Available online at www.sciencedirect.com
ScienceDirect
Behavior Therapy 42 (2011) 676–688
Behavior Therapy
www.elsevier.com/locate/bt

Preliminary Psychometric Properties of the Acceptance and Action Questionnaire–II: A Revised Measure of Psychological Inflexibility and Experiential Avoidance

Frank W. Bond
Goldsmiths, University of London

Steven C. Hayes
University of Nevada, Reno

Ruth A. Baer
University of Kentucky

Kenneth M. Carpenter
Columbia University, College of Physicians and Surgeons

Nigel Guenole
Goldsmiths, University of London

Holly K. Orcutt
Northern Illinois University

Tom Waltz
University of Nevada, Reno

Robert D. Zettle
Wichita State University

The present research describes the development and psychometric evaluation of a second version of the Acceptance and Action Questionnaire (AAQ-II), which assesses the construct referred to as, variously, acceptance, experiential avoidance, and psychological inflexibility. Results from 2,816 participants across six samples indicate the satisfactory structure, reliability, and validity of this measure. For example, the mean alpha coefficient is .84 (.78–.88), and the 3- and 12-month test–retest reliability is .81 and .79, respectively. Results indicate that AAQ-II scores concurrently, longitudinally, and incrementally predict a range of outcomes, from mental health to work absence rates, that are consistent with its underlying theory. The AAQ-II also demonstrates appropriate discriminant validity. The AAQ-II appears to measure the same concept as the AAQ-I (*r*=.97) but with better psychometric consistency.

Authorship order was alphabetically determined for Kenneth M. Carpenter, Nigel Guenole, Holly K. Orcutt, Tom Waltz, and Robert D. Zettle. We should like to thank the following people for their important contribution to this research: Laura Curtiss, Eric J. Fox, Magda Kulesza, Susan M. Orsillo, Lizabeth Roemer, and Kelly G. Wilson.

Address correspondence to Frank W. Bond, Ph.D., Department of Psychology, Goldsmiths, University of London, New Cross, London SE14 6NW, United Kingdom; e-mail: F.Bond@gold.ac.uk.

0005-7894/xx/xxx-xxx/$1.00/0

Keywords: psychological flexibility; experiential avoidance; acceptance; AAQ; psychological inflexibility

There is a broad and growing body of evidence that mental health and behavioral effectiveness are influenced more by how people relate to their thoughts and feelings than by their form (e.g., how negative they are). This basic finding has been shown in many specific areas. For example, in chronic pain, psychosocial disability is predicted more by the experiential avoidance of pain than by the degree of pain (McCracken, 1998). A number of concepts central to modern empirical clinical methods have emerged with this same basic theme, including distress tolerance (e.g., Brown, Lejuez, Kahler, & Strong, 2002; Schmidt, Richey, Cromer, & Buckner, 2007), thought suppression (e.g., Wenzlaff & Wegner, 2000), and mindfulness (Baer, 2003). This core insight is key to a number of the newer contextual cognitive behavior therapy (CBT) approaches to treatment such as mindfulness based cognitive therapy (MBCT; Segal, Williams, & Teasdale, 2001), dialectical behavior therapy (DBT; Linehan, 1993), metacognitive therapy (Wells, 2000), and acceptance and commitment therapy (ACT; Hayes, Strosahl, & Wilson, 1999). The purpose of the present paper is to examine the measurement of a concept that developed originally within ACT, and that seems to apply to other forms of contextual CBTs (e.g., see Rüsch et al., 2008).

The Acceptance and Action Questionnaire

The Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004) is the most widely used measure of experiential avoidance and psychological inflexibility. The original item pool for this short (9 to 16 items, depending on the version) Likert-style scale was generated by ACT therapists and researchers to represent the kind of phenomena that constitutes this unidimensional construct. As such, the final scale contained items on negative evaluations of feelings (e.g., "Anxiety is bad"), avoidance of thoughts and feelings (e.g., "I try to suppress thoughts and feelings that I don't like by just not thinking about them"), distinguishing a thought from its referent (e.g., "When I evaluate something negatively, I usually recognize that this is just a reaction, not an objective fact"), and behavioral adjustment in the presence of difficult thoughts or feelings (e.g., "I am able to take action on a problem even if I am uncertain what is the right thing to do.").

