1992).

**Table 1** Descriptive statistics (mean, *SD*, skewness, and kurtosis of research variables)

	SDH1	SDH2	SDH3	SDH4	SDH5	SDH6	WB	Resilience	Stress
Mean	3.02	3.17	3.39	3.07	3.25	3.37	13.36	21.64	21.92
<i>SD</i>	0.86	0.73	0.76	0.77	0.73	0.89	3.48	4.68	5.23
Skewness	-0.64	-0.52	-1.16	-0.50	-0.62	-1.29	-0.69	-0.28	0.31
Kurtosis	-0.22	-0.24	0.89	-0.20	-0.32	0.65	0.04	-0.04	0.17

**Table 2** Individual item reliability statistics if item dropped and item-total correlation

If item dropped				
Item	McDonald's $\omega$	Cronbach's $\alpha$	Greatest lower bound ( <i>glb</i> )	Item-rest correlation
SDH1	0.815	0.817	0.859	0.641
SDH2	0.814	0.813	0.848	0.665
SDH3	0.814	0.816	0.856	0.645
SDH4	0.818	0.821	0.861	0.615
SDH5	0.807	0.804	0.833	0.714
SDH6	0.844	0.846	0.879	0.505

Test–retest reliability was computed based on a subsample of 125 participants who agreed to complete again the SDHS at 2-week interval. Intraclass correlation coefficient (ICC), i.e., 0.83, 95% CI [0.82, 0.84],  $p < 0.001$  (with no significant difference between overall scores obtained in the two assessment sessions,  $t = 0.43$ , *ns*), proved a good internal reliability, in accordance with the cut-off recommended by Koo and Li (2016).

### Construct Validity

The original single factor model proposed by Joseph et al. (2004) was tested to analyze the psychometric structure of the SDHS in the current study. CFA revealed that the unidimensional structure fitted very well the data. More specifically,  $\chi^2(9) = 52.43$  ( $p < 0.001$ ), CFI = 0.987, TLI = 0.978, RMSEA = 0.060; 90% CI [0.045, 0.076], RMSEA  $p$ -value = 0.33, and SRMR = 0.048. All factor loadings were above the cut-off ( $> 0.50$ ) recommended by Hair et al. (2018). They ranged from 0.53 to 0.81 (see all standardized factor loadings and  $R^2$  values for each item in Table 3).

### Convergent Validity

Convergent validity was highlighted by computing AVE and CR. Both coefficients were above the cutoff recommended in the literature (i.e., AVE  $> 0.50$ , according to Hair et al. (2018), respectively, CR  $> 0.70$ , as mentioned by Cho et al. (2016)). More precisely, AVE = 0.51 and CR = 0.76 were obtained.

### Concurrent Validity

Positive correlation between overall score of SDHS and resilience,  $r(1324) = 0.51$ ,  $p < 0.001$ , as well negative association with perceived stress,  $r(1324) = -0.59$ ,  $p < 0.001$ , provided evidence for the concurrent validity of the SDHS.

### Item Response Theory analysis

Conducting the IRT (GRM) analysis required testing two assumptions: unidimensionality and local independence. The CFA analysis supported unidimensionality. To test local

**Table 3** Standardized factor loadings and  $R^2$  for the SDHS items

Item	$\lambda$	$R^2$
SDH1	0.69	0.47
SDH2	0.76	0.57
SDH3	0.68	0.46
SDH4	0.70	0.48
SDH5	0.81	0.66
SDH6	0.53	0.28

independence, the inter-item residual correlation matrix was computed. The results ranged from 0.01 to 0.09. Since these values were significantly lower than the 1.30 cut-off recommended by Christensen et al. (2017), the second assumption of local independence was proved.

### Discrimination Parameters

According to the discrimination parameters, which ranged from 1.43 to 3.85 (as shown in Table 4), all the SDHS items can distinguish between contiguous latent trait levels. Applying Pelfrene et al.'s (2001) recommendation, item 6 (“I felt that life was meaningless”) had a high discrimination level ( $1.35 < \alpha < 1.69$ ), and the remaining items had an even higher level ( $\alpha > 1.70$ ). Arranging the discrimination parameters in descending order (5, 2, 4, 3, 1, 6), the first three items were positively worded selected to assess happiness. More specifically, item 5 (“I felt that life was enjoyable”) had the highest discrimination parameter ( $\alpha > 3$ ), followed by item 2 (“I felt happy”) and item 4 (“I felt pleased with the way I am”), both with  $\alpha$  ranging from 2 to 3. Item 6 (“I felt life was meaningless”) resulted in the lowest  $\alpha$  value.

