Journal of Contextual Behavioral Science 33 (2024) 100822

Contents lists available at ScienceDirect

Journal of Contextual Behavioral Science

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

Development of the Southampton Experiential Avoidance and Acceptance Scale (SEAAS)

Zoe McAndrews [a,b,*], Claire M. Hart [a], Lusia Stopa [a]

[a] *University of Southampton, University Road, Southampton, SO17 1BJ, UK*
[b] *Leeds and York Partnership NHS Foundation Trust, Trust Headquarters, St Mary's House, Main House, St Mary's Road, Potternewton, Leeds, LS7 3JX, UK*

ARTICLE INFO

Keywords:
Acceptance
Avoidance. psychometrics
Scale development
ACT
Assessment

ABSTRACT

Acceptance is an important construct in various psychological models seeking to describe psychological distress and emotion regulation. Existing measures either focus on broader constructs like psychological flexibility or lack proper validation. There are no established tools which measure acceptance as a general process, as defined within Acceptance and Commitment Therapy (ACT). The absence of valid and reliable measures impedes research on acceptance processes in clinical change. To address this, we developed the Southampton Experiential Avoidance and Acceptance Scale (SEAAS). Across four studies an item pool was generated, evaluated, and refined into an 18-item scale with a theoretically coherent two-factor structure. The SEASS demonstrates strong psychometric properties, including excellent internal consistency, convergent, concurrent, and discriminant validity, and test-retest reliability. This scale is a promising new tool for assessing acceptance and experiential avoidance.

1. Introduction

For many years, acceptance has been central to various domains of psychological inquiry, and consequently features in several psychological therapies. Third wave cognitive behavioural therapies (CBT) have grown in popularity in the past 20-years, with a greater emphasis on the person's relationship to thought and emotion, rather than on their content as is typical of earlier manifestations of CBT. Third-wave therapies like Acceptance and Commitment Therapy (ACT) propose new models and interventions for working with distress (Hayes & Hofmann, 2017). Acceptance is defined as the "adoption of an intentionally open, receptive, and flexible posture with respect to moment-to-moment experience", and is considered vital for psychological well-being (Hayes et al., 2013). The ACT model postulates that psychological suffering is a ubiquitous human experience arising from psychological inflexibility: "the rigid dominance of psychological reactions over chosen values and contingencies in guiding action" (Bond et al., 2011, p. 678). Experiential avoidance, a key process of psychological inflexibility, involves attempts to control or avoid unwanted private experiences. Thus, resulting in behaviour incongruent with one's values, preventing the pursuit of a meaningful life (Hayes et al., 1999).

Within the ACT model, each of the six identified pathological processes contributing to psychological inflexibility has a counterpart promoting flexibility; this model is sometimes described as the 'hexaflex'. Acceptance is the antithesis of experiential avoidance; an active therapeutic process involving moving towards offered experiences with a willingness to embrace them without attempting to alter their form, frequency, or intensity (Hayes et al., 2006, 2013). ACT does not advocate acceptance of all experiences, only those not readily changeable, or where non-acceptance would incur behavioural costs, such as increased psychological harm or decreased quality of life. It views acceptance as a means to live a valued and meaningful life, rather than an end goal (Blackledge & Hayes, 2001).

Clinical and research interest in acceptance, especially within ACT, has surged, with over 1025 randomised control trials (RCTs) published across various mental and physical health conditions (Hayes, 2023). A meta-analysis revealed ACT's efficacy across diverse populations and target issues, spanning a broad range of mental and physical health difficulties, such as depression, psychosis, and chronic pain (Gloster et al., 2020). However, the flaws with extant assessment tools present challenges with measuring specific processes within the ACT model and therefore it is not possible to determine mechanisms of change within interventions.

A systematic review identified significant methodological flaws in studies developing and evaluating existing acceptance measures (McAndrews et al., 2019). Acceptance is not always a well-defined

* Corresponding author. University of Southampton, University Road, Southampton, SO17 1BJ, UK.
E-mail address: zoe.mcandrews@nhs.net (Z. McAndrews).

https://doi.org/10.1016/j.jcbs.2024.100822
Received 8 April 2024; Received in revised form 15 August 2024; Accepted 15 August 2024
Available online 16 August 2024

construct, with multiple possible interpretations. Psychoeducation about the construct is a key component of ACT, as acceptance is often conflated with passivity, tolerance, and resignation, or with condoning and liking (Harris, 2009). Measuring acceptance through self-report outside a therapeutic context, or with individuals unfamiliar with the construct, presents challenges due to diverse semantic interpretations. For example, an extant measure of emotion regulation includes a scale named 'tolerance' which the authors equate to acceptance, that is, employing non-defensive responses to emotional experiences (Hofmann & Kashdan, 2010). Involving target populations in selecting scale items is crucial to ensure accurate representation of the intended construct and helps researchers understand the elements that matter to scale respondents (Terwee et al., 2007). Absence of non-expert input in developing acceptance scales has led to poor content validity, raising concerns about their ability to effectively measure the construct (McAndrews et al., 2019). This is especially pertinent for widely applicable concepts like acceptance, used by individuals across diverse populations and demographics.

Another challenge in measuring processes arises from the evolution of models and descriptions over time. Originally, acceptance and experiential avoidance were terms used to conceptualise the overarching concepts in the ACT model. Psychological flexibility and inflexibility are now preferred descriptors due to their broader applications (Bond et al., 2011). The widely used Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004), faced issues of poor reliability and unnecessary complexity, prompting a revision to address these shortcomings (AAQ-II; Bond et al., 2011). The AAQ-II was redefined as a measure of psychological inflexibility, departing from its original focus on experiential avoidance, aligning with the evolving ACT model. Possibly due to its predecessor, the AAQ-II is widely used for assessing acceptance, despite not being a dedicated measure of this process.

Despite initial factor analyses indicating concurrent validity, the AAQ-II has been criticised for its inability to differentiate outcome from process. Wolgast (2014) suggests the emphasis on distress within items in the AAQ-II may result in multiple interpretations for scale completers, potentially confounding process with outcome. Similarly, Gámez et al. (2011) contend the AAQ-II's focus on distress causing dysfunction may introduce item contamination, making it unclear whether responses reflect levels of experiential avoidance, distress, or both. Tyndall et al. (2019) found the AAQ-II more highly correlated with measures of distress than of experiential avoidance, raising further concerns about the scales ability to measure the higher order construct of experiential avoidance. A recent review recommended the AAQ-II should not be used as a measure of psychological flexibility due to significant shortcomings, and instead should be used as a measure of distress alone (Cherry et al., 2021).

Dimensionality of acceptance and non-acceptance is an important but overlooked consideration. During the development of the Multidimensional Experiential Avoidance Questionnaire (MEAQ), Gámez et al. (2011) discovered reverse-scored acceptance items did not directly load onto non-acceptance scales, suggesting these concepts may not be opposing ends of the same construct. Instead, they could be distinct but related constructs, similar to positive and negative affect. Recently, Thomas et al. (2022) examined the factor structure of the Multidimensional Psychological Flexibility Inventory (MPFI; Rolffs et al., 2018) and found a two-factor model (flexibility and inflexibility) was a better fit than a single factor. This aligns with research indicating that approach and avoidance regulatory systems are distinct across various domains (Gable et al., 2003). This has significant implications for existing acceptance scales, as their items predominantly focus on non-acceptance and/or inflexibility, which are then reverse scored. Consequently, there is a paucity of appropriate tools to appropriately measure constructs relating to psychological flexibility, such as acceptance (Cherry et al., 2021).

