



Published in final edited form as:

Psychol Assess. 2023 May ; 35(5): 430–442. doi:10.1037/pas0001214.

Confirmatory Measurement Modeling and Longitudinal Invariance of the CompACT-15: A Short-Form Assessment of Psychological Flexibility

Ti Hsu¹, Lesa Hoffman², Emily B. K. Thomas¹

¹Department of Psychological and Brain Sciences, University of Iowa

²Department of Psychological and Quantitative Foundations, University of Iowa

Abstract

The Comprehensive assessment of Acceptance and Commitment Therapy processes (CompACT; Francis et al., 2016) is a recently developed measure of psychological flexibility (PF) possessing several advantages over other measures of PF, including multidimensional assessment and relative brevity. Unfortunately, previous psychometric evaluations of the CompACT have been limited by their use of exploratory factor analysis to assess dimensionality, coefficient alpha to assess reliability, and a lack of attention to measurement equivalence in assessing change over time. The current study used confirmatory factor analysis (CFA) and item factor analysis (IFA) to examine the dimensionality, factor-specific reliability, longitudinal measurement invariance, and construct validity of the CompACT items in a longitudinal online sample of U.S. adults ($N = 523$). Converging evidence across CFA and IFA confirmatory latent variable measurement models provides support for the reduction of the 23-item CompACT to a 15-item short form with a more stable factor structure, acceptable reliability over large ranges of its three latent factors, and measurement equivalence of its items in assessing latent change over time. Results also support the construct validity of the CompACT-15 items based on its relations with theoretically relevant measures. Overall, the CompACT-15 appears to be a psychometrically sound instrument with the potential to contribute to research and intervention efforts.

Keywords

Psychological Flexibility; The Comprehensive assessment of Acceptance and Commitment Therapy processes; confirmatory factor analysis; item factor analysis

Correspondence concerning this article should be addressed to Emily B.K. Thomas, Department of Psychological and Brain Sciences, University of Iowa, G60 Psychological and Brain Sciences Building, Iowa City, Iowa 52242-1407. Phone: 319-467-1691. emily-kroska@uiowa.edu.

Ti Hsu: Conceptualization, Data Curation, Formal Analysis, Methodology, Visualization, Writing-Original Draft, Writing-Review and Editing. Lesa Hoffman: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Supervision, Visualization, Writing-Original Draft, Writing-Review and Editing. Emily B.K. Thomas: Conceptualization, Funding Acquisition, Investigation, Project Administration, Resources, Supervision, Writing-Original Draft, Writing-Review and Editing.

Preliminary findings from the study were included in a poster presentation at the 2021 Virtual World Conference of the Association for Contextual Behavioral Science (ACBS WC), and a symposium talk at the 2022 ACBS WC.

This study was not preregistered. Syntax and additional results are available as supplementary materials.

Psychological flexibility (PF) is a set of abilities by which one can focus on the present moment and maintain behaviors in pursuit of one's values, even when experiencing difficult thoughts and emotions (Hayes et al., 2006). The six-factor Hexaflex model (Hayes et al., 1999) divided PF into six subprocesses: *acceptance*, *defusion*, *contact with the present moment*, *self-as-context*, *values*, and *committed action*. Strosahl et al. (2012) collapsed these six subprocesses into three abilities: *openness* (i.e., *acceptance* and *defusion*), taking a non-judgmental stance toward unpleasant thoughts and emotions; *awareness* (i.e., *contact with the present moment* and *self-as-context*), experiencing the present moment from one's own perspective; and *engagement* (i.e., *values* and *committed action*), identifying and engaging in self-valued actions.

PF is thought to be related to many constructs, including executive function, self-control, self-regulation, and emotion regulation (Doorley et al., 2020). Other studies have reported positive associations among PF and personality traits such as extraversion, emotionality, and agreeableness (Pyszkowska, 2020). In a narrative review examining different definitions and measurement of flexibility and rigidity constructs in the psychological literature, Cherry and colleagues (2021) described conceptual overlaps between PF and cognitive flexibility, a well-studied construct in the neuropsychological literature. They also noted that the various flexibility constructs all shared features of handling distress, taking action, and pursuing values and goals. Of note, concerns regarding PF's status as a construct often focus on whether PF is truly distinct from related constructs. Some studies have addressed the uniqueness of PF above other constructs by demonstrating that some measurements of PF show incremental validity above and beyond related constructs (e.g., Kollman et al., 2019; Gloster et al., 2011). However, there is also evidence that the most common measures of PF do not provide robust evidence of discriminant validity (Ong et al., 2019a). Due to these issues, PF's uniqueness as a construct still warrants investigation, especially in the context of PF measurement refinement and development.

Importantly, PF skills are posited to increase with treatment during Acceptance and Commitment Therapy (ACT), which in turn is theorized to reduce negative psychological and physical outcomes (Brandon et al., 2021; Hayes et al., 2013). This conceptualization of PF as a mechanism of change—that is, as a modifiable ability, and not just a personality trait—is empirically supported by research findings that improvements in PF relate to increased wellbeing (Wersebe et al., 2018), as well as concurrent symptom reductions in many conditions, ranging from depression and anxiety (Fledderus et al., 2013) to chronic pain (Hughes et al., 2017). Studies demonstrating that PF mediates associations between ACT and treatment outcomes (for a meta-analysis, see Stockton et al., 2019) also provide evidence for this conceptualization. However, some researchers have considered this existing evidence as weak by arguing that its quality is dampened by the same methodological issue that affects PF's standing as a unique construct: flawed measurement (Arch et al., 2022). Most importantly, the PF measures that are traditionally used as mechanisms of change have not yet been examined for longitudinal invariance, a critical gap in the literature. As a result, it cannot be determined that change observed in PF is not due to artifacts of measurement and is in fact related to therapeutic change.

Unidimensional Measurement of PF

PF is most often measured indirectly with the Acceptance and Action Questionnaire-II (AAQ-II; Bond et al., 2011), whose seven items instead directly assess experiential avoidance, or psychological *inflexibility* (i.e., a lack of PF; Hayes et al., 1999). In describing the AAQ-II's development, Bond et al. (2011) found promising psychometric evidence across six samples, including a well-fitting one-factor structure (after adding a residual covariance), good alpha reliability, and good test-retest reliability after three and twelve months. With respect to evidence for construct validity, they found that greater psychological inflexibility was related one year later to greater concurrent depression, anxiety, stress, and distress.

Fledderus et al. (2012) examined the psychometric properties of a ten-item version of the AAQ-II (including three items dropped by Bond et al., 2011) in a clinical adult sample. Their confirmatory factor analyses (CFA) replicated the established one-factor structure and showed that an AAQ-II sum score had incremental validity beyond mindfulness in predicting anxiety, depression, and mental health. They also examined measurement equivalence (i.e., a lack of differential item functioning, or DIF) and found that most items functioned equivalently by age and sex. However, Ong et al. (2019b) reported a greater amount of DIF across clinical and nonclinical samples using the seven-item AAQ-II, indicating that item usage may need to differ based on respondent type. Flynn et al. (2016) conducted a CFA on the seven-item AAQ-II in a Hispanic sample and achieved marginal fit for a one-factor structure only after adding residual covariances between two other pairs of items. They also showed that an AAQ-II sum score predicted distress and life satisfaction beyond mindfulness and thought suppression and was correlated with lower life satisfaction and with greater anxiety, depression, and stress.

However, other studies have reported problems for the AAQ-II's discriminant validity with measures of psychological distress. In an exploratory factor analysis of AAQ-II items combined with additional items, Wolgast (2014) found that the AAQ-II items were more strongly related to items measuring distress than items measuring nonacceptance or acceptance. Likewise, CFAs conducted by Tyndall et al. (2019) indicated that the AAQ-II factor was more strongly correlated with factors for depression, anxiety, and stress than with a second factor of psychological inflexibility (i.e., the Brief Experiential Avoidance Questionnaire items; Gámez et al., 2014). Others have suggested that AAQ-II items actually measure neuroticism and negative affect (Rochefort et al., 2018). Thus, it appears that the AAQ-II items likely measure overall distress and internalizing symptoms more strongly than (or in addition to) their targeted construct of psychological inflexibility, warranting the development of alternative measures of PF.

Multidimensional Measurement of PF

Beyond insufficient discriminant validity, the AAQ-II also faces two other potential barriers to serving as a comprehensive measure of PF. First, as mentioned earlier, the wording of the AAQ-II items directly assesses psychological *inflexibility*, not PF. Such usage may be problematic given recent debates as to whether psychological flexibility and inflexibility in

fact lie along the same continuum (Ciarrochi et al., 2014), and whether a positive construct can be measured through the absence of a negative construct (Johnson & Wood, 2017). Second, the AAQ-II items measure (a lack of) PF as a unidimensional construct, whereas PF was originally theorized to be multidimensional. As a result, subsequent to the development of the original AAQ-II and its many contextually-sensitive variants (see Ong et al., 2019a), a number of other measures have tried to address its limitations as a unidimensional measure of PF. For instance, the 62-item Multidimensional Experiential Avoidance Questionnaire (Gámez et al., 2011) assesses six dimensions of experiential avoidance (i.e., psychological inflexibility). Also, the 60-item Multidimensional Psychological Flexibility Inventory (MPFI; Rolffs et al., 2018) has six subscales that assess psychological flexibility and six subscales that assess psychological inflexibility. However, the response burden created from including such large numbers of items may make the use of these multidimensional measures impractical in many settings.