### Difficulty Parameters

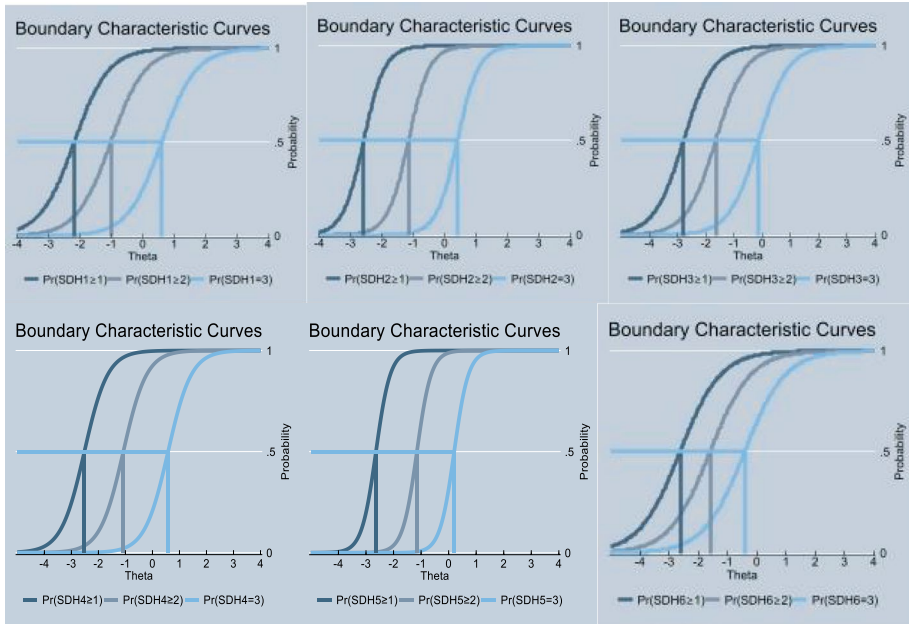
In terms of difficulty parameters, the  $\beta$  values ranged between  $-2.78$  and  $0.60$  (as seen in Table 4 and Fig. 1). According to the various thresholds recommended in the literature (Martínez et al., 2018), these difficulty values are indicative of both easy levels ( $\beta < -2$ ) and average levels ( $-2 < \beta < 2$ ). It should be mentioned that two of the negatively worded items (i.e., item 3: “I felt cheerless” and item 6: “I felt that life was meaningless”) have thresholds that are all negative ( $\beta_1 - \beta_3$ ), which means they are very easy to endorse. In addition, these values indicate that choosing “never” for these negatively worded items (which are reverse scored) reflects lower levels of WB than does selecting “often” for all the positively worded items.

The ascending ranking of the SDHS items according to their location on the difficulty scale (more specifically on the upper threshold, i.e.,  $\beta_3$ ) was 6, 3, 5, 2, and 4. The upper thresholds for the positively worded items ranged from 0.20 to 0.60 *SD* from the mean of the latent trait ( $\theta$ ). Thus, the upper threshold of the negatively worded items indicated lower levels of WB as compared to the positively worded items. Although the upper thresholds of the positively worded items were positive, they remained close to the mean of  $\theta$ . This suggests that they reflect the latent trait to a greater extent than the negatively worded items, but other factors impact the latent trait as well.

**Table 4** Discrimination and difficulty parameters for the SDHS items

Item	$\alpha$	$\beta_1$	$\beta_2$	$\beta_3$
SDH1	1.83	-2.18	-1.01	0.60
SDH2	2.99	-2.58	-1.15	0.40
SDH3	1.91	-2.78	-1.64	-0.13
SDH4	2.31	-2.51	-1.07	0.57
SDH5	3.85	-2.63	-1.15	0.20
SDH6	1.44	-2.62	-1.59	-0.41





**Fig. 1** The BCC graphs for the SDHS items

### The SDHS Items Precision to Measure WB or IIF

As the IIF graphs illustrate (see Fig. 2), the positively worded items contributed to the latent trait to a great extent because they peaked from  $+1.40 SD$  to  $+3.80 SD$  from the mean of  $\theta$ . In contrast, the negatively worded items were less reliable. More precisely, their amount of information peaked from  $+0.85 SD$  to  $+1.05 SD$  from the mean of  $\theta$ . In addition, two of them had an apex lower than 1 (i.e., item “1” and item “6”).