Beyond ACT-specific measures, various tools contain acceptance and non-acceptance scales, including mindfulness (e.g., Five Facet Mindfulness Questionnaire; Baer et al., 2008), and emotion regulation measures (e.g., Difficulties in Emotion Regulation Scale; Gratz & Roemer, 2004). However, there is insufficient evidence regarding whether these scales share adequate construct validity aligned with ACTs definition of acceptance. Furthermore, many of these are also non-acceptance scales that are reverse-scored, likely sharing the aforementioned dimensionality issues.

The absence of a valid and reliable acceptance measure hinders researchers from assessing the efficacy of acceptance-based interventions and understanding the potential role of acceptance as a mechanism of change within therapy. Many existing measures assess psychological inflexibility more broadly, like the AAQ-II and the MEAQ, or combine aspects of the hexaflex, such as the CompACT (Francis et al., 2016) which combines acceptance and defusion into one subscale. Given this significant limitation, we sought to develop a new acceptance measure aligned with the ACT model, suitable for use across diverse populations; the Southampton Experiential Avoidance and Acceptance Scale (SEAAS).

Four studies were conducted to develop, refine and examine the psychometric properties of the scale. Study 1 involved creating and refining a unique item pool. Study 2 focused on testing the factor structure and reducing item count to create a useable scale. Study 3 employed confirmatory methods to validate the factor structure. Study 4 assessed the measure's temporal stability. Approval for this research, across all four studies, was obtained from the appropriate institutional Research Ethics Committee.

2. Study 1

Study 1 sought to develop an initial item pool to form the basis of the new measure, containing items relating to acceptance and experiential avoidance.

2.1. Method

2.1.1. Preliminary item pool

2.1.1.1. Operational definition of acceptance. For the purposes of the new measure, acceptance was defined as the process of "actively contacting psychological experiences – directly, fully, and without needless defense" (Hayes et al., 1996, p. 1163). Acceptance was distinguished from passive tolerance, and defined as an active process of embracing experiences as they are now (Hayes, 2004). In other words, it is the process by which individuals actively make room for thoughts, feelings, and memories to be as they are, to come and go naturally, without trying to control or change them, even if they are unpleasant.

2.1.1.2. Item pool development. Based on theoretical understanding of acceptance, the first author reviewed existing measures of acceptance and psychological flexibility, creating a pool of 120-items deemed relevant to acceptance. To ensure broad applicability, items focussed on various internal events: thoughts, memories, feelings, and emotions. Recognising the challenge of accepting negative events, especially in therapy (Hayes et al., 2013), specific items were designed to inquire about unpleasant thoughts and emotions. Negatively worded items were generated and reverse-coded to mitigate response acquiescence (DeVellis, 2003).

Following scale development recommendations (DeVellis, 2003), the initial item pool underwent verification and feedback from ACT experts, as clinicians or researchers meeting specific criteria: i) five or more years using ACT as a primary model in clinical practice, and/or ii) provided teaching or training in ACT, and/or iii) had two or more ACT-related publications. Experts were recruited through local NHS Trusts, and through social media special interest sites for ACT practitioners and researchers. Experts rated each item on its relevance to the defined construct (as per the definition above; 1 = *does not describe this at all,* 10 = *completely describes this*), and provided feedback on clarity, conciseness, acceptability, and any other aspects of the item. Consensus criteria were defined a priori (Diamond et al., 2014), considering items with ratings of seven or higher as concurrent with the provided definition of acceptance and items with a congruence of 75% or greater between raters as strongly related to the construct. We also sought feedback on the clarity, accessibility, and understanding of items from an opportunity sample of individuals known to the first author from a non-clinical population. No incentives were offered for participation.

2.1.2. Participants

Five ACT experts reviewed the item pool. All were Clinical Psychologists, and all provided teaching or training in ACT. Experts had a mean age of 37 years ($SD = 9.03$, range 29–52), with 60% of the sample identifying as female. Five non-experts, who were unfamiliar with ACT, also reviewed the item pool. Non-experts had a mean age of 37 years ($SD = 13.76$, range 23–54), with 40% of the sample identifying as female.

2.2. Results

Fifty-two items were considered highly congruent by 75% or more of the experts and were retained for further consideration, resulting in a fifty-two-item scale. Qualitative feedback from experts prompted a refinement in item wording by distinguishing between emotions and physical sensations instead of using the term "feelings". Non-experts had no recommendations or concerns regarding item clarity or understanding.

3. Study 2

Study 2 aimed to further refine the item pool, reducing the fifty-two items into a theoretically consistent, parsimonious, and practical scale for versatile use across clinical and research settings. We utilised exploratory factor analysis (EFA) to identify factor structure, anticipating acceptance and non-acceptance items would load onto different factors, thus aligning with the distinct nature of approach and avoidance regulatory systems at a higher order level (Gable et al., 2003), and with contemporary psychological flexibility measures such as the MPFI (Rolffs et al., 2018).

3.1. Method

3.1.1. Participants

An online sample of 268 (215 female, 50 male, 2 gender fluid, 1 unidentified) participants was recruited for this study. The mean age of participants was 27.81 years ($SD = 11.86$, range 18–73). Participants were from the United Kingdom (UK; $n = 225$, 83.7%), United States of America (USA; $n = 24$, 8.9%), Canada ($n = 4$, 1.5%), Asia ($n = 4$, 1.5%), Ireland ($n = 3$, 1.1%), Australia ($n = 3$, 1.1%), Europe ($n = 2$, .7%), and not stated ($n = 2$, .7%).

3.1.2. Materials and procedure

Participants were recruited through the university student research portal, social media and online research participation interest groups (e.g., http://www.onlinepsychresearch.co.uk/) to take part in a study on developing a new questionnaire. Participants recruited at the university ($n = 109$) received research credits for their participation. Those recruited via online platforms ($n = 159$) were not offered compensation. After providing written informed consent, which detailed the right to withdraw, and demographic information, participants rated each of the 52-items using a 7-point scale (1 = *never true*, 7 = *always true*). Participants took an average of 8 min, 27 s to complete the study.

