A Psychometric Comparison of Psychological Inflexibility Measures:

Discriminant Validity and Item Performance

Clarissa W. Ong, M.S.

Benjamin G. Pierce, Ph.D.

Julie M. Petersen, B.S.

Jennifer L. Barney, M.S.

Jeremiah E. Fruge, M.S.

Michael E. Levin, Ph.D.

Michael P. Twohig, Ph.D.

Corresponding author:
Clarissa W. Ong
Department of Psychology
Utah State University
2810 Old Main Hill
Logan, UT 84322-2810
Office: (435) 797-8303
Email: clarissa.ong@usu.edu

Abstract

Psychological inflexibility is a rigid behavioral pattern that interferes with engagement in personally meaningful activities; it is the hypothesized root of suffering in acceptance and commitment therapy (ACT). Thus, the quality of its measurement affects the research, theory, and practice of ACT. The current study aimed to evaluate the discriminant validity and item performance of four measures of psychological inflexibility: the Acceptance and Action Questionnaire—II (AAQ-II), a revised version of the AAQ-II (AAQ-3), the Brief Experiential Avoidance Questionnaire (BEAQ), and the Comprehensive assessment of Acceptance and Commitment Therapy processes (CompACT). We analyzed data from community ($n = 253$), student ($n = 261$), and treatment-seeking samples ($n = 140$) using exploratory factor analysis and multigroup graded-response models. The CompACT had the strongest discriminant validity followed by the AAQ-3, whereas items in the CompACT Behavioral Awareness and Valued Action subscales performed most consistently across groups. No single measure emerged as clearly superior to others; rather, appropriate selection of measures depends on the goals and context of assessment. Scientific and clinical implications are discussed.

Keywords: psychological inflexibility, psychometric, discriminant validity, item response theory, measurement, assessment

A Psychometric Comparison of Psychological Inflexibility Measures:

Discriminant Validity and Item Performance

Psychological inflexibility refers to patterns of behavior dominated by experiential

avoidance and cognitive control at the expense of personal values and contact with direct

experience (Hayes, Luoma, Bond, Masuda, & Lillis, 2006). It is characterized by rigid responses

(e.g., persistent avoidance) to internal and external stimuli that interfere with engagement in

meaningful activities. In the model of acceptance and commitment therapy (ACT), psychological

inflexibility is the linchpin of psychopathology or psychological suffering whereas its inverse,

psychological flexibility, is the hypothesized process of change and target of ACT (Hayes et al.,

2006). Psychological flexibility is defined as the ability to mindfully observe experiences

occurring in the present moment while intentionally choosing actions in line with self-chosen

values (Hayes et al., 2006).

Psychological inflexibility has been assessed in many domains including anxiety and

obsessive-compulsive disorders (Bluett, Homan, Morrison, Levin, & Twohig, 2014), parenting

and children (Brassell et al., 2016), stigma (Krafft, Ferrell, Levin, & Twohig, 2017), and chronic

pain (Feinstein et al., 2011). It is most commonly evaluated with the Acceptance and Action

Questionnaire—II (AAQ-II), a general measure of psychological inflexibility that has been

administered in various contexts (Bond et al., 2011) including cross-sectional surveys (e.g.,

Levin et al., 2016), laboratory experiments (e.g., Prins, Decuypere, & Van Damme, 2014), and

clinical trials examining ACT and related interventions (e.g., Twohig et al., 2015).

Studies focused on overall scale performance generally support the psychometric validity

of the AAQ-II (e.g., Bond et al., 2011; Fledderus, Oude Voshaar, ten Klooster, & Bohlmeijer,

2012; Flynn, Berkout, & Bordieri, 2016). However, researchers have disputed the discriminant

validity of the AAQ-II with evidence indicating the AAQ-II may be measuring distress rather than psychological inflexibility (Tyndall et al., 2019; Wolgast, 2014). If the AAQ-II assesses overall distress rather than psychological inflexibility, then research and theory development based on AAQ-II data may not be reliable or valid. Thus, there is a need to evaluate whether the AAQ-II specifically measures psychological inflexibility or a related construct like emotional distress to clarify the quality of the evidence base constructed with AAQ-II data.

The AAQ-II may also have limitations with respect to item functioning within itself and across groups (Ong, Pierce, Woods, Twohig, & Levin, 2019). For instance, certain AAQ-II items have been found to be more sensitive to different levels of psychological inflexibility than others and to perform differently across clinical and nonclinical samples (Ong et al., 2019). Differential item functioning means the same score reflects different levels of the construct depending on which item in the scale is being answered or the population of interest, making it difficult to interpret what a total score reflects, what a given score means across populations, or to make between-group comparisons. For example, it would be unclear if a score suggests levels of psychological inflexibility that warrant intervention or is normative for a given group. In addition, differential item performance as a function of responder characteristics undermines the ability of a measure to accurately detect changes in psychological inflexibility over the course of therapy as the presentation of people who have completed treatment may be different from their presentation at baseline. It is possible people at posttreatment interpret the same wording differently from at pretreatment, making AAQ-II scores before and after treatment incomparable. If this is the case, then the AAQ-II cannot reliably measure changes in psychological inflexibility over the course of treatment. These scenarios illustrate how inconsistent item functioning weakens our evidence base and may lead to misleading findings.

A more recent assessment of psychological inflexibility that was developed following the AAQ-II , the Comprehensive assessment of Acceptance and Commitment Therapy processes (CompACT; Francis, Dawson, & Golijani-Moghaddam, 2016) also has its own limitations. Although the CompACT explains more variance in current functioning and showing greater treatment sensitivity than the AAQ-II, it is not as comprehensive with respect to evaluating psychological inflexibility as the newer 60-item Multidimensional Psychological Flexibility Inventory (MPFI; Rogge, Daks, Dubler, & Saint, 2019; Rolffs, Rogge, & Wilson, 2016). It is also worth considering how the widely used Brief Experiential Avoidance Questionnaire (BEAQ; Gámez et al., 2014) compares to the AAQ-II given that the AAQ-II has commonly been referenced as a measure of experiential avoidance (e.g., Moroz & Dunkley, 2019; Ojalehto, Abramowitz, Hellberg, Buchholz, & Twohig, 2020), with correlational evidence supporting this characterization (Francis et al., 2016).

From a functional contextual perspective, accuracy in measurement and consistent psychometric performance and item functioning by responder characteristics are not about elucidating "true" scores on a given instrument. Rather, the purpose is to measure variables related to meaningful change in wellbeing in various populations and to detect individual or group differences that will ultimately inform prevention and intervention efforts (i.e., measures that help meet the analytic goals of prediction and influence). For example, it may be more important to assess *changes* in psychological inflexibility over the course of treatment to know whether an intervention works as hypothesized than an accurate "level" of psychological inflexibility. Given the significant clinical and scientific consequences of inaccurate and unreliable measurement of a core concept to ACT like psychological inflexibility, it is imperative

that we determine the psychometric merit of its measures, which has implications for how much weight we should place on the knowledge base built with findings from them.

Our aims in this study were to evaluate (1) whether psychological inflexibility measures assess a construct distinct from emotional distress (discriminant validity) and (2) whether different populations show different behavior or sample-specific variability in response to items (item performance). If there is variability in item response properties across populations, comparing scores between samples from different populations would be specious. Conversely, even if a measure performs identically across groups, the consistency is meaningless if the measure is not actually assessing the construct of interest.

We compared the AAQ-II, BEAQ, CompACT, and AAQ-3, a revised version of the AAQ-II (more information on the AAQ-3 is provided in the Method section). These were all the measures of psychological inflexibility of which we were aware when the study was designed; the BEAQ was included to make our comparison more comprehensive. Measures of psychological inflexibility that were developed since the inception of this study, the Multidimensional Psychological Flexibility Inventory (MPFI; Rolffs et al., 2016), Open and Engaged State Questionnaire (OESQ; Benoy et al., 2019), and Everyday Psychological Inflexibility Checklist (EPIC; Thompson, Bond, & Lloyd, 2019), were not included our survey. The overall goal of these analyses is to provide meta-data on psychological inflexibility measurement and, by extension, shed light on the quality of our current evidence base.

Method

Recruitment

Eligibility criteria for the present study were: (1) at least 18 years old and (2) ability to complete the letter of information and measures in English. Our current sample comprised three

groups: undergraduate college students enrolled in psychology classes in the western United

States, individuals currently seeking mental health treatment, and community members from

Amazon Mechanical Turk (MTurk).

College students were recruited using fliers on campus and online postings on university

websites and compensated with course credit. Treatment-seeking individuals were self-identified

and recruited through online postings using Facebook posts and Reddit forums ("subreddits")

specifically intended for survey participants (r/SampleSize, r/Assistance), therapy and mental

health support (r/therapy, r/TalkTherapy, r/mentalhealth), local groups (r/Logan, r/SaltLakeCity),

and academic psychology (r/psychology). Treatment-seeking participants who completed the

survey were entered into a raffle where they had the chance to obtain one of 20 $25 Amazon gift

cards. Community participants were recruited on Amazon's Mechanical Turk (MTurk) platform,

an online marketplace where "requestors" (task creators) can post tasks (e.g., surveys, writing,

experiments) to be completed by MTurk "workers" (community members). Participants were

paid $2.00 for survey completion. MTurk has been found to be a reliable means of data

collection, producing representative samples similar to other forms of diverse sampling methods

(Buhrmester, Kwang, & Gosling, 2016; Mullinix, Leeper, Druckman, & Freese, 2015)

Participants

A flowchart depicting elimination of careless or insufficient effort responders is

presented in Figure 1. The final sizes for our samples were 253, 261, and 140 for community

members, college students, and treatment-seeking clients respectively ($N = 654$). The majority of

the community sample identified as heterosexual, White, and not religious. The number of

female- and male-identifying people in the community sample was approximately equal, and the

mean age was 39.9 years ($SD = 10.9$). Most students identified as female, heterosexual, White,

and members of The Church of Jesus Christ of Latter-day Saints. The mean age in the student

sample was 20.2 years (*SD* = 4.2). Most people in the treatment-seeking group identified as

female, heterosexual, White, and not religious. Their mean age was 29.8 years (*SD* = 10.4).

Demographic information for our samples is reported in Table 1.

Procedures

All procedures were approved by a university institutional review board. After indicating

they had read and understood the letter of information for the current study and were at least 18

years old, participants completed an online survey battery that included demographic items and

measures of psychological constructs (described in the Measures section below). The survey was

accessed via an anonymous link on Qualtrics, a secure survey and data collection platform. The

order in which each measure was presented within the survey was randomized to minimize order

effects; items within each measure were presented in the same order. Completion of the survey

took approximately 30 minutes.

Measures

Psychological inflexibility.

Acceptance and Action Questionnaire – II (AAQ-II; Bond et al., 2011). The AAQ-II is

a seven-item measure designed to assess psychological inflexibility. Each item is rated on a

seven-point scale ranging from 1 (*never true*) to 7 (*always true*). Items include "I'm afraid of my

feelings" and "Emotions cause problems in my life." Responses are summed for a total score

ranging from 7 to 49; higher scores indicate higher levels of psychological inflexibility. The

AAQ-II has been used to assess psychological inflexibility in both clinical and community

samples and has demonstrated adequate reliability and validity (Bond et al., 2011). Internal

consistency for the AAQ-II in the present study was excellent ($\alpha = .94$).

Acceptance and Action Questionnaire 3 (AAQ-3). The AAQ-3 is a modified seven-item measure created for the current study (see Appendix A). Modifications were based on a qualitative examination of findings from an item response theory (IRT) analysis of the AAQ-II (Ong et al., 2019). The wording of all items was adjusted to increase clarity and improve item-level functioning (i.e. generate more consistent item performance across nonclinical and clinical samples); poorer performing items identified by the IRT analysis were revised more extensively. We removed references to possession of internal experiences (e.g., "my painful memories") in items 1 and 4, clarified references to valued living in items 1 and 7 (e.g., replaced "a life I would value" with "a meaningful life"), linked experiential avoidance to disengagement from values in items 2 and 5 (e.g., added "…that I don't do things I care about"), specified responses to emotions rather than emotions per se as problematic in items 5 and 6 (e.g., replaced "emotions" with "how I react to emotions"), and added examples of internal experiences in items 1, 3, and 7 (e.g., "worries, feelings"). A comparison of wording used in the AAQ-II and AAQ-3 is provided in Appendix B. The AAQ-3 uses the same anchors as the AAQ-II: 1 (*never true*) and 7 (*always true*). Like the AAQ-II, higher scores reflect more psychological inflexibility. Internal consistency for the AAQ-3 was excellent (α = .94).