Another detail emphasized in the IIF graphs is that the curves were bi- or tri-modal. However, this pattern is not unusual when considering that each category of responses (from 0 to 3) has its own precision for measuring the latent trait, which may peak over a different trait range.

### The SDHS Precision as Whole to Measure WB or TIF

The TIF graph (as shown in Fig. 3) highlights relatively uniform information about individuals between  $-3.5 SD$  and  $+1 SD$  from the mean of the latent trait ( $\theta$ ). A decline beyond those points in both directions was noticed. In addition, there is a consistent amount of scale information at around  $-1.5$  and  $-1 SD$  from the mean, peaking at around  $-1.2 SD$  from the mean. The curve of standard error illustrates that the average score on the SDHS is under one-third of one standard error. In addition, the lowest standard error registered between  $-2.5SD$  and  $+0.5SD$  from the mean of  $\theta$  suggests highest reliability on this range. The reliability decreases for higher and very higher levels of WB, since standard error increases for  $+2SD$  and  $+3SD$  (as shown in the TIF graph).

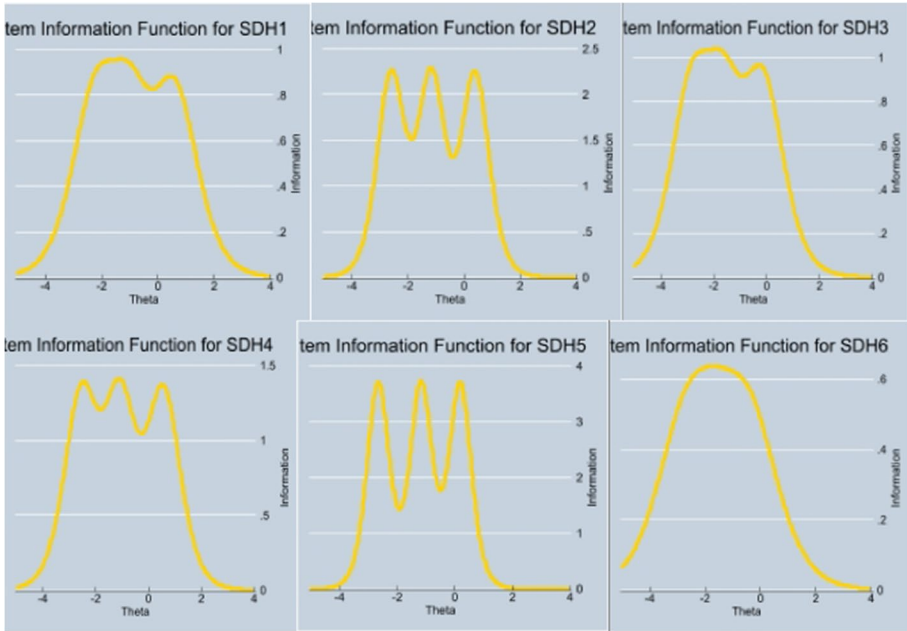
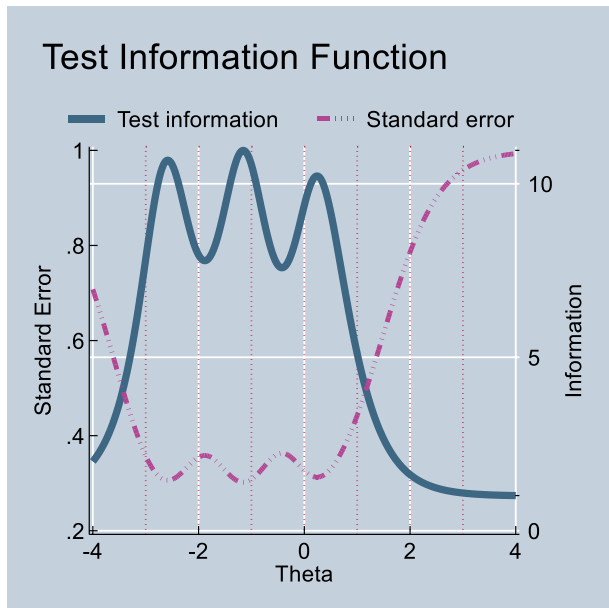