3.2. Results

A Missing Completely at Random test (MCAR; Little & Rubin, 2002) was shown to be significant, $x^2 = 767.609$, $df = 657$, $p < 001$. Review of the data evidenced fewer than 5% of cases contained missing values, and as no pattern was evident, it was deemed appropriate to continue with the analyses using listwise deletion (Little & Rubin, 2014). All items in the correlation matrix demonstrated at least one correlation greater than .30, fulfilling the prerequisite for conducting factor analysis (Tabachnick & Fidell, 2014). Observation of communalities and number of factors in line with proposals made by MacCallum, Widaman, Zhang, & Hong (1999) indicated our sample would have sufficient power with a ratio of 5 participants per variable. Principle axis factoring with Oblique rotation loadings was used to explore the factor structure of the item pool, with loadings below .3 suppressed. Scree plots and tables indicated potential for a two- or four-factor structure. Upon examination of loadings and items, the two-factor structure, labelled "openness and willingness" (OW) and "avoidance and control" (AC), appeared most coherent with the ACT model, explaining 47.06% of the variance with 48-items loading across the two factors. Cronbach's alpha for these subscales were α = .95 and α = .92, respectively. Negatively worded items were excluded due to poor factor loadings. Items containing the phrase "I do not try to change these feelings" cross loaded onto both factors but did not load strongly onto either.

Given repetition and redundancy remained in the item pool, an iterative process by which items were removed based on theory, repetition, and lowest loadings, with reassessment of factor structure and internal consistency, was undertaken. For example, "I allow myself to experience unpleasant emotions, no matter how long they last" was removed as this is similar to item 9 "I allow myself to experience emotions for as long as they remain" and had a poorer factor loading. This process, in conjunction with recommendations for optimum scale length (8–10 items per subscale; Netemeyer et al., 2003), culminated in an 18-item questionnaire, which remained consistent with the two identified factors (factor loadings are summarised in Table 1). Internal consistency was α = .92 for OW and α = .90 for AC.

4. Study 3

In scale development, it is common practice to follow up EFA with a confirmatory factor analysis (CFA) model to determine whether the hypothesised factor structure is retained. Expected relationships are defined a priori in a model and statistically tested (Byrne, 2010). Three analytic strategies were employed to develop and confirming a theoretically consistent model (Jöreskog, 1993). Firstly, a purely confirmatory approach tested the fit of the 18-item two-factor model proposed in Study 2. Secondly, a model-generating strategy was used with the aim of preserving the two-factor structure, whilst improving model fit through refining the model. Thirdly, an alternative model strategy was used to test whether the two-factor solution was a better fit than a single-factor structure.

Next, as we sought to develop a measure which could be used with diverse populations, we assessed whether the 18-items of the SEAAS functioned in a similar way across adults of different age groups. Participants were grouped into three age groups: young (18–30 years), middle (31–64), and older adults (65 and over). To ensure sufficient power to proceed with CFA modelling and subsequent invariance testing, a sample of over three hundred was required for the present study (Tabachnick & Fidell, 2014). In addition to confirming the hypothesised factor structure of the new scale, in Study 3 we also sought to test the scales convergent, discriminant and incremental validity. Convergent validity was assessed by exploring the relationship of the SEAAS to the wider ACT model and psychological flexibility using the AAQ-II (Bond et al., 2011). The AAQ-II was chosen as this is a well-established ACT measure of psychological inflexibility, often used to measure acceptance. Higher AAQ-II scores (indicating greater

Table 1
Factor loadings for Exploratory Factor Analysis with Direct Oblimin Rotation for Refined Item Pool.

Item	Openness and Willingness	Avoidance and control
Item 9 - I allow myself to experience emotions for as long as they remain	.794	
Item 7 - I allow myself to experience unpleasant thoughts even if they get worse	.790	
Item 14 - I allow my thoughts to be there for as long as they remain	.783	
Item 3 - I am willing to fully experience all thoughts that come to into my mind	.770	
Item 4 - I allow myself to experience unpleasant emotions even if they get worse	.760	
Item 10 - I am willing to fully experience all emotions that arise	.741	
Item 15 - I allow my thoughts to come and go freely	.706	
Item 18 - I welcome all emotions	.706	
Item 2 - I allow myself to experience all of my emotions	.623	
Item 16 - I work hard to keep out unpleasant emotions		.867
Item 1 - I try to put problems out of my mind		.758
Item 13 - I often do things to avoid unpleasant emotions		.749
Item 17 - I tell myself I should not feel certain things		.731
Item 6 - I try to distract myself when I feel unpleasant emotions		.713
Item 8 - I work hard to keep out unpleasant thoughts		.688
Item 12 - It is important to keep emotions under control		.668
Item 11 - When I experience unpleasant emotions, I try to change or get rid of them		.660
Item 5 - I tell myself that I should not have certain thoughts		.547

inflexibility) were expected to correlate positively with the AC subscale, but negatively with the OW subscale. Small-to-moderate correlations were expected as the AAQ-II measures wider psychological inflexibility which includes other ACT processes in addition to acceptance/experiential avoidance.

The relationship between the SEAAS and two well-established emotion regulation strategies (emotional suppression and cognitive reappraisal) was assessed using the Emotion Regulation Questionnaire (ERQ; Gross & John, 2003). Given emotional suppression is a form of experiential avoidance, this was expected to correlate positively with the AC subscale. However, as the ERQ only explores emotional suppression, a small correlation was expected. A small negative correlation of suppression with OW was expected. The reappraisal subscale was not expected to correlate with either SEAAS subscale.

Concurrent validity was assessed using the Positive and Negative Affect Scale (PANAS; Watson et al., 1988) and Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS; Stewart-Brown et al., 2009) as acceptance and avoidance are considered important determinants of psychological affect and wellbeing. OW was expected to correlate positively with positive affect (PA) and wellbeing, and negatively with negative affect (NA). Conversely, the AC subscale was expected to correlate positively with NA, and negatively with PA and wellbeing. Small-to-moderate correlations were expected for all variables.

The impression management subscale of the Balanced Inventory of Desired Responding – Short form (BIDR-16; Hart et al., 2015) was used to assess for possible desirable responding in this self-report measure. As impression management is considered a theoretically distinct concept, no association was anticipated with either subscale. However, significant correlations of impression management with any of the key variables would be grounds for including it as a covariate in subsequent analyses.

Incremental validity was explored using methods proposed by Haynes and Lench (2003). The SEAAS subscales were added to the second step of a series of regression models, after the AAQ-II, to predict PA, NA and wellbeing. It was predicted that the SEAAS would add to the variance explained by the AAQ-II in predicting outcomes relating to affect and wellbeing, thus demonstrating incremental validity.

4.1. Method

4.1.1. Participants

A sample of 513 community dwelling adults from a non-clinical population (407 female, 99 male, 4 not stated, 1 queer and 2 ‘other’) with a mean age of 43.61 years (SD = 24.55, range 18–91) were recruited through a variety of methods, including a university research portal, social media, online research groups, and a volunteer database. Participants were from the UK (n = 482, 95.4%), USA (n = 10, 1.9%), Ireland (n = 5, 1.0%), Australia (n = 3, .5%), Europe (n = 2, .4%), Canada (n = 1, .2%), or not stated (n = 3, .6%). Almost half (n = 228, 44.4%) of the sample identified having had previously experienced mental health difficulties. Of these, 148 (65.3%) had received treatment for these difficulties, with 98 (43.0%) having received some psychological or talking therapy. Fifty-seven participants (25.0%) reported previous CBT, 11 (4.8%) had received mindfulness, 3 (1.3%) had attended DBT, and no participants reported having ACT.