The Comprehensive assessment of Acceptance and Commitment Therapy processes (CompACT) was created by Francis et al. (2016) as yet another multidimensional measure of PF. Relative to the other measures, the CompACT has fewer items and assesses PF using the same three-factor model employed in Focused Acceptance and Commitment Therapy (FACT; Strosahl et al., 2012), a brief, evidence-based therapy based on ACT. FACT and similar approaches address efficient treatment access and other important mental healthcare needs, and the CompACT provides a therapeutically compatible measure of change. Using exploratory factor analysis (EFA), Francis et al. (2016) refined their initial item pool into 23 items proposed to measure three latent factors (as subscales)—8 for *openness to experience*, 5 for *behavioral awareness*, and 8 for *valued action*. Acceptable coefficient alpha reliability was reported for each subscale, as well as the expected patterns for PF construct validity (e.g., negative relations with depression and anxiety, positive relations with health and well-being). In recent reviews, Ong et al. (2020) commended the CompACT's strong discriminant validity, and Cherry et al. (2020) praised its evidence for content validity, internal consistency, and discriminant validity.

Although the CompACT is a significant step forward in the multidimensional assessment of PF, three limitations of its psychometric evaluation thus far should be addressed. First, its dimensionality has been assessed largely through EFAs in which all 23 items were specified as measuring all three latent factors and the fit of the hypothesized factors was not tested. Yet scale evaluation using confirmatory latent variable measurement models is essential, particularly at later stages of measurement validation. Such models provide empirical tests of theoretically driven latent factor structures (i.e., those in which items are specified as measuring only certain factors) capturing common content or other common characteristics (e.g., wording direction).

Second, the CompACT's reliability has largely been assessed via coefficient alpha, which is merely a function of the average inter-item correlation and the number of items. Although commonly used (see McNeish, 2018), alpha reliability makes three untested assumptions about the items: unidimensionality in measuring a single latent factor, equal discrimination (i.e., their strength in measuring that factor), and no residual correlations (i.e., no local dependency, as might result from common wording direction). In contrast,

confirmatory measurement models can provide model-based estimates of reliability that more accurately reflect the characteristics of the items in measuring a unidimensional latent factor (McDonald, 1999).

Third, the stability over time of the CompACT items has not yet been examined. Of the few randomized controlled trials that have evaluated changes in PF using the CompACT (e.g., Lappalainen et al., 2021; Levin et al., 2019; Petersen et al., 2021), two assessed change over time in a total score of the three subscales. But this strategy does not account for unreliability, confounding changes over time in the way in which items measure their respective factors, or differential change across the three PF dimensions. Measurement stability over time is critical when assessing longitudinal changes in PF over the course of treatment—measurement validity hinges upon the items measuring the same constructs equivalently across time. Accordingly, a logical precursor to measuring systematic improvement is a demonstration of construct stability over time in a non-treatment sample (in which spontaneous improvements are unlikely).

Current Study

The present study reports a psychometric examination of four aspects of the CompACT items—their dimensionality, reliability, stability over time, and construct validity—in an observational longitudinal sample of adults. We propose to reduce the current CompACT 23-item measure to a 15-item short form that follows a more straightforward factor structure yet maintains good reliability of each of the three latent dimensions. Given the 7-point ordinal item response format, two distinct types of confirmatory measurement models could be used for these purposes. For maximal application of our findings to researchers from different backgrounds, we report integrated results across both types of measurement models, as introduced below.

One model type, *confirmatory factor analysis* (CFA), specifies a linear relationship of one or more latent factors in predicting each item response. Because CFA uses a normal conditional distribution for the item responses (which are thus assumed to be continuous), it can be less appropriate for ordinal items (Fernando, 2009). Fortunately, CFA models can be modified to predict categorical responses, creating a second type of model, known as either *item factor analysis* (IFA; i.e., confirmatory factor analysis for categorical outcomes) or *item response theory* (IRT), depending on the form of the item parameters the model provides (as elaborated later; for original developments see Muthén, 1984; Takane & de Leeuw, 1987). IFA/IRT models specify a nonlinear prediction of each item response by the latent factor(s) through the use of link (transformation) functions, as well as a multinomial (instead of normal) conditional response distribution. That is, CFA is to linear regression as IFA/IRT is to ordinal regression (each using a latent factor as a predictor instead of an observed variable predictor). In the present work, we use both CFA and IFA/IRT models to assess the fit of the hypothesized factor structure and each factor's reliability (which necessarily varies over the latent factor in IFA/IRT models).

In addition, both types of models provide powerful analytic frameworks for evaluating measurement stability across occasions (or groups of individuals). This concern, known

by the general term *measurement equivalence*, is often referred to as *measurement* (or *factorial*) *invariance* within CFA models or *differential item functioning* within IFA/IRT models (Millsap, 2012). An examination of measurement equivalence over time addresses the extent to which longitudinal changes reflect true changes in the respondents' latent factors, rather than artifactual variation created by changes in how the items relate to their factors over time. In the present study we sought to demonstrate stability of the latent factors in the absence of treatment change.

Finally, given the refinement of the CompACT into a 15-item short form, we also report evidence of its construct validity as a measure of PF with other theoretically related measures. We expected the three CompACT dimensions to each relate positively to resilience but negatively to psychological inflexibility, intolerance of uncertainty, and psychological distress.

Method

Participants

Participants ($N = 523$) were recruited from Amazon Mechanical Turk (MTurk) in May 2020. After these baseline data were examined, only participants who successfully completed all attention check items (as described below) were retained and contacted to participate in two follow-up surveys. At baseline, participants were predominantly White (78.1%; African American or Black 12%; American Indian or Alaska Native 1.4%; Asian 5.6%; Multiracial 2.7%; Did not disclose 0.2%), non-Hispanic (82.9%), male (59.6%), heterosexual (86.5%; lesbian, gay, or homosexual 1.9%; bisexual 11.6%), and employed full-time (76.3%; employed part-time 10.4%; unemployed 11%). On average, participants were 37.42 years of age ($SD = 11.46$) and had received 15.09 years of education ($SD = 2.93$).

Participants could complete online surveys at three occasions: baseline (T0; $n = 485$), one month later (T1; $n = 360$), and two months later (T2; $n = 266$). We conducted bivariate analyses to examine differences between T1 and T2 completers and those who did not return. Age did not differ between T1 completers ($M = 37.66$, $SD = 11.27$) and T1 non-completers ($M = 36.74$, $SD = 12.00$), $t(482) = -.78$, $p = .44$, but T2 completers ($M = 38.48$, $SD = 11.40$) were older than T2 non-completers ($M = 36.14$, $SD = 11.42$), $t(482) = -2.25$, $p < .05$, $d = -.005$. T1 completers had more years of education ($M = 15.31$, $SD = 2.70$) than T1 non-completers ($M = 14.46$, $SD = 3.46$), $t(176.22) = -2.48$, $p < .05$, $d = -.014$, as did T2 completers ($M = 15.35$, $SD = 2.53$) relative to T2 non-completers ($M = 14.47$, $SD = 3.34$), $t(391.02) = -2.48$, $p < .05$, $d = -.006$. Completion was not related to: race at T1, $\chi^2(4) = 2.12$, $p = .71$, or T2, $\chi^2(4) = 2.06$, $p = .73$, sex at T1, $\chi^2(1) = 0.56$, $p = .46$, or T2, $\chi^2(1) = 0.001$, $p = .99$, employment at T1, $\chi^2(2) = 0.93$, $p = .63$, or T2, $\chi^2(2) = 1.52$, $p = .47$, or sexual orientation at T1, $\chi^2(1) = 0.14$, $p = .76$. Completers were more likely to identify as Hispanic/Latino/a/x than non-completers at T1, $\chi^2(1) = 16.77$, $p < .001$, $\phi = .19$, and T2, $\chi^2(1) = 17.47$, $p < .001$, $\phi = .19$, and completers were more likely to identify as heterosexual than lesbian, gay, or bisexual at T2 $\chi^2(1) = 6.53$, $p < .05$, $\phi = .12$.

Procedure

The current data were obtained as part of a larger longitudinal study examining coping strategies, risk factors, and psychiatric symptoms during the COVID-19 pandemic. All study procedures were approved by and conducted in compliance with the University of Iowa Institutional Review Board. Identical to procedures described in Krokso et al. (2020), participants were recruited through the MTurk platform CloudResearch, which ensures quality responses by screening for automated responding. Inclusion criteria were being a U.S. resident; fluency in English; age 18 years; and previous completion of 100 Human Intelligence Tasks (HITs) with an approval rate 95%. Participants were given \$3.50 for the baseline survey and \$3.00 for each follow-up survey via Amazon Payments within three days of survey completion.

All items were administered through a Qualtrics survey, which included several validity checks to reduce inattentive, straight-line, or automated responding. First, a reCAPTCHA response was required immediately after consent and the survey would not proceed without a valid response. Furthermore, an open arithmetic question (e.g., What is 3 + 4?) to which participants could provide numerical or text-based answers was used to filter out automated bots. Attention checks that requested participants to “please click strongly agree if you’re reading this” were also placed throughout the survey to ensure that questions were read. A total of 38 participants who did not pass validity checks were excluded from our analyses. This study was not preregistered. Syntax and additional results are included as supplementary materials.

