Personality and Individual Differences 207 (2023) 112151

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

The Hedonic and Arousal Affect Scale (HAAS): A brief adjective checklist to assess affect states

Pablo Roca [a,*], Carmelo Vázquez [b], Daniel Ondé [c]

[a] Faculty of Health Sciences, Universidad Villanueva, Spain
[b] School of Psychology, Complutense University of Madrid, Spain
[c] Department of Psychobiology & Behavioral Sciences Methods, Complutense University of Madrid, Spain

A R T I C L E I N F O

Keywords:
Affect
Mood
Emotion
Scale
Valence
Arousal

A B S T R A C T

The present study aims to develop a brief instrument to assess self-reported affective experiences, the Hedonic and Arousal Affect Scale (HAAS), rooted in the valence-arousal model of affect. Throughout four different studies, we found that: (1) the 12-item version showed a better goodness-of-fit than an initial longer version (Study 1; n = 259); (2) the two-dimensional model of affect (i.e., four-factor model: positive affect and high arousal, positive affect and low arousal, negative affect and high arousal, and negative affect and low arousal) showed the best fit to our data (Study 2; n = 525); (3) the HAAS showed evidence of concurrent validity with related measures in the field (Study 3; n = 480); and (4) it showed partial support for temporal invariance (Study 4; n = 262). The content and psychometric qualities of the HAAS make it a suitable brief scale to measure affect and could be particularly useful for repeated measures designs such as psychological interventions, experimental studies, or ecological momentary assessment studies.

Understanding how affect is structured in an individual's subjective experience is a fundamental challenge in psychology (Diener, 1999). However, the map of distinct affective experiences is varied, and still relatively unknown (Cowen & Keltner, 2017). Over the past decades, researchers turned to dimensional affect models in which affective experience can be categorized by a small number of fundamental dimensions. Valence and arousal have consistently emerged in most models of affect (Larsen & Diener, 1992; Russell, 1980; Watson et al., 1999). The valence dimension captures the hedonic tone of the experience (i.e., whether something is a source of pleasure or displeasure), whereas arousal is related to the amount of activation associated with a particular state. These two dimensions cut across all different affective experiences, and include transitory emotional states (that have a clear onset and offset), mood states (that may have some persistence over time), and enduring traits or dispositions (Boyle et al., 2015).

Self-report is still the most direct and feasible way to access subjective experiences (LeDoux & Hofmann, 2018), and in the last decades, dozens of assessment tools have been developed (Boyle et al., 2015; Dalal & Credé, 2013). Measurements differ regarding the number of items covered, the time frame (e.g., right now, last week, or in general), the type of measurement scale (e.g., intensity, frequency, or duration), the technology used to report affect (e.g., paper-and-pencil, computer-administered tests, or phone apps), or the nature of the assessed responses (e.g., moods that are tonic vs. emotions, that are 'phasic'). Instruments also differ regarding their underlying conceptual framework. Whereas some instruments just focus on two constructs (e.g., positive and negative affect), others use broader inclusion criteria (Heuchert & McNair, 2012; Lubin & Zuckerman, 1999). Furthermore, while it is relatively easy to develop a reliable and valid measure of positive and negative affect, the assessment of the arousal dimension is more problematic (Watson & Vaidya, 2012).

A good example of the intrinsic difficulties of building sound assessment tools to assess affective experiences is the case of the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988), one of the standard measures in affect literature. However, despite being a widely used instrument, the scale is not free of limitations (Pressman et al., 2019; Watson & Vaidya, 2012). An important shortcoming of the PANAS is that it is restricted to high-arousal positive (e.g., active, excited, enthusiastic) and negative states (e.g., nervous, scare, irritable), which exclude frequent emotions like calmness or sadness, although there are more modern versions of this instrument that include some low arousal items (Watson & Clark, 1999). Other questionnaires have been developed to assess affect, such as the Mood Adjective Checklist (MACL; Nowlis, 1965), the Multiple Affect Adjective Checklist-R (MAACL;

* Corresponding author: Faculty of Health Sciences, Universidad Villanueva, 28034 Madrid, Spain.
E-mail address: pablo.roca@villanueva.edu (P. Roca).

https://doi.org/10.1016/j.paid.2023.112151
Received 26 September 2022; Received in revised form 29 January 2023; Accepted 21 February 2023
Available online 9 March 2023

Zuckerman et al., 1983), the Profile of Mood States (POMS; McNair et al., 1971), the Differential Emotions Scale (DES; Blumberg & Izard, 1985), or the Scale of Positive and Negative Experiences (SPANE; Diener et al., 2009; Diener et al., 2010), among others. However, most of these measures suffer from certain limitations in terms of their validity (Watson et al., 2017). Most of these scales overrepresent negative affect (underrepresenting low-arousal positive affective experiences, such as 'calm' or 'relaxed'), include long and time-consuming sets of adjectives, and include items that are not often experienced (e.g., 'inspired') or have to do with motivations (e.g., 'determined') rather than emotions (Harmon-Jones et al., 2016). Furthermore, some categories of emotions are represented by similar adjectives (e.g., nervous, jittery, scared, afraid…), artificially increasing the consistency of the instruments (Bech, 1990; Boyle et al., 2015; Little et al., 1999).

1. Hedonic and Arousal Affect Scale (HAAS) proposal

Although the existing instruments have made enormous contributions to the research and applied fields, there is still room for improvement. The development of a brief scale assessing valence-arousal affective experiences in a balanced manner is important and necessary for several reasons. First, both well-being and emotional disorders are characterized by the presence of emotional states (e.g., boredom, sadness, calmness) that are not represented in scales using high-arousal descriptors (Diener et al., 2009). Second, although brief measures are not generally as psychometrically robust as the long versions (Credé et al., 2012), brief measures are especially useful when time resources are limited and when the construct is measured multiples times, as is the case of momentary ecological assessments (Shiffman et al., 2008) or mood induction procedures (Joseph et al., 2020), reducing participants' burden. Third, there are new psychological interventions (e.g., mindfulness or well-being therapies) in which expected low-arousal positive and negative emotional changes are not easily captured by standard instruments (Zeng et al., 2015), which, incidentally, may hide the true magnitude of changes (Moskowitz et al., 2021).