4.1.2. Materials and procedure

For Study 3, we employed analysis of data collected by the first author for another research project. Ethical approval was granted for secondary data analysis. Participants completed the 18-item SEAAS questionnaire and several scales to test convergent and discriminant validity, as well as social desirability biases in responses.

The Acceptance and Action Questionnaire II (AAQ-II; Bond et al., 2011) is a 7-item questionnaire measuring psychological flexibility and inflexibility. Items are rated using a 7-point scale from 1 = *never true* to 7 = *always true*. Scores range from 7 to 49, with lower scores indicating greater psychological flexibility, and higher scores indicating greater inflexibility. The AAQ-II demonstrates good reliability (ranging from α = .78 to .88), and test-retest reliability (r = .79; Bond et al., 2011). Internal consistency for the current study was comparable with published values (α = .85).

The Positive and Negative Affect Scale (PANAS; Watson et al., 1988) is a 20-item measure, comprising two 10-item mood scales measuring positive and negative affect. Respondents rate the extent to which they have experienced particular emotions within the past two-weeks, on a 5-point scale (1 = *very slightly or not at all* to 5 = *very much*). The PANAS has good internal consistency when used with the general population (α = .89 and α = .85 for PA and NA respectively; Crawford & Henry, 2004), as well as good test-retest reliability (r= .68 and .71, respectively; Watson et al., 1988). Internal consistency for the current study was comparable to published examples for both subscales; PA α = .90 and NA α = .88.

The Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS; Stewart-Brown & Janmohamed, 2008) is a 7-item self-report measure of psychological well-being and functioning. Items pertain to subjective positive affect and psychological functioning over the past two-weeks and are rated using a 5-point scale ranging from 1 = *none of the time* to 5 = *all of the time*. Possible scores range from 7 to 35, with higher scores indicating greater levels of psychological well-being. The scale has good test-retest reliability (r = .83; Stewart-Brown et al., 2009) and reliability for the present study was α = .88.

The Impression Management subscale of the *Balanced Inventory of*

Desired Responding Short Form (BIDR-16; Hart et al., 2015) is an 8-item measure used to control for socially desirable responding when self-report measures are employed in studies. Participants respond to items on a 7-point scale (from 1 = *strongly disagree*, 7 = *strongly agree*). The scale has good internal consistency (ranging from $\alpha = .63$ to $\alpha = .82$; Hart et al., 2015). The scale had acceptable reliability in this study; $\alpha = .72$.

4.1.3. Analytic strategy

Goodness of fit for each CFA model was assessed using Maximum Likelihood chi-square (χ^2), comparative fit index (CFI), and root square mean error of approximation (RMSEA). A good-fitting model is indicated by a non-significant chi-square test, CFI indices of.95 and RMSEA indices below .60 (Hu and Bentler, 1998). The model chi-square test is often criticised for being overly conservative and sensitive to sample size (Bentler & Bonnet, 1980; Shi et al., 2019), thus a non-significant chi-square test was unlikely in the present study.

4.2. Results

Data exploration indicated less than 1% missing data, which was missing at random; MCAR test (Little & Rubin, 2002) was non-significant ($x^2 = 1188.58$, df = 1170, $p = .346$). Full Information Maximum Likelihood (FIML) estimations in AMOS were employed to handle missing data, offering advantages over deletion methods by reducing bias, preserving power, while minimally impacting model fit (Byrne, 2010). FIML imputations were consistently used in subsequent analyses.

4.2.1. A priori measurement model

The 18-item, two-factor solution was subject to CFA using AMOS 25 (Arbuckle, 2017). Goodness of fit indicators suggested a poor fit to the data; $\chi^2(134, N = 513) = 822.18, p < .001$ CFI = .83, RMSEA = .10 (90% confidence interval [CI] = [.094, .107]. A model generating strategy was adopted to refine the SEAAS. A series of two-factor models were tested.

4.2.2. Model generating

All indicators in the model loaded significantly in the expected direction onto the specified latent factors, with no cross loadings. As expected, the two latent factors demonstrated a small, negative covariance, confirming their distinct nature (−.35). However, modification indices indicated misspecified error covariances, likely due to content overlap in items (Byrne, 2010). Consequently, the model was respecified with covariances between identified error terms added where theoretically relevant. The model was tested for comparable fit after each respecification in line with best practice (these stages are summarised in Table 2). While a non-significant chi-square difference was not obtained, no further covariances were added beyond model seven as there was no theoretical justifications for respecifying the model. The final model is depicted in Fig. 1.

4.2.3. Alternative model: assessing dimensionality

An 18-item, one-factor model was tested, which demonstrated a poorer fit to the data than the initial two-factor model; $\chi^2(135, N = 513) = 1826.82, p < .001$, CFI = .59, RMSEA = .156 (90% confidence interval [CI] = [.150, .163]. Chi-square difference between the models was 1004.64, with 1*df* difference. Given this value is in excess of 1000, it was not possible to compute a *p* value; however, the Akaike information criterion (AIC) values of the one- and two-factor models (1934.82 and 932.18, respectively) demonstrated the two-factor solution was a better fit to the data as the AIC value was smaller (Tabachnick & Fidell, 2014).

Table 2
Model comparisons for respecification of CFA model.

Model	χ^2	df_M	$\Delta\chi^2$	CFI	RMSEA [CI]
0. Baseline	822.18	134	–	.83	.100 [.094, .107]
1. Covariance e4-e7	673.64	133	148.54**	.87	.089 [.083, .096]
2. Covariance e17-e5	582.55	132	91.09**	.89	.082 [.075, .089]
3. Covariance e14-e15	511.26	131	162.38**	.91	.075 [.069, .082]
4. Covariance e2-e3	441.06	130	70.20**	.92	.068 [.061, .075]
5. Covariance e17-e9	402.40	129	38.66**	.93	.064 [.057, .072]
6. Covariance e3-e9	380.39	128	22.01**	.94	.062 [.055, .069]
7. Covariance e2-e4	364.93	127	15.461**	.94	.061 [.053, .069]

Note. **significant at $p < .001$.

4.2.4. Measurement invariance: age

Given the current study's sample included a significant number of older adults, which is uncommon with many early scale development papers relying on university student populations (Clark & Watson, 2016), we investigated whether the SEAAS performed consistently across diverse age groups, to ensure utility of the scale. As a well-fitting measurement model had been identified for the pooled data set, it was possible to proceed and determine whether individual items performed differently as a function of age, using a multigroup approach. The sample was divided into three age groups: younger adults ($n = 246$), middle-aged adults ($n = 81$), and older adults ($n = 177$). Testing for invariance across groups is conducted in multiple stages, whereby a baseline model of the two groups in which all parameters are free to vary (i.e., take on different values for each group), is first estimated. Constraints are then systematically added to different parts of the model to test for invariance at each level, and these nested models are compared to the baseline model to determine whether model fit significantly decreases using the chi-square difference statistic ($\Delta\ \chi^2$; Byrne, 2010).