The AAQ has proven to be broadly useful. A meta-analysis of 27 studies that used this measure found that it predicted a wide range of quality-of-life outcomes (e.g., depression, anxiety, general mental health, job satisfaction, future work absence, and future job performance), with an average effect size of $r = .42$ (Hayes, Luoma, Bond, Masuda, & Lillis, 2006; see also Chawla & Ostafin, 2007). The AAQ shows these effects even after controlling for one or more individual characteristics, such as emotional intelligence, negative affectivity, thought suppression, social desirability, and locus of control (see Bond, Hayes, & Barnes-Holmes, 2006, for a review). Importantly, the AAQ does not just correlate with quality-of-life indices. Studies have shown that the AAQ mediates the impact of other coping processes such as cognitive reappraisal (Kashdan, Barrios, Forsyth, & Steger, 2006), moderates the effect of treatment (Masuda et al., 2007), and in some studies mediates the impact of ACT (Bond & Bunce, 2000; Flaxman & Bond, 2010). The AAQ also predicts dropout from DBT (Rüsch et al., 2008); in addition, reductions in experiential avoidance, as measured by the AAQ, predict corresponding reductions in depression among DBT patients seeking treatment for borderline personality disorder (Berking, Neacsiu, Comtois, & Linehan, 2009). Thus, the AAQ appears more broadly applicable to modern contextual CBT methods, not just ACT.

The success of the AAQ has led to a growing number of versions that are tailored to particular applied areas or specific populations, such as pain (McCracken, Vowles, & Eccleston, 2004), smoking (Gifford et al., 2004), diabetes management (Gregg, Callaghan, Hayes, & Glenn-Lawson, 2007), tinnitus (Westin, Andersson, & Hayes, 2008), weight (Lillis & Hayes, 2008), coping with epilepsy (Lundgren, Dahl, & Hayes, 2008), and coping with psychotic symptoms (Shawyer et al., 2007), among several others. So far, all of these specific versions work well in predicting outcomes within their respective areas and have been particularly effective as mediators of ACT interventions that target these specific problems (e.g., Gifford et al., 2004; Gregg et al., 2007; Lillis & Hayes, 2008; Lundgren et al., 2008). However, a more general AAQ that can be used in a wide variety of contexts remains important for studying this theoretical model and the processes that underlie therapeutic and behavioral change.

The Achilles' Heel of the AAQ-I: Comprehension and Reliability

In many studies, the internal consistency of the AAQ (which from here forward we will term the AAQ-I) has often been a problem. In an early validation study (Hayes et al., 2004), the alpha coefficient of this unidimensional measure was a just satisfactory .70, and its test–retest reliability was .64 over 4 months. In subsequent studies, alpha levels have sometimes been lower, especially with community samples and certain subpopulations (e.g., the less well educated, those who use English as a second language). The low alpha problem appears to result, at least in part, from unnecessary item complexity and the subtlety of the concepts addressed. For example, the AAQ-I item "I rarely worry about getting my anxieties, worries, and feelings under control," rated from 1 (*never true*) to 7 (*always true*), approaches a double negative. The item, "When I evaluate something negatively, I usually recognize that this is just a reaction, not an objective fact," can seem incomprehensible to persons not exposed to ACT or other contextual CBTs.

Perhaps as a result, the factor structure of the AAQ-I has been somewhat unstable. The original validation study identified 9- and 16-item single-factor versions (Hayes et al., 2004), but other research identified a two-factor 16-item version (Bond & Bunce, 2003). Thus, there is a need for the development of a more stable and psychometrically sound instrument.

THE ACT MODEL

In that context, it is important briefly to discuss the underlying theory driving the development of the instrument. When ACT was originally conceived, the overarching term for its model of psychological ill-health was "experiential avoidance"—the attempt to alter the form, frequency, or situational sensitivity of difficult private events (i.e., thoughts, feelings, and physiological sensations), *even when doing so* leads to actions that are inconsistent with one's values and goals (e.g., avoiding anxiety even when doing so prevents people from pursuing a long-held goal) (Hayes, Wilson, Gifford, Follette, & Strosahl, 1996). "Acceptance" was the term used to positively describe this model and was defined as the willingness to experience (i.e., not alter the form, frequency, or sensitivity of) unwanted private events, *in order to pursue* one's values and goals (e.g., being willing to feel fear in pursuit of a long-held goal; Hayes et al.).