Based upon the extant literature and scales, the present study aimed to contribute to the field by developing a brief instrument to assess affective experiences that could assist researchers and practitioners in measuring affect within the traditional bidimensional affect space (i.e., valence and arousal). Efficiency and basic coverage of affective experiences is the main goal of this scale. Study 1 aimed to compare the internal structure of the two-factor model (i.e., positive-negative affect) and the four-factor model (i.e., positive-negative affect × high-low arousal) between the long (24 items) and the short version (12 items) of the HAAS. Study 2 evaluated the two-dimension structure in a new sample (i.e., cross-validation study) by using the 12-item brief scale to assess the valence-arousal model of affect. Study 3 analyzed the concurrent validity of the scale with other related constructs in the literature of affect. Study 4 tested the temporal invariance in a 2-month interval using weekly assessments.

2. Study 1: pilot study

This study aimed to compare the internal structure of the two-factor model (i.e., positive-negative affect) and the four-factor model (i.e., positive-negative affect × high-low arousal) between the long (24 items) and the short version (12 items) of the HAAS.

To develop the scale, a large pool of adjectives was selected from various sources. First, for high arousal items, we used the adjectives with higher factor loadings in the PANAS (López-Gómez et al., 2015). Also, based on the literature reviewed, we introduced further affective experiences that are not included in the PANAS, such as 'vigorous', 'lively', and 'energetic' for the high positive affect dimension (Cohen et al., 2003; Gregg & Shepherd, 2009; Matthews et al., 1990; Norcross et al., 1984), or 'anxious' and 'uptight' for the high negative affect dimension (Cohen et al., 2003; Fernández et al., 2002).

Second, for low arousal items, we included adjectives used in the circumplex model of affect (Posner et al., 2005; Watson & Tellegen, 1985), the four dimensions of dispositional mood (Huelsman et al., 1998), the Activation and Safe/Content Affect Scale (Gilbert et al., 2008), the Cohen's emotions styles (Cohen et al., 2003; Cohen et al., 2006), the Four-Dimension Mood Scale (Gregg & Shepherd, 2009), the Profile of Mood States (Fernández et al., 2002; Norcross et al., 1984), the Mood Adjective Checklist (Matthews et al., 1990), the Activation-Deactivation Adjective Check List (Thayer, 1986), the Modified Differential Emotions Scale (Fredrickson et al., 2003; Galanakis et al., 2016), the Multiple Affect Adjective Checklist (Lubin et al., 2001), the Mood Adjective Check List (Nowlis, 1965), and the Brief Mood Introspection Scale (Mayer & Gaschke, 2013). Furthermore, to assess the items' content validity, two authors with experience in experimental psychopathology (initials hidden for review) reviewed the fit between the adjectives and the low positive and negative affect constructs. First, each author selected a list of representative adjectives for low positive and negative affect constructs. In this selection process, priority was given to those adjectives that were already published in other validated affect scales and that showed good psychometric properties. Secondly, the authors exchanged their list of adjectives and classified them into low positive and negative affect, choosing those adjectives that appeared in both lists. The scale was developed and validated in Spanish with Spanish-speaking participants. Following Wild et al.'s (2005) recommendations, when no version of the item was available in a published instrument, an adjective translation-back translation procedure was carried out by the same authors, keeping those adjectives in which there was an agreement.

From the initial pool of items, 24 adjectives were finally selected (6 for each affect × arousal category) with consensus from the authors by following these criteria: (1) exclusion of synonym adjectives (e.g., scare-afraid-fearful) to avoid artificially inflated correlations within a factor (Bech, 1990; Boyle et al., 2015), which is a problem that has been detected in several well-known scales of depression (Fried, 2017); (2) exclusion of antonym adjectives (e.g., happy-unhappy/sad or tranquil-nervous) to avoid correlated error terms; (3) exclusion of extreme adjectives in the continuum (e.g., we used 'tired' instead of 'exhausted', or 'active' instead of 'euphoric') as they are less common in daily experiences; (4) exclusion of items not expressing affect but domains like cognitive and meta-cognitive processes of affect (e.g., positive-negative, good-bad, or pleasant-unpleasant), or personality constructs (e.g., 'shyness', 'optimistic', 'pessimistic', etc.); (5) inclusion of clinically-relevant adjectives (e.g., 'guilty', 'sad', 'calm') as well as adjectives relevant to experimental tasks (e.g., 'tired' or 'bored'), some of them less frequent affective experiences. As our goal was to build a scale as short as possible, with the requirement of having a minimum of three items per factor (Brown, 2015; Kline, 2015), based upon the initial six items per factor from the 24-item version, we selected the three items per factor that we agreed to represent frequent emotional experiences using colloquial expressions. The resulting 12-item version was then compared to the 24-item version using a confirmatory approach.

2.1. Method

2.1.1. Participants

A total of 259 participants filled out the initial set of 24 items. The administration format was paper and pencil. The mean age was 45.4 years ($SD = 10.2$), 72.6 % were women, 86.5 % had university studies, 68 % were married, and 11.2 % were unemployed. Participants were individuals from the general population interested in continuous education courses offered by a university-associated research center. They filled out basic demographic questions and the 24-affective adjectives scale.

2.1.2. Procedure

A random number generator software was used to create a list with 24 adjectives. Individuals were requested to respond on a 5-point Likert scale from 0 to 4 (not at all, slightly, moderately, very, extremely), indicating to what extent each adjective described how they felt (either right now -Studies 1 and 2- or in the last week -Studies 3 and 4). These formats of response have been widely used in previous studies in the field (Cohen et al., 2006; Watson & Tellegen, 1985). The study was conducted between January 2018 and November 2019, just before the COVID-19 pandemic.

2.1.3. Data analysis

To validate the arousal-valence model based on previous studies on the field (e.g., Kuppens et al., 2013; Larsen & Diener, 1992; Russell, 1980), we conducted a confirmatory strategy comparing the two-factor model (positive vs negative affect) vs the four-factor model (positive-negative affect × high-low arousal). We followed a confirmatory instead of an exploratory approach because we wanted to test a specific theoretical approach and not simply explore how items are grouped. Two Confirmatory Factor Analysis (CFA) models were run using LISREL 8.8 (Jöreskog & Sörbom, 2006), first for a two-factor model (positive vs. negative affect: Pos and Neg factors) and second for a four-factor model (positive-negative affect × high-low arousal: Pos High, Pos Low, Neg High, and Neg Low factors). The four-factor structure is the hypothesized model and the two-factor structure (a nested model) is used for comparison purposes.

Parameter estimation was performed using Robust Maximum Likelihood (RML), using a polychoric correlation matrix as input. For model fit evaluation, we used Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI-NNFI). We used a χ^2 difference test to compare the model fit of the two and four-factor solutions. Complementarily, given that the extraction of additional factors can produce overfitting, we have also used the comparative indices from the CFA with polychoric correlation matrix (LISREL analysis) as Expected Cross Validation Index (ECVI) and Akaike Information Criterion (AIC), which are more conservative approaches. See further details in Supplementary Materials.