Fig. 2 The IIF graphs for the SDHS items

Fig. 3 The TIF graph or the SDHS precision (test information function, blue line; standard errors, magenta line)



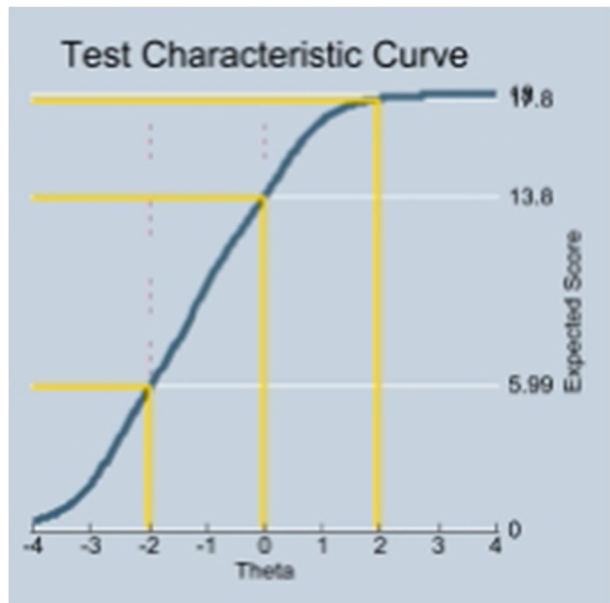
The maximum amount of information (high precision) scored approximately 11 of the latent trait estimates. Though the SDHS provides a precise estimation of scores for a relatively broad range of the latent trait continuum, it also includes more negative *SD* from the mean of  $\Theta$  than positive (as shown in Fig. 3). This result reinforces the aforementioned pattern related to the upper end of the scale.

### Test Characteristic Curves (TCC)

The TCC graph (see Fig. 4) shows that in the 3 to 18 range of total scores, the information curve increases steeply and monotonically. This steep curve highlights the probability that endorsing a correct response monotonically increases as the WB of the respondent increases. In addition, results from the TCC emphasize that the average score was approximately 14.

Using the 95% critical values from the standard normal distribution (1.96 and 1.96), this plot also tells us that we expect about 95% of participants to score from 6 to 18, which can be interpreted that most people have medium and high levels of WB. The remaining 5% represents respondents with extremely low WB levels (i.e., scores  $\leq 6$ ). Considering that the person location on 1.96 *SD* below the mean of the latent trait indicates an expected score of 6, then a preliminary (before clinical assessment confirmation) cut-off for clinically WB is overall score  $\leq 6$ . Descriptive statistics confirmed the results of the TCC. More specifically, 5.1% of participants obtained an overall score  $\leq 6$ .

Fig. 4 The TCC graph



## Discussion

The current study is the first to use IRT to examine the precision of SDHS measurements, overcoming the limits of the more commonly used CFA that considers only the raw data and ignores item-person relationships. The findings of this research, which are based on CTT, provide evidence for the sound psychometric quality of the Romanian version of the SDHS. Similar to previous studies, the SDHS in this study was unidimensional. Researchers proved high internal consistency of the SDHS using various techniques: (i) McDonald's  $\omega$ , (ii) CR, and (iii) bottom-up perspective of reliability shown by various IIF graphs computed during the IRT analysis. McDonald's  $\omega$  was calculated because the use of Cronbach's  $\alpha$  is not appropriate due to the impossibility of satisfying the tau-equivalence assumption. Additionally, to compare with the original study, Cronbach's  $\alpha$  coefficient was also calculated. Similar results were found. These findings proved a good level of inter-relationship among items.

CFA highlights that the unidimensional structure of the SDHS fits the data well, as all items were significantly loaded on the latent trait (WB). Although all factor loadings were above the recommended cutoff, item 6 ("I felt that life was meaningless") had both the  $\lambda$  value and the coefficient of determination ( $R^2$ ), the latter of which was lower than the other remaining items. It is not at all counterintuitive that this negative, reverse scored item contributes less to the overall WB score compared to the other items.

The positive association between the overall resilience score of the SDHS, as well as the negative correlation with perceived stress, confirms the concurrent validity of the SDHS. These results align with previous research that emphasized resilient people have higher levels of subjective happiness (Benada & Chowdhry, 2017; Brailovskaia et al., 2019; T. Hu et al., 2015; Pourkord et al., 2020) and WB inversely related to perceived stress (Cusinato et al., 2020; Li & Hasson, 2020; Ryan, 2022).