Measures

Psychological Flexibility—As described previously, the 23-item CompACT (Francis et al., 2016) measures three factors: *openness to experience* (10 items), *behavioral awareness* (5 items), and *valued action* (8 items). The items and descriptive statistics at T0 are given in Table 1. Participants responded to each item on a 7-point ordinal scale from 0 = *strongly disagree* to 6 = *strongly agree*, and 12 of the 23 items were reverse-coded so that higher values always indicated greater PF. The 7-item AAQ-II (Bond et al., 2011) measures psychological *inflexibility*, or a lack of PF (e.g., “Emotions cause problems in my life”). Participants responded to each item on a 7-point ordinal scale from 1 = *never true* to 7 = *always true*, such that higher values always indicated greater *inflexibility*.

Construct Validity—The 6-item Brief Resilience Scale (BRS; Smith et al., 2008) measures *resilience* as the ability to recover from stress (e.g., “I tend to bounce back quickly after hard times”). Participants responded to each item on a 5-point ordinal scale from 1 = *strongly disagree* to 5 = *strongly agree*, and 3 of the 6 items were reverse-coded so that higher values always indicated greater resilience. The 12-item Intolerance of Uncertainty Scale-12 (IUS-12; Carleton et al., 2007) measures *prospective intolerance of uncertainty* (7 items; e.g., “Unforeseen events upset me greatly”) and *inhibitory intolerance of uncertainty* (5 items; e.g., “Uncertainty keeps me from living a full life”). Participants responded to each item on a 5-point ordinal scale from 1 = *not at all characteristic of me* to 5 = *entirely characteristic of me*, such that higher values indicated greater intolerance. Lastly, the 10-item Kessler Psychological Distress Scale (K10; Kessler et al., 2002) measures

psychological distress via recent depression and anxiety symptoms (e.g., “During the last 30 days, about how often did you feel worthless”). Participants responded to each item on a 5-point ordinal scale from 1 = *none of the time* to 5 = *all of the time*, such that higher values always indicated greater distress.

Analytic Strategy

Two types of confirmatory latent variable measurement models—CFA and IFA/IRT—were used to examine the CompACT items’ hypothesized three-factor dimensionality, reliability per factor, measurement equivalence across three occasions, and correlational evidence for construct validity. Model–data fit was assessed using the model χ^2 test, Tucker-Lewis index (TLI), comparative fit index (CFI), root mean squared error of approximation (RMSEA) estimate [and 90% confidence interval], and standardized root-mean-square residual (SRMR). Acceptable global fit was indicated by a nonsignificant χ^2 test, TLI and CFI $\geq .95$, RMSEA $\leq .06$, and SRMR $\leq .08$ (Hu & Bentler, 1999). Local misfit was indicated by discrepancies between the model-predicted and data-estimated correlations for each pair of items. *Mplus* v. 8.7 (Muthén & Muthén, 1998–2017) was used to estimate both types of models, as described next.

First, analogous to linear regression, CFA models are designed to predict a continuous, normally distributed item response (y_{is} for item i and subject s) from at least one latent factor score (F_s for subject s) as: $y_{is} = \mu_i + \lambda_i F_s + e_{is}$. A CFA item’s parameters include an intercept (μ_i , the expected response at $F_s = 0$), a factor loading (λ_i , a slope for the difference in response per unit F_s), and a residual variance (of e_{is} across subjects). Each latent factor must be given a mean and variance across subjects to identify the model. In the present study, we fixed factor means to 0 and factor variances to 1 (except where noted) and estimated all item parameters.

We used robust full-information maximum likelihood (MLR) to estimate all CFA models, which yields the parameters that best recreate the observed item response means and Pearson covariance matrix. Accordingly, CFA model comparisons were conducted using rescaled likelihood ratio tests (i.e., the scaled difference in $-2 \times \text{LL}$ with degrees of freedom as the number of added parameters). Notably, although “robust” refers to the use of a scaling factor by which to adjust model fit statistics and parameter standard errors for multivariate nonnormality, it cannot overcome the inherent unsuitability of a CFA model for ordinal responses (see Fernando, 2009). That is, a linear slope for the latent factor can predict item responses that exceed the possible range of options, particularly in skewed response distributions, in which case the corresponding assumption of constant reliability across the latent factor may also not hold.

We addressed these limitations of CFA models by also examining results from IFA/IRT models (i.e., confirmatory factor analysis for categorical outcomes), in which the IFA version was estimated directly using diagonally weighted least squares (WLSMV with a THETA parameterization that fixes item residual variances to 1). Accordingly, IFA/IRT model comparisons were conducted using the required DIFFTEST procedure in *Mplus*. We chose this limited-information estimator—which yields the model parameters that best recreate the observed response category probabilities and data-estimated polychoric

correlation matrix—for three reasons. First, it provides the same model–data fit statistics as for traditional linear CFA models. Second, it allows the direct inclusion of residual covariances, as needed in the invariance models (as reported below). Third, it avoids integration over the unknown factor scores, which becomes computationally intractable for multidimensional models (see Muthén et al., 2015).

Analogous to ordinal regression, IFA/IRT models predict the probability of each of K total response options using a cumulative link function and $K - 1$ binary submodels (that each predict the response probability of $y_{is} > k$). WLSMV estimation uses a cumulative probit link function, such that each submodel's outcome was the z-score for the predicted probability that corresponds to the area to the left under a standard normal distribution. Like CFA models, both IFA and IRT models specify the prediction of item responses from latent factor(s), but IFA and IRT models differ from CFA models—and from each other—by their item parameters.

The IFA parameterization is: $\text{probit}[p(y_{is} > k)] = -\tau_{ik} + \lambda_j F_s$, in which all submodels for an item share one factor loading (λ_j), but each submodel has its own threshold (τ_{ik}) location parameter. Given $K = 7$ response options (from $k = 0$ to 6) requiring six submodels, the first predicts $\text{probit}[p(y_{is} > 0)] = -\tau_{i0} + \lambda_j F_s$, the second predicts $\text{probit}[p(y_{is} > 1)] = -\tau_{i1} + \lambda_j F_s$, and so on. As indicated by each minus sign, a threshold is the opposite of an intercept: Whereas an intercept μ_{ik} (indexing easiness) would give the predicted probit of $y_{is} > k$ at $F_s = 0$, a threshold τ_{ik} (indexing difficulty) gives the predicted probit of $y_{is} > k$ at $F_s = 0$ instead.

Thresholds create directional correspondence of IFA parameters to those of an equivalent graded response (normal ogive) IRT model. In this case, the corresponding IRT parameterization is: $\text{probit}[p(y_{is} > k)] = a_j(F_s - b_{ik})$, in which the IRT item discrimination a_j replaces the IFA factor loading λ_j , and the submodel-specific IRT item difficulty b_{ik} replaces the IFA threshold τ_{ik} . Given a factor mean = 0 and factor variance = 1, we converted the IFA parameters into IRT parameters (via *Mplus* MODEL CONSTRAINT) as $a_j = \lambda_j$ and $b_{ik} = \tau_{ik} / \lambda_j$ (see also Paek et al., 2018). An IRT b_{ik} difficulty gives the latent factor at which the predicted probability $p(y_{is} > k) = .50$, a more useful parameter than an IFA threshold. In addition, the b_{ik} difficulty values can quantify how useful the response options are in differentiating respondents across the latent factor, a consideration that cannot be addressed using CFA item intercept parameters.

Results

Dimensionality and Scale Revision

We first examined the extent to which the CompACT-23 item covariances at T0 could be predicted by the hypothesized three correlated factors of *openness to experience*, *behavioral awareness*, and *valued action*. Unfortunately, we found poor model fit by every index using both CFA, $\chi^2(227) = 1028, p < .001, \text{TLI} = .813, \text{CFI} = .832, \text{RMSEA} = .085 [.080, .091], \text{SRMR} = .128$, and IFA, $\chi^2(227) = 3964, p < .001, \text{TLI} = .769, \text{CFI} = .793, \text{RMSEA} = .184 [.179, .189], \text{SRMR} = .116$. Two design issues appear to have contributed to the model misfit (as evidenced by more positive correlations among certain items than were

predicted by the factors). First is the uneven use of positive or negative wording across factors: All 5 behavioral awareness items are negatively-worded, and all 8 valued action items are positively worded, but of the 10 openness to experience items, 7 are negative and 3 are positive. Second is redundancy in content, such as between item 13 (“I am willing to fully experience whatever thoughts, feelings and sensations come up for me, without trying to change or defend against them”) and item 22 (“I can take thoughts and feelings as they come, without attempting to control or avoid them”).

In supplemental analyses, we attempted ad-hoc model revisions to improve local misfit. Unfortunately, it appeared that acceptable model fit for all 23 items could only be salvaged at the expense of parsimony and interpretability (i.e., after adding 6 cross-loadings and 4 residual covariances to capture unintended multidimensionality; see supplemental materials Table S1 including *Mplus* syntax). We opted *not* to pursue these data-driven and potentially sample-idiosyncratic model modifications. Instead, we elected to pursue an alternative strategy—we developed a more cohesive short form by eliminating items with redundant content, inconsistent wording direction within a factor, or that required familiarity with ACT concepts (e.g., defusion, self-as-context). Specifically, we removed items 6, 18, 13, 20, and 22 so that openness to experience would be measured only by negative items, we did not remove any items from behavioral awareness, and we removed items 5, 14, and 21 from valued action. In the resulting short-form (named the *CompACT-15*; items are provided in supplementary materials Table S2), each factor is now measured by five items worded in the same direction (and all negatively-worded items are still reverse-coded so that higher factor scores always represented greater PF).