2.2. Results

Table 1 shows descriptive statistics of the scores of the 24 items and the estimated factor loadings (λ_{ik}* standardized) for both the two-factor and the four-factor CFA models. With 24 items, the fit indices showed slightly better values for the four-factor CFA model (RMSEA = 0.097 (L_i = 0.090; L_s = 0.104), SRMR = 0.107, CFI = 0.960 and TLI = 0.955) than for the two-factor CFA model (RMSEA = 0.132 (L_i = 0.125; L_s = 0.139), SRMR = 0.123, CFI = 0.925 and TLI = 0.918). Moreover, the χ^2_{diff} test showed that the four-factor CFA model fit the data significantly better than the two-factor CFA model (1381.43–845.99 = 535.44, $p < .0005$). Comparative fit indices also showed better values (the lower the better) for the four-factor CFA model (ECVI = 3.70, AIC = 953.99) than for the two-factor CFA model (ECVI = 5.73, AIC = 1479.43). All λ_{ik}* were >0.50 in both CFA models, and the correlations between factors were: $\varphi_{21} = -0.673$ for the two-factor CFA model, and $\varphi_{21} = 0.720$, $\varphi_{31} = -0.437$, $\varphi_{41} = -0.642$, $\varphi_{32} = -0.776$, $\varphi_{42} = -0.596$ and $\varphi_{43} = 0.795$ for the four-factor CFA model.[1] The raw correlations between factors were: $r_{21} = -0.594$ (two-factor model), and $r_{21} = 0.623$, $r_{31} = -0.312$, $r_{41} = -0.496$, $r_{32} = -0.653$, $r_{42} = -0.461$ and $r_{43} = 0.633$ (four-factor model).

With the selected 12 items of the short form, the fit indices showed slightly better values for the four-factor CFA model (RMSEA = 0.085 (L_i = 0.069; L_s = 0.102), SRMR = 0.083, CFI = 0.970 and TLI = 0.959) than for the two-factor CFA model (RMSEA = 0.087 (L_i = 0.072; L_s = 0.103), SRMR = 0.085, CFI = 0.966 and TLI = 0.957). The χ^2_{diff} test showed that the four-factor CFA model fit the data significantly better than the two-factor CFA model (157.35–138.35 = 19, p = .002). Comparative fit indices also showed better values for the four-factor CFA model (ECVI = 0.77, AIC = 198.14) than for the two-factor CFA model (ECVI = 0.80, AIC = 207.35). The correlations between factors were similar to those obtained in the 24-item form (two-factor CFA model: $\varphi_{21} = -0.657$; four-factor CFA model: $\varphi_{21} = 0.899$, $\varphi_{31} = -0.518$, $\varphi_{41} = -0.567$, $\varphi_{32} = -0.670$, $\varphi_{42} = -0.617$ and $\varphi_{43} = 0.898$)[1]. Again, all λik* were >0.50 in both CFA models. The raw correlations between factors were: $r_{21} = -0.536$ (two-factor model), and $r_{21} = 0.618$, $r_{31} = -0.347$, $r_{41} = -0.432$, $r_{32} = -0.518$, $r_{42} = -0.392$ and $r_{43} = 0.548$ (four-factor model).

2.3. Study 1 conclusions

Study 1 results show a better degree of empirical consistency for the two-dimension structure (i.e., a better degree of approximation to observed data). For the 24-item form, although the overall degree of approximation obtained is good enough for a pilot study, some goodness-of-fit indices are not optimal (i.e., RMSEA around 0.1, SRMR >0.1). Better results were obtained from the 12-item form, where the four-factor CFA model fits significantly better than the two-factor model.

3. Study 2: cross-validation study

To evaluate the stability of the two-dimension structure of the 12 items short form of the scale analyzed in Study 1, we tested this CFA model in a new sample of participants.

3.1. Method

3.1.1. Participants

A total of 525 participants filled out the 12-item scale obtained in Study 1. The mean age was 45.4 years (*SD* = 9.8), 72 % were women, 84.8 % had university studies, 60.4 % were married, and 9 % were unemployed. The procedure for recruiting participants was the same as the one used in Study 1. Participants filled out basic demographic questions and the 12-affective adjectives scale.

3.1.2. Data analysis

The two and four-factor CFA models (12 items) were fit into the responses of the cross-validation sample (*N* = 525) following the same analyses as indicated in Study 1.

3.2. Results

Table 2 shows descriptive score statistics of the 12 items and λ_{ik}* for both the two-factor and the four-factor CFA models. All the items showed a relative symmetric response pattern, with a skewness index between ±2.

As in Study 1, all λ_{ik}* were >0.5 in both CFA models (with the exception of "bored" in the two-factor model: λ_{42}* = 0.483, and "tired" in the four-factor model: λ_{34}* = 0.499). The correlations between factors, for the two-factor CFA model, were: $\varphi_{21} = -0.722$, and for the four-factor CFA model: $\varphi_{21} = 0.798$, $\varphi_{31} = -0.383$, $\varphi_{41} = -0.676$, $\varphi_{32} = -0.642$. $\varphi_{42} = -0.795$ and $\varphi_{43} = 0.871$). The raw correlations between factors were: r21 = −0.567 (two-factor model), and $r_{21} = 0.629$, $r_{31} = -0.281$, $r_{41} = -0.421$, $r_{32} = -0.549$, $r_{42} = -0.545$ and $r_{43} = 0.550$ (four-factor model).

The fit indices showed slightly better values for the four-factor CFA model (RMSEA = 0.068 (L_i = 0.056; L_s = 0.079), SRMR = 0.063, CFI = 0.981 and TLI = 0.973) than for the two-factor CFA model (RMSEA = 0.092 (L_i = 0.082; L_s = 0.103), SRMR = 0.079, CFI = 0.960 and TLI =

[1] 1 – Pos High, 2 – Pos Low, 3 – Neg High, 4 – Neg Low.

Table 1
Descriptive statistics of the scores in the 24 items and estimated factor loadings (two and four-factor models) of the HAAS.