As the residual error covariances were important to overall model fit for all groups, these were treated as equal across groups for all analyses, including the baseline model in line with recommendations (Byrne, 2010). The baseline model, in which all groups are estimated simultaneously, had a good fit to the data $\chi^2(df = 396) = 783.83, p < .001$; CFI = .908; RMSEA .044, C.I. [.040, .049], $p < .001$. Next, a measurement weights model was tested, in which all unstandardised factor loadings were constrained equal across age groups. This model had no significant impact on model fit, demonstrating invariance across groups (i.e., items of the SEAAS designed to measure acceptance and avoidance operate equally across age groups), as demonstrated by a non-significant chi-square difference statistic (Table 3). The structural aspect of the model was tested for invariance. In the structural covariances model, the covariance between the two identified factors were constrained to be equal, which did not detriment model fit, suggesting covariance between acceptance and avoidance was invariant across groups. These results are summarised in Table 3.

4.2.5. Construct validity

A series of bivariate correlations were used to examine relationships between theoretically related and unrelated constructs (see Table 4). Given the number of comparisons, false discovery rate (FDR) analyses were undertaken to control for family-wise error rates. FDR was selected over traditional Bonferroni corrections as FDR-based comparisons are more powerful and values have a monotonic linear increase, as opposed to the Bonferroni which has a fixed, and often overly conservative, threshold (Pike, 2011).

As expected, Openness and Willingness (OW) correlated negatively with AAQ-II and was positively related to wellbeing and positive affect. Avoidance and Control (AC) correlated positively with the AAQ-II, the suppression scale of ERQ, and negative affect as predicted. While OW negatively correlated with negative affect, and AC was negatively correlated with positive affect and wellbeing, these relationships were less strong.

4.2.6. Incremental validity

Three multiple regression models were run to test the incremental validity of the SEAAS on predicting wellbeing, PA and NA. As predicted, the SEAAS showed additional incremental validity beyond the AAQ-II

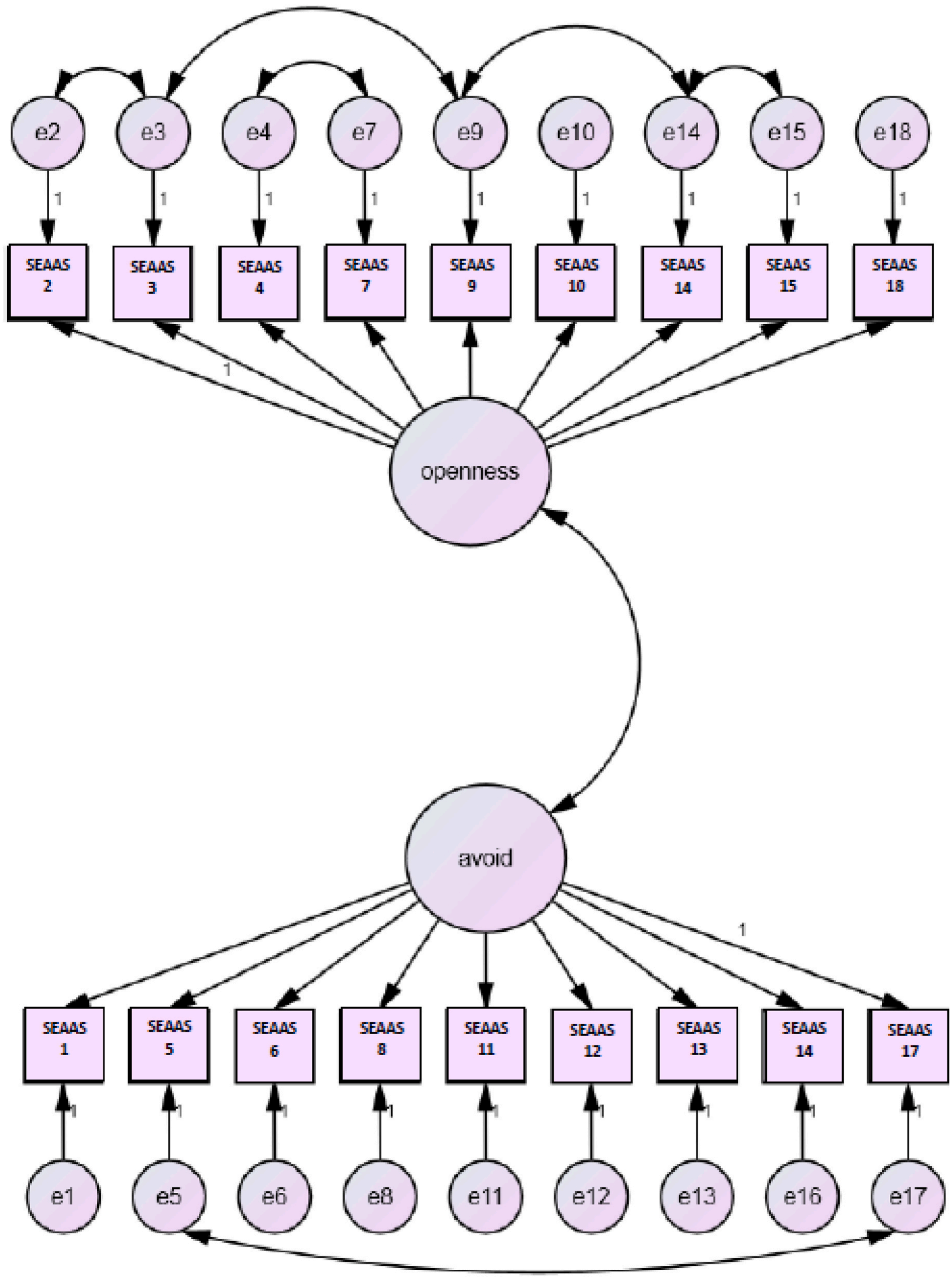

Fig. 1. Respecified model of factorial structure for the Southampton Experiential Avoidance and Acceptance Scale (SEAAS) Model 7.

(see Table 5). The SEAAS predicted an additional 5% of the variance for wellbeing and 4% for PA. While the SEAAS also predicted further variance for NA, this was only an additional 1%.

5. Study 4

Study 4 sought to evaluate the test-retest reliability of the SEAAS using an independent sample. Test-retest reliability is defined as a tools ability to consistently measure phenomena in the same participants at different time points (Patel & Joseph, 2016), and is a desired property

Table 3
Goodness of fit statistics for tests of multigroup invariance across age groups.

Constraints	χ^2 (*df*)	*p*	$\Delta\ \chi^2$ (*df*)	*p*	CFI	RMSEA [C.I.]
1. Configural Model	783.83 (396)	.0000	–	–	.908	.044 [.040, .049]
2. Measurement Weights	810.63 (428)	.0000	5.2864 (11)	.728	.909	.042 [.038, .047]
3. Structural covariances	821.47 (434)	.0000	53.5656 (33)	.486	.908	.042 [.038, .047]

for psychometric measures over short intervals in the absence of any intervention or other changes. Consequently, the SEAAS was expected to demonstrate good test-retest reliability. Different emotion regulation strategies can be considered antecedents or emotion modulation responses, suggesting a bidirectional relationship between emotions and regulation approaches (Gross & John, 2003). As such, anxiety and depression were measured in the current study to evidence stability over time between measurements.