Brief Experiential Avoidance Questionnaire (Gámez et al., 2014). The BEAQ is a 15-item measure of experiential avoidance based on the MEAQ (Gámez, Chmielewski, Kotov, Ruggero, & Watson, 2011). Respondents answer items using a six-point Likert scale from 1 (*strongly disagree*) to 6 (*strongly agree*). Items include "I feel disconnected from my emotions" and "I would give up a lot not to feel bad." A total score is calculated by summing all 15 items; higher scores indicate more experiential avoidance. Data from three independent samples suggest

the BEAQ has good internal consistency and validity (Gámez et al., 2014). Internal consistency

for the BEAQ in the current study was good (α = .89).

Comprehensive Assessment of Acceptance and Commitment Therapy (CompACT;

Francis et al., 2016). The CompACT is a 23-item measure used to assess psychological

inflexibility across three subscales: (1) Openness to Experience (10 items), (2) Behavioral

Awareness (five items), and (3) Valued Action (eight items). The Openness to Experience

subscale corresponds to acceptance and defusion (or the "open" pillar of ACT), Behavioral

Awareness captures present moment and mindfulness ("aware" pillar), and Valued Action

includes values and committed action ("engaged" pillar). Items are rated on a seven-point scale

from 0 (*strongly disagree*) to 6 (*strongly agree*) with certain items reverse-scored. Items include

"I work hard to keep out upsetting feelings" (Openness to Experience subscale), "I find it

difficult to stay focused on what's happening in the present" (Behavioral Awareness subscale),

and "I behave in line with my personal values" (Valued Action subscale). Scores for each

subscale are summed and higher scores are associated with greater psychological flexibility.

Preliminary evidence suggests the CompACT has good internal consistency and convergent and

divergent validity (Francis et al., 2016). Internal consistency was excellent for the full scale (α

= .91) and good to excellent for its subscales (α = .84 for Openness to Experience, .89 for

Behavioral Awareness, and .90 for Valued Action).

Emotional distress.

Depression Anxiety Stress Scales-21 (DASS-21; Henry & Crawford, 2005). The DASS-

21 measures three categories of emotional distress: depression, anxiety, and stress. The DASS-21

comprises three seven-item subscales (one for each category), and participants rate how much

each item applied to them over the last week on a four-point Likert scale from 0 (*did not apply to*

me at all) to 3 (*applied to me very much or most of the time*). Items include "I felt downhearted

and blue" (Depression subscale), "I felt I was close to panic" (Anxiety subscale), and "I tended

to overreact to situations" (Stress subscale). Higher scores on each subscale indicate more

emotional distress in that category. The DASS-21 has been used in various populations and has

consistently been found to have good to excellent reliability and validity (Henry & Crawford,

2005). Internal consistency was excellent for the full scale ($\alpha = .95$) and good to excellent for the

subscales (α = .94 for Depression, .86 for Anxiety, and .89 for Stress).

Statistical Analyses

Analyses were performed using R in RStudio (R Core Team, 2019; RStudio Team, 2019)

with the following packages: lavaan (Rosseel, 2012), psych (Revelle, 2018), tidyverse

(Wickham, 2017), furniture (Barrett & Brignone, 2017), careless (Yentes & Wilhelm, 2018), and

lubridate (Grolemund & Wickham, 2011). Our analytic plan was preregistered at

https://osf.io/7bcnf.

Careless responding. We removed cases of careless or insufficient effort responding

based on response time and long-string analysis (Curran, 2016). Given a recommended cutoff of

2s per item (Huang, Curran, Keeney, Poposki, & DeShon, 2012), we set a minimum response

time of 156 items × 2s = 312s. In addition, because our longest scale (CompACT) has 23 items,

we chose 23/2 = 11.5 (rounded up to 12) as the upper acceptable limit for consecutive

responding (Curran, 2016). That is, data from individuals who gave the same response for 13

consecutive items were deleted from analyses based on the assumption that careless responders

may simply select the same answer to every question (Curran, 2016).

Discriminant validity. Exploratory principal axis factor analyses (EFAs) with direct

oblimin rotation were used to evaluate discriminant validity of the psychological inflexibility

measures in each sample. Each psychological inflexibility measure and the DASS-21 were

included in an EFA to determine overlap with emotional distress given previous critiques on this

aspect of the AAQ-II (Tyndall et al., 2019; Wolgast, 2014). Unique factor loadings of items from

each scale (e.g., AAQ-II items load on to a factor that is distinct from the factor on to which

DASS-21 items load) would support discriminant validity (i.e., the items from each scale are

measuring different latent constructs).

The number of extracted factors in the final model was based on parallel analysis and

model fit indices (i.e., Tucker Lewis index [TLI], Root Mean Square Residual [RMSR], Root

Mean Square Error of Approximation [RMSEA]). Parallel analysis compares the scree plot of

factors of the observed data with that of a randomly generated data set with the same properties

as the observed data and recommends the number of factors to extract. Once the number of

factors to be extracted was determined, a principal factor solution with an oblimin (oblique)

transformation of the factor axes was specified for all models as we expected correlation among

factors.

Variability in item performance. Graded response models (GRM; Samejima, 1997)

were used to examine variability in the performance of items of each scale across samples. The

GRM provides information on item performance through extending the item discrimination and

"difficulty" parameters from the 2-parameter logistic item response theory model to ordinal

categories. Item discrimination refers to an item's overall sensitivity to variability in the

underlying construct. Mathematically, this parameter is represented as a constant value applied to

the estimation of each individual's probability of responding above each threshold between

adjacent categories (e.g., between categories 2 and 3 on a 5-point scale); as such, the

discrimination parameter provides an assessment of how neatly categories distinguish among

varying levels of the underlying construct. Item "difficulty" refers to the level of the underlying

construct or latent variable associated with a 50% chance of responding above or below a

threshold between categories of response on a given item. Unlike the discrimination parameter,

separate difficulty values are estimated for each categorical threshold for each item.

The item discrimination and difficulty parameters provided specific information about

differences in the performance of items of each psychological inflexibility scale by sample type.

Differences in the item discrimination parameters across samples were interpreted as reflecting

variability in the item's sensitivity to fluctuations in psychological inflexibility across samples.

Conversely, the median difficulty values of the items in each sample provided information on

how strongly participants tended to endorse higher responses. Variability in the difficulty

parameters across samples may indicate that the same score on an item reflects different levels of

latent psychological inflexibility, depending on the group being assessed. Specifically, if all

difficulty parameters are shifted in a consistent direction, such that all items and scores have

greater difficulty in one sample compared with another, this may simply reflect differences in the

average levels or variability of the underlying construct in that sample. However, if difficulty

parameters shift inconsistently across items or scores from one sample to the next, then estimates

of underlying construct may similarly reflect inconsistent levels of the construct across groups;

in other words, the same inflexibility score may indicate different experiences from community

to student to clinical samples.

Multiple-group GRM's were estimated for each scale to obtain difficulty and

discrimination parameter estimates for the student, community, and treatment-seeking samples.

Each GRM used a standardized parametrization where the variances of the latent factors (i.e., the

underlying measurement of psychological inflexibility) were fixed to equivalence across

samples, while discrimination and difficulty parameters were permitted to vary. This allowed

discrimination and difficulty parameters to be compared in relation to a common metric for

psychological inflexibility. Estimation was performed using robust weighted least squares and

theta parametrization to facilitate interpretation of the residuals relative to a standardized metric.

Fit was assessed using the Comparative Fit Index (CFI), TLI, and RMSEA as global fit indices

and the Standardized Root Mean Squared Residual (SRMR; Asparouhov & Muthén, 2018) as a

residual-specific fit index. The SRMR is especially informative as it can detect sources of local

misfit such as correlated residuals that may violate assumptions of the GRM.

Results

Discriminant Validity

Community sample.

AAQ-II. Parallel analysis suggested extraction of four factors and model fit indices

showed good fit for the four-factor model (TLI = .938, RMSR = .023, RMSEA = .068). Factor

loadings from the EFA of the AAQ-II and DASS-21 are presented in Table 2. AAQ-II items

loaded on to one factor with one AAQ-II item cross-loading (i.e., loading ≥ .30 for more than

one factor) with a factor corresponding to the DASS-21 Depression subscale. That is, the latent

constructs measured by the items from each scale were not clearly distinct.

Factor correlations between the AAQ-II and DASS-21 were moderate to high (*r*s = .51

to .66); the AAQ-II and DASS-21 Stress factor had the weakest correlation.

AAQ-3. Parallel analysis suggested extraction of four factors and model fit indices

showed good fit for the four-factor model (TLI = .935, RMSR = .023, RMSEA = .070). Factor

loadings from the EFA of the AAQ-3 and DASS-21 are presented in Table 3. The pattern of

loadings shows all AAQ-3 items loaded on to one factor, and there were no cross-loadings. That

is, the AAQ-3 and DASS-21 items respectively loaded on to distinct factors.

Factor correlations between the AAQ-3 and DASS-21 were high (rs = .63 to .69); the

AAQ-3 and DASS-21 Stress factor had the weakest correlation.

BEAQ. Parallel analysis suggested extraction of four factors and model fit indices

showed adequate fit for the four-factor model (TLI = .895, RMSR = .034, RMSEA = .071).

Factor loadings from the EFA of the BEAQ and DASS-21 are presented in Table 4. The pattern

of loadings shows BEAQ items loaded on to two factors in our study, though the BEAQ was

represented by a single factor in the original validation study (Gámez et al., 2014). The second

BEAQ factor also included two DASS-21 Anxiety items; together, this factor appears to

represent awareness of feelings. One BEAQ item ("I won't do something until I absolutely have

to") cross-loaded with both these factors. Thus, the latent constructs measured by the items from

the BEAQ and DASS-21 respectively were not distinct.

The correlations between the two BEAQ factors was .23. Factor correlations between the

BEAQ and DASS-21 ranged from weak to moderate (rs ranged from .28 to .45), suggesting

stronger associations between the BEAQ and DASS-21 measures than within the two BEAQ

factors. This suggests the BEAQ shows some overlap with the DASS-21 and heterogeneity in

constructs assessed within itself.

CompACT. Parallel analysis suggested extraction of five factors and model fit indices

showed good fit for the five-factor model (TLI = .922, RMSR = .029, RMSEA = .056). Factor

loadings from the EFA of the CompACT and DASS-21 are presented in Table 5. CompACT

items loaded on to three factors that were approximately consistent with the subscales identified

in the original validation study with the exception of items 6 and 20 that did not load on to their

corresponding subscale, Openness to Experience. In addition, items 4, 13, 18, and 22 from the

Openness to Experience subscale cross-loaded on to one of the other two factors in our factor

analysis (see Table 5). Several CompACT items had cross-loadings with other factors within the

scale but none with DASS-21 factors, indicating the latent constructs measured by each scale

were distinct.

The CompACT factors had weak to moderate correlations with the DASS-21 factors (*r*s

ranged from -.35 to -.17). Correlations within CompACT factors were weak (*r*s ranged from .15

to .23). Similar to the BEAQ, this suggests the constructs measured within the CompACT may

be heterogeneous.

Student sample.

AAQ-II. Parallel analysis suggested extraction of four factors and model fit indices

showed good fit for the four-factor model (TLI = .922, RMSR = .036, RMSEA = .057). Factor

loadings from the EFA of the AAQ-II and DASS-21 are presented in Table 2. AAQ-II items

loaded on to two factors and had no cross-loadings with DASS-21 items. That is, AAQ-II items

showed a bifactorial structure in the student sample unlike the unidimensional structure in the

community sample. In addition, items from each scale respectively loaded on to distinct factors,

which means the scales were assessing distinct latent constructs.