These two terms were, and still are, very useful in highlighting how people's actions can be inflexibly, overly, and detrimentally determined by the avoidance of undesirable internal events at the expense of situational opportunities for pursuing one's values and goals (Hayes et al., 2006). However, when taken to represent the entire ACT model "acceptance" and "experiential avoidance" have unwanted features. For one, the focus of these terms is on how people respond to difficult thoughts, feelings, and physiological sensations, but these can include positive emotions (as when people avoid feelings of joy for fear of future disappointment); the term "experiential avoidance" can easily disguise that possibility. Behavioral effectiveness and living a vital life can also be inhibited when neutral or pleasant internal events decrease people's sensitivity to values-related contingencies that exist in a given situation. For example, believing that one is wonderful can reduce behavioral flexibility when mistakes are made; likewise, daydreaming about an upcoming holiday can decrease people's ability to respond to goal-related contingencies that are more important or pressing. Under such circumstances, people are not necessarily avoiding their internal events but their actions are disproportionally under the control of such events at the expense of values-related contingencies.

The ACT model has always maintained that depending upon the values-related opportunities afforded in a given situation people need to be *flexible* as to the degree to which they base their actions on current contingencies or their internal events—no matter whether those events are unwanted, wanted, or neutral. In order to highlight ACT's emphasis on flexibility, its underlying model has, over the past few years, been increasingly referred to as *psychological flexibility* or simply *flexibility* (e.g., Hayes et al., 1999). It is defined as the ability to fully contact the present moment and the thoughts and feelings it contains without needless defense, and, depending upon what the situation affords, persisting in or changing behavior in the pursuit of goals and values (Hayes et al., 2006). In contrast, *psychological inflexibility* (or *inflexibility*) entails the rigid dominance of psychological reactions over chosen values and contingencies in guiding action; this often occurs when people fuse with evaluative and self-descriptive thoughts and attempt to avoid experiencing unwanted internal events, which has the "ironic" effect of enhancing people's distress (e.g., Wenzlaff & Wegner, 2000), reducing their contact with the present moment, and decreasing their likelihood of taking values-based actions. In such a context, people feel buffeted by their uncontrollable and feared internal experiences.

Acceptance and experiential avoidance are examples of psychological flexibility and inflexibility, respectively, and it is still appropriate to use those terms; they refer to psychological stances and actions that people take when the present moment contains thoughts and feelings that people may not wish to contact; as a result, they are often used when discussing psychopathology and psychotherapy. However, ACT techniques are increasingly used to maximize behavioral effectiveness; for example, to facilitate job performance and sporting skills (e.g., Bond, Flaxman, & Bunce, 2008; Bond, Flaxman, van Veldhoven, & Biron, 2010). In many of these circumstances (but certainly not all), the avoidance of unwanted internal events is not necessarily ACT's main focus, rather it may be on identifying team values, improving problem solving, or enhancing contingency sensitivity and the like; in such cases, it is more appropriate to refer to ACT's attempts to increase psychological flexibility. For these reasons we primarily use this more general, overarching term to refer to ACT's model, but when we use acceptance and experiential avoidance, they can be understood as examples of psychological flexibility and inflexibility, respectively.

OVERVIEW OF THE PRESENT STUDIES

The overall aim of the three studies presented here was to address the shortcomings of the AAQ-I, as a measure of psychological flexibility, by developing a second version, which we will term the AAQ-II. The studies describe how we reexamined the measurement of this construct and investigated the initial psychometric properties of the AAQ-II. In Study 1, a panel of experts generated items that assessed psychological flexibility and inflexibility. We then eliminated those with low corrected item-total correlations because the goal was to create a theoretically derived, unidimensional scale that assessed flexibility and inflexibility. This item elimination process promotes satisfactory internal consistency, which is a primary reason for developing the AAQ-II, but it does not necessarily produce a unidimensional measure (Nunnally & Bernstein, 1994); therefore, we then carried out an exploratory factor analysis in order to establish the measure's structure. Based upon its results, in Study 2 we specified and tested a latent factor measurement model for the AAQ-II in three new samples from very different populations. Once the structure and reliability of the AAQ-II was established, in Study 3 we examined its predictive, concurrent, discriminant, convergent, and incremental validities using data obtained from six different samples.

Method

Study 1: Item Generation, Selection, and Exploratory Factor Analysis

AAQ-II Item Generation

A panel of 12 ACT researchers and practitioners from Australia, Europe, and the United States who had been key to the development of ACT and the AAQ-I, was established in order to generate items that followed from the domain of psychological flexibility/inflexibility. Specifically, panel members developed statements that stemmed from either the likely dominance or nondominance of internal events over contingencies in determining values-directed actions. (Dominance and nondominance of internal events represent inflexibility and flexibility, respectively.) These statements reflected an unwillingness to experience unwanted emotions and thoughts (e.g., "I'm afraid of my feelings"), the ability to be in the present moment (e.g., "I am in control of my life"), and commitment to flexible values-directed actions when experiencing psychological events that could undermine them (e.g., "My thoughts and feelings do not get in the way of how I want to live my life"). ACT's underlying theory (Hayes et al., 1999) and research on the AAQ-I (e.g., Hayes et al., 2004) suggest that these statements should be homogeneous with respect to the content of psychological flexibility.