We found good fit for a three-factor model with simple structure (i.e., without any cross-loadings or residual covariances) for the CompACT-15 items by most indices using both CFA, $\chi^2(87) = 175, p < .001, TLI = .964, CFI = .970, RMSEA = .046 [.036, .055], SRMR = .046$, and IFA, $\chi^2(87) = 383, p < .001, TLI = .973, CFI = .978, RMSEA = .084 [.075, .092], SRMR = .039$. Standardized loadings ranged from .66 to .89; all parameters are reported in the supplemental materials Table S3 for the CFA model and Table S4 for the IFA model. The openness to experience and behavioral awareness factors were strongly related ($r \approx .7$), but valued action was less related to openness to experience ($r \approx .1$) or behavioral awareness ($r \approx .4$), indicating that the factors appeared to be practically distinguishable in the present sample.

Reliability and Item Parameters

Model-based reliability of the CompACT-15 was examined in three ways, two of which were in the context of a CFA model that assumes constant reliability across each latent factor, and one in the context of an IFA/IRT model in which reliability varies across each latent factor. First, we computed omega reliability for each dimension (Brown, 2015; McDonald, 1999). Like alpha, omega also indexes the reliability of a unidimensional sum score. But unlike alpha, omega includes differences in the item–factor relations via the item factor loadings and residual variances. Omega reliability was .850 for openness to experience, .929 for behavioral awareness, and .851 for valued action, indicating good reliability despite only five items per dimension. Second, we computed factor score

reliability as the proportion of true trait variance for each factor score over the total variance (that also includes unreliability). Given the factor variance fixed to 1 and the constant estimated factor score standard error (SE), factor score reliability can be computed as $1 / (1 + SE^2)$. The result was reliability = 0.886 (SE = 0.359) for openness to experience, reliability = 0.940 (SE = 0.253) for behavioral awareness, and reliability = 0.878 (SE = 0.372) for valued action, indicating good factor score reliability as well.

However, closer inspection indicates that these constant reliability estimates may be too optimistic for respondents with extreme factor scores. Figure 1 shows CFA-predicted item responses for factor scores within ± 3 SD of the mean = 0 (see also Fernando, 2009). A linear slope for the latent factor does not appear adequate for many items, as evidenced by their impossible predicted responses (i.e., below 0 or above 6), which occurred for openness to experience factor scores outside approximately ± 2 SD, for behavioral awareness factor scores below -2 SD or above 1.3 SD, and for valued action factor scores above 1.6 SD.

To address this concern, we turn to the results of the IFA/IRT models, whose nonlinear slopes constrain the predicted item responses to their possible discrete options. Figure 2 depicts each item's IRT a_j discrimination estimate (top panel), coupled with the range of the latent factor covered by its IRT b_{jk} item difficulty estimates (bottom panel). In general, a finding of response categories with well-differentiated difficulty levels would indicate that each option contributes unique information to the measurement of the latent factor. In the present case, the lack of differentiation of some response categories suggests that fewer categories may be sufficient, but differentially so across dimensions. For instance, little differentiation was found between responses of 3 and 4 for the negatively-worded openness to experience and behavioral awareness items, but between 0 and 1 instead for the positively-worded valued action items.

These item characteristics in turn can be used to compute the total amount of information provided by the items for a more fine-grained picture of reliability across each latent factor (McDonald, 1999). Figure 3 (top panel) depicts total (test) information using a reliability metric of 0–1 as computed as: information / (information + 1); the bottom panel shows standard errors computed as: $1 / \text{SQRT}(\text{information})$ at each factor score. Each dimension had reliability $> .80$ for persons with mid-range factor scores (i.e., -2.6 SD to 1.7 SD), with relatively greater reliability for behavioral awareness in this mid-range. Reliability was relatively higher at lower factor scores for valued action, as well as at higher factor scores for openness to experience.

Measurement Stability over Time

We then examined the stability of the CFA and IFA model parameters across three one-month intervals (T0, T1, and T2) by testing longitudinal measurement invariance, in which the unstandardized factor loadings, intercepts or thresholds, and residual variances were constrained equal over time in successive models. We examined the resulting decreases in global fit using nested model comparisons (via rescaled likelihood ratio tests for CFA models with MLR estimation, or via DIFFTEST for IFA models with WLSMV estimation), as well as changes in the other global model fit statistics (Chen, 2007). We also examined modification indices at $\alpha = .01$ for potential noninvariance of individual parameters

(i.e., to prevent salient differences from being washed out in omnibus comparisons). We have summarized the primary findings below and provide complete model fit results in supplemental materials Table S5, as well as *Mplus* syntax for the final invariance models in supplemental materials Table S7.

We began with a configural model that included nine correlated factors (i.e., three factors at each of three occasions) and residual covariances for the same item across occasions; all other parameters were estimated as described for our T0 data above. The configural invariance model had adequate fit by most indices using both CFA, $\chi^2(864) = 1271$, $p < .001$, TLI = .951, CFI = .957, RMSEA = .031 [.027, .035], SRMR = .056, and IFA, $\chi^2(864) = 1899$, $p < .001$, TLI = .966, CFI = .970, RMSEA = .050 [.047, .053], SRMR = .049. We then estimated a metric (weak) invariance model in which all factor loadings were constrained equal over time; factor variances were fixed to 1 at T0 for identification and estimated at T1 and T2. The CFA metric invariance model did not fit worse than the configural invariance model, indicating equivalent relations of the items to their factors over time; the same was true of the IFA metric invariance model after allowing the loading for item 11 to be less related to openness to experience at T1.

We then estimated a scalar (strong) invariance model in which the CFA intercepts or IFA thresholds for loading-invariant items were also constrained equal over time; factor means were fixed to 0 at T0 for identification and estimated at T1 and T2. Both the CFA and IFA scalar invariance models did not fit worse than their respective metric invariance models, indicating equivalent expected item responses at $F_s = 0$ over time. Finally, we estimated a strict invariance model in which the residual variances were also constrained equal over time; in IFA this required a prior comparison model in which the residual variances of the invariant items were estimated at T1 and T2 instead of fixed to 1. The CFA residual invariance model did not fit worse than the scalar invariance model, indicating equivalent amounts of unique item variance over time; the same was true of the IFA residual invariance model after allowing item 15 (measuring openness to experience) to have greater unique variation at T1 and T2 than at T0. In summary, our results indicated that most of the item parameters functioned equivalently across three occasions—such stability is expected in a non-treatment setting in which true change or reactivity is unlikely.

Construct Validity

Finally, we used CFA and IFA to examine the extent to which the CompACT-15 latent factors showed the expected relations with the latent factors measured by the BRS, AAQ-II, IUS-12, and K10 in our T0 data. As detailed in supplemental materials Table S6, we first examined the dimensionality of each measure's items and implemented theoretically interpretable remedies of misfit. More specifically, BRS resilience was best captured by separate factors for positively- and negatively-worded items ($r \approx .7$). AAQ-II psychological inflexibility was best captured by a single factor and two item residual covariances. IUS-12 prospective and inhibitory intolerance of uncertainty were best captured by separate factors ($r \approx .8$) and an uncorrelated random intercept factor (i.e., with equal loadings) for three items' residuals. K10 psychological distress was best captured by a single factor and five item residual covariances; an alternative two-factor structure also fit well but yielded factors

with $r > .9$ (which were thus not practically different). The combined nine-factor model had adequate fit by most indices using both CFA, $\chi^2(1131) = 1918, p < .001, TLI = .945, CFI = .950, RMSEA = .038 [.035, .041], SRMR = .058$, and IFA, $\chi^2(1131) = 2646, p < .001, TLI = .972, CFI = .974, RMSEA = .053 [.050, .055], SRMR = .046$. We provide *Mplus* syntax for this nine-factor model in supplemental materials Table S8.

Table 2 provides the latent factor correlations using a CFA measurement model in the upper diagonal and using an IFA measurement model in the lower diagonal. As predicted, the CompACT-15's three PF factors correlated positively with resilience, the extent of which was related to congruence in wording. That is, PF openness to experience and behavioral awareness (as measured by negatively-worded items) had stronger relations with the factor of negatively-worded resilience items, whereas PF valued action (as measured by positively-worded items) had a stronger relation with the factor of positively-worded resilience items. Also as predicted, PF correlated negatively with psychological inflexibility, prospective and inhibitory intolerance of uncertainty, and psychological distress; the strongest correlations were again found for the congruently negatively-worded PF factors for openness to experience and behavioral awareness.

Discussion

Summary of Findings

The CompACT-23 is a promising multidimensional measure of PF whose evaluation has thus far been limited to exploratory factor analysis, coefficient alpha reliability, and the use of sum scores in evaluating its construct validity. The present study's psychometric evaluation expands on prior work by using two types of confirmatory latent variable measurement models (CFA and IFA/IRT) in a non-clinical longitudinal sample of adults. To address the item design issues that may have led to a lack of fit of the hypothesized factor structure, we propose a new short-form, the CompACT-15. It measures PF with five items worded in the same direction within each of three dimensions: *openness to experience* (negative wording, reverse-coded), *behavioral awareness* (negative wording, reverse-coded), and *valued action* (positive wording).