5.1. Method

5.1.1. Participants

An online sample of 24 individuals (were recruited via a university student research portal and online research participation interest groups (e.g., http://www.onlinepsychresearch.co.uk/). G*power analysis indicated a sample of 23 would be required to detect a large effect size (Faul et al., 2009). The mean age of the sample was 35.76 years (*SD* = 13.88, range 18–69), and 76.2% were female, with 71.4% living in the UK, 19.0% in the USA and the remainder in Europe.

5.1.2. Materials and procedure

Participants completed the 18-item SEAAS, as well as a measure of mood and anxiety two weeks apart (T1 and T2). These measures were selected to assess stability during the interim period, and participants were asked about any changes or difficulties experienced in the interval. Informed consent, including the right to withdraw was provided at T1. Demographic information was captured at T1. Internal consistency of OW was .94 and .95, and AC was .92 and .94.

The *Patient Health Questionnaire* (PHQ-9; Kroenke et al., 2001) is a 9-item self-report measure of low mood over the past two weeks. Items are rated on a 4-point scale from 0 = *not at all* to 3 = *more than half the days*, and possible scores range from 0 to 27, with clinical levels of depression represented by scores ≥10. The PHQ-9 has been shown to have good internal consistency (α = .89; Kroenke et al., 2001). Reliable change on the PHQ-9 is evidenced by a change in score of ≥6 (National Collaborating Centre for Mental Health [NCCMH], 2018). Internal consistency in the present study was .93 and .92.

The *Generalised Anxiety Disorder Assessment* (GAD-7; Spitzer et al., 2006) is a 7-item self-report measure of generalised anxiety over the past two weeks. Items are rated on a 4-point scale, 0 = *not at all* – 3 = *nearly every day*, and possible scores range from 0 to 21, with clinical levels of depression represented by scores ≥8. The GAD-7 has been shown to have good internal consistency (α = .92) and test-retest reliability (intraclass correlation = .83; Spitzer et al., 2006). Reliable change on the

Table 4
Correlations between the SEAAS subscales with existing scales in study 3.

Measure		SEAAS				AAQ-II	
		Openness and Willingness		Avoidance and Control			
	Mean (*SD*)	*r*	R^2	*r*	R^2	*r*	R^2
AAQ-II	21.86 (*10.24*)	−.25[a]	.062	.41[a]	.170	–	–
ERQ							
Reappraisal	28.84 (*6.37*)	−.02	.004	.21[a]	.044	−.18[a]	.032
Suppression	14.13 (*4.96*)	−.25[a]	.062	.22[a]	.048	.18[a]	.032
PANAS							
Positive Affect	33.30 (*7.80*)	.26[a]	.065	−.10[b]	.010	.48[a]	.230
Negative Affect	21.91 (*8.06*)	−.19[a]	.037	.34[a]	.114	.61[a]	.372
SWEMWBS	35.00 (*6.75*)	.34[a]	.114	−.19[a]	.038	.61[a]	.372
BIDR-IM	22.52 (*4.51*)	.02	.004	.06	.004	.002	.000

Note. AAQ-II= Acceptance and Action Questionnaire-Revised; ERQ = Emotion Regulation Questionnaire; SWEMWBS= Short Warwick-Edinburgh Mental Wellbeing Scale; BIDR-IM= Balanced Inventory of Desired Responding Short Form- Impression Management Subscale. All significance levels are FDR adjusted.
[a] Significant at $p < .001$.
[b] Significant at $p < .05$.

Table 5
Incremental validity of the SEAAS.

Outcome	Step	Variable	β	*t*	*p*	R^2	R^2 change	*p*
Wellbeing	1	AAQ-II	−.628	−18.20	<.001	.40	.39	<.001
	2	AAQ-II	−.633	−17.23	<.001	.44	.05	<.001
		OW	.218	6.11	<.001			
		AC	.142	3.72	<.001			
Positive Affect	1	AAQ-II	−.483	−12.32	<.001	.23	.23	<.001
	2	AAQ-II	−.505	−12.00	<.001	.27	.04	<.001
		OW	.188	.19	<.001			
		AC	.171	.17	<.001			
Negative Affect	1	AAQ-II	.612	17.34	<.001	.37	.37	<.001
	2	AAQ-II	.567	14.59	<.001	.38	.01	.028
		OW	−.014	−.37	<.001			
		AC	.098	2.45	<.001			

Note. Method: Enter. AAQ-II= Acceptance and Action Questionnaire-Revised; OW = SEAAS Openness and Willingness subscale; AC = SEAAS Avoidance and Control subscale.

GAD-7 is evidenced by a change in score of ≥4 (NCCMH, 2018). Internal consistency in the present study was .94 and .96.

5.2. Results

Three individuals' scores on the PHQ-9 or GAD-7 demonstrated reliable change, indicating their mood had changed in the intervening weeks, and as such were not included in the analysis, resulting in a sample of 21 (16 males, 5 females). No individuals reported any significant changes during the interim and none were accessing psychological therapy which may have influenced their responses. Data was bootstrapped using 1000 resamples due to skew in the data. Scores on the SEAAS were stable over the two-week period for both subscales, with test-retest reliability for AC $r = .91$ ($p = .01$) and $r = .84$ ($p = .01$) for OW.

6. General discussion

No established measures currently assess acceptance as a general process within the ACT model. Existing ACT measures which assess acceptance either do so with regards to acceptance of a particular difficulty, or in the context of wider conceptualisations of psychological flexibility. Concerns have been raised regarding the validity of extant measures. As such, the present study sought to develop and validate a new measure of acceptance.

The measure development process yielded an 18-item, two-factor scale: capturing acceptance processes and experiential avoidance. This factor structure was confirmed in Study 3 using a diverse sample of adults. In contrast to many published scales, the SEAAS does not contain reverse-scored items; these were excluded following factor analyses due to poor loadings. Research indicates positive and negatively worded items are likely to load onto distinct factors due to method effects, rather than to substantial differences in construct validity (DiStefano & Motl, 2006). However, during the initial EFA when no factors were specified, this was not the case. Items of both positive and negative valence loaded onto the same factor, although they demonstrated differential factor loading strengths. This replicates Gamez et al.'s (2011) findings that reverse-scored acceptance items did not satisfactorily load onto the non-acceptance scale and is in keeping with Thomas et al.'s (2022) findings of psychological flexibility and inflexibility loading onto separate factors. As such, while method effects may be present, the study's findings suggest there may be important differences in the latent structure these items measure. As the SEAAS is the only scale of acceptance which currently distinguishes and separately measures avoidance and acceptance, rather than relying on reverse scoring, further work is needed to fully understand this relationship.