Factor correlations between the AAQ-II and DASS-21 were weak to strong (*r*s = .19

to .55); the correlation coefficient between the two AAQ-II factors was .41.

AAQ-3. Parallel analysis suggested extraction of four factors and model fit indices

showed good fit for the four-factor model (TLI = .916, RMSR = .036, RMSEA = .058). Factor

loadings from the EFA of the AAQ-3 and DASS-21 are presented in Table 3. AAQ-3 items

loaded on to one factor with one item cross-loading with a second AAQ-3 factor. AAQ-3 items

had no cross-loadings with DASS-21 items. In other words, items from each scale respectively

loaded on to distinct factors, which means they were assessing distinct latent constructs.

Factor correlations between the AAQ-3 and DASS-21 were weak to strong (rs = .05

to .59); the correlation coefficient between the two AAQ-3 factors was .07.

BEAQ. Parallel analysis suggested extraction of four factors and model fit indices

showed adequate fit for the four-factor model (TLI = .846, RMSR = .045, RMSEA = .060).

Factor loadings from the EFA of the BEAQ and DASS-21 are presented in Table 4. The pattern

of loadings shows most BEAQ items loaded on to two factors; items 2 ("I'm quick to leave any

situation that makes me feel uneasy"), 4 ("I feel disconnected from my emotions"), and 9 ("It's

hard for me to know what I'm feeling") did not load on to any factor (with a cutoff of ≥ .30). As

with the community sample, the unidimensional structure of the BEAQ was not replicated in our

student sample. The second factor also included two DASS-21 items: one each from the

Depression and Stress subscales. Thus, the latent constructs measured by the items from the

BEAQ and DASS-21 respectively were not distinct, replicating findings from the community

sample.

The correlations between the two BEAQ factors was .34. Factor correlations between the

BEAQ and DASS-21 ranged from weak to moderate (rs ranged from .26 to .31), suggesting

stronger associations between measures than within the BEAQ. Similar to in the community

sample, this suggests the BEAQ shows some heterogeneity in constructs assessed within itself.

CompACT. Parallel analysis suggested extraction of five factors and model fit indices

showed good fit for the five-factor model (TLI = .879, RMSR = .040, RMSEA = .051). Factor

loadings from the EFA of the CompACT and DASS-21 are presented in Table 5. The pattern of

loadings shows the majority of CompACT items loaded on their corresponding subscales with

the exception of five Openness to Experience items (2, 4, 6, 11, and 18). In addition, item 15

from the Openness to Experience subscale cross-loaded on to the Behavioral Awareness factor

(see Table 5). CompACT items had no cross-loadings with DASS-21 factors, which means the

latent constructs measured by each scale were distinct.

The CompACT factors had weak to moderate correlations with the DASS-21 factors (rs

ranged from -.40 to -.10). Correlations within CompACT factors were weak (rs ranged from .15

to .22). Similar to the BEAQ in both samples and the CompACT in the community sample, this

suggests the constructs measured within the CompACT were heterogenous and possibly less

strongly related to each other than to the DASS-21.

Treatment-seeking sample.

AAQ-II. Parallel analysis suggested extraction of three factors and model fit indices

showed adequate fit for the three-factor model (TLI = .810, RMSR = .055, RMSEA = .090).

Factor loadings from the EFA of the AAQ-II and DASS-21 are presented in Table 2. AAQ-II

item 1 cross-loaded with the DASS-21 Depression factor. Multiple DASS-21 items also loaded

on to the AAQ-II factor, suggesting both measures were assessing the same latent construct in

our treatment-seeking sample. Factor correlations between the AAQ-II and DASS-21 were

moderate (rs = .44 to .48).

AAQ-3. Parallel analysis suggested extraction of three factors and model fit indices

showed adequate fit for the three-factor model (TLI = .825, RMSR = .054, RMSEA = .088).

Factor loadings from the EFA of the AAQ-3 and DASS-21 are presented in Table 3. The AAQ-3

and several DASS-21 items loaded on to one factor, which means these items from both scales

were assessing the same latent construct. Factor correlations between the AAQ-3 and DASS-21

were moderate (rs = .40 to .47).

BEAQ. Parallel analysis suggested extraction of four factors and model fit indices showed adequate fit for the four-factor model (TLI = .814, RMSR = .052, RMSEA = .077). Factor loadings from the EFA of the BEAQ and DASS-21 are presented in Table 4. The pattern of loadings shows that BEAQ items loaded on to two factors; this pattern was consistent with that in the community and student samples but inconsistent with the BEAQ's hypothesized single-factor structure. One of the factors also included an item from the DASS-21 Anxiety subscale and the second factor included two items from the DASS-21 Stress subscale. Thus, the latent constructs measured by the BEAQ and DASS-21 respectively were not distinct, replicating findings from our other samples.

The correlations between the two BEAQ factors was .11. Factor correlations between the BEAQ and DASS-21 ranged from weak to moderate (*r*s ranged from .07 to .35), suggesting stronger associations between measures than within the BEAQ, replicating the intra-measure heterogeneity we observed in the community and student samples.

CompACT. Parallel analysis suggested extraction of five factors and model fit indices showed adequate fit for the five-factor model (TLI = .820, RMSR = .049, RMSEA = .072). Factor loadings from the EFA of the CompACT and DASS-21 are presented in Table 5. The pattern of loadings shows CompACT items loaded on to three factors that were approximately consistent with their corresponding subscales; item 18 from the Openness to Experience subscale loaded on to the Valued Action subscale and item 20 from the Openness to Experience subscale cross-loaded on to both the Openness to Experience and Valued Action subscales (see Table 5). CompACT items had no cross-loadings with DASS-21 items and did not load on to any of the DASS-21 factors, which means the latent constructs measured by each scale were distinct.

The CompACT factors had weak to moderate correlations with the DASS-21 factors (*r*s

ranged from -.43 to -.13). Correlations within CompACT factors were weak (*r*s ranged from .10

to .24). Similar to the BEAQ across samples and the CompACT in the community sample, this

suggests the constructs measured within the CompACT were heterogenous and possibly less

strongly related to each other than to the DASS-21.

Item Performance

The multiple-group GRMs were used to assess variability in item performance across

samples. Prior to running these models, two adjustments were made to accommodate frequencies

of zero for certain response categories on the AAQ-3 and CompACT Valued Action.

Specifically, the highest and second-highest response categories of the AAQ-3 were collapsed

due to zero responses of 7 on any item in the student sample, whereas the lowest and second-

lowest response categories of the CompACT Valued Action were collapsed due to no responses

of 0 in the community sample. Given these adjustments, the results pertaining to these scales

should be interpreted with reference to a reduced number of response categories.

The GRMs showed adequate fit according to the global (CFI, TLI) and residuals-based

(SRMR) fit indices for the AAQ-II (CFI = .995, TLI = .992, SRMR = .056), AAQ-3 (CFI = .997,

TLI = .996, SRMR = .048), CompACT Behavioral Awareness (CFI = .998, TLI = .995, SRMR

= .037), and CompACT Valued Action (CFI = .994, TLI = .992, SRMR = .053) models. Fit was

marginal to poor based on these indices for the multi-group GRMs estimated for the CompACT

Openness to Experience (CFI = .924, TLI = .903, SRMR = .108) and BEAQ (CFI = .969, TLI

= .964, SRMR = .089). In addition, the RMSEA indicated marginal to poor fit for all models,

with RMSEA = .150 for the AAQ-II, .110 for the AAQ-3, .186 for the CompACT Openness to

Experience, .088 for the CompACT Behavioral Awareness, .088 for the CompACT Valued

Action, and .116 for the BEAQ. This discrepancy between the RMSEA and other fit indices may be due to a dependency of the RMSEA on χ^2 goodness-of-fit statistics, which tend to be inflated with small samples and GRM item response theory models (Studts, 2012).

The discrimination and difficulty parameters of each item in each of the GRMs are presented in Figure 2 and Figure 3, respectively. As displayed in Figure 2, items of the AAQ-II and AAQ-3 showed evidence of stronger discrimination among varying levels of psychological inflexibility in the community sample, versus the student or treatment-seeking samples. As such, the items of both measures appeared to detect variations in inflexibility with more precision in the community sample than in the treatment-seeking or student samples. The CompACT subscales and BEAQ showed comparable levels of item discrimination, with the lowest overall discrimination across items in the GRMs on the CompACT Openness to Experience and BEAQ. However, evidence of misfit in the CompACT Openness to Experience and BEAQ may undermine the reliability of results for these scales; the underlying construct may not reflect a unidimensional scale, therefore, assessing the ability of items to discriminate among levels of that construct is questionable.

Figure 3 plots the "difficulty" parameters by response threshold (e.g., the threshold between a response of 1 versus a 2, 2 versus 3, etc.) for each item of each inflexibility measure. Considering the difficulty parameters presented in Figure 3, items tended to be more "difficult" for students and community members in scales that assessed inflexibility. This shift in difficulty is likely due to lower average levels of inflexibility among students and community members compared to people seeking treatment for whom it may be fairly "easy" to score highly on inflexibility. However, for both versions of the AAQ, item 4 appeared to be somewhat more "difficult" for students, as evidenced by a separation between the lines plotting student and

community members' difficulty parameters for this item in Figure 3. In both the AAQ-II and

AAQ-3, this item referred to "painful memories," suggesting community members may more

strongly endorse having problems with painful or traumatic memories than students.

In contrast to measures of inflexibility, items of the CompACT Behavioral Awareness

and Valued Action scales assessed *flexibility* and tended to be most "difficult" for the treatment-

seeking sample, followed by the student and the community samples. Items of the CompACT

Behavioral Awareness subscale were most difficult for the treatment-seeking sample, moderately

difficult among students, and "easiest" for the community sample. Conversely, items of the

CompACT Valued Action subscale showed similar difficulty estimates across the student and

community samples. One exception to this pattern was Item 7 of the Valued Action subscale,

which assessed taking actions that were important despite feelings of stress; this item appeared

somewhat more difficult for community members compared with students, as evidenced by a

separation between the lines plotting community and student difficulties by threshold for this

item in Figure 3.

Paneling across items of the CompACT Openness to Experience and BEAQ depicted a

fair amount of variability in the difficulty parameters. Given poor overall fit of a unidimensional

GRM, this may be a function of the misfit of a unidimensional model to these scales, sample-

specific factors, or both. As displayed in Figure 3, certain items of these scales appeared to be

nearly equally difficult across samples (e.g., items 8 and 13 of the BEAQ), suggesting

comparable rates of endorsement relative to the overall scale irrespective of sample. Conversely,

other items differed in terms of the ordering and shapes of the group-specific difficulty lines,

suggesting unmeasured factors may be impacting group-specific responding for certain items.

Discussion

The aims of the present study were to examine two aspects of psychometric validity of psychological inflexibility measures: discriminant validity and item-specific performance. The measures selected for the current study were the AAQ-II, a revised AAQ-II (AAQ-3), BEAQ, and CompACT.

Discriminant Validity

A common criticism of the AAQ-II—the most widely used measure of psychological inflexibility—is it does not sufficiently discriminate between psychological inflexibility as a behavioral pattern of responding (e.g., context-insensitive avoidance of distressing stimuli) and the experience of emotional distress (e.g., anxiety; Rochefort, Baldwin, & Chmielewski, 2018; Tyndall et al., 2019; Wolgast, 2014). The problem with this limitation is the AAQ-II may not actually be assessing ability to respond effectively to inner experiences but is instead measuring those inner experiences to which people are responding. This differentiation is particularly critical in clinical settings where ACT researchers and clinicians are more interested in whether people are changing how they respond to difficult thoughts and feelings than whether people are experiencing difficult thoughts and feelings.

Given this background, we sought to examine the relationship between measures of psychological inflexibility and the DASS-21 using EFAs in nonclinical (community and student) and clinical samples. Specifically, we wanted to evaluate the discriminant validity of psychological inflexibility measures to determine which measures were most effective at distinguishing psychologically inflexible responding from experiencing distress—and if these measures were similarly effective across populations.