Items	Spanish translation[a]	Affect/arousal	Mean (SD)	Skewness	Kurtosis	Factor loadings (24 items)		Factor loadings (12 items)	
						2-Factor model	4-Factor model	2-Factor model	4-Factor model
Interested	Interesado por las cosas	Pos high	3.0 (0.89)	−0.92	0.94	0.625	0.642	0.630	0.646
Active	Activo/a	Pos high	2.7 (0.83)	−0.24	−0.22	0.562	0.654	0.562	0.636
Vigorous	Vigoroso/a	Pos high	2.0 (1.08)	−0.28	−0.49	0.602	0.696		
Lively	Animado/a	Pos high	2.8 (0.89)	−0.39	−0.38	0.849	0.851		
Energetic	Lleno/a de energía	Pos high	2.3 (1.05)	−0.23	−0.50	0.857	0.903		
Strong	Fuerte	Pos high	2.3 (0.97)	−0.32	−0.08	0.784	0.807	0.741	0.785
Calmed	Calmado/a	Pos low	2.4 (0.95)	−0.22	−0.22	0.569	0.799	0.537	0.565
Tranquil	Tranquilo/a	Pos low	2.5 (1.01)	−0.38	−0.14	0.538	0.764		
Peaceful	En paz	Pos low	2.4 (1.18)	−0.46	−0.61	0.725	0.777		
At ease	A gusto	Pos low	2.8 (0.90)	−0.43	−0.08	0.844	0.798	0.842	0.866
Content	Contento/a	Pos low	2.5 (0.92)	−0.38	0.08	0.850	0.685	0.862	0.844
Serene	Sereno/a	Pos low	2.6 (1.00)	−0.50	−0.11	0.717	0.885		
Jittery	Agitado/a.	Neg high	1.2 (1.08)	0.67	−0.39	0.562	0.716		
Anxious	Ansioso/a.	Neg high	0.9 (1.03)	0.97	0.25	0.684	0.777	0.664	0.690
Irritable	Irritable.	Neg high	0.5 (0.83)	2.18	4.84	0.804	0.801	0.834	0.861
Scared	Temeroso/a.	Neg high	0.5 (0.81)	1.55	2.31	0.730	0.756		
Uptight	Tenso/a.	Neg high	0.8 (0.95)	1.00	0.23	0.813	0.910		
Guilty	Culpable.	Neg high	0.4 (0.87)	2.21	4.31	0.807	0.759	0.808	0.797
Bored	Aburrido/a.	Neg low	0.3 (0.71)	2.52	6.24	0.523	0.586	0.521	0.584
Exhausted	Fatigado/a.	Neg low	0.9 (1.00)	0.83	−0.19	0.633	0.682		
Sad	Triste.	Neg low	0.4 (0.76)	2.05	4.58	0.840	0.844	0.833	0.887
Dull	Apagado/a.	Neg low	0.5 (0.84)	1.82	3.07	0.793	0.883		
Tired	Cansado/a.	Neg low	1.3 (1.05)	0.64	−0.34	0.576	0.628	0.510	0.502
Downcast	Abatido/a.	Neg low	0.4 (0.75)	2.49	6.53	0.879	0.865		

[a] To maximize participants' identification with the adjectives, the format of the HAAS items includes gender ending as needed.

Table 2
Descriptive statistics of the 12 items and estimated factor loadings (two and four-factor models) of the HAAS. We have ordered the table according to the factors corresponding to each model. The numbering of the items corresponds to the order in which they were applied, and the numbering in brackets corresponds to the numbering used in Study 1.

Items	Spanish translation	Affect/arousal	Mean (SD)	Skewness	Kurtosis	Factor loadings (12 items)	
						2-Factor model	4-Factor model
Interested	Interesado por las cosas	Pos high	2.6 (0.81)	−0.33	0.18	0.591	0.693
Active	Activo/a	Pos high	2.5 (0.81)	−0.33	0.40	0.591	0.749
Strong	Fuerte	Pos high	2.0 (0.90)	−0.37	0.01	0.644	0.742
Calm	Calmado/a	Pos low	2.2 (0.86)	−0.36	0.16	0.657	0.662
Placid	A gusto	Pos low	2.5 (0.83)	−0.57	0.35	0.929	0.939
Content	Contento/a	Pos low	2.4 (0.84)	−0.35	0.06	0.906	0.903
Anxious	Ansioso/a	Neg high	1.2 (0.98)	0.61	−0.10	0.678	0.740
Irritable	Irritable	Neg high	0.9 (0.86)	0.93	0.72	0.673	0.728
Guilty	Culpable	Neg high	0.6 (0.88)	1.54	1.96	0.680	0.680
Bored	Aburrido/a	Neg low	0.4 (0.72)	1.81	2.92	0.483	0.514
Sad	Triste	Neg low	0.8 (0.93)	1.14	0.87	0.734	0.741
Tired	Cansado/a	Neg low	1.6 (1.04)	0.29	−0.62	0.511	0.499

0.950). The χ^2_{diff} test showed that the four-factor CFA model fit significantly better with the data than the two-factor CFA model (288.28–162.78 = 125.50, $p < .0005$). Comparative fit indices showed better values for the four-factor CFA model (ECVI = 0.43, AIC = 222.78) than for the two-factor CFA model (ECVI = 0.65, AIC = 338.28). The omega reliability coefficients (McDonald, 2013) were as follows: Pos High = 0.73, Pos Low = 0.81, Neg High = 0.83, and Neg Low = 0.71.

3.3. Study 2 conclusions

The four-factor CFA model showed adequate goodness-of-fit values and a statistically significantly better fit than the two-factor CFA model. Thus, the favorable results for the hypothesized two-dimension model found in the pilot study have been successfully replicated in a new and larger cross-validation sample.

4. Study 3: concurrent validity

Correlation analyses were conducted to explore the concurrent validity of the 12-item short-form scale with relevant clinical and psychological measures, including stress, anxiety, depression, rumination, life satisfaction, and psychological well-being.

4.1. Method

4.1.1. Participants

A total of 480 participants from Study 2 (91.4 %) responded to a set of different instruments, in addition to the HAAS. The mean age was 45.4 years (*SD* = 9.7), 76.7 % were women, 90.6 % had university studies, 64.2 % were married, and 9.5 % were unemployed.

4.1.2. Measures

The set of instruments included: 1) *Depression Anxiety Stress Scale* (DASS-21; Lovibond & Lovibond, 1995) to measure stress, anxiety, and depression symptoms; 2) *Ruminative Response Style* (RRS; Nolen-Hoeksema & Morrow, 1991) to measure rumination with two factors (cognitive reflection and brooding); 3) *Pemberton Happiness Index* (PHI; Hervás & Vázquez, 2013) to measure psychological well-being; and 4) *Satisfaction with Life Scale* (SWLS, Diener et al., 1985) to measure global life satisfaction. See Supplementary Materials for a detailed description of the instruments.