Theoretically, acceptance and experiential avoidance are considered as active processes; acceptance involves actively approaching private experiences with openness and curiosity, whereas experiential avoidance entails attempts to avoid, control, or escape such events. The absence of one approach to private experiences does not necessarily imply the presence of the other, i.e., low avoidance does not equate to high acceptance. Consequently, acceptance and experiential avoidance may represent distinct, albeit related constructs, much like positive and negative affect. Since this reconceptualisation of affect, models recognised the need to address both the downregulation of negative affect and up-regulation of positive emotions. While some strategies can facilitate both processes, this is not always the case. Consequently, if experiential avoidance and acceptance are distinct constructs, it is important for clinical research to explore whether strategies which seek to foster acceptance also serve to reduce avoidance, or if further strategies are required to explicitly address avoidance.

The SEAAS displayed excellent internal reliability across samples, demonstrating invariance with adults at different life stages, suggesting scale items are not susceptible to cohort effects. Predicted patterns of relationships with theoretically related and distinct constructs were also found, providing evidence of the SEAAS's validity. Incremental validity was also demonstrated for all three outcomes beyond psychological inflexibility/experiential avoidance as measured. Evidence for the temporal stability of the SEAAS was also demonstrated.

We showed acceptance was not associated with reappraisal, whereas avoidance was. Many items on the 'avoidance and control' scale refer to change agendas which could be considered synonymous with cognitive change advocated in reappraisal and may explain some of the shared variance. Both acceptance and reappraisal require individuals to step back from and observe private events. However, the response to these differs (allowing versus changing), with both approaches having clinical merit. It has been suggested it is the act of 'decentering' before responding to events that is significant, with both acceptance and reappraisal requiring mindful awareness of events prior to responding (Teasdale, 1999). The lack of association between the two measures would suggest the subscales distinguish responses to private events from the 'decentering' process. Other published ACT measures have not directly compared acceptance and reappraisal, so further work is needed to understand this relationship.

The ACT model posits symptom reduction is not the aim of acceptance, rather it promotes living in line with values alongside private experiences, although many individuals do report some symptom reduction. Despite affect modulation not being the aim of acceptance, relationships between affect and acceptance were found. Acceptance demonstrated a stronger relationship with psychological wellbeing than with affect. Given links between acceptance and elements pertinent to valued living, this stronger relationship is unsurprising.

Correlations of SEAAS subscales with all variables were smaller than those reported for some published ACT process measures. There are several possible reasons for this. Firstly, the SEAAS seeks to measure two specific processes within the ACT model, whereas measures such as the CompACT measures the model in its entirety, and subscales combine multiple processes. Consequently, the proportion of variance explained is smaller. Other process measures, such as the Valued Living Questionnaire (VLQ; Wilson et al., 2010) demonstrate comparable relationships with two of the ACT processes - affect and wellbeing - as does the SEAAS. Secondly, scales such as the CompACT were developed using items from existing measures, resulting in shared items which artificially increases shared variance and strengthens the relationships between constructs.

Age may also have influenced the strength of relationships between variables in the present study. The sample in Study 3 contained substantially more older adults than are commonly used in such studies. For example, the VLQ was developed and validated using a sample where 93.2% of adults were aged 18–22. The SEAAS demonstrated measurement invariance and so items can be considered to operate equally across the lifespan. However, relationships between emotion regulation strategies and affect have been shown to differ as a function of age (Allen & Windsor, 2017). As such, the strength of relationships between the SEAAS and other scales may be a result of age-related differences in these constructs.

The SEAAS benefitted from the involvement of both experts and non-experts in developing the item pool. Experts contributed their specialised knowledge to ensure items aligned with the intended construct, while non-experts provided valuable perspectives to ensure items were relevant and comprehensible to the broader population. This collaborative approach helped establish a more comprehensive and accurate representation of the intended construct, increasing the overall validity of the study's measurement tool.

A limitation pertains to the samples used to validate our measure – all of whom were recruited from a non-clinical population. While 45% of the sample identified themselves as having experienced mental health difficulties, it is not possible to draw conclusions as to how the SEAAS would perform in a clinical population, and this requires further testing. Further research is also required to assess the sensitivity of the SEAAS to treatment effects, particularly in the context of acceptance-based interventions. This will help investigate hypotheses regarding mechanisms of change. Given acceptance and experiential avoidance are theorised to predict committed action, which in turn predicts psychological well-being/distress, future studies should examine a predictive model of these processes.

The present study sought to develop a functional measure of acceptance and explore the dimensionality of acceptance and experiential avoidance. While our results suggest these are two distinct constructs, there is not enough evidence to demonstrate this finding is free of method effects. The SEAAS is the only acceptance scale to not contain reverse scored items. Further research is needed to explore the dimensionality and impact of reverse scoring and negatively valanced items of extant measures.

The SEAAS is a theoretically coherent measure which demonstrates strong psychometric properties. The brevity and simplicity of the SEAAS makes it well suited for use in clinical and research settings. As the measure is generic to all private events, it is applicable to a wide range of populations and contexts. The SEAAS has the advantage of being able to measure both acceptance and experiential avoidance independently, which would support further exploration of process dimensionality within the ACT model. While the present study demonstrates the SEAAS shows promise as a measure of acceptance, further research is needed to empirically determine its robustness across other populations, further explore psychometric properties, and examine its clinical utility.

Data availability

Data is available upon reasonable request.

CRediT authorship contribution statement

Zoe McAndrews: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Claire M. Hart:** Writing – review & editing, Supervision, Methodology. **Lusia Stopa:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research was carried out and funded as part of a Doctorate in Clinical Psychology course, undertaken by the first author.

References

Allen, V. C., & Windsor, T. D. (2017). Age differences in the use of emotion regulation strategies derived from the process model of emotion regulation: A systematic review. *Ageing & Mental Health*. https://doi.org/10.1080/13607863.2017.1396575

Arbuckle, J. L. (2017). *Amos*. Chicago: IBM SPSS [Computer Program] Version 25.0. .

Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., … Williams, J. M. G. (2008). Construct validity of the five Facet mindfulness questionnaire in meditating and nonmeditating samples. *Assessment, 15*(3), 329–342. https://doi.org/10.1177/1073191107313003

Blackledge, J. T., & Hayes, S. C. (2001). Emotion regulation in acceptance and commitment therapy. *Journal of Clinical Psychology, 57*(2), 243–255. https://doi.org/10.1002/1097-4679(200102)57:2<243::AID-JCLP9>3.0.CO;2-X

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., … Zettle, R. D. (2011). Preliminary psychometric properties of the acceptance and action questionnaire-II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy, 42*(4), 676–688. https://doi.org/10.1016/j.beth.2011.03.007

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. New York, New York: Routledge.