The EFAs partially corroborated research demonstrating the poor discriminant validity of the AAQ-II (Rochefort et al., 2018; Tyndall et al., 2019; Wolgast, 2014). AAQ-II items did not

measure a distinct latent construct relative to DASS-21 items in the community and treatment-seeking samples but did so among students. Items from the AAQ-3, a measure developed from a previous IRT analysis of the AAQ-II (Ong et al., 2019) and for this survey, assessed a distinct latent construct from those measured by DASS-21 items in both the community and student samples, suggesting it may have stronger discriminant validity than the AAQ-II. However, it still showed poor discriminant validity in the treatment-seeking sample. The AAQ-3 was designed to retain the structure of the AAQ-II while modifying item wording in an attempt to increase clarity. Our findings indicate this attempt was partially successful as the AAQ-3 only improved discriminant validity in the community sample.

At the same time, we note both the AAQ-II and AAQ-3 had high factor correlations with the DASS-21 Depression and Anxiety subscales, meaning the latent constructs assessed by each scale—hypothesized to be psychological inflexibility and emotional distress respectively—are closely related. In addition, the AAQ-II produced a two-factor structure in the student sample, undermining the reliability of its hypothesized unidimensional structure.

The BEAQ has a strong track record of discriminant validity (Rochefort et al., 2018; Tyndall et al., 2019), yet among the psychological inflexibility scales tested in our analyses, BEAQ items showed the most overlap with DASS-21 items in terms of loading on to the same factors. These results suggest that BEAQ items did not measure a construct clearly distinct from emotional distress. In addition, the BEAQ showed intra-scale factor correlations of weaker or similar magnitude than those with the DASS-21, which suggests it may not be measuring a homogeneous construct. This may be expected given its items were drawn from different subscales on the MEAQ (Gámez et al., 2011).

The CompACT appeared to represent a distinct latent construct from the DASS-21 in all our samples, indicating stronger discriminant validity than the AAQ-II, AAQ-3, and BEAQ. The CompACT also showed a similar pattern of factor correlations to the BEAQ; although factor correlations with DASS-21 subscales were weak to moderate, factor correlations with its own subscales were weak, pointing to some heterogeneity within the CompACT. Furthermore, items did not perfectly load on to their corresponding subscales. Thus, although CompACT showed good discriminant validity, its structural validity may be less stable.

Inconsistent validity of the AAQ-II across samples, failure of BEAQ items to load on to a distinct factor, and moderate correspondence between factors identified in current and original analyses on the CompACT underscore the difficulty of reliably and concisely measuring a construct like psychological inflexibility, which demands evaluation with reference to a specific context (Hayes et al., 2006; Kashdan & Rottenberg, 2010). That is, we can only determine if a behavior is psychologically flexible or inflexible with an understanding of the context in which it occurs (Hayes, Barnes-Holmes, & Wilson, 2012). Unfortunately, it is difficult to integrate this dynamic aspect of psychological inflexibility into standardized self-report measures, and it is unlikely any comprehensive measure of psychological inflexibility would show the hypothesized unidimensional pattern (Rolffs et al., 2016). Our findings highlight the complexity of constructing a measure of psychological inflexibility that is simultaneously comprehensive enough to capture this contextually sensitive construct, precise enough to be differentiated from other highly related constructs, and coherent enough that its items still hang together in a theoretically sensible way.

Item Performance

Our second aim concerned item-specific performance or whether items are similarly related to their corresponding scales across populations. Differences in the performances of items across populations contributes to misleading conclusions when items are simply summed to compute a total score. Our findings based on unidimensional graded response models (GRM; Samejima, 1997) suggest there may be differences in sensitivity of items to individual differences in inflexibility, and in the levels of psychological inflexibility reflected by different response options across samples. In other words, the same individual differences in item scores may not reflect an equivalent difference in levels of latent psychological inflexibility between respondents from different samples (varying sensitivity) and the same score may not reflect the same level of psychological inflexibility across samples (varying difficulty).

Additionally, the poor fit of such models to the CompACT Openness to Experience subscale and BEAQ raise questions about whether these scales can or should be summed to form a one-dimensional composite value in certain groups or to compare levels of psychological inflexibility between groups drawn from different populations. The standardized residual variance across items in the student group was larger in both of these poorly fitting models, further supporting the interpretation that unidimensional, additive scaling may not be appropriate for these measures.

Item performance analyses indicated items on the AAQ-II and AAQ-3 showed greater sensitivity to individual differences within the community sample compared with the clinical or student groups. As such, differences among scores on these scales may be less well-defined in student or treatment-seeking populations, as compared with community samples reflecting a broader population. This may be due to a restricted range of responses in either group, with students possibly tending to have a range of scores restricted at the lower-end and treatment-

seeking individuals tending to have a restriction in range at the upper-end of the scales.

However, such findings also suggest that students and treatment-seeking populations respond

differently to the items on the basis of specific group factors such as education or experience

with therapy. Therefore, differences between scores are not equivalent across samples and

interpretations that rely on between-sample comparisons of score differences are likely to be

tenuous if they are based on the assumption of equivalent variability across community, student,

and clinical samples.

The item difficulty parameters also indicated variation in the degree to which participants

endorsed items on each scale across samples. Participants in the treatment-seeking sample tended

to endorse items of the AAQ-II and AAQ-3 more strongly, such that it was "easier" for an

individual from this sample to obtain a higher psychological inflexibility score compared with

participants in the student or community samples. While this finding may be expected due to the

higher severity of concerns likely observed in the treatment-seeking sample, it does raise

questions about the interpretation of scores relative to a participant's psychosocial context. For

instance, students endorsing mid-range scores on the AAQ-II may be experiencing a fair amount

of impairment or distress relative to other students, while clients endorsing similar scores may be

showing improvement relative to other clients. Additionally, there was evidence of divergence

between the community and student samples in the difficulty associated with an item of the

AAQ-II and AAQ-3 which referred to "painful memories." This result suggests unmeasured,

context-specific factors, such as differential rates of past trauma, may impact how this item is

interpreted. Such findings highlight the importance of considering the context of individual

scores and intra-individual change over using static, context-insensitive clinical cutoffs as

markers of severity. These results discourage comparing levels of psychological inflexibility

using the AAQ-II across samples.

Consistent with findings for measures of psychological inflexibility, items within the

scales measuring psychological *flexibility* were somewhat more "difficult" for participants in the

treatment-seeking sample. This may be due to lower average levels of flexibility among

treatment-seeking participants, such that it may be rare to observe higher responses to items

asking about clarity in personal values, taking values-consistent action, or perceived alignment

between behaviors and values. Consequently, increments in behavioral awareness or valued

living skills in a clinical or therapy-seeking sample may represent significant growth, whereas

similar incremental increases in a student or community group may be "easier" on these

CompACT subscales. In addition, there was evidence that responses to Behavioral Awareness

items were generally "easier" for community members, suggesting higher scores may not reflect

the same extent of flexibility compared with students or treatment-seeking individuals. This may

variously reflect differences in the interpretation of such items, contextual differences in

exposure to language endemic to clinical psychology, or actual sample differences in the range

of flexibility represented. Finally, students appeared to more "easily" endorse being able to take

valued action in the presence of stress, as compared with community members, which may again

reflect differences in the experience of stress itself (e.g., students may experience academic

stressors more frequently), in flexibility in response to stress, or in the interpretation of "stress"

relative to one's experience and linguistic communities.

Taken together, the results of our analyses investigating differential item performance

suggest that we need to carefully examine how the content of inflexibility measures is

interpreted. While inflexibility as a construct is understood in functional terms, our findings

suggest the ways people respond to items may be influenced by their content or form. Consequently, differences in people's experiences, interpretations of items, and social communities may impact the (in)flexibility scores yielded from these assessments. This was most evident in cases where a single item or subset of items diverged from a pattern of group differences in the difficulty parameters observed relative to the other items of the scale. In these cases, the differences in item responding may not parallel real differences in average (in)flexibility or variability in a sample but represent item-specific deviations in interpretation and in participant responding. For instance, students often exposed to other students talking about academic "stress" may interpret an item asking about "taking valued action despite stress" differently from a community member whose livelihood involves agricultural activity and taking care of their family. Such findings highlight the need to investigate differential item responding more specifically, relative to people's distress levels, symptom experiences (e.g., having traumatic memories), socioeconomic circumstances, treatment experiences, and multiple social identities often underrepresented in psychometric and treatment research studies.

Conclusions

Overall, we found that (1) the CompACT performed most consistently in terms of discriminant validity followed by the AAQ-3, and (2) the Behavioral Awareness and Valued Action subscales were most robust in terms of having consistent sensitivity to individual differences in psychological (in)flexibility. However, the CompACT produced a different factor structure from that in its psychometric development analyses (Francis et al., 2016), potentially pointing to poor structural validity, and showed inconsistency in item difficulty across samples. At the same time, the measures framed in terms of positive psychological skills tended to be the most consistent, corroborating evidence suggesting that psychological flexibility and

psychological inflexibility should be considered separately (Rogge et al., 2019), which has salient implications for how therapeutic progress is conceptualized on the basis of such instruments.

Because none of the measures tested demonstrated especially strong discriminant validity or consistency in item performance, we cannot recommend a single measure for general use. Instead, our findings point to relative utility of these measures given specific goals. For example, if the goal is to assess psychological inflexibility independently of emotional distress *within a sample*, then the AAQ-3 and CompACT may be suitable for nonclinical samples, whereas the CompACT may be more valid for clinical samples. If the goal is to compare present-moment awareness or values-consistent behavior *across samples*, the CompACT Behavioral Awareness and Valued Action subscales are the most useful options.

The need to reliably and validly assess psychological inflexibility in ACT research in various populations is paramount if the goal is to develop a more adequate science of human behavior. Our results suggest not all current measures of psychological inflexibility can meet this challenge. The reliability and validity of newer measures of psychological inflexibility like the MPFI (Rolffs et al., 2016) and Everyday Psychological Inflexibility Checklist (Thompson et al., 2019) should also be similarly evaluated and replicated across samples and studies to bolster confidence in conclusions based on data generated from these measures.

Limitations

First, our student and treatment-seeking samples primarily identified as White and female, and the mean age of all our samples was below 40. Thus, despite a range of presentations in our overall sample, current results may not extend to marginalized ethnic groups, male-identifying people, and older adults. Second, while we had adequate power to conduct our

primary analyses, using larger samples would allow for cross-validation of findings and increase

confidence in their replicability, especially for the treatment-seeking sample. This limitation

bears especially strongly for the GRM analyses, which, although informative, should be

replicated in much larger groups; the ideal replication would include close to 1,000 participants

in each group, given most models entailed estimating 5 or more difficulty parameters for 5 or

more items, plus the discrimination parameters. Third, using other measures of distress in the

EFAs would have clarified the relative discriminant validity of the included measures. For

example, it is possible that the AAQ-II is more strongly differentiated from a measure of

symptom-specific distress than one of general feelings of depression and anxiety. Finally,

inclusion of more recently developed measures of psychological inflexibility like the MPFI and

OESQ would have made this a more comprehensive psychometric comparison.

References

Asparouhov, T., & Muthén, B. (2018, 2018). SRMR in Mplus. Retrieved from

http://www.statmodel.com/download/SRMR2.pdf

Barrett, T. S., & Brignone, E. (2017). Furniture for quantitative scientists. *R Journal, 9*, 142-148.

Benoy, C., Knitter, B., Knellwolf, L., Doering, S., Klotsche, J., & Gloster, A. T. (2019). Assessing psychological flexibility: Validation of the Open and Engaged State Questionnaire. *Journal of Contextual Behavioral Science, 12*, 253-260. doi:10.1016/j.jcbs.2018.08.005

Bluett, E. J., Homan, K. J., Morrison, K. L., Levin, M. E., & Twohig, M. P. (2014). Acceptance and commitment therapy for anxiety and OCD spectrum disorders: an empirical review. *Journal of Anxiety Disorders, 28*(6), 612-624. doi:10.1016/j.janxdis.2014.06.008

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. doi:10.1016/j.beth.2011.03.007

Brassell, A. A., Rosenberg, E., Parent, J., Rough, J. N., Fondacaro, K., & Seehuus, M. (2016). Parent's psychological flexibility: Associations with parenting and child psychosocial well-being. *Journal of Contextual Behavioral Science, 5*(2), 111-120. doi:10.1016/j.jcbs.2016.03.001

Buhrmester, M., Kwang, T., & Gosling, S. D. (2016). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data?