4.1.3. Data analysis

First, a CFA using LISREL 8.8 (Jöreskog & Sörbom, 2006) was conducted based on the responses of the participants to each instrument, specifying the structural model recommended by previous research. Second, to assess concurrent validity, we calculated Pearson's correlations between each theoretical subscale (or total scale) of these instruments and each factor of the HAAS resulting from the two-dimension model. See further details in Supplementary Materials.

4.2. Results

Table 3 shows the fit indices of each measurement model of related constructs. We found an adequate fit for the DASS-21 and the SWLS models, an unacceptable fit for the RRS model, and an ambiguous fit for the PHI model (i.e., a poor result for RMSEA but a good result for SRMR, CFI, and TLI). However, we decided to maintain these models and their corresponding correlations with the four factors of HAAS, although highlighting this note of caution regarding the lack of fit of the RRS model. The patterns of correlations were congruent with the negative or positive content of the scales. Apart from the correlation between reflection and the Pos High factor, all the correlations were statistically significant ($p < .05$).

4.3. Study 3 conclusions

The four analyzed subscales have shown statistically significant relationships with other related variables (DASS-21, RRS, PHI, and SWLS). Also, the direction of the correlations between the hedonic contents of the HAAS items and the rest of the measures was completely congruent. However, in the case of PHI and RSS, these results should be considered with caution since the theoretical structure of the models underlying these variables has not obtained a sufficient degree of empirical consistency in our data. Furthermore, the pattern of correlation better discriminates the dimension of valence (i.e., positive vs negative) than the dimension of arousal (i.e., low vs high).

5. Study 4: temporal invariance

Temporal measurement invariance of the 12-item short-form was tested to validate the temporal stability of the two-dimension structure over a 2-month interval and 7 time-points measures (i.e., one per week).

5.1. Method

5.1.1. Participants

A total of 262 participants from Study 2 filled out the 12-item version of the HAAS as part of an inter-session assessment along with an 8-week psychological program to reduce stress (Kabat-Zinn, 2013). Participants completed the scale assessing how they had felt during the last week seven times, once per week, the day before each face-to-face session. The mean age was 45.8 ($SD = 9.8$), 71.8 % were women, 91.8 % had university studies, 63.9 % were married, and 9.1 % were unemployed. All 525 participants evaluated in Study 2 were offered to participate in the intervention, of which 455 participants decided to start the intervention (86.6 %). Of those, 262 participants (57.6 %) completed all weekly measures. No significant differences were found between completers and drop-out cases in gender ($\chi^2 = 3.85, p = .15$), age ($t_{(491)} = -0.91, p = .36$), education ($\chi^2 = 5.86, p = .44$), marital status ($\chi^2 = 8.08, p = .15$), and employment status ($\chi^2 = 5.36, p = .80$).

5.1.2. Data analysis

Temporal invariance of the four-factor CFA model (12 items) was tested by simultaneously analyzing seven repeated measures of the HAAS (once a week) (Svetina et al., 2020; Wu & Estabrook, 2016). We used the lavaan (Rosseel, 2012) and the semTools (Jorgensen et al., 2018) R-packages. We tested equivalence between nested invariance models (configural, metric and scale) by a nonsignificant probability level of the χ^2 difference test (χ^2_{diff}; p-value > .05), RMSEA, CFI and TLI. See further details in Supplementary Materials.

5.2. Results

For model 1 (configural invariance) the following fit indices were obtained: RMSEA = 0.100, CFI = 0.972 and TLI = 0.962. Model fit improved across the three successive models. RMSEA decreased from baseline to model 2 (RMSEA = 0.080) and to model 3 (RMSEA = 0.069). Furthermore, CFI and TLI increased ($\Delta CFI_{model1\text{-}model2} = 0.003$ and $\Delta CFI_{model\ 2\text{-}model\ 3} = 0.005$; $\Delta TLI_{model1\text{-}model\ 2} = 0.014$ and $\Delta TLI_{model\ 2\text{-}model\ 3} = 0.006$). χ^2_{diff} between model 1 and model 2 ($\chi^2_{(144)} = 130.21; p = .79$) and between model 2 and model 3 ($\chi^2_{(48)} = 43.7; p = .65$) showed no significant fit changes.

5.3. Study 4 conclusions

The temporal evaluation of the HAAS showed partial support for parameter invariance. CFI, TLI and the χ^2_{diff} tests indicated empirical consistency of temporal invariance, but RMSEA showed a lack of fit of the baseline model. However, as Svetina et al. (2020) pointed out, the implications of this lack of agreement between fit indices regarding temporal invariance evaluation are not clear. The partial lack of fit in our results may be due to the relatively small sample size and the relatively high dropout rate (almost half of the cross-validation sample), given the difficulty of obtaining seven weekly measurements in a longitudinal study. In addition, it is convenient to assess whether other temporary measures are more appropriate to reflect temporal stability of the HAAS (i.e., biweekly, monthly) since a weekly time-point series could imply some kind of progressive errors.

Table 3
CFA fit indices of measurement models of related constructs and Pearson's correlation between constructs and the four factors of the HAAS (affect positive (+) and negative (−), arousal high and low).

Measures	Model fit				Correlations			
	RMSEA ($L_i - L_s$)	SRMR	CFI/TLI	Scales	Factor 1 pos high	Factor 2 pos low	Factor 3 neg high	Factor 4 neg low
DASS-21	0.051 (0.044–0.058)	0.065	0.990/0.988	Stress	−0.369**	−0.443**	0.498**	0.505**
				Anxiety	−0.441**	−0.473**	0.431**	0.506**
				Depression	−0.312**	−0.367**	0.412**	0.424**
RRS	0.099 (0.085–0.112)	0.096	0.940/0.920	Brooding	−0.233**	−0.314**	0.400**	0.399**
				Reflection	−0.066	−0.122*	0.181**	0.176**
PHI[a]	0.114 (0.103–0.126)	0.044	0.968/0.960	Total	0.433**	0.440**	−0.373**	−0.447**
SWLS	0.081 (0.047–0.119)	0.027	0.992/0.985	Total	0.368**	0.380**	−0.324**	−0.405**

[a] ML estimation method (skewness of all the 11 items was <1.25).
* Correlation coefficient statistically significant at 0.05.
** Correlation coefficient statistically significant at 0.001.