In the IRT analysis, six important patterns emerged. First, the SDHS items had a high discrimination capacity, meaning that they effectively reflect the relationship between the position of each participant on the latent trait continuum and the probability of endorsing the items of the scale. The obtained high and very high  $\alpha$  prove that the SDHS items can easily distinguish between contiguous latent trait levels.

Second, contrary to  $\alpha$  parameters,  $\beta$  parameters indicated very low and respectively low levels of SDHS items' difficulty. This finding suggests that low and respectively moderate WB levels are required to endorse the items of SDHS, since previous literature (De Ayala, 2008) mentions that high difficulty indicates high levels of the latent trait. Two negatively worded items (item 3 "I felt cheerless" and item 6 "I felt life was meaningless") were the easiest to endorse since they had the lowest difficulty parameters on all thresholds. Thus, choosing "never" for these items indicates lower levels of WB than does selecting the same response category for the remaining negative item (i.e., item 1), respectively than choosing "often" for positively worded items (i.e., items 2, 4, and 5). More specifically, the difference between the highest difficulty parameter (item 5) and the lowest difficulty parameters (items 6 and 3) suggests that the response to item 5 is more likely due to an individual's level of WB than the response to the other two items.

Furthermore, the upper thresholds ( $\beta_3$ ) of difficulty parameters for items 3 and 6 were negative, and all remaining items were located a little over a half  $SD$  above the mean of  $\theta$ . These results again indicate that supplementary factors contribute to an individual's high and very high WB levels more so than those captured by the SDHS items. This pattern seems plausible for two reasons: (i) the SDHS is built on the depression-happiness

continuum and (ii) researchers must consider that the latent trait can be explained not only by the presence of positive emotions and the absence of negative ones but also by other factors. This is why items of the SDHS are located on the latent trait continuum mostly below the mean or around one-half ( $+0.6SD$ ) above the mean of the latent trait, thus indicating the very low and moderate levels of WB. In addition, the Likert-type scale used in the SDHS may cause biases in favor of low WB rather than high WB. For example, for positively worded items, there is only one category higher than “sometimes,” namely, “often,” but two categories less than “sometimes,” that is, “rarely” and “never.” Considering this, respondents can have tendency to perceive the SDHS as a measure of lower WB than higher WB.

Third, regarding the IIF graph, the trace lines indicated the positive items contributed with more information to the latent trait, that is, WB, compared to the negatively worded items. Item “I felt that the life was enjoyable” had highest discrimination parameter.

Fourth, the reliability analysis at item level confirms the robustness of the SDHS’s internal construct and item validity, but mostly on the negative range of the latent trait continuum. The TIF graph, in which information function and standard error lines were plotted together, highlights that high values for information are associated with low values of SE, proving that the SDHS Romanian version is highly precise in measuring WB between  $-3.5SD$  and  $+1SD$  from the mean of the latent trait. Though the SDHS provides excellent precision according to TIF and SE values (in the aforementioned range), it does not provide enough information to reliably identify individuals who have higher levels of WB (namely, those who are located on  $+2SD$  and  $+3SD$  from the mean).

Fifth, the TCC graph highlights that the scores measured by the SDHS increase monotonically. Though the expected score was 14, the peak score of information provided by the scale was approximately 11, which suggests a moderate level of WB.

Sixth, IRT provided the opportunity to rank the SDHS items based on their psychometric properties. Slopes and item locations highlight that the positive items are psychometrically more robust than the negative items because they provide more reliable information on the latent trait. Additionally, scores of 6 and lower are indicative of clinically low WB, as can be seen on the  $x$ -axis in Fig. 4. Thus, the preliminary cut-off (before clinical assessment confirmation) is  $<7$ . Though in previous studies the proposed cut-off was  $<10$ , in the current research, scores less than 10 indicate a person location at  $1SD$  below the mean, which theoretically suggests that around 16% of participants were clinically depressed. Considering the statistics from the WHO (2022), however, this percentage seems to be implausible and unrealistic. According to the obtained cut-off in this study, 5.1% of participants had extremely low or clinically low WB levels.

## Strengths of the Current Study

The contributions of this study are six-fold. This is the first study conducted on the SDHS using the advanced psychometric approach of IRT. The advantage of this method is that IRT evaluates not only raw scores (as CTT does) but also score patterns and item properties or item-person relationships.