Converging evidence from CFA and IFA/IRT models indicated good fit of a correlated three-factor model with simple structure—without any cross-loadings or residual covariances. However, the CFA-predicted item responses indicated that a linear relation with the latent factors may not be adequate for these ordinal items, in which case the high levels of reliability obtained from the CFA model (i.e., through omega and factor score reliability) would be too optimistic. We addressed this limitation by using nonlinear IRT/IFA models, which constrain the predicted item responses to their discrete options and generate corresponding test information functions.

The IRT/IFA models yielded additional results not applicable in CFA models. First, test information functions describe how reliability varies across levels of the latent factor based on the item parameters. As shown in Figure 3, our results suggest that the openness to experience and behavioral awareness items will best measure respondents of mid-range abilities, whereas the valued action items will be comparatively more sensitive

for respondents of lower ability. These functions provide direction as to how items might be modified or added in order to improve reliability in targeted areas (i.e., by adding items of greater difficulty for valued action). Second, the IRT item difficulty parameters also indicate the utility of each response category. As shown by the closely-spaced symbols in Figure 2, our results suggest that seven categories may not be needed to adequately differentiate responses across latent factor levels. However, how to best modify the response options remains an open question, given that a different set of choices would likely be needed for the negatively-worded items (for which lower responses were more prevalent) than for the positively-worded options (for which higher responses were more prevalent). Future research would have to weigh the relative benefits of a simpler response format for a given item against the additional complexity of using different formats across items.

The vast majority of CompACT-15 item parameters demonstrated measurement stability (i.e., measurement invariance, or lack of differential item functioning) across three occasions over a two-month period. While we did not expect systematic changes in PF in the present non-treatment sample, such measurement stability is a necessary prerequisite for the measurement of true changes in PF over time (e.g., in response to ACT or other treatments targeting PF). Lastly, in structural equation models with nine factors at the baseline occasion, we found the expected correlations to support construct validity (positive relations with resilience; negative relations with psychological inflexibility, prospective and inhibitory intolerance of uncertainty, and psychological distress), although the size of these relations was related to wording congruence.

Strengths and Limitations

A notable strength of the present study is its use of two different types of confirmatory measurement models (CFA and IFA/IRT). This strategy ensures the robustness of our findings across model specification (linear versus nonlinear factor prediction). One drawback of WLSMV estimation for IFA models is that it assumes incomplete item responses are missing completely at random (MCAR). This is in contrast to MLR estimation for CFA models, which assumes item responses are missing at random (MAR) instead. Given that missingness rates per occasion were 0–1% across items, missing data were not likely to have biased any within-occasion results. In contrast, nonrandom attrition at the T1 and T2 occasions may not have been adequately captured in the invariance analyses. We found that persons who identified as Hispanic/Latino(a/x) or as lesbian, gay, or bisexual were less likely to complete follow-up surveys than those who identified as non-Hispanic/Latino(a/x) or heterosexual, which may reflect the disproportionate impact of COVID-19 on these communities (Fortuna et al., 2020; Salerno et al., 2020). Thus, while it appears an MCAR assumption is untenable, we note that it is not possible to assess the influence of unmeasured differences in non-returning participants (i.e., missing not at random instead).

A possible weakness of the current study is its recruitment of respondents using MTurk, a platform recently subject to scrutiny due to increased activity from “bots” (computer programs completing HITs) and “farmers” (people using server farms to bypass location restrictions). Our study followed recommendations for best practices (Chmielewski & Kucker, 2020) to safeguard against invalid responding. However, we also acknowledge the

reported influx of new MTurk respondents during the COVID-19 pandemic who are less cognitively reflective, White, Democratic, and experienced compared to respondents prior to the pandemic (Arechar & Rand, 2021). Although it is unclear what effects this more diverse but less attentive subject pool may have on data quality, demographic diversity has also been named a significant strength of MTurk relative to student or other online samples (Buhrmester et al., 2016).

Notably, the CompACT-15 was less related to distress than the seven-item AAQ-II, which had very high associations with latent distress (CFA $r = .93$, IFA $r = .95$), in line with previous AAQ-II research. We also note that several recent measures (Psy-Flex, Gloster et al., 2021; Open and Engaged State Questionnaire, Benoy et al., 2019; Personalized Psychological Flexibility Index, Kashdan et al., 2021) were not available at the time of our data collection, and thus were not examined as part of the CompACT-15's nomological network. Finally, the brevity of the CompACT-15, relative to longer multidimensional PF assessments (e.g., the 60-item MPFI), is an important strength. We reduced the original items into a smaller, less redundant set without sacrificing reliability while improving model fit. Nevertheless, the current evidence suggests that the CompACT-15 can be used for efficient yet reliable measurement of PF.

Future Directions

As is always recommended (Brown, 2015), it will be important to evaluate the extent to which the fit of the revised factor structure (15-item short form) replicates in additional similar samples. Further, given our non-clinical sample of U.S. adults, it will also be important to examine to what extent our findings about the CompACT-15's dimensionality, reliability, measurement equivalence, and construct validity may generalize to other populations. For example, measurement equivalence of the CompACT-15 item properties should be examined across groups of persons receiving psychotherapy (especially ACT) and those who are not, as well as across groups just starting psychotherapy and those who have been in treatment long-term. Such research may better elucidate whether and how prior exposure to PF-related content may affect one's items responses and thus how PF item characteristics may differ across clinical/nonclinical samples. Furthermore, given that we examined measurement stability of the CompACT-15 at relatively brief one-month intervals, future studies should explore measurement equivalence of the CompACT-15 items across longer temporal contexts.

Another open question is to what extent our findings of the CompACT-15's three-factor dimensionality, factor reliability, and construct validity replicate across cultural groups, in which items may perform differently based on different cultural norms or values. Given that ACT was developed in the United States but draws from Eastern ideologies such as Buddhism, future studies could examine the CompACT-15's measurement stability of across groups of White Americans and East Asians. Interestingly, some cross-cultural work was reported by Trindade et al. (2021), who independently proposed a similar CompACT short form using a sample of Portuguese respondents who completed a translation of the original 23 items. Based on EFA, they proposed dropping the same five items measuring openness to experience as we did in the present study. They did not remove any valued

action items, in contrast to our removal of three items (5, 14, 21) on the basis of problematic redundancy in content. Using CFA, they also reported partial metric invariance across a second sample of British respondents, although they did not report tests of scalar or residual invariance. Finally, they examined subscale correlations for several measures and found similar magnitudes of negative relations with the AAQ-II. Future studies would benefit from the use of confirmatory measurement models to more rigorously examine the psychometric properties of CompACT items across additional populations.

The results of the current study also highlight the importance of wording direction in items that measure PF—in particular, the assumption that items written to measure *in*flexibility will instead measure flexibility after being reverse-coded as is assumed for the AAQ-II items. Negatively-worded items have been theorized to reduce acquiescence and mindless responding (DeVellis, 2003), resulting in a better measure of a single construct when combined with positively-worded items. In contrast, the present results bolster previous findings indicating that negatively-worded items do not function as the unidimensional opposite of positively-worded items (DiStefano & Motl, 2006). Within the CompACT-15, we observed lower correlations of the positively-worded valued action factor with the negatively-worded openness to experience and behavioral awareness factors. We also observed larger correlations of the CompACT-15 factors with other factors measured by similarly worded items. Thus, our results do not support the conceptualization of psychological inflexibility as the inverse of PF, and so this remains a potential criticism of the CompACT-15 item structure. Future work should assess the extent PF and psychological inflexibility differ in their prediction of psychological outcomes.

Conclusion

Findings from two types of confirmatory measurement models in a nonclinical U.S. adult sample support the psychometric utility of the proposed CompACT-15 short-form as a measure of PF. We found support for its three-factor dimensionality, acceptable reliability over large ranges of its three latent factors, and measurement equivalence of its item characteristics over time. We also found evidence for the validity of the CompACT-15 as a measure of PF processes that relate as expected to—but are still distinguishable from—theoretically relevant constructs. As clinical science continues to move towards transdiagnostic and process-based therapies, the CompACT-15 appears to be a concise yet psychometrically sound instrument with strong potential to contribute to research, treatment, and prevention efforts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Arch JJ, Fishbein JN, Finkelstein LB, & Luoma JB (2022). Acceptance and Commitment Therapy (ACT) Processes and Mediation: Challenges and How to Address Them. *Behavior Therapy*, S0005789422000892. 10.1016/j.beth.2022.07.005
- Arechar AA, & Rand DG (2021). Turking in the time of COVID. *Behavior Research Methods*, 53(6), 2591–2595. 10.3758/s13428-021-01588-4 [PubMed: 33963495]