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4-19. doi:10.1016/j.jesp.2015.07.006

Feinstein, A. B., Forman, E. M., Masuda, A., Cohen, L. L., Herbert, J. D., Moorthy, L. N., &

    Goldsmith, D. P. (2011). Pain intensity, psychological inflexibility, and acceptance of

    pain as predictors of functioning in adolescents with juvenile idiopathic arthritis: a

    preliminary investigation. *Journal of Clinical Psychology in Medical Settings 18*(3), 291-

    298. doi:10.1007/s10880-011-9243-6

Fledderus, M., Oude Voshaar, M. A., ten Klooster, P. M., & Bohlmeijer, E. T. J. P. a. (2012).

    Further evaluation of the psychometric properties of the Acceptance and Action

    Questionnaire–II. *24*(4), 925.

Flynn, M. K., Berkout, O. V., & Bordieri, M. J. (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.

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.

    doi:10.1016/j.jcbs.2016.05.003

Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., Suzuki, N., & Watson, D. (2014). The

    Brief Experiential Avoidance Questionnaire: Development and initial validation.

    Psychological Assessment, 26(1), 35-45. doi:10.1037/a0034473

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. doi:10.1037/a0023242

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software, 40*(3), 1-25. Retrieved from http://www.jstatsoft.org/v40/i03/

Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual Behavioral Science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science, 1*(1-2), 1-16. doi:10.1016/j.jcbs.2012.09.004

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. doi:10.1016/j.brat.2005.06.006

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44*(Pt 2), 227-239. doi:10.1348/014466505X29657

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99-114. doi:10.1007/s10869-011-9231-8

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review, 30*(7), 865-878. doi:10.1016/j.cpr.2010.03.001

Krafft, J., Ferrell, J., Levin, M. E., & Twohig, M. P. (2017). Psychological inflexibility and stigma: A meta-analytic review. *Journal of Contextual Behavioral Science*. doi:10.1016/j.jcbs.2017.11.002

Levin, M. E., Luoma, J. B., Vilardaga, R., Lillis, J., Nobles, R., & Hayes, S. C. (2016). Examining the role of psychological inflexibility, perspective taking, and empathic

concern in generalized prejudice. *Journal of Applied Social Psychology, 46*(3), 180-191.

doi:10.1111/jasp.12355

Moroz, M., & Dunkley, D. M. (2019). Self-critical perfectionism, experiential avoidance, and

depressive and anxious symptoms over two years: A three-wave longitudinal study.

Behav Res Ther, 112, 18-27. doi:10.1016/j.brat.2018.11.006

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. J. J. o. E. P. S. (2015). The

generalizability of survey experiments. *2*(2), 109-138.

Ojalehto, H. J., Abramowitz, J. S., Hellberg, S. N., Buchholz, J. L., & Twohig, M. P. (2020).

Adherence to exposure and response prevention as a predictor of improvement in

obsessive-compulsive symptom dimensions. *Journal of Anxiety Disorders, 72*, 102210.

doi:10.1016/j.janxdis.2020.102210

Ong, C. W., Pierce, B. G., Woods, D. W., Twohig, M. P., & Levin, M. E. (2019). The

Acceptance and Action Questionnaire – II: An item response theory analysis. *Journal of

Psychopathology and Behavioral Assessment, 41*, 123-134. doi:10.1007/s10862-018-

9694-2

Prins, B., Decuypere, A., & Van Damme, S. J. E. J. o. P. (2014). Effects of mindfulness and

distraction on pain depend upon individual differences in pain catastrophizing: An

experimental study. *18*(9), 1307-1315.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria:

R Foundation for Statistical Computing.

Revelle, W. (2018). psych: Procedures for psychological, psychometric, and personality

research. R package version 1.8.10. Retrieved from https://CRAN.R-

project.org/package=psych

Rochefort, C., Baldwin, A. S., & Chmielewski, M. (2018). Experiential avoidance: An

    examination of the construct validity of the AAQ-II and MEAQ. *Behavior Therapy,*

    49(3), 435-449. doi:10.1016/j.beth.2017.08.008

Rogge, R. D., Daks, J. S., Dubler, B. A., & Saint, K. J. (2019). It's all about the process:

    Examining the convergent validity, conceptual coverage, unique predictive validity, and

    clinical utility of ACT process measures. *Journal of Contextual Behavioral Science, 14*,

    90-102. doi:10.1016/j.jcbs.2019.10.001

Rolffs, J. L., Rogge, R. D., & Wilson, K. G. (2016). Disentangling components of flexibility via

    the hexaflex model: Development and validation of the Multidimensional Psychological

    Flexibility Inventory (MPFI). *Assessment, 25*(4), 458-482.

    doi:10.1177/1073191116645905

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of*

    Statistical Software, 48(2), 1-36. doi:10.18637/jss.v048.i02

RStudio Team. (2019). RStudio: Integrated Development for R. Retrieved from

    http://www.rstudio.com/

Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton

    (Eds.), *Handbook of Modern Item Response Theory* (pp. 85-100). New York, NY:

    Springer.

Studts, C. R. (2012). *Utility of a goodness-of-fit index for the graded response model with small*

    sample sizes: A Monte Carlo investigation. University of Louisville, Retrieved from

    https://ir.library.louisville.edu/etd/1397/

Thompson, M., Bond, F. W., & Lloyd, J. (2019). Preliminary psychometric properties of the

      Everyday Psychological Inflexibility Checklist. *Journal of Contextual Behavioral*

      Science, 12, 243-252. doi:10.1016/j.jcbs.2018.08.004

Twohig, M. P., Abramowitz, J. S., Bluett, E. J., Fabricant, L. E., Jacoby, R. J., Morrison, K.

      L., . . . Smith, B. M. (2015). Exposure therapy for OCD from an acceptance and

      commitment therapy (ACT) framework. *Journal of Obsessive-Compulsive and Related*

      Disorders, 6, 167-173. doi:10.1016/j.jocrd.2014.12.007

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. doi:10.1016/j.jcbs.2018.09.005

Wickham, H. (2017). tidyverse: Easily install and load 'Tidyverse' packages. R package version

      1.1.1. Retrieved from http://CRAN.R-project.org/package=tidyverse

Wolgast, M. (2014). What does the Acceptance and Action Questionnaire (AAQ-II) really

      measure? *Behavior Therapy, 45*, 831-839. doi:10.1016/j.beth.2014.07.002

Yentes, R. D., & Wilhelm, F. (2018). careless: Procedures for computing indices of careless

      responding. R package version 1.1.3.

Table 1
Means and Standard Deviations or Frequencies for Demographic Variables in Samples

	Community *n* = 253	Student *n* = 261	Treatment-seeking *n* = 140
Age	39.9 (10.9)	20.2 (4.2)	29.8 (10.4)
Sex			
Female	120 (47.4%)	197 (75.5%)	104 (74.3%)
Male	132 (52.2%)	64 (24.5%)	36 (25.7%)
Intersex	1 (0.4%)	0 (0%)	0 (0%)
Gender identity			
Female	120 (47.4%)	197 (75.5%)	100 (71.4%)
Male	132 (52.2%)	64 (24.5%)	32 (22.9%)
Transgender	1 (0.4%)	0 (0%)	1 (0.7%)
Not listed	0 (0%)	0 (0%)	7 (5%)
Sexual orientation			
Asexual	7 (2.8%)	25 (9.6%)	3 (2.1%)
Bisexual	25 (9.9%)	7 (2.7%)	26 (18.6%)
Gay or lesbian	3 (1.2%)	3 (1.1%)	13 (9.3%)
Heterosexual	215 (85%)	216 (82.8%)	80 (57.1%)
Queer	1 (0.4%)	1 (0.4%)	7 (5%)
Pansexual	0 (0%)	4 (1.5%)	7 (5%)
Not listed	2 (0.8%)	4 (1.5%)	4 (2.9%)
Ethnicity			
Native American/Indigenous	6 (2.4%)	4 (1.5%)	0 (0%)
Asian	63 (24.9%)	1 (0.4%)	9 (6.4%)
Black	15 (5.9%)	1 (0.4%)	3 (2.1%)
Latinx	11 (4.3%)	10 (3.8%)	12 (8.6%)
Middle Eastern	3 (1.2%)	0 (0%)	2 (1.4%)
Pacific Islander	0 (0%)	3 (1.1%)	0 (0%)
White	165 (65.2%)	249 (95.4%)	114 (81.4%)
Multiracial	1 (0.4%)	1 (0.4%)	3 (2.1%)
Not listed	2 (0.8%)	0 (0%)	3 (2.1%)
Religion			
Mormon/LDS	0 (0%)	205 (78.5%)	2 (1.4%)
Catholic	33 (13%)	3 (1.1%)	8 (5.7%)
Methodist	6 (2.4%)	0 (0%)	1 (0.7%)
Protestant	45 (17.8%)	2 (0.8%)	7 (5%)
Lutheran	1 (0.4%)	0 (0%)	1 (0.7%)
Jewish	7 (2.8%)	0 (0%)	8 (5.7%)
Muslim	3 (1.2%)	0 (0%)	1 (0.7%)

Buddhist	3 (1.2%)	1 (0.4%)	2 (1.4%)
Hindu	43 (17%)	0 (0%)	0 (0%)
Not religious	99 (39.1%)	30 (11.5%)	83 (59.3%)
Not listed	13 (5.1%)	20 (7.7%)	27 (19.3%)

Note. LDS = The Church of Jesus Christ of Latter-day Saints.

Table 2

Standardized Loadings (Pattern Matrix) and Factor Correlations for the AAQ-II and DASS-21

Sub-scale	Variable	Community				Student				Treatment-Seeking		
		AAQ-II	D	A	S	AAQ-II-1	AAQ-II-2	D	A/S	AAQ-II	D	A/S
	Worries get in the way of my success	**.87**	.05	-.09	.04	**.83**	-.03	.09	-.03	**.68**	.04	.03
	Emotions cause problems in my life	**.84**	-.13	.06	.16	**.74**	.16	-.09	.09	**.86**	-.01	-.06
	I worry about not being able to control my worries and feelings	**.78**	.06	-.06	.16	**.66**	.18	-.04	.14	**.74**	-.18	.17
	My painful experiences and memories make it difficult for me to live a life that I would value	**.75**	.08	.18	-.03	.25	**.65**	.15	.03	**.50**	**.34**	-.02
	My painful memories prevent me from having a fulfilling life	**.72**	.08	.27	-.10	.26	**.62**	.15	.09	**.63**	.17	.05
	It seems like most people are handling their lives better than I am	**.63**	**.31**	-.13	.02	**.69**	.04	.15	-.02	**.64**	.26	-.08
	I'm afraid of my feelings	**.63**	.20	.11	.05	**.56**	.20	.01	.07	**.62**	.08	.10
D	I felt that life was meaningless	.10	**.85**	.06	-.15	-.09	.06	**.83**	.01	.07	**.82**	-.16
D	I was unable to become enthusiastic about anything	-.01	**.82**	.02	.06	-.08	.08	**.78**	.09	-.09	**.84**	.10
D	I felt I wasn't worth much as a person	.06	**.78**	.05	.02	.17	.03	**.67**	-.04	.13	**.70**	.04
D	I felt that I had nothing to look forward to	.13	**.73**	.06	.04	.07	-.07	**.81**	-.04	-.05	**.83**	.01
D	I felt downhearted and blue	-.03	**.72**	-.06	.27	.15	-.11	**.54**	.17	.18	**.64**	.03
D	I couldn't seem to experience any positive feeling at all	-.01	**.68**	.17	.11	.00	.12	**.65**	.08	.00	**.73**	.13
D	I found it difficult to work up the initiative to do things	.04	**.54**	.03	.27	**.34**	-.14	.26	.23	**.35**	**.39**	.02
A	I experienced breathing difficulty	-.08	.19	**.68**	.11	-.07	.07	.08	**.62**	-.10	-.01	**.73**
A	I felt scared without any good reason	.09	.21	**.59**	.06	-.04	-.02	.02	**.67**	.22	-.01	**.51**
A	I experienced trembling	.02	.11	**.57**	.25	-.12	.21	-.01	**.65**	-.08	.00	**.68**
A	I was worried about situations in which I might panic and make a fool of myself	.24	.08	**.56**	.04	.19	.04	.04	**.47**	**.42**	-.07	**.41**
A	I felt I was close to panic	.15	.14	**.55**	.13	.08	.08	.13	**.58**	.16	.02	**.64**