6. Discussion

The present study aimed to contribute to the field by developing a brief instrument to assess self-reported affective experiences. The HAAS is a 12-item scale rooted in the broadly accepted bi-dimensional theoretical framework of affect (i.e., affect, arousal) (Kuppens et al., 2013). The HAAS includes affective adjectives that can describe, depending on the time frame used in the scale, both short-lived moods and more dispositional traits.

Our study showed that, compared to the pool of adjectives included in the initial list, the 12-item version of the HAAS showed better goodness-of-fit indices (Study 1). Following Clark and Watson (2019) and Kruyen et al. (2013) recommendations, we used a new and larger sample to evaluate the psychometric properties of the short form (Study 2). The HAAS showed that a classic two-dimension model of affect (i.e. positive-negative affect × high-low arousal) fit the data better than a one-dimension model (i.e., positive and negative affect). Regarding the concurrent validity of the study, we used instruments that cover a relatively wide and theoretically pertinent array of constructs (Study 3). In general, the HAAS showed a consistent pattern of correlations with other related variables. The magnitude of the correlations varied from 0.51 (the Stress and Anxiety subscales from the DASS-21 with negative-low arousal factor) to −0.31 (the Depression subscale from the DASS-21 and positive adjectives with high arousal). Also, interestingly, both the negative and positive adjectives of the scale did significantly correlate, with opposite signs, with this set of related variables. Thus, the scale is not redundant with any of the scales or subscales tapping constructs related to affective experiences and both the positive and negative contents show consistent patterns of correlations with the positive (e.g., life satisfaction) and negative constructs (e.g., rumination, depression) covered in the selected scales. It must be noted that the instruments used in the validation process (Study 3) cover different time frames. Nevertheless, the HAAS showed a similar magnitude of correlations to the instruments regardless of their respective time frames. Finally, the HAAS showed partial support for temporal invariance (Study 4), so it may be useful to assess changes in repeated measures designs, such as clinical trials, experimental procedures, and experience sampling studies. Importantly, the results of this study are restricted to the between-person level, and future studies should analyze the structure of HAAS at within-person level (Rush & Hofer, 2014), even acknowledging that item selection within-person studies face similar conceptual challenges to the ones found in between-person studies (e.g., Eisele et al., 2021; Horstmann & Ziegler, 2020).

The high correlations observed within each pair of affect factors (i.e., positive low vs high and negative low vs high), as well as the pattern of correlations found concerning the concurrent validity of the HAAS (showing that it has a better discrimination between positive and negative affect than between high and low arousal), suggest that, in its current version, the scale has some limitations to discriminate between high and low arousal dimensions. Therefore, we should be cautious when interpreting the high and low arousal factors of the HAAS. For instance, in clinical settings where high-arousal emotions need to be assessed (e.g., euphoria, panic, despair…), the HAAS may not be sensitive enough to detect extreme emotional states. In such cases, clinicians and researchers should choose alternative instruments that fit better those types of emotions.

Meanwhile, we are designing further studies to refine the arousal dimension of the HAAS. Finding a robust arousal dimension is always more challenging than finding a robust positive and negative affect dimension (Watson & Vaidya, 2012). Future studies should continue to explore the contribution of HAAS regarding other brief measures of affect (e.g., I-PANAS-SF; Thompson, 2007), as well as its potential use in other contexts, including studies using psychophysiological measures (e.g., EEG, EKG, EMG…), interventions aimed at reducing arousal (e.g., relaxation and mindfulness), and affect induction procedures or priming experimental tasks that can be used to induce high vs low arousal states (Joseph et al., 2020).

Nevertheless, this first version of the HAAS scale has several strengths. First, it includes both positive and negative affective experiences of low and high arousal following a bidimensional affect space. Second, the selection of adjectives was driven by consolidated theories in the field of affect (Larsen & Diener, 1992; Posner et al., 2005; Russell, 2003). We carefully selected words that represent affective experiences (Watson & Vaidya, 2012), excluding synonym and antonym adjectives (e.g., scare-afraid-fearful or happy-unhappy) and extreme adjectives as they are less common in daily experiences (e.g., tired instead of exhausted), also avoiding meta-cognitive processes of affect (e.g., positive-negative or good-bad), personality constructs (e.g., optimistic or pessimistic) and words reflecting motivations rather than affect (Harmon-Jones et al., 2016). Third, we tested the temporal invariance of the HAAS, which is a major gap in the standard procedures in the validation of affect scales (Galinha et al., 2013), even more so, considering the widespread use of affect scales in longitudinal and time-series designs. Future studies should examine the within-person structure in longitudinal settings, analyzing the emotional changes measured by the HAAS in different psychological interventions. Of note, the study was conducted between January 2018 and November 2019, just before the COVID-19 pandemic, which avoids the emotional impact that the pandemic has had on the general population's emotional responses (Santomauro et al., 2021).

Regarding limitations of the study, as in any other brief instrument, the HAAS cannot adequately reflect, with 12 items, the broad spectrum of emotions that individuals can identify in their experiences (Harmon-Jones et al., 2016). Furthermore, the HAAS factors showed not very high-reliability coefficients (from 0.71 to 0.83), which is a common limitation when developing brief instruments (Credé et al., 2012). Perhaps a more important concern has to do with the multidimensionality of affect scales (Clark & Watson, 2019). In our case, the correlations between factors were high in general, especially in the case of negative high and negative low arousal. Therefore, future studies with new samples should explore the potential hierarchical relationship (e.g., bifactor analysis) between affect and arousal of the adjectives included in the HAAS. Another limitation was that only concurrent validity with relevant clinical and psychological measures was tested, so future studies should explore convergent, discriminant, and incremental validity of the HAAS compared to other affect scales, such as the PANAS or Russell's affect grid. Our confirmatory approach guided by theory might incur a risk of confirmation bias if other alternatives are not considered. Furthermore, future studies might use lexical databases to control for the frequency of use of each adjective (e.g., Armstrong et al., 2012). A final challenge for the HAAS will be to confirm its utility to assess affect experiences in a variety of contexts (e.g., clinical studies in real-world settings), as well as measurement invariance across countries and languages (e.g., English version) given the often ignored potential cross-cultural differences when constructing instruments to assess affective experiences (Wang et al., 2019).

Open Science

Supplementary Materials and the data are available at https://github.com/nirakara-lab/HAAS.

Methodological disclosure

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. The research was approved by the university ethics committee prior to participant recruitment (Ref 2016/17-016).

Funding source

This work was supported by Ministry of Science PID2019-108711GB-

I00 to Carmelo Vázquez and Real Colegio Complutense at Harvard/Santander Bank grant (CT27/16-CT28/16) to Pablo Roca.