Cherry, K. M., Hoeven, E. V., Patterson, T. S., & Lumley, M. N. (2021). Defining and measuring "psychological flexibility": A narrative scoping review of diverse flexibility and rigidity constructs and perspectives. *Clinical Psychology Review, 84* (March 2020). https://doi.org/10.1016/j.cpr.2021.101973

Clark, L. A., & Watson, D. (2016). Constructing validity: Basic issues in objective scale development. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 187–203). American Psychological Association. https://doi.org/10.1037/14805-012.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*(3), 245–265. https://doi.org/10.1348/0144665031752934

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA, US: Sage Publications.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology, 67*(4), 401–409. https://doi.org/10.1016/j.jclinepi.2013.12.002

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling, 13*(3), 440–464. https://doi.org/10.1207/s15328007sem1303_6

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Francis, A. W., Dawson, D. L., & Golijani-Moghaddam, N. (2016). The development and validation of the comprehensive assessment of acceptance and commitment therapy processes (CompACT). *Journal of Contextual Behavioral Science, 5*(3), 134–145. https://doi.org/10.1016/j.jcbs.2016.05.003

Gable, S. L., Reis, H. T., & Elliot, A. J. (2003). Evidence for bivariate systems: An empirical test of appetition and aversion across domains. *Journal of Research in Personality, 37*(5), 349–372. https://doi.org/10.1016/S0092-6566(02)00580-9

Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., & Watson, D. (2011). Development of a measure of experiential avoidance: The multidimensional experiential avoidance questionnaire. *Psychological Assessment, 23*(3), 692–713. https://doi.org/10.1037/a0023242

Gloster, A. T., Walder, N., Levin, M. E., Twohig, M. P., & Karekla, M. (2020). The empirical status of acceptance and commitment therapy: A review of meta-analyses. *Journal of Contextual Behavioral Science, 18*, 181–192. https://doi.org/10.1016/j.jcbs.2020.09.009

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment, 26*(1), 41–54. https://doi.org/10.1007/s10862-008-9102-4

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*, 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Harris, R. (2009). *ACT made simple*. Oakland, CA: New Harbinger Publications, Inc.

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *Sage Open, 5*(4), 1–9. https://doi.org/10.1177/2158244015621113

Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behavior Therapy, 35*, 639–665. https://doi.org/10.1016/S0005-7894(04)80013-3

Hayes, S. C. (2023). State of the ACT evidence. Retrieved 3rd July 2023 from https://contextualscience.org/state_of_the_act_evidence.

Hayes, S. C., & Hofmann, S. G. (2017). The third wave of cognitive behavioral therapy and the rise of process-based care. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 16*(3), 245–246. https://doi.org/10.1002/wps.20442

Hayes, S. C., Levin, M. E., Plumb-Vilardaga, J., Villatte, J. L., & Pistorello, J. (2013). Acceptance and commitment therapy and contextual behavioral science: Examining the progress of a distinctive model of behavioral and cognitive therapy. *Behavior Therapy, 44*(2), 180–198. https://doi.org/10.1016/J.BETH.2009.08.002

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy, 44*(1), 1–25. https://doi.org/10.1016/j.brat.2005.06.006

Hayes, S. C., Strohsal, K. D., & Wilson, K. G. (1999). *Acceptance and Commitment Therapy: An experiential approach to behavior change*. New York, NY: Guilford Press.

Hayes, S. C., Strosahl, K., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., … McCurry, S. M. (2004). Measuring experiential avoidance: A preliminary test of a working model. *Psychological Record, 54*(4), 553–578. https://doi.org/10.1007/BF03395492

Hayes, S. C., Wilson, K. W., Gifford, E. V., Follette, V. M., & Strosahl, K. (1996). Experiential avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology, 64*(6), 1152–1168. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8991302.

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15*(4), 456. https://doi.org/10.1037/1040-3590.15.4.456

Hofmann, S. G., & Kashdan, T. B. (2010). The affective style questionnaire: Development and psychometric properties. *Journal of Psychopathology and Behavioral Assessment, 32*(2), 255–263. https://doi.org/10.1007/s10862-009-9142-4

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Sage.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606–613. https://doi.org/10.1016/S0165-0327(02)00237-9

Little, R., & Rubin, D. (2002). *Statistical Analysis with Missing Data*. New York: Wiley. https://doi.org/10.1002/9781119013563.

Little, R. J. A., & Rubin, D. B. (2014). *Incomplete data*. John Wiley & Sons, Inc.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychol. Methods, 4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

McAndrews, Z., Richardson, J., & Stopa, L. (2019). Psychometric properties of acceptance measures: A systematic review. *Journal of Contextual Behavioral Science, 12*, 261–277. https://doi.org/10.1016/j.jcbs.2018.08.006

National Collaborating Centre for Mental Health. (2018). The improving access to psychological therapies manual: Appendices and helpful resources. Retrieved from https://www.rcpsych.ac.uk/docs/default-source/improving-care/nccmh/iapt/nccmh-iapt-manual-appendices-helpful-resources-v2.pdf?sfvrsn=a607ef5_4.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA, US: Sage Publications.

Patel, H. R., & Joseph, J. M. (2016). Questionnaire designing process: A review. *Journal of Clinical Trials, 6*(2), 2–7. https://doi.org/10.4172/2167-0870.1000255

Pike, N. (2011). Using False Discovery Rates for multiple comparisons in ecology and evolution. *Methods in Ecology and Evolution, 2*(3), 278–282. https://doi.org/10.1111/j.2041-210X.2010.00061.x

Rolffs, J. L., Rogge, R. D., & Wilson, K. G. (2018). Disentangling components of flexibility via the hexaflex model: Development and validation of the multidimensional psychological flexibility inventory (MPFI). *Assessment, 25*(4), 458–482. https://doi.org/10.1177/1073191116645905

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement, 79*(2), 310–334. https://doi.org/10.1177/0013164418783530

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166* (10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stewart-Brown, S., & Janmohamed, K. (2008). *Warwick-Edinburgh mental well-being scale*. User guide. Version, 1(10.1037).

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh mental well-being scale (WEMWBS): A Rasch analysis using data from the Scottish health education population survey. *Health and Quality of Life Outcomes, 7*(1), 15. https://doi.org/10.1186/1477-7525-7-15

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics: Pearson new international edition* (6th ed.). Harlow, Essex: Pearson Education Ltd.

Teasdale, J. D. (1999). Metacognition, mindfulness and the modification of mood disorders. *Clinical Psychology & Psychotherapy, 6*(2), 146–155. https://doi.org/10.1002/(SICI)1099-0879(199905)6:2<146::AID-CPP195>3.0.CO;2-E

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., … de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Thomas, K. N., Bardeen, J. R., Witte, T. K., Rogers, T. A., Benfer, N., & Clauss, K. (2022). An examination of the factor structure of the multidimensional psychological flexibility inventory. *Assessment, 29*(8), 1714–1729. https://doi.org/10.1177/107319112110243

Tyndall, I., Waldeck, D., Pancani, L., Whelan, R., Roche, B., & Dawson, D. L. (2019). The Acceptance and Action Questionnaire-II (AAQ-II) as a measure of experiential avoidance: Concerns over discriminant validity. *Journal of Contextual Behavioral Science, 12*, 278–284. https://doi.org/10.1016/j.jcbs.2018.09.005

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wilson, K. G., Sandoz, E. K., Kitchens, J., & Roberts, M. (2010). The Valued Living Questionnaire: Defining and measuring valued action within a behavioral framework. *Psychological Record, 60*, 249–272. https://doi.org/10.1007/BF03395706

Wolgast, M. (2014). What does the acceptance and action questionnaire (AAQ-II) really measure? *Behavior Therapy, 45*(6), 831–839. https://doi.org/10.1016/j.beth.2014.07.002