A	I was aware of the action of my heart in the absence of physical exertion	.11	-.03	**.55**	.16	-.06	.09	.05	**.58**	-.19	-.07	**.53**
A	I was aware of dryness of my mouth	.22	-.13	**.48**	.21	-.18	.11	.07	**.33**	.06	-.13	**.34**
S	I felt that I was rather touchy	-.02	.10	-.01	**.78**	.18	-.21	.05	**.36**	**.32**	.13	.23
S	I was intolerant of anything that kept me from getting on with what I was doing	.03	.04	.01	**.73**	.00	.05	.24	**.35**	-.01	**.41**	.27
S	I tended to overreact to situations	.13	.01	.07	**.72**	**.42**	-.19	.14	.26	**.42**	.02	**.34**
S	I found myself getting agitated	.16	.08	.01	**.70**	.17	-.22	.15	**.55**	**.31**	.13	**.48**
S	I felt that I was using a lot of nervous energy	-.09	.04	.27	**.67**	.11	.04	-.07	**.70**	.15	-.02	**.63**
S	I found it difficult to relax	.17	.05	.07	**.65**	.06	-.15	.08	**.71**	.03	.19	**.64**
S	I found it hard to wind down	.10	.06	.14	**.58**	.00	.02	-.06	**.62**	-.05	.18	**.52**

Factor correlations

	1. AAQ-II	2. D	3. A	4. S	1. AAQ-II 1	2. AAQ-II 2	3. D	4. A/S	1. AAQ-II	2. D	3. A/S
1	1	.66	.51	.61	1	.41	.55	.53	1	.48	.44
2		1	.58	.64	.41	1	.24	.19		1	.26
3			1	.64	.55	.24	1	.65			1
4				1	.53	.19	.65	1			

Note. Factor labels (in topmost row) were assigned based on standardized loadings. The subscale to which each item actually belongs is indicated in the leftmost column. Standardized loadings greater than 0.3 are bolded. Items are presented in descending order of factor loadings within each latent factor and measure in the community sample. AAQ-II = Acceptance and Action Questionnaire—II; DASS-21 = Depression Anxiety Stress Scales-21; D = Depression; A = Anxiety; S = Stress.

Table 3

Standardized Loadings (Pattern Matrix) and Factor Correlations for the AAQ-3 and DASS-21

Sub-scale	Variable	Community				Student				Treatment-Seeking		
		AAQ-3	D	A	S	AAQ-3-1	AAQ-3-2	D	A/S	AAQ-3	D	A/S
	Worries, feelings, or memories keep me from moving toward my goals	**.90**	.04	-.07	.02	**.79**	-.07	.05	-.04	**.76**	.10	-.02
	How I react to emotions causes problems in important areas of my life	**.83**	-.17	.11	.12	**.73**	-.03	-.02	.05	**.86**	-.07	-.04
	Painful worries, feelings, or memories make it impossible for me to live a meaningful life	**.80**	.08	.12	-.08	**.56**	.30	.16	.07	**.70**	.08	-.02
	Painful memories prevent me from having a fulfilling life	**.77**	.06	.12	-.05	**.53**	**.42**	.15	.09	**.71**	.08	.06
	I'm so afraid of my feelings that I don't do things I care about	**.73**	.26	.03	-.06	**.68**	.13	.05	.07	**.80**	.02	.02
	I worry about losing control of my thoughts, feelings, or memories	**.71**	.07	.12	.08	**.63**	.17	-.05	.18	**.68**	-.08	.19
	I do not handle my emotions well	**.69**	.08	-.17	.27	**.79**	-.20	.02	-.02	**.86**	.03	-.08
D	I felt that life was meaningless	.01	**.88**	.10	-.13	-.07	.09	**.85**	-.02	.04	**.84**	-.16
D	I was unable to become enthusiastic about anything	.04	**.78**	.01	.07	-.06	.05	**.79**	.08	-.07	**.83**	.06
D	I felt I wasn't worth much as a person	.11	**.75**	.01	.03	.18	.02	**.70**	-.07	.07	**.72**	.06
D	I felt that I had nothing to look forward to	.11	**.73**	.04	.07	.02	-.11	**.81**	-.02	-.01	**.82**	-.04
D	I felt downhearted and blue	-.02	**.70**	-.06	.28	.13	-.12	**.52**	.17	.09	**.67**	.06
D	I couldn't seem to experience any positive feeling at all	.07	**.63**	.17	.10	.02	-.03	**.67**	.07	-.04	**.76**	.12
D	I found it difficult to work up the initiative to do things	.05	**.52**	.04	.27	.28	-.19	.25	.23	.26	**.44**	.05
A	I experienced breathing difficulty	.07	.11	**.66**	.06	-.01	.07	.07	**.60**	-.13	-.03	**.74**
A	I felt scared without any good reason	.05	.22	**.60**	.06	-.04	.16	-.01	**.68**	.14	.03	**.55**
A	I was aware of the action of my heart in the absence of physical exertion	.11	-.04	**.59**	.11	-.02	.18	.04	**.58**	-.02	-.15	**.45**
A	I experienced trembling	.04	.10	**.58**	.22	.02	.21	.00	**.59**	-.14	-.01	**.71**

A	I felt I was close to panic	.08	.17	**.57**	.12	.05	.08	.16	**.58**	.11	.03	**.67**
A	I was aware of dryness of my mouth	.10	-.09	**.52**	.21	.03	.05	.01	.27	-.04	-.10	**.39**
A	I was worried about situations in which I might panic and make a fool of myself	.28	.06	**.52**	.03	.23	-.09	.04	**.45**	**.31**	-.03	**.49**
S	I felt that I was rather touchy	.03	.09	0	**.76**	.07	-.08	.03	**.38**	**.34**	.12	.23
S	I tended to overreact to situations	.13	-.01	.07	**.73**	.25	-.25	.14	**.31**	**.43**	.03	**.36**
S	I was intolerant of anything that kept me from getting on with what I was doing	-.05	.07	.08	**.71**	.04	-.15	.22	**.34**	.08	**.36**	.23
S	I found myself getting agitated	.08	.11	.05	**.70**	.06	-.27	.11	**.61**	**.31**	.13	**.50**
S	I found it difficult to relax	.19	.06	.03	**.66**	.01	-.11	.05	**.73**	.02	.18	**.64**
S	I felt that I was using a lot of nervous energy	-.06	.03	**.31**	**.62**	.03	-.06	-.02	**.71**	.10	-.03	**.67**
S	I found it hard to wind down	.04	.10	.14	**.59**	.05	-.05	-.08	**.61**	-.06	.18	**.52**

Factor correlations

	1. AAQ-3	2. D	3. A	4. S	1. AAQ-3 1	2. AAQ-3 2	3. D	4. A/S	1. AAQ-3	2. D	3. A/S
1	1	.69	.63	.62	1	.07	.58	.59	1	.47	.40
2		1	.59	.62	.07	1	.05	.06		1	.31
3			1	.65	.58	.05	1	.66			1
4				1	.59	.06	.66	1			

Note. Factor labels (in topmost row) were assigned based on standardized loadings. The subscale to which each item actually belongs is indicated in the leftmost column. Standardized loadings greater than 0.3 are bolded. AAQ-3 = Acceptance and Action Questionnaire 3; DASS-21 = Depression Anxiety Stress Scales-21; D = Depression; A = Anxiety; S = Stress.

Table 4

Standardized Loadings (Pattern Matrix) and Factor Correlations for the BEAQ and DASS-21

Sub-scale	Variable	Community				Student				Treatment-Seeking			
		BEAQ1	BEAQ2	D	A/S	BEAQ1	BEAQ2	D	A/S	BEAQ1	BEAQ2	D	A/S
	I go out of my way to avoid uncomfortable situations	**.85**	-.15	.04	.04	**.34**	**.44**	-.09	.08	**.71**	.01	.00	.05
	I work hard to keep out upsetting feelings	**.85**	-.06	-.13	.08	**.62**	-.01	-.11	.16	**.64**	.17	-.06	.10
	One of my big goals is to be free from painful emotions	**.79**	.04	-.11	.06	**.67**	.06	.11	.02	**.58**	-.04	.14	-.01
	I rarely do something if there is a chance that it will upset me	**.75**	.01	.05	.06	**.41**	**.31**	.02	-.04	**.81**	-.18	.02	-.06
	I'm quick to leave any situation that makes me feel uneasy	**.73**	-.03	.03	-.05	.28	.19	-.17	.02	**.65**	.03	-.06	.12
	If I have any doubts about doing something, I just won't do it	**.73**	.03	.15	-.08	**.39**	**.34**	.02	-.03	**.59**	.14	.10	-.03
	I would give up a lot not to feel bad	**.68**	-.01	.08	.04	**.45**	.06	.11	.08	**.61**	-.18	.18	.09
	Pain always leads to suffering	**.65**	.11	.05	.08	**.44**	.08	.16	-.02	**.54**	-.05	.18	.05
	The key to a good life is never feeling any pain	**.60**	.26	-.13	.15	**.44**	-.17	.21	-.08	**.45**	.09	-.03	.11
	I won't do something until I absolutely have to	**.59**	**.30**	.18	-.09	.00	**.55**	.11	-.08	**.63**	.25	-.01	-.08
	When unpleasant memories come to me, I try to put them out of my mind	**.59**	-.07	-.08	-.07	**.36**	.04	-.15	-.03	**.51**	**.31**	-.21	.03
	I try to put off unpleasant tasks for as long as possible	**.58**	.08	.30	-.11	.13	**.62**	-.01	.09	**.67**	.11	-.03	-.07
	It's hard for me to know what I'm feeling	.25	**.50**	.10	.18	.23	.30	.20	.15	**.40**	**.40**	.16	-.17
	I feel disconnected from my emotions	.29	**.45**	.19	-.04	.26	.21	.27	.03	.28	**.58**	.19	-.13
	Fear or anxiety won't stop me from doing something important	.21	-.23	**.33**	-.15	.05	**.37**	.19	.05	**.46**	.01	.12	.18
D	I felt that life was meaningless	-.01	.07	**.91**	-.10	.04	.03	**.80**	-.04	.02	-.04	**.86**	-.16
D	I felt I wasn't worth much as a person	-.01	.03	**.85**	.03	-.03	.11	**.72**	.01	.02	.06	**.75**	.06

D	I was unable to become enthusiastic about anything	.04	.03	**.80**	.04	.06	-.08	**.75**	.09	-.07	.08	**.81**	.04
D	I felt that I had nothing to look forward to	.06	-.01	**.77**	.11	-.04	.02	**.80**	.00	-.02	-.05	**.82**	-.03
D	I felt downhearted and blue	.00	-.16	**.73**	.22	-.01	.06	**.53**	.23	.09	-.01	**.69**	.08
D	I couldn't seem to experience any positive feeling at all	.02	.03	**.65**	.24	.06	-.09	**.70**	.09	-.05	.04	**.76**	.10
D	I found it difficult to work up the initiative to do things	.18	-.07	**.53**	.24	-.18	**.33**	.30	**.31**	.23	.05	**.47**	.09
S	I tended to overreact to situations	.03	-.11	.03	**.85**	.08	.07	.22	**.36**	.26	-.24	.15	**.50**
S	I felt that I was using a lot of nervous energy	.01	-.00	-.02	**.85**	.02	.03	-.05	**.74**	.09	.05	-.01	**.67**
S	I felt that I was rather touchy	.00	-.14	.06	**.81**	-.11	**.32**	-.01	**.38**	.15	-.12	.22	**.35**
S	I found myself getting agitated	.06	-.06	.11	**.76**	-.11	-.02	.13	**.66**	.23	-.16	.21	**.59**
S	I was intolerant of anything that kept me from getting on with what I was doing	.10	-.12	-.01	**.75**	.06	-.05	.22	**.37**	-.03	.03	**.39**	.28
S	I found it difficult to relax	.09	-.03	.10	**.73**	.04	-.01	.02	**.75**	.02	**.48**	.11	**.63**
S	I found it hard to wind down	.03	.05	.10	**.68**	.15	-.18	-.06	**.65**	-.05	**.46**	.12	**.47**
A	I experienced trembling	-.01	.24	.12	**.64**	.03	.00	.03	**.58**	-.23	.00	.01	**.70**
A	I was aware of dryness of my mouth	.06	.27	-.07	**.59**	-.10	-.02	.00	**.33**	.00	.03	-.12	**.37**
A	I felt I was close to panic	.03	.29	.16	**.55**	.09	-.02	.16	**.60**	.12	-.02	.04	**.68**
A	I experienced breathing difficulty	-.08	**.34**	.17	**.54**	.01	-.01	.05	**.61**	-.14	.02	-.03	**.71**
A	I was aware of the action of my heart in the absence of physical exertion	.03	**.32**	.01	**.53**	-.05	.05	.03	**.56**	-.06	.07	-.17	**.46**
A	I felt scared without any good reason	.02	.26	.22	**.50**	-.04	.16	-.03	**.61**	.05	-.06	.07	**.59**
A	I was worried about situations in which I might panic and make a fool of myself	.03	.28	.16	**.49**	.04	.13	.09	**.49**	**.39**	-.22	.00	**.56**