- Benoy C, Knitter B, Knellwolf L, Doering S, Klotsche J, Gloster AT (2019) Assessing psychological flexibility: Validation of the Open and Engaged State Questionnaire. *Journal of Contextual Behavioral Science*, 12, 253–260. 10.1016/j.jcbs.2018.08.005
- Bond FW, Hayes SC, Baer RA, Carpenter KM, Guenole N, Orcutt HK, Waltz T, & Zettle RD (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. 10.1016/j.beth.2011.03.007 [PubMed: 22035996]
- Brandon S, Pallotti C, & Jog M (2021). Exploratory Study of Common Changes in Client Behaviors Following Routine Psychotherapy: Does Psychological Flexibility Typically Change and Predict Outcomes? *Journal of Contemporary Psychotherapy*, 51(1), 49–56. 10.1007/s10879-020-09468-2
- Brown TA (2015). *Confirmatory factor analysis for applied research*. Guilford.
- Buhrmester M, Kwang T, & Gosling SD (2016). Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality data? In Kazdin AE (Ed.), *Methodological issues and strategies in clinical research* (pp. 133–139). American Psychological Association. 10.1037/14805-009
- Carleton RN, Norton MAPJ, & Asmundson GJG (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, 21(1), 105–117. 10.1016/j.janxdis.2006.03.014 [PubMed: 16647833]
- Chen FF (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. 10.1080/10705510701301834
- Chmielewski M, & Kucker SC (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, 11(4), 464–473. 10.1177/1948550619875149
- Cherry KM, Hoeven EV, Patterson TS, & Lumley MN (2021). Defining and measuring “psychological flexibility”: A narrative scoping review of diverse flexibility and rigidity constructs and perspectives. *Clinical Psychology Review*, 84, 101973. 10.1016/j.cpr.2021.101973 [PubMed: 33550157]
- Ciarrochi J, Bilich L, & Godsell C (2010). Psychological flexibility as a mechanism of change in acceptance and commitment therapy. *Assessing mindfulness and acceptance processes in clients: Illuminating the theory and practice of change*, 2010, 51–75.
- DeVellis RF (2016). *Scale development: Theory and applications* (Vol. 26). Sage publications.
- DiStefano C, & Motl RW (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(3), 440–464. 10.1207/s15328007sem1303_6
- Doorley JD, Goodman FR, Kelso KC, & Kashdan TB (2020). Psychological flexibility: What we know, what we do not know, and what we think we know. *Social and Personality Psychology Compass*, 14(12), 1–11. 10.1111/spc3.12566
- Ferrando PJ (2009). Difficulty, discrimination, and information indices in the linear factor analysis model for continuous item responses. *Applied Psychological Measurement*, 33(1), 9–24.
- Fledderus M, Bohlmeijer ET, Fox J-P, Schreurs KMG, & Spinhoven P (2013). The role of psychological flexibility in a self-help acceptance and commitment therapy intervention for psychological distress in a randomized controlled trial. *Behaviour Research and Therapy*, 51(3), 142–151. 10.1016/j.brat.2012.11.007 [PubMed: 23337183]
- Fledderus M, Oude Voshaar MAH, ten Klooster PM, & Bohlmeijer ET (2012). Further evaluation of the psychometric properties of the Acceptance and Action Questionnaire–II. *Psychological Assessment*, 24(4), 925–936. 10.1037/a0028200 [PubMed: 22545700]
- Flynn MK, Berkout OV, & Bordieri MJ (2016). Cultural considerations in the measurement of psychological flexibility: Initial validation of the Acceptance and Action Questionnaire–II among Hispanic individuals. *Behavior Analysis: Research and Practice*, 16(2), 81–93. 10.1037/bar0000035
- Fortuna LR, Tolou-Shams M, Robles-Ramamurthy B, & Porche MV (2020). Inequity and the disproportionate impact of COVID-19 on communities of color in the United States: The need for a trauma-informed social justice response. *Psychological Trauma: Theory, Research, Practice, and Policy*, 12(5), 443–445. 10.1037/tra0000889

- Francis AW, Dawson DL, & 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. 10.1016/j.jcbs.2016.05.003
- 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. 10.1037/a0023242 [PubMed: 21534697]
- Gloster AT, Block VJ, Klotsche J, Villanueva J, Rinner MTB, Benoy C, Walter M, Karekla M, & Bader K (2021). Psy-Flex: A contextually sensitive measure of psychological flexibility. *Journal of Contextual Behavioral Science*, 22, 13–23. 10.1016/j.jcbs.2021.09.001
- Gloster AT, Klotsche J, Chaker S, Hummel KV, & Hoyer J (2011). Assessing psychological flexibility: What does it add above and beyond existing constructs? *Psychological Assessment*, 23(4), 970–982. 10.1037/a0024135 [PubMed: 21767025]
- Hayes SC, Levin ME, Plumb-Villardaga J, Villatte JL, & 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. 10.1016/j.beth.2009.08.002 [PubMed: 23611068]
- Hayes SC, Luoma JB, Bond FW, Masuda A, & Lillis J (2006). Acceptance and Commitment Therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, 44(1), 1–25. 10.1016/j.brat.2005.06.006 [PubMed: 16300724]
- Hayes SC, Strosahl KD, Wilson KG (1999). *Acceptance and Commitment Therapy: An experiential approach to behavior change*. New York, NY: Guilford Press.
- Hu L, & Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. 10.1080/10705519909540118
- Hughes LS, Clark J, Colclough JA, Dale E, & McMillan D (2017). Acceptance and Commitment Therapy (ACT) for chronic pain: A systematic review and meta-analyses. *The Clinical Journal of Pain*, 33(6), 552–568. 10.1097/AJP.0000000000000425 [PubMed: 27479642]
- Johnson J, & Wood AM (2017). Integrating positive and clinical psychology: Viewing human functioning as continua from positive to negative can benefit clinical assessment, interventions and understandings of resilience. *Cognitive Therapy and Research*, 41(3), 335–349. 10.1007/s10608-015-9728-y
- Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand S-LT, Walters EE, & Zaslavsky AM (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32(6), 959–976. 10.1017/S0033291702006074 [PubMed: 12214795]
- Kollman DM, Brown TA, & Barlow DH (2009). The construct validity of acceptance: A multitrait-multimethod investigation. *Behavior Therapy*, 40(3), 205–218. 10.1016/j.beth.2008.06.002 [PubMed: 19647522]
- Kroska EB, Roche AI, Adamowicz JL, & Stegall MS (2020). Psychological flexibility in the context of COVID-19 adversity: Associations with distress. *Journal of Contextual Behavioral Science*, 18, 28–33. [PubMed: 32837889]
- Lappalainen P, Pakkala I, Strömmer J, Sairanen E, Kaipainen K, & Lappalainen R (2021). Supporting parents of children with chronic conditions: A randomized controlled trial of web-based and self-help ACT interventions. *Internet Interventions*, 24, 100382. 10.1016/j.invent.2021.100382 [PubMed: 33816128]
- Levin ME, Haeger J, & Cruz RA (2019). Tailoring acceptance and commitment therapy skill coaching in the moment through smartphones: Results from a randomized controlled trial. *Mindfulness*, 10(4), 689–699. 10.1007/s12671-018-1004-2
- McDonald RP (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McNeish D (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. 10.1037/met0000144 [PubMed: 28557467]
- Millsap RE (2012). *Statistical approaches to measurement invariance*. Routledge.
- Muthén BO (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115–132. 10.1007/BF02294210

- Muthén LK & Muthén BO (1998–2017). Mplus user's guide. (8th ed.). Muthén & Muthén.
- Muthén BO & Muthén LK, & Asparouhov T (2015). Estimator choices with categorical outcomes. Retrieved from <http://www.statmodel.com/download/EstimatorChoices.pdf>.
- Ong CW, Lee EB, Levin ME, & Twohig MP (2019). A review of AAQ variants and other context-specific measures of psychological flexibility. *Journal of Contextual Behavioral Science*, 12, 329–346. 10.1016/j.jcbs.2019.02.007
- Ong CW, Pierce BG, Petersen JM, Barney JL, Fruge JE, Levin ME, & Twohig MP (2020). A psychometric comparison of psychological inflexibility measures: Discriminant validity and item performance. *Journal of Contextual Behavioral Science*, 18, 34–47. 10.1016/j.jcbs.2020.08.007
- Ong CW, Pierce BG, Woods DW, Twohig MP, & Levin ME (2019). The Acceptance and Action Questionnaire – II: An item response theory analysis. *Journal of Psychopathology and Behavioral Assessment*, 41(1), 123–134. 10.1007/s10862-018-9694-2
- Paek I, Cui M, Gübes NO, & Yang Y (2018). Estimation of an IRT model by Mplus for dichotomously scored responses under different estimation methods. *Educational and Psychological Measurement*, 78(4), 569–588. [PubMed: 30147117]
- Petersen JM, Krafft J, Twohig MP, & Levin ME (2021). Evaluating the open and engaged components of Acceptance and Commitment Therapy in an online self-guided website: Results from a pilot trial. *Behavior Modification*, 45(3), 480–501. 10.1177/0145445519878668 [PubMed: 31550904]
- Pyzskowska A (2020). Personality predictors of self-compassion, ego-resiliency and psychological flexibility in the context of quality of life. *Personality and Individual Differences*, 161, 109932. 10.1016/j.paid.2020.109932
- Rocheft C, Baldwin AS, & Chmielewski M (2018). Experiential avoidance: An examination of the construct validity of the AAQ-II and MEAQ. *Behavior Therapy*, 49(3), 435–449. 10.1016/j.beth.2017.08.008 [PubMed: 29704971]
- Rolffs JL, Rogge RD, & Wilson KG (2018). Disentangling components of flexibility via the Hexaflex model: Development and validation of the Multidimensional Psychological Flexibility Inventory (MPFI). *Assessment*, 25(4), 458–482. 10.1177/1073191116645905 [PubMed: 27152011]
- Salerno JP, Williams ND, & Gattamorta KA (2020). LGBTQ populations: Psychologically vulnerable communities in the COVID-19 pandemic. *Psychological Trauma: Theory, Research, Practice, and Policy*, 12(S1), S239–S242.
- Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, & Bernard J (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine*, 15(3), 194–200. 10.1080/10705500802222972 [PubMed: 18696313]
- Stockton D, Kellett S, Berrios R, Sirois F, Wilkinson N, & Miles G (2019). Identifying the Underlying Mechanisms of Change During Acceptance and Commitment Therapy (ACT): A Systematic Review of Contemporary Mediation Studies. *Behavioural and Cognitive Psychotherapy*, 47(3), 332–362. 10.1017/S1352465818000553 [PubMed: 30284528]
- Strosahl KD, Robinson PJ, & Gustavsson T (2012). *Brief interventions for radical change: Principles and practice of focused acceptance and commitment therapy*. New Harbinger Publications.
- Takane Y, & de Leeuw J (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 393–408. 10.1007/BF02294363
- Trindade IA, Ferreira NB, Mendes AL, Ferreira C, Dawson D, & Golijani-Moghaddam N (2021). Comprehensive assessment of Acceptance and Commitment Therapy processes (CompACT): Measure refinement and study of measurement invariance across Portuguese and UK samples. *Journal of Contextual Behavioral Science*, 21, 30–36. 10.1016/j.jcbs.2021.05.002
- Tyndall I, Waldeck D, Pancani L, Whelan R, Roche B, & Dawson DL (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. 10.1016/j.jcbs.2018.09.005
- Wersebe H, Lieb R, Meyer AH, Hofer P, & Gloster AT (2018). The link between stress, well-being, and psychological flexibility during an Acceptance and Commitment Therapy self-help intervention. *International Journal of Clinical and Health Psychology*, 18(1), 60–68. 10.1016/j.ijchp.2017.09.002 [PubMed: 30487911]