Second, this study calculated the reliability of the SDHS using both classical and modern techniques. More precisely, in terms of CTT, the McDonald’s  $\omega$  coefficient was computed. In the framework of modern approach, the IIF and TIF graphs were generated. While the CTT shows only the average reliability of all SDHS items, the IRT has the advantage of providing a bottom-up picture of reliability through the IIF graph. The graph

captures variations depending on the pattern of each item, providing an extensive picture of the internal consistency of the SDHS.

Third, CTT analysis evaluated three types of validity: construct, convergent, and concurrent. All proved good psychometric properties of the Romanian version of the SDHS. Construct validity replicated the unidimensionality found in the original study. Convergent validity indicated that the variance of the SDHS scores was captured mainly by the latent trait and less by measurement error.

Fourth, the current findings identified the SDHS's most difficult and discriminant items. The items that provided the highest levels of information for the latent trait (i.e., WB) were those that were positively worded. Less information was registered in the case of one negatively worded item (i.e., item 6 "I felt that life was meaningless").

Fifth, based on the TIF graph, a preliminary (before clinical assessment confirmation) cut-off ( $<7$ ) was identified for the clinically low WB. The cut-off highlighted in the current research is lower than that proposed in the original study (i.e.,  $<10$ ). Different cut-offs obtained are worthy of more investigation in future research. As aforementioned, these results must be considered preliminary since they were obtained using participants only from the general population.

Sixth, this is the first validation study of the SDHS conducted in Romanian culture. Previous validation studies of adapted versions of the SDHS were conducted only in Spanish (Martínez et al., 2018) and Arabic (Yildirim & Balahmar, 2020). Thus, this study expands the extant empirical body of research on the psychometric properties of the SDHS.

## Limitations and Future Studies Directions

Beyond these strengths, this study has some limitations. First, researchers obtained the data using a cross-sectional design; future longitudinal studies are needed to capture the dynamics of the latent trait (WB). Second, no data were collected from a clinical sample. Future studies using clinical samples are necessary to confirm the cut-off for the clinically low WB measured by the SDHS. Another suggestion for further research is to integrate the analysis of a disputable topic, namely the presence or absence of gender differences that might influence WB. Such investigations must start with measuring the invariance of the SDHS using CTT framework and the differential item functioning (DIF) based on IRT.

## Conclusion

The authors of the SDHS (Joseph et al., 2004) explicitly encouraged future studies to validate the SDHS on various populations. To date, no study has been conducted in Romania. In addition, all previous studies were performed using CTT framework. The current research aimed to fill this gap. Construct, convergent, and concurrent validities proved good psychometric properties of the SDHS. IRT revealed excellent scale precision mostly on a negative range from the mean ( $-3.5 SD, +1 SD$ ). All items had higher and respectively very high discrimination, indicating that each item can distinguish between contiguous levels of WB. Negatively worded items were the easiest to be endorsed, suggesting that positively worded items bring more information to the latent trait. The current findings demonstrate that the SDHS has excellent accuracy in measuring clinically low and moderate levels of WB. Thus, professionals can use the scale to measure WB dynamics in the context of therapeutic interventions, more precisely to highlight whether the symptomatic

scores for clinically low levels of WB have improved. In the case of the Warwick–Edinburgh mental well-being scale (WEMWBS), which is also a measure WB, Marmara et al. (2022) obtained a similar pattern of less reliability for high levels of WB. However, the WEMWBS also does not capture low levels of WB. Since the SDHS is built on the depression-happiness continuum, it can capture extremely low (i.e., clinically low) levels of WB, but neither the SDHS nor the WEMWBS can capture high WB levels. Possible future developments of the WB scales should consider adding items of high difficulty to increase measurement precision in the upper end of the latent trait continuum.

**Funding sources** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declarations

**Conflict of Interest** The author declares no competing interests.

**Ethics** All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation. The study adhered to the tenets of the Helsinki Declaration 1975 as revised in 2000. Approval was obtained from the ethics committee of University of Bucharest (Reg.No.CEC: 063/27.04.2020).