Factor correlations

	1. BEAQ1	2. BEAQ2	3. D	4. A/S	1. BEAQ1	2. BEAQ2	3. D	4. A/S	1. BEAQ1	2. BEAQ2	3. D	4. A/S
1	1	.23	.45	.42	1	.34	.21	.26	1	.11	.35	.22
2		1	.29	.28		1	.29	.31		1	.12	.07

| | 3 | 1 | .70 | | 1 | .67 | | 1 | .32 |
| | 4 | | 1 | | | 1 | | | 1 |

Note. Factor labels (in topmost row) were assigned based on standardized loadings. The subscale to which each item actually belongs is indicated in the leftmost column. Standardized loadings greater than 0.3 are bolded. BEAQ = Brief Experiential Avoidance Questionnaire; DASS-21 = Depression Anxiety Stress Scales-21; D = Depression; A = Anxiety; S = Stress.

Table 5

Standardized Loadings (Pattern Matrix) and Factor Correlations for the CompACT and DASS-21

Sub-scale	Variable	Community					Student					Treatment-Seeking				
		OE	BA	VA	D	A/S	OE	BA	VA	D	A/S	OE	BA	VA	D	A/S
OE	I work hard to keep out upsetting feelings	**.68**	.17	-.14	-.13	-.05	**.33**	**.40**	-.29	.08	.00	**.74**	.14	-.01	.08	-.03
OE	I go out of my way to avoid situations that might bring difficult thoughts, feelings, or sensations	**.64**	.22	.02	-.12	-.01	.26	**.47**	-.13	.15	-.10	**.58**	**.07**	-.04	-.21	-.06
OE	One of my big goals is to be free from painful emotions	**.59**	.20	-.11	-.13	.03	.10	**.37**	-.14	-.11	.02	**.55**	.09	.08	.04	-.01
OE	I try to stay busy to keep thoughts or feelings from coming	**.57**	**.37**	.01	.04	-.11	.28	**.38**	-.03	-.12	-.07	**.63**	.24	-.25	-.08	-.04
OE	I tell myself that I shouldn't have certain thoughts	**.44**	.23	.01	-.17	-.11	**.45**	.29	-.18	-.02	-.11	**.71**	.09	-.11	.06	-.06
OE	I am willing to fully experience whatever thoughts, feelings and sensations come up for me, without trying to change or defend against them	**.41**	-.26	**.62**	.04	-.15	**.65**	-.04	.18	.02	-.01	**.61**	-.15	.29	-.07	.02
BA	I rush through meaningful activities without being really attentive to them	.13	**.80**	.02	.00	-.01	-.05	**.59**	.15	.00	-.04	-.02	**.56**	.18	-.02	-.04
BA	Even when doing the things that matter to me, I find myself doing them without paying attention	.06	**.77**	.08	.03	-.10	-.08	**.76**	.05	-.06	.06	.03	**.79**	.17	.05	.02
BA	It seems I am "running on automatic" without much awareness of what I'm doing	.00	**.77**	.07	.01	-.10	-.07	**.59**	.11	-.18	-.05	.09	**.70**	.00	-.21	.00
BA	I do jobs or tasks automatically, without being aware of what I'm doing	.02	**.75**	.10	.07	-.11	-.06	**.60**	-.13	-.02	-.02	.12	**.67**	-.09	-.18	.00

BA	I find it difficult to stay focused on what's happening in the present	.19	**.65**	.11	-.14	.02	.08	**.55**	.10	-.02	-.17	.19	**.40**	.12	-.02	-.25
OE	I get so caught up in my thoughts that I am unable to do the things that I most want to do	.16	**.57**	.13	-.23	-.02	.13	**.51**	.08	-.22	-.02	**.46**	.20	.17	-.09	-.21
OE	Even when something is important to me, I'll rarely do it if there is a chance it will upset me	**.36**	**.49**	.04	-.17	-.06	-.05	**.48**	.29	.12	-.09	.22	.13	**.42**	-.08	-.09
VA	I can keep going with something when it's important to me	-.03	.23	**.74**	.03	-.07	.15	.03	**.65**	.05	-.02	.16	.03	**.61**	-.11	.02
VA	I undertake things that are meaningful to me, even when I find it hard to do so	.01	.08	**.73**	-.01	-.07	.21	.01	**.64**	.06	-.04	.16	-.03	**.54**	-.08	.05
VA	I act in ways that are consistent with how I wish to live my life	-.19	.12	**.70**	-.18	.05	-.07	.00	**.58**	-.04	-.04	-.04	-.04	**.65**	-.19	-.08
VA	I am able to follow my long terms plans including times when progress is slow	.02	.05	**.67**	-.09	.00	.09	-.05	**.57**	-.14	.03	.05	-.04	**.70**	.02	-.04
VA	I can identify the things that really matter to me in life and pursue them	-.04	.09	**.65**	-.29	.12	.07	.05	**.55**	-.07	.01	-.04	.08	**.54**	-.17	.09
OE	I can take thoughts and feelings as they come, without attempting to control or avoid them	**.36**	-.20	**.63**	.01	-.08	**.62**	-.10	.17	-.10	-.06	**.64**	-.14	.26	-.05	-.03
VA	I make choices based on what is important to me, even if it is stressful	-.13	.05	**.62**	-.27	.07	0	.06	**.77**	.06	-.08	.06	.09	**.74**	.06	-.07
VA	My values are really reflected in my behavior	-.21	.14	**.61**	-.02	-.03	-.01	.05	**.72**	-.07	.06	-.18	.15	**.64**	-.18	-.04
VA	I behave in line with my personal values	-.21	.27	**.60**	-.04	-.11	-.09	-.01	**.71**	-.10	.00	-.11	.15	**.65**	-.08	.08
OE	Thoughts are just thoughts-- they don't control what I do	.22	-.17	**.52**	.00	-.02	**.41**	-.03	.03	-.13	.24	**.44**	-.06	**.47**	.02	.09

D	I felt that life was meaningless	-.04	-.05	-.03	**.88**	-.13	.02	-.03	-.04	**.73**	.02	-.02	.00	-.07	**.80**	-.14
D	I felt I wasn't worth much as a person	.00	.05	-.17	**.79**	.05	-.04	-.03	-.09	**.70**	.02	-.09	-.01	-.12	**.66**	.07
D	I felt that I had nothing to look forward to	-.09	.01	-.07	**.74**	.10	.01	-.01	.03	**.85**	-.03	.08	.01	-.07	**.80**	-.02
D	I was unable to become enthusiastic about anything	-.09	.03	-.12	**.72**	.09	-.01	.02	.01	**.76**	.09	.08	-.10	.02	**.81**	.04
D	I couldn't seem to experience any positive feeling at all	-.06	-.01	-.07	**.64**	.21	-.03	-.07	-.01	**.65**	.07	-.05	-.10	-.10	**.65**	.09
D	I felt downhearted and blue	-.07	.14	-.15	**.60**	.29	-.06	-.01	-.01	**.55**	.21	-.09	-.02	.02	**.68**	.10
A	I felt scared without any good reason	.03	-.16	.19	**.55**	.28	.05	.03	-.03	-.04	**.70**	-.19	.07	.11	.10	**.57**
D	I found it difficult to work up the initiative to do things	-.09	.00	-.10	**.48**	.29	-.02	-.11	.00	**.36**	.27	-.08	-.04	.02	**.52**	.15
A	I felt I was close to panic	.06	-.20	.16	**.45**	**.35**	-.07	-.02	-.02	.15	**.60**	-.09	.11	.03	.10	**.69**
A	I experienced breathing difficulty	.13	-.22	.15	**.44**	**.33**	-.11	.08	.01	.08	**.60**	.07	-.12	.09	-.09	**.68**
A	I was worried about situations in which I might panic and make a fool of myself	-.01	-.19	.11	**.41**	.30	-.03	-.21	-.07	.07	**.45**	-.18	.22	-.18	.03	**.61**
S	I tended to overreact to situations	.00	.03	-.05	-.01	**.88**	-.00	-.22	.00	.19	**.34**	-.06	.18	-.13	.19	**.55**
S	I felt that I was rather touchy	0	.04	-.03	.02	**.81**	.06	-.03	-.12	.00	**.42**	-.14	.16	.05	**.32**	.35
S	I felt that I was using a lot of nervous energy	.00	-.04	.02	.03	**.79**	-.10	-.02	.04	-.02	**.71**	-.09	-.03	-.02	-.06	**.69**
S	I found myself getting agitated	-.05	-.01	-.04	.06	**.79**	.07	.02	-.04	.12	**.64**	-.01	.05	-.09	.23	**.64**
S	I was intolerant of anything that kept me from getting on with what I was doing	-.05	.04	.06	.01	**.78**	.09	0	-.01	.23	**.38**	.17	-.04	-.02	**.41**	.31
S	I found it difficult to relax	-.03	-.08	-.06	.04	**.76**	.04	-.01	-.02	.05	**.73**	-.04	-.26	.00	.07	**.62**
S	I found it hard to wind down	.02	-.13	-.05	.03	**.69**	.06	-.05	0	-.08	**.62**	-.08	-.29	.05	.04	**.45**
A	I experienced trembling	.10	-.26	.13	**.32**	**.45**	-.04	.00	.02	-.01	**.61**	.20	-.11	.05	-.06	**.68**
A	I was aware of dryness of my mouth	.04	-.22	.14	.19	**.41**	.05	.02	.02	.05	.28	.19	-.06	-.26	-.25	**.43**

A	I was aware of the action of my heart in the absence of physical exertion	.05	**-.30**	.09	.21	**.34**	.02	.02	.04	.04	**.59**	.13	.03	.02	-.13	**.46**

Factor correlations

	1. OE	2. BA	3. VA	4. D	5. A/S	1. OE	2. BA	3. VA	4. D	5. A/S	1. OE	2. BA	3. VA	4. D	5. A/S
1	1	.23	.15	-.22	-.25	1	.22	.15	-.10	-.20	1	.19	.24	-.24	-.24
2		1	.15	-.47	-.48		1	.14	-.33	-.40		1	.10	-.31	-.15
3			1	-.35	-.17			1	-.37	-.22			1	-.43	-.13
4				1	.70				1	.65				1	.32
5					1					1					1