Wolgast M (2014). What does the Acceptance and Action Questionnaire (AAQ-II) really measure? Behavior Therapy, 45(6), 831–839. 10.1016/j.beth.2014.07.002 [PubMed: 25311291]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Public Significance Statement

Psychological flexibility (PF) involves present-moment awareness and the pursuit of one's values even when facing difficult situations and emotions. Using two kinds of confirmatory factor measurement models, this study examined the psychometric properties of a measure of PF: the Comprehensive assessment of Acceptance and Commitment Therapy processes (Francis et al., 2016). Results suggest a more psychometrically sound short form—the CompACT-15—has strong potential to contribute to research and intervention efforts.

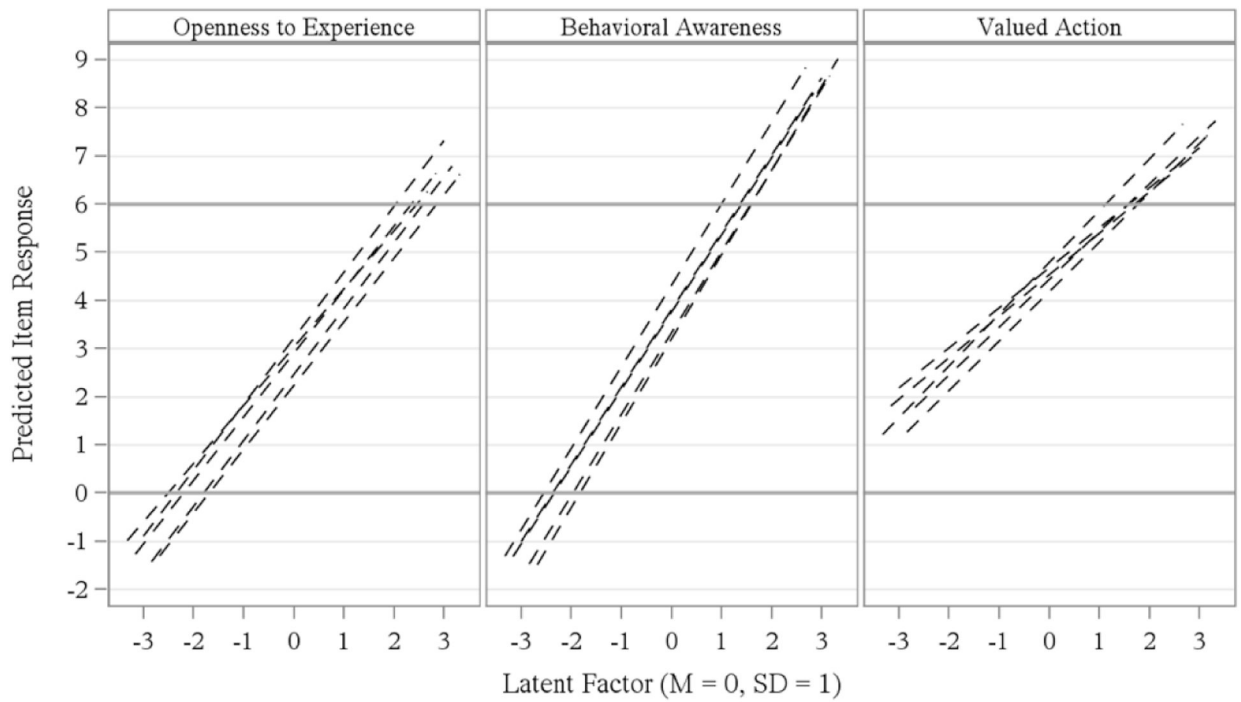


Figure 1.
Effect of CompACT-15 Item Intercept and Slope Parameters on Item Responses Predicted from Confirmatory Factor Analysis

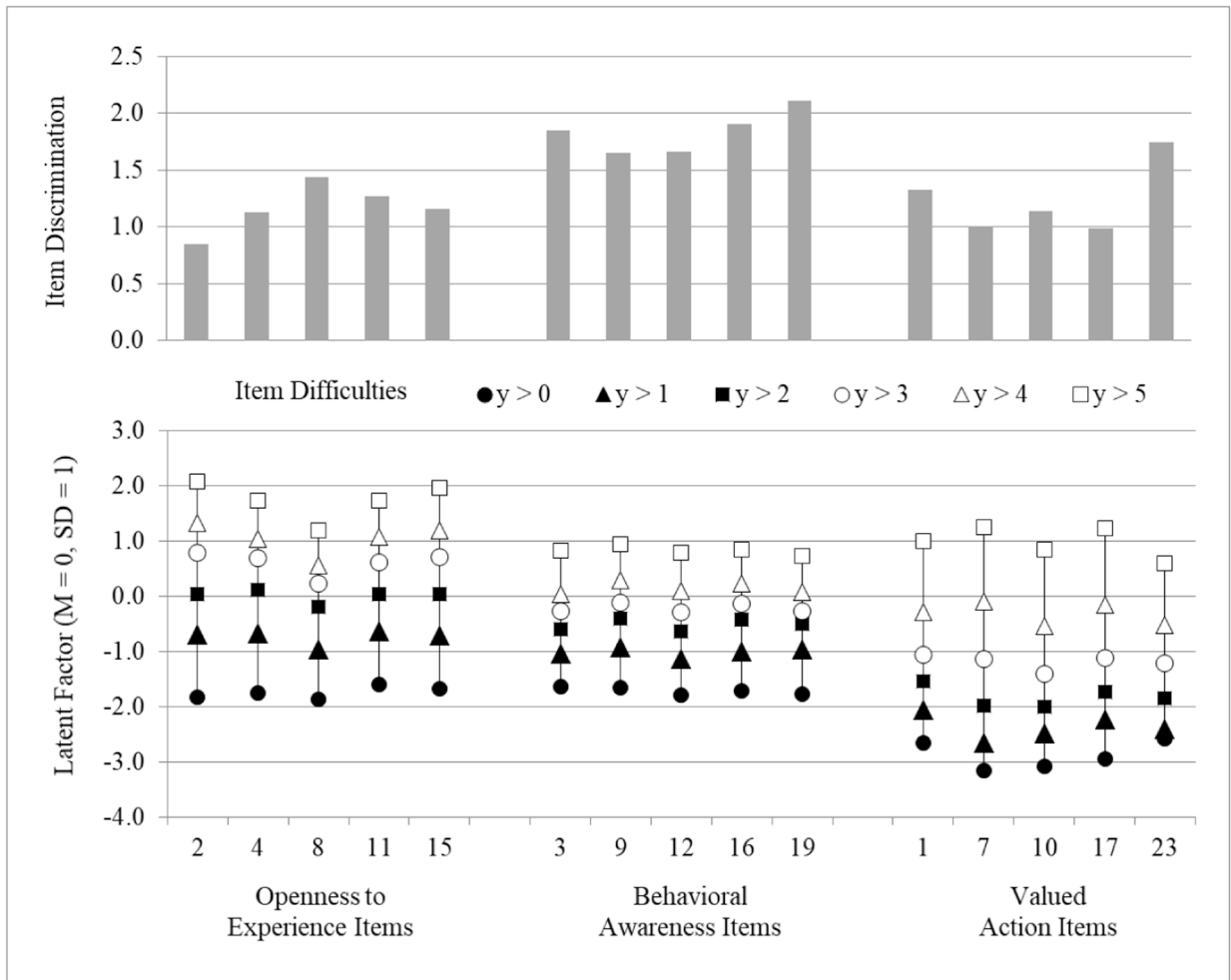


Figure 2. Illustration of CompACT-15 Item Discrimination and Difficulty Parameters Obtained Through Item Response Theory Modeling