CRediT authorship contribution statement

PR and CV developed the study conception and design. Testing and data collection were performed by PR. Data analysis and interpretation were performed by DO. The first draft of the manuscript was written by PR and DO, under the supervision of CV. All the authors revised and drafted the manuscript. All the authors approved the final version of the manuscript for submission.

Declaration of competing interest

None.

Data availability

The data of this study are available at https://github.com/nirakara-lab/HAAS

Acknowledgments

The authors want to thank all participants for their generosity to voluntarily participate in the study. We also thank Ausiàs Cebolla and Nirakara Lab for their help and inspiration throughout the project.

References

Armstrong, B. C., Tokowicz, N., & Plaut, D. C. (2012). eDom: Norming software and relative meaning frequencies for 544 english homonyms. *Behavior Research Methods, 44*(4), 1015–1027. https://doi.org/10.3758/S13428-012-0199-8/FIGURES/3

Bech, P. (1990). Measurement of psychological distress and well-being. *Psychotherapy and Psychosomatics, 54*(2–3), 77–89. https://doi.org/10.1159/000288382

Blumberg, S. H., & Izard, C. E. (1985). Affective and cognitive characteristics of depression in 10- and 11-year-old children. *Journal of Personality and Social Psychology, 49*(1), 194–202. https://doi.org/10.1037/0022-3514.49.1.194

Boyle, G. J., Helmes, E., Matthews, G., & Izard, C. E. (2015). Measures of affect dimensions. In *Measures of personality and social psychological constructs* (pp. 190–224). Academic Press. https://doi.org/10.1016/B978-0-12-386915-9.00008-5.

Brown, T. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Publications.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Cohen, S., Alper, C. M., Doyle, W. J., Treanor, J. J., & Turner, R. B. (2006). Positive emotional style predicts resistance to illness after experimental exposure to rhinovirus or influenza a virus. *Psychosomatic Medicine, 68*(6), 809–815. https://doi.org/10.1097/01.psy.0000245867.92364.3c

Cohen, S., Doyle, W. J., Turner, R. B., Alper, C. M., & Skoner, D. P. (2003). Emotional style and susceptibility to the common cold. *Psychosomatic Medicine, 65*(4), 652–657. https://doi.org/10.1097/01.PSY.0000077508.57784.DA

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences of the United States of America, 114*(38), E7900–E7909. https://doi.org/10.1073/pnas.1702247114

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the big five personality traits. *Journal of Personality and Social Psychology, 102*(4), 874–888. https://doi.org/10.1037/a0027403

Dalal, R. S., & Credé, M. (2013). Job satisfaction and other job attitudes. In K. Geisinger, B. Bracken, J. Carlson, I. Hansen, N. Kuncel, S. Reise, & M. Rodriguez (Eds.), *APA handbooks in psychology®. APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 675–691). American Psychological Association. https://doi.org/10.1037/14047-037.

Diener, E. (1999). Introduction to the special section on the structure of emotion. *Journal of Personality and Social Psychology, 76*(5), 803–804. https://doi.org/10.1037/0022-3514.76.5.803

Diener, E., Derrick, A. E., Ae, W., Tov, W., Chu, A. E., Ae, K.-P., Choi, D.-W., Shigehiro, A. E., Ae, O., Biswas-Diener, R., Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., & Oishi, S. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97*(2), 143–156. https://doi.org/10.1007/s11205-009-9493-y

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., Sandvik, E., & Pavot, W. (2009). Happiness is the frequency, not the intensity, of positive versus negative affect. In *Assessing well-being* (pp. 213–231). Dordrecht: Springer. https://doi.org/10.1007/978-90-481-2354-4_10.

Eisele, G., Lafit, G., Vachon, H., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2021). Affective structure, measurement invariance, and reliability across different experience sampling protocols. *Journal of Research in Personality, 92*, Article 104094. https://doi.org/10.1016/J.JRP.2021.104094

Fernández, E., Fernández, C., & Pesqueira, G. (2002). Adaptación al español del cuestionario "Perfil de los estados de Ánimo" en una muestra de deportistas. *Psicothema*, 708–713.

Fredrickson, B., Tugade, M., Waugh, C., & Larkin, G. (2003). What good are positive emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on september 11th, 2001. *Journal of Personality and Social Psychology, 84*(2), 365.

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders, 208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Galanakis, M., Stalikas, A., Pezirkianidis, C., & Karakasidou, I. (2016). Reliability and validity of the modified differential emotions scale (mDES) in a greek sample. *Psychology, 07*(01), 101–113. https://doi.org/10.4236/psych.2016.71012

Galinha, I. C., Pereira, C. R., & Esteves, F. G. (2013). Confirmatory factor analysis and temporal invariance of the Positive and Negative Affect Schedule (PANAS). *Psicologia: Reflexao e Critica, 26*(4), 671–679. https://doi.org/10.1590/S0102-79722013000400007

Gilbert, P., McEwan, K., Mitra, R., Franks, L., Richter, A., & Rockliff, H. (2008). Feeling safe and content: A specific affect regulation system? Relationship to depression, anxiety, stress, and self-criticism. *Journal of Positive Psychology, 3*(3), 182–191. https://doi.org/10.1080/17439760801999461

Gregg, V. H., & Shepherd, A. J. (2009). Factor structure of scores on the state version of the four dimension mood scale. *Educational and Psychological Measurement, 69*(1), 146–156. https://doi.org/10.1177/0013164408322002

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. *PLoS One, 11*(8), Article e0159915. https://doi.org/10.1371/journal.pone.0159915

Hervás, G., & Vázquez, C. (2013). Construction and validation of a measure of integrative well-being in seven languages: The Pemberton happiness index. *Health and Quality of Life Outcomes, 11*(1), 66. https://doi.org/10.1186/1477-7525-11-66

Heuchert, J., & McNair, D. (2012). *Profile of mood states 2nd EditionTM. (POMS) [Database record]*. APA PsycTests.

Horstmann, K. T., & Ziegler, M. (2020). In *, 34(6)*. *Assessing personality states: What to consider when constructing personality state measures* (pp. 1037–1059). https://doi.org/10.1002/PER.2266. Doi:10.1002/per.2266.

Huelsman, T. J., Nemanick, R. C., & Munz, D. C. (1998). Scales to measure four dimensions of dispositional mood: Positive energy, tiredness, negative activation, and relaxation. *Educational and Psychological Measurement, 58*(5), 804–819. https://doi.org/10.1177/0013164498058005006

Jöreskog, K., & Sörbom, D. (2006). *LISREL 8.8 for Windows. [Computer software]*. IL: Scientific Software International.