Note. Factor labels (in topmost row) were assigned based on standardized loadings. The subscale to which each item actually belongs is indicated in the leftmost column. Standardized loadings greater than 0.3 are bolded. CompACT = Comprehensive assessment of Acceptance and Commitment Therapy processes; DASS-21 = Depression Anxiety Stress Scales-21; VA = Valued Action; BA = Behavioral Awareness; OE = Openness to Experience; D = Depression; A = Anxiety; S = Stress.
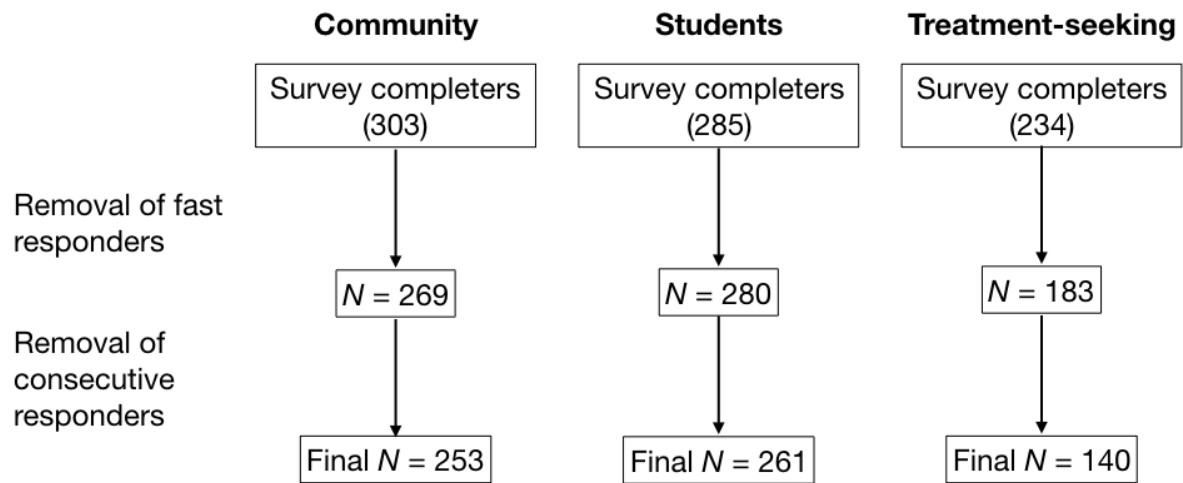
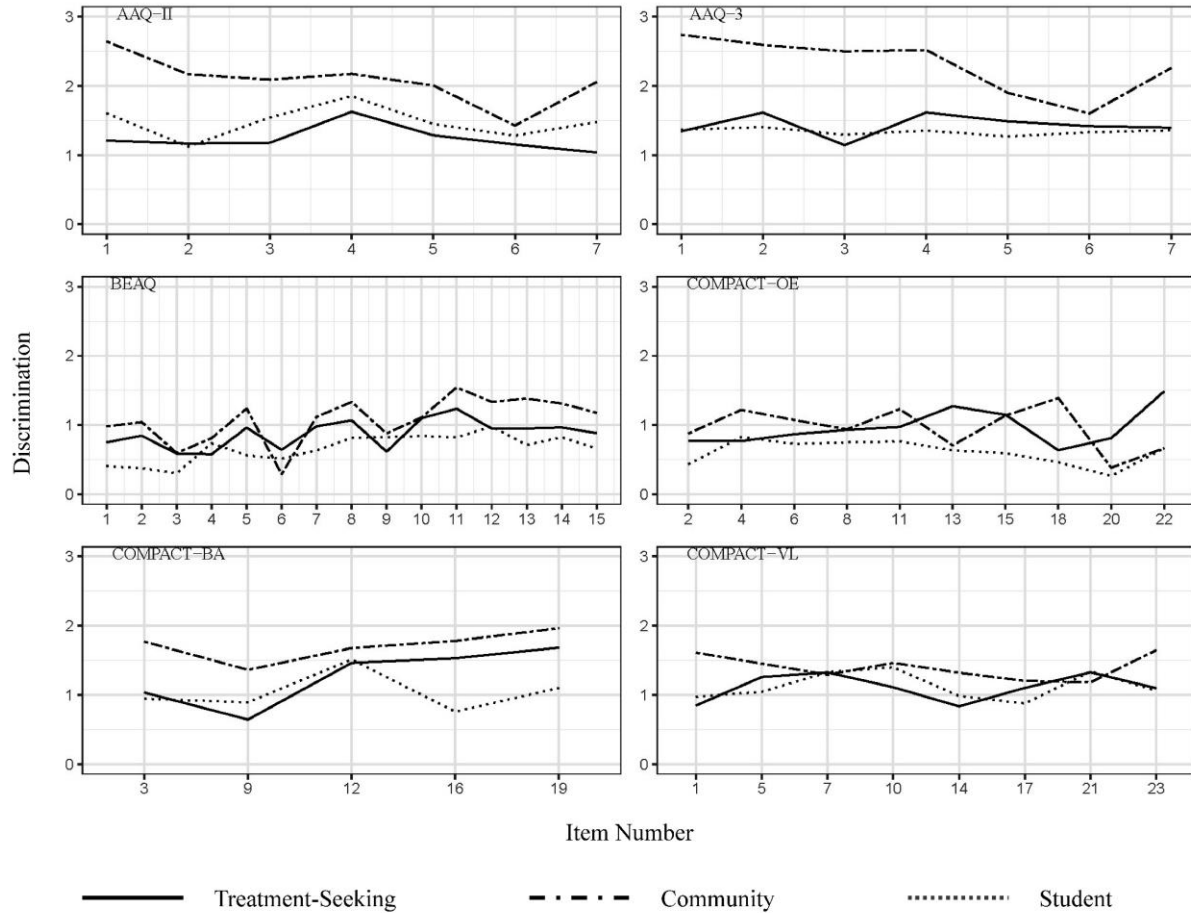*Figure 1*. Flowchart of removal of careless or insufficient effort responders from study samples.

Figure 2. Item discrimination parameters estimated based on the graded response model for each scale. Items are plotted along the x-axis and discrimination parameters along the y-axis. Larger discrimination values indicate greater differentiation of psychological (in)flexibility across adjacent item response categories.
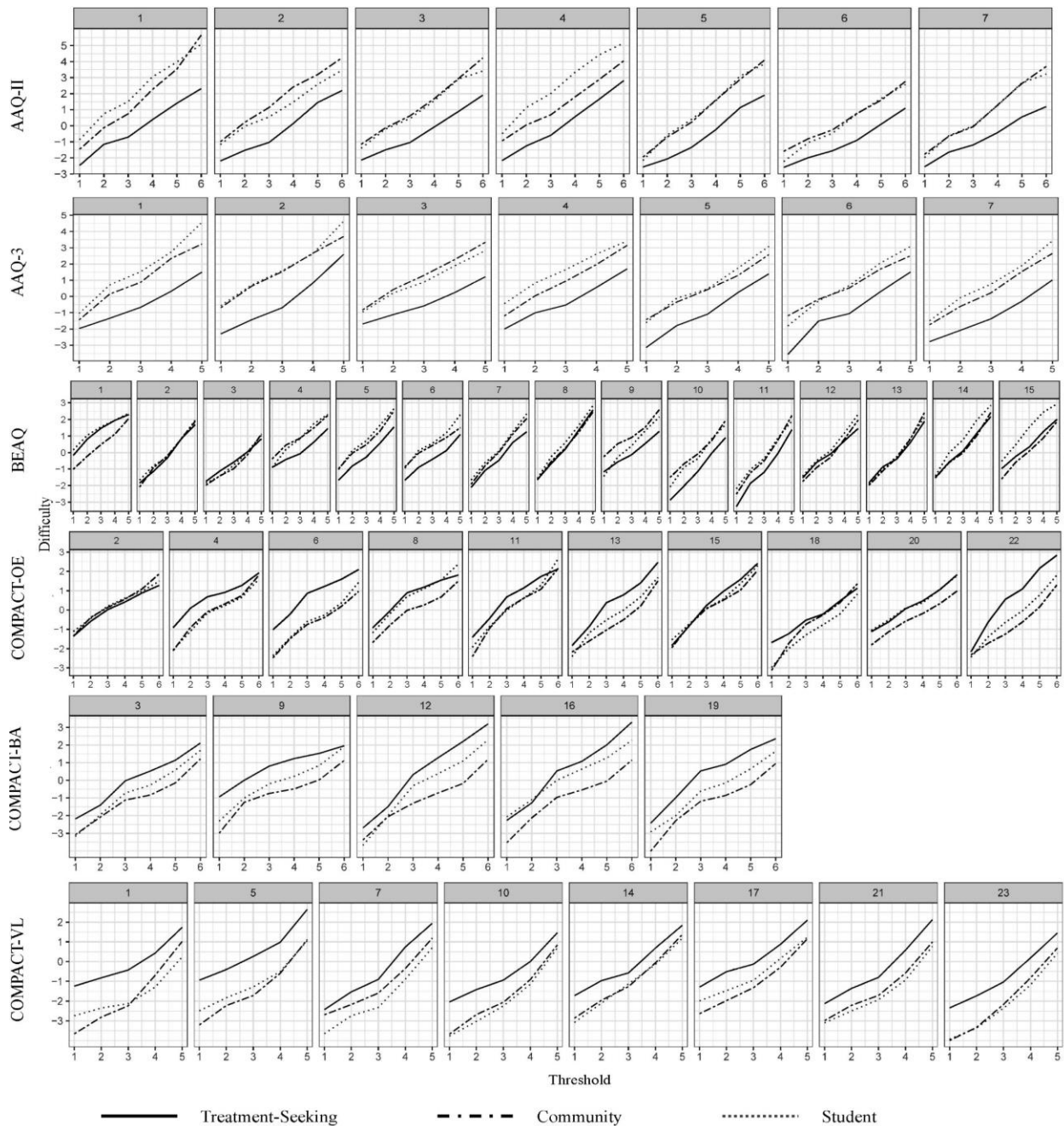
Figure 3. Item difficulty parameters estimated based on the graded response models for each scale. Item thresholds are plotted along the x-axis and item difficulties are plotted along the y-axis; figures are paneled by item of each scale. Thresholds refer to rgw junction between two response options (i.e., between responses of 1 and 2, 2 and 3, etc.), and the difficulties refer to the value of psychological (in)flexibility associated with a 50% probability of responding above the threshold. The difficulty parameters are assessed relative to each sample; for instance, a response of "4" to an item of the AAQ-II in the student sample may indicate high psychological inflexibility relative to other students, but the same response in the clinical sample may reflect low psychological inflexibility relative to other treatment-seeking people. In this case, the item would be more "difficult" for students versus people seeking treatment.

Appendix A
Acceptance and Action Questionnaire-3 (AAQ-3)

Below you will find a list of statements. Please rate how true each statement is for you by circling a number next to it. Use the scale below to make your choice.

1	2	3	4	5	6	7
never true	very seldom true	seldom true	sometimes true	frequently true	almost always true	always true

1. Painful worries, feelings, or memories make it impossible for me to live a meaningful life.	1 2 3 4 5 6 7
2. I'm so afraid of my feelings that I don't do things I care about.	1 2 3 4 5 6 7
3. I worry about losing control of my thoughts, feelings, or memories.	1 2 3 4 5 6 7
4. Painful memories prevent me from having a fulfilling life.	1 2 3 4 5 6 7
5. How I react to emotions causes problems in important areas of my life.	1 2 3 4 5 6 7
6. I do not handle my emotions well.	1 2 3 4 5 6 7
7. Worries, feelings, or memories keep me from moving toward my goals.	1 2 3 4 5 6 7

This is a one-factor measure of psychological inflexibility or experiential avoidance. Score the scale by summing the seven items. Higher scores equal greater levels of psychological inflexibility.

Appendix B
Comparison of AAQ-II and AAQ-3

Wording changes are italicized in the AAQ-3 column.

	AAQ-II	AAQ-3
1	My painful experiences and memories make it difficult for me to live a life that I would value.	Painful *worries, feelings*, or memories make it *impossible* for me to live a *meaningful life*.
2	I'm afraid of my feelings.	I'm so afraid of my feelings *that I don't do things I care about*.
3	I worry about not being able to control my worries and feelings.	I worry about *losing control* of my thoughts, feelings, or memories.
4	My painful memories prevent me from having a fulfilling life.	Painful memories prevent me from having a fulfilling life.
5	Emotions cause problems in my life.	*How I react to* emotions causes problems *in important areas of my life*.
6	It seems like most people are handling their lives better than I am.	*I do not handle my emotions well*.
7	Worries get in the way of my success.	Worries, *feelings, or memories keep me from moving toward my goals*.