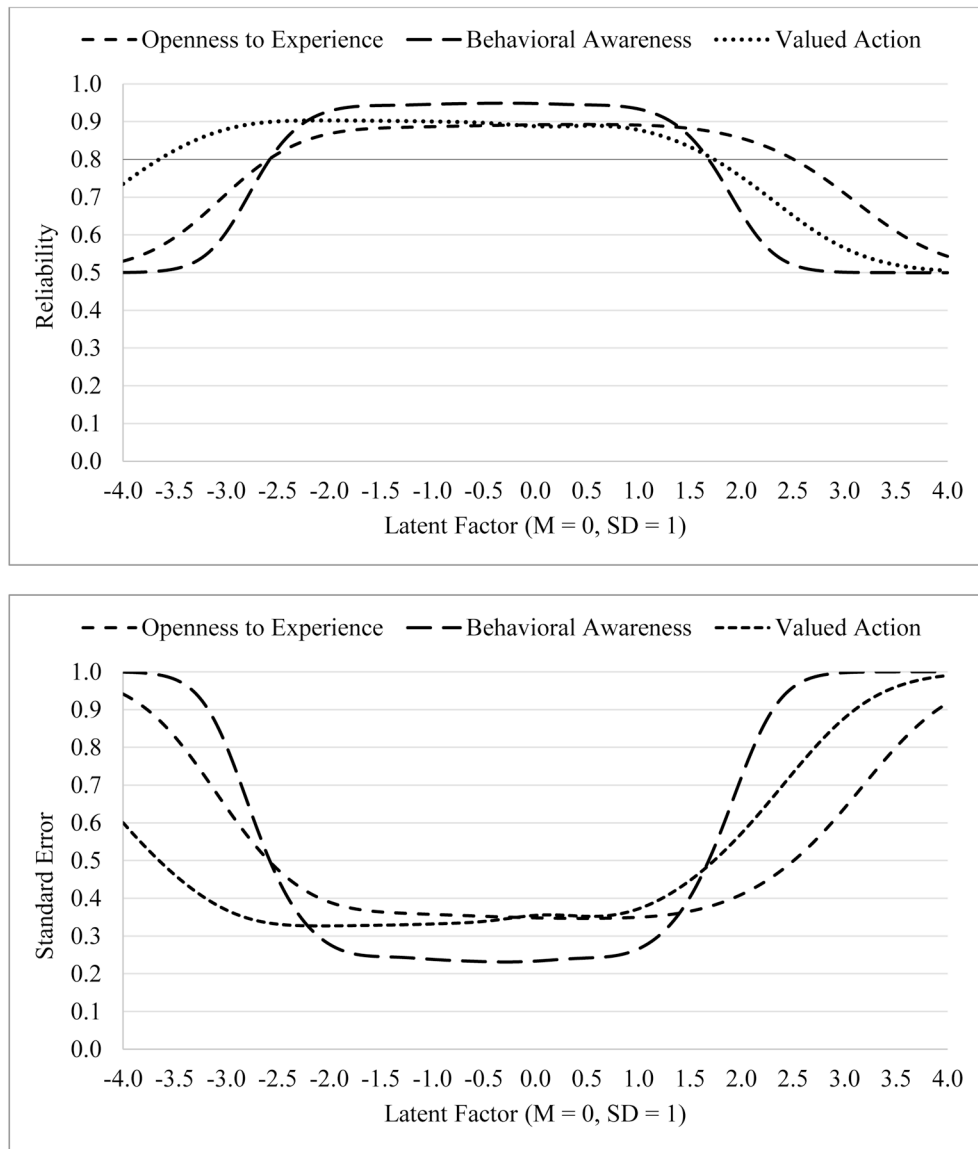


Figure 3. Test Information Functions (Top) and Factor Score Standard Errors (Bottom) by Latent Factor for the CompACT-15 Items

Table 1
 Descriptive Statistics for CompACT-23 Item Responses at T0 (* Indicates Item Removed in Creating the CompACT-15)

CompACT-23 Item Number and Factor	N	Mean	SD	Median	% Response Categories 1-2	% Response Categories 3-5	% Response Category 6-7
Openness to Experience							
2. One of my big goals is to be free from painful emotions	482	3.37	1.82	2	32.78	47.51	19.71
4. I try to stay busy to keep thoughts or feelings from coming	483	3.32	1.81	2	30.43	47.83	21.74
6. I get so caught up in my thoughts that I am unable to do the things that I most want to do*	482	2.65	2.00	3	23.44	37.97	38.59
8. I tell myself that I shouldn't have certain thoughts	481	2.79	1.90	3	21.62	45.74	32.64
11. I go out of my way to avoid situations that might bring difficult thoughts, feelings, or sensations	484	3.32	1.79	2	30.79	49.38	19.83
13. I am willing to fully experience whatever thoughts, feelings and sensations come up for me, without trying to change or defend against them*	484	3.96	1.47	4	6.40	52.27	41.32
15. I work hard to keep out upsetting feelings	483	3.36	1.72	2	29.61	51.97	18.43
18. Even when something is important to me, I'll rarely do it if there is a chance it will upset me*	484	2.66	1.93	3	21.90	42.77	35.33
20. Thoughts are just thoughts - they don't control what I do*	483	3.78	1.64	4	11.59	48.65	39.75
22. I can take thoughts and feelings as they come, without attempting to control or avoid them*	485	4.00	1.53	4	8.66	45.98	45.36
Behavioral Awareness							
3. I rush through meaningful activities without being really attentive to them	484	2.24	1.96	4	17.77	33.68	48.55
9. I find it difficult to stay focused on what's happening in the present	482	2.50	1.99	4	21.58	37.76	40.66
12. Even when doing the things that matter to me, I find myself doing them without paying attention	485	2.20	1.93	4	16.29	36.70	47.01
16. I do jobs or tasks automatically, without being aware of what I'm doing	480	2.41	1.96	4	18.75	38.96	42.29
19. It seems I am "running on automatic" without much awareness of what I'm doing	483	2.25	1.98	4	19.05	33.54	47.41
Valued Action							
1. I can identify the things that really matter to me in life and pursue them	483	4.43	1.39	5	4.97	35.82	59.21
5. I act in ways that are consistent with how I wish to live my life*	482	4.60	1.27	5	3.53	34.65	61.83
7. I make choices based on what is important to me, even if it is stressful	483	4.39	1.27	5	2.90	44.31	52.80
10. I behave in line with my personal values	484	4.67	1.26	5	3.10	30.99	65.91
14. I undertake things that are meaningful to me, even when I find it hard to do so*	483	4.24	1.45	5	6.00	42.24	51.76
17. I am able to follow my long term plans including times when progress is slow	483	4.34	1.41	5	5.80	39.54	54.66
21. My values are really reflected in my behavior*	483	4.50	1.36	5	4.14	34.78	61.08
23. I can keep going with something when it's important to me	485	4.75	1.23	5	1.86	30.72	67.42

Note. The frequency of response to each item was summarized using three columns for brevity only; no responses were collapsed across categories for any analyses. Items reprinted from “The development and validation of the Comprehensive assessment of Acceptance and Commitment Therapy processes (CompACT).” Journal of Contextual Behavioral Science, 5(3), Francis, A. W., Dawson, D. L., & Golijani-Moghaddam, N., 134–145, 2016, with permission from Elsevier.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Latent Variable Correlations using Confirmatory Factor Analysis (Upper Diagonal) and Item Factor Analysis (Lower Diagonal)

Table 2

Latent Factum	1	2	3	4	5	6	7	8	9
1. Openness to Experience (CompACT-15)	–	.72	.12	.25	.67	–.73	–.49	–.67	–.65
2. Behavioral Awareness (CompACT-15)	.74	–	.35	.16	.66	–.78	–.32	–.64	–.75
3. Valued Action (CompACT-15)	.16	.45	–	.60	.37	–.32	–.13	–.40	–.36
4. Brief Resilience Scale Positively Worded Items (BRS)	.25	.23	.63	–	.67	–.34	–.52	–.51	–.33
5. Brief Resilience Scale Negatively Worded Items (BRS)	.68	.69	.43	.70	–	–.74	–.56	–.76	–.71
6. Acceptance and Action Questionnaire-II (AAQ-II)	–.76	–.81	–.39	–.37	–.77	–	.45	.73	.93
7. Prospective Intolerance of Uncertainty (IUS-12)	–.50	–.34	–.14	–.49	–.58	.49	–	.80	.46
8. Inhibitory Intolerance of Uncertainty (IUS-12)	–.68	–.67	–.44	–.52	–.78	.76	.81	–	.71
9. Kessler Psychological Distress Scale (K-10)	–.69	–.77	–.39	–.35	–.74	.95	.52	.76	–

Note. All CFA correlations were significant at the $p < .001$ level except between Prospective Anxiety and Valued Action ($p = .033$), between positively-worded BRS items and Behavioral Awareness ($p = .003$), and between Openness to Experience and Valued Action ($p = .042$). All IFA correlations were significant at the $p < .001$ level.