Jorgensen, T., Pornprasertmanit, S., Schoemann, A., & Rosseel, Y. (2018). semTools: Useful tools for structural equation modeling. R package version 0.5-1. Retrieved from https://CRAN.R-project.org/package=semTools.

Joseph, D. L., Chan, M. Y., Heintzelman, S. J., Tay, L., Diener, E., & Scotney, V. S. (2020). The manipulation of affect: A meta-analysis of affect induction procedures. *Psychological Bulletin, 146*(4), 355–375. https://doi.org/10.1037/bul0000224

Kabat-Zinn, J. (2013). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness (Revised ed.)*. Hachette UK.

Kline, R. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing, 13*(3), 223–248. https://doi.org/10.1080/15305058.2012.703734

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin, 139*(4), 917–940. https://doi.org/10.1037/a0030811

Larsen, R., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Review of personality and social psychology* (Vol. 13, pp. 25–59). Sage Publications Inc.

LeDoux, J. E., & Hofmann, S. G. (2018). The subjective experience of emotion: A fearful view. *Current Opinion in Behavioral Sciences, 19*, 67–72. https://doi.org/10.1016/j.cobeha.2017.09.011

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4*(2), 192–211. https://doi.org/10.1037/1082-989X.4.2.192

López-Gómez, I., Hervás, G., & Vazquez, C. (2015). Adaptación de la "Escala de afecto positivo y negativo" (PANAS) en una muestra general española. *Psicología Conductual, 23*(3), 529–548.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the depression anxiety stress scales (DASS) with the Beck depression and anxiety inventories. *Behaviour Research and Therapy, 33*(3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-U

Lubin, B., Van Whitlock, R., Reddy, D., & Petren, S. (2001). A comparison of the short and long forms of the multiple affect adjective check list-revised (MAACL-R). *Journal of Clinical Psychology, 57*(3), 411–416. https://doi.org/10.1002/jclp.1023

Lubin, B., & Zuckerman, M. (1999). *Manual for the MAACL-R: Multiple affect adjective check list-revised*. EdITS/Educational and Industrial Testing Service.

Matthews, G., Jones, D. M., & Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST mood adjective checklist. *British Journal of Psychology, 81*(1), 17–42. https://doi.org/10.1111/j.2044-8295.1990.tb02343.x

Mayer, J., & Gaschke, Y. (2013). *Brief Mood Introspection Scale (BMIS). Measurement instrument database for the social science*.

McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press.

McNair, D., Lorr, M., & Droppleman, L. (1971). *Manual profile of mood states. Educational and industrial testing service*.

Moskowitz, J. T., Cheung, E. O., Freedman, M., Fernando, C., Zhang, M. W., Huffman, J. C., & Addington, E. L. (2021). Measuring positive emotion outcomes in positive psychology interventions: A literature review. *Emotion Review, 13*(1), 60–73. https://doi.org/10.1177/1754073920950811

Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 of Loma prieta earthquake. *Journal of Personality and Social Psychology, 61*(1), 115–121.

Norcross, J. C., Guadagnoli, E., & Prochaska, J. O. (1984). Factor structure of the profile of mood states (POMS): Two partial replications. *Journal of Clinical Psychology, 40*(5), 1270–1277. https://doi.org/10.1002/1097-4679(198409)40:5<1270::AID-JCLP2270400526>3.0.CO;2-7

Nowlis, V. (1965). *Factor analytic studies of the mood adjective checklist*. Rochester University NY department of psychology.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17*(3), 715–734. https://doi.org/10.1017/S0954579405050340

Pressman, S. D., Jenkins, B. N., & Moskowitz, J. T. (2019). Positive affect and health: What do we know and where next should we go? *Annual Review of Psychology, 70*(1), 627–650. https://doi.org/10.1146/annurev-psych-010418-102955

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Rush, J., & Hofer, S. M. (2014). Differences in within- and between-person factor structure of positive and negative affect: Analysis of two intensive measurement studies using multilevel structural equation modeling. *Psychological Assessment, 26*(2), 462–473. https://doi.org/10.1037/a0035666

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Santomauro, D. F., Herrera, A. M. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., … Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet, 398*(10312), 1700–1712. https://doi.org/10.1016/S0140-6736(21)02143-7

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using M plus and the lavaan/semtools packages. *Structural Equation Modeling, 27*(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776

Thayer, R. (1986). Activation-deactivation adjective check list: Current overview and structural analysis. *Psychological Reports, 58*(2), 607–614. https://doi.org/10.2466/pr0.1986.58.2.607

Thompson, E. R. (2007). In *, 38(2). Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS)* (pp. 227–242). https://doi.org/10.1177/0022022106297301. doi:10.1177/0022022106297301.

Wang, C. D. C., Hayslip, B., Sun, Q., & Zhu, W. (2019). Grandparents as the primary care providers for their grandchildren: A cross-cultural comparison of Chinese and US samples. *International Journal of Aging and Human Development, 89*(4), 331–355. https://doi.org/10.1177/0091415018824722

Watson, D., & Clark, L. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*. The University of Iowa. https://doi.org/10.17077/48vt-m4t2

Watson, D., Clark, L., & 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.

Watson, D., Stanton, K., & Clark, L. (2017). Self-report indicators of negative valence constructs within the research domain criteria (RDoC): A critical review. *Journal of Affective Disorders, 216*, 58–69. https://doi.org/10.1016/j.jad.2016.09.065

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*(2), 219–235. https://doi.org/10.1037/0033-2909.98.2.219

Watson, D., & Vaidya, J. (2012). Mood measurement: Current status and future directions. In *Handbook of psychology* (2nd ed., pp. 349–375). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118133880.hop202013.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*(5), 820–838. https://doi.org/10.1037/0022-3514.76.5.820

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health, 8*(2), 94–104. https://doi.org/10.1111/j.1524-4733.2005.04054.x

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika, 81*(4), 1014–1045. https://doi.org/10.1007/s11336-016-9506-0

Zeng, X., Chiu, C. P. K., Wang, R., Oei, T. P. S., & Leung, F. Y. K. (2015). The effect of loving-kindness meditation on positive emotions: A meta-analytic review. *Frontiers in Psychology, 6*, 1693. https://doi.org/10.3389/fpsyg.2015.01693

Zuckerman, M., Lubin, B., & Rinck, C. M. (1983). Construction of new scales for the multiple affect adjective check list. *Journal of Behavioral Assessment, 5*(2), 119–129. https://doi.org/10.1007/BF01321444