## References

- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.
- Benada, N., & Chowdhry, R. (2017). A correlational study of happiness, resilience and mindfulness among nursing student. *Indian Journal of Positive Psychology, 8*(2), 105–107.
- Brailovskaia, J., Rohmann, E., Bierhoff, H.-W., Schillack, H., & Margraf, J. (2019). The relationship between daily stress, social support and Facebook Addiction Disorder. *Psychiatry Research, 276*, 167–174.
- Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research, 14*(2), 189–217.
- Cho, S.-D., Heo, S.-E., & Moon, D. H. (2016). A convergence study on the hospital nurse's perception of patient safety culture and safety nursing activity. *Journal of the Korea Convergence Society, 7*(1), 125–136.
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q 3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*(3), 178–194.
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98–101.
- Cohen, S., & Williamson, G. M. (1991). Stress and infectious disease in humans. *Psychological Bulletin, 109*(1), 5.
- Cusinato, M., Iannattone, S., Spoto, A., Poli, M., Moretti, C., Gatta, M., & Miscioscia, M. (2020). Stress, resilience, and well-being in Italian children and their parents during the COVID-19 pandemic. *International Journal of Environmental Research and Public Health, 17*(22), 8297.
- De Ayala, R. (2008). Multivariate and mixture distribution Rasch models by von Davier & Carstensen: Review of Mathias von Davier & Claus H. Carstensen (Eds.) (2007). *Multivariate and mixture distribution Rasch models*. Springer. *Applied Measurement in Education, 21*(2), 83–186. <https://doi.org/10.1080/08957340801934692>
- Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist, 55*(1), 34–43. <https://doi.org/10.1037/0003-066X.55.1.34>
- 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., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, 97(2), 143–156.
- Ellison, C. W. (1983). Spiritual well-being: Conceptualization and measurement. *Journal of Psychology and Theology*, 11(4), 330–338.
- Ellison, C. W., & Smith, J. (1991). Toward an integrative measure of health and well-being. *Journal of Psychology and Theology*, 19(1), 35–45.
- Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.
- Fisher, J. W., Francis, L. J., & Johnson, P. (2000). Assessing spiritual health via four domains of spiritual wellbeing: The SH4DI. *Pastoral Psychology*, 49(2), 133–145.
- Gomez, R., & Fisher, J. W. (2003). Domains of spiritual well-being and development and validation of the Spiritual Well-Being Questionnaire. *Personality and Individual Differences*, 35(8), 1975–1991.
- Graham, C. (2017). *Happiness for all? In Happiness for All?* Princeton University Press.
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85(2), 348.
- Guidi, J., Rafanelli, C., Roncuzzi, R., Sirri, L., & Fava, G. A. (2013). Assessing psychological factors affecting medical conditions: Comparison between different proposals. *General Hospital Psychiatry*, 35(2), 141–146.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis* (8th ed.). Cengage.
- Holmqvist, F., Guan, N., Zhu, Z., Kowey, P. R., Allen, L. A., Fonarow, G. C., Hylek, E. M., Mahaffey, K. W., Freeman, J. V., & Chang, P. (2015). Impact of obstructive sleep apnea and continuous positive airway pressure therapy on outcomes in patients with atrial fibrillation—Results from the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF). *American Heart Journal*, 169(5), 647–654.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Hu, T., Zhang, D., & Wang, J. (2015). A meta-analysis of the trait resilience and mental health. *Personality and Individual Differences*, 76, 18–27. <https://doi.org/10.1016/j.paid.2014.11.039>
- Huta, V., & Waterman, A. S. (2014). Eudaimonia and its distinction from hedonia: Developing a classification and terminology for understanding conceptual and operational definitions. *Journal of Happiness Studies*, 15(6), 1425–1456.
- Joseph, S., Linley, P. A., Harwood, J., Lewis, C. A., & McCollam, P. (2004). Rapid assessment of well-being: The short depression-happiness scale (SDHS). *Psychology and Psychotherapy: Theory, Research and Practice*, 77(4), 463–478.
- Kahneman, D., Diener, E., & Schwarz, N. (Eds.). (1999). *Well-being: Foundations of hedonic psychology*. Russell Sage Foundation. <http://www.jstor.org/stable/10.7758/9781610443258>
- Keyes, C. L. M. (1998). Social well-being. *Social Psychology Quarterly*, 61(2), 121–140. <https://doi.org/10.2307/2787065>
- Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology*, 73(3), 539.
- Keyes, C. L. (2013). Promoting and protecting positive mental health: Early and often throughout the lifespan. In C. L. M. Keyes (Ed.), *Mental well-being* (pp. 3–28). Springer.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Li, Z.-S., & Hasson, F. (2020). Resilience, stress, and psychological well-being in nursing students: A systematic review. *Nurse Education Today*, 90, 104440. <https://doi.org/10.1016/j.nedt.2020.104440>
- Longo, Y., Coyne, I., Joseph, S., & Gustavsson, P. (2016). Support for a general factor of well-being. *Personality and Individual Differences*, 100, 68–72.
- Longo, Y., Coyne, I., & Joseph, S. (2017). The scales of general well-being (SGWB). *Personality and Individual Differences*, 109, 148–159.
- Longo, Y., Coyne, I., & Joseph, S. (2018). Development of the short version of the scales of general well-being: The 14-item SGWB. *Personality and Individual Differences*, 124, 31–34.
- Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology*, 71(3), 616.
- Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology*, 9(2), 111–131. <https://doi.org/10.1037/1089-2680.9.2.111>



- Marmara, J., Zarate, D., Vassallo, J., Patten, R., & Stavropoulos, V. (2022). Warwick Edinburgh Mental Well-Being Scale (WEMWBS): Measurement invariance across genders and item response theory examination. *BMC Psychology*, *10*(1), 1–17.
- Martínez, D. L., Muñoz, J. J. F., & Navarro-Pardo, E. (2018). Spanish version of the short depression-happiness scale (SDHS): A reliability and validity study in a sample of Spanish elderly people. *Acción Psicológica*, *15*(1), 17–24.
- Mayordomo-Rodríguez, T., Meléndez-Moral, J. C., Viguer-Segui, P., & Sales-Galán, A. (2015). Coping strategies as predictors of well-being in youth adult. *Social Indicators Research*, *122*(2), 479–489. <https://doi.org/10.1007/s11205-014-0689-4>
- McDonald, C. J. (1999). New tools for yield improvement in integrated circuit manufacturing: Can they be applied to reliability? *Microelectronics Reliability*, *39*(6–7), 731–739.
- Pavot, W., & Diener, E. (2008). The satisfaction with life scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology*, *3*(2), 137–152.
- Pelfrene, E., Vlerick, P., Mak, R. P., De Smet, P., Kornitzer, M., & De Backer, G. (2001). Scale reliability and validity of the Karasek' Job Demand-Control-Support' model in the Belstress study. *Work & Stress*, *15*(4), 297–313.
- Pourkord, M., Mirdrikvand, F., & Karami, A. (2020). Predicting resilience in students based on happiness, attachment style, and religious attitude. *Health, Spirituality and Medical Ethics*, *7*(2), 27–34.
- Ryan, M. (2022). *The happiness makeover: Overcome stress and negativity to become a hopeful, happy person*. Mango Media Inc.
- Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, *52*, 141.
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*(6), 1069.
- Ryff, C. D. (1995). Psychological well-being in adult life. *Current Directions in Psychological Science*, *4*(4), 99–104.
- Ryff, C. D. (2018). Well-being with soul: Science in pursuit of human potential. *Perspectives on Psychological Science*, *13*(2), 242–248.
- Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*(4), 719–727. <https://doi.org/10.1037/0022-3514.69.4.719>
- Ryff, C. D., & Singer, B. H. (2008). Know thyself and become what you are: A eudaimonic approach to psychological well-being. *Journal of Happiness Studies*, *9*(1), 13–39.
- Seligman, M. E. (2011). *Flourish: A visionary new understanding of happiness and well-being*. Free Press.
- Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine*, *15*(3), 194–200. <https://doi.org/10.1080/10705500802222972>
- Sousa, V. D., & Rojjanasirart, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, *17*(2), 268–274.
- Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology*, *53*(1), 80–93. <https://doi.org/10.1037/0022-0167.53.1.80>
- Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*, *57*(8), 1358–1362.
- Waterman, A. S., Schwartz, S. J., Zamboanga, B. L., Ravert, R. D., Williams, M. K., Bede Agocha, V., Yeong Kim, S., & Brent Donnellan, M. (2010). The Questionnaire for Eudaimonic Well-Being: Psychometric properties, demographic comparisons, and evidence of validity. *The Journal of Positive Psychology*, *5*(1), 41–61.
- World Health Organization. (2022). *Depression*. <https://www.who.int/news-room/fact-sheets/detail/depression>. Accessed 3 July 2022
- Yildirim, M., & Balahmar, N. B. (2020). Adaptation and validation of the Arabic version of the short depression-happiness scale. *Current Psychology*. <https://doi.org/10.1007/s12144-020-01214-0>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.