As noted above, psychological flexibility is a subtle construct that can be difficult to convey in short statements that are understandable to people uninitiated in ACT, or contextual CBTs related to it. Thus, even though we wanted to produce a brief scale that could be used in settings in which time is limited (e.g., the workplace), we believed that we would need to test a relatively large number of items in order to eliminate those that failed to correlate adequately with the overall content domain due to problems of wording and the like. Following Nunnally and Bernstein (1994), we generated items in order to ensure content homogeneity (i.e., each item followed from the single domain of psychological flexibility/inflexibility); however, we also ensured that items were methodologically heterogeneous; to this end, we developed very similar items that differed only in how they were "keyed."

In the end, the panel produced 48 items that, it agreed, well represented the content of psychological flexibility. A smaller, subpanel of five ACT experts, including two originators of ACT (S. C. Hayes and K. G. Wilson), then rated each item in terms of its clarity and the degree to which the items sufficiently represented this construct. As a result of this process, the subpanel decided to reword three items and, in order to improve content validity, remove one item and add two others. This resulted in 49 items, all of which used a Likert-type scale that ran from 1 (*never true*) to 7 (*always true*), with higher scores indicating greater levels of psychological inflexibility.

We then asked 26 postgraduate students at Goldsmiths, University of London, who were familiar with the psychological flexibility construct, and 18 adults from a community sample in the United Kingdom, who were not familiar with it, to complete the 49-item measure and provide feedback with regard to the clarity and readability of the items. Based upon this feedback, we made grammatical changes to three items.

ITEM SELECTION AND FACTOR STRUCTURE

Participants and Procedure

Participants in the first study were 206 students from the University of Nevada, Reno; their mean age was 19 years ($SD = 3.57$), 65% were female, and 67% identified as white or Caucasian. The 49 items of the trial version of the AAQ-II was part of a larger packet of questionnaires for another project. In exchange, they received credit for a research participation requirement.

Results and Discussion

As the item pool was generated to reflect the single domain of psychological flexibility/inflexibility, we followed Nunnally and Bernstein's (1994) guidelines and first examined the corrected item-total correlations of the 49 items and eliminated those with a coefficient below .30. Using this criterion, we removed 22 questions. We then conducted an exploratory factor analysis on the remaining 27 items in order to examine their factor structure because even though a scale has content homogeneity, it does not follow that it is also unidimensional (Kline, 2005; Nunnally, 1978).

As the goal of this exploratory analysis was to identify one or more latent variables underlying the observed variables, we first conducted a common factor analysis (Floyd & Widaman, 1995) and determined the number of factors to extract through parallel analysis (Horn, 1965). Here, the number of factors selected is equal to the number of eigenvalues obtained that have values greater than those produced by random, uncorrelated data based on the same number of observations and variables as the original data set. Research indicates that parallel analysis is an accurate factor extraction procedure (Zwick & Velicer, 1986), and based upon it we retained three factors. We used an oblique rotation (Direct Oblimin), as we expected that these factors would be elements of a higher-order one, psychological flexibility (Nunnally, 1978), and therefore should be significantly correlated. To define the factors, we inspected the pattern matrix and eliminated any item that had a loading below .4 on all three factors (6 variables) or a loading of .4 or above on more than one factor (11 variables; Ferguson & Cox, 1993). We then reran the same extraction and rotation procedures on the remaining 10 items, and as can be seen in Table 1, we identified two distinct factors. The first had an eigenvalue of 4.64 and it accounted for 41.47% of the variance; the second had an eigenvalue of 1.06 and it accounted for 4.94% of the variance.

This ratio of the first to the second eigenvalue (i.e., 4.38) suggests that this is probably a unidimensional measure and, thus, is represented by only one factor (Reise, Morizot, & Hays, 2007; Robins, Fraley, & Kruger, 2007). This conclusion is supported by a number of other indices; specifically, all of the items on the first factor are "negatively" worded, and the three items comprising the second factor are all positively worded. As there does not appear to be any theoretical difference between the items on each factor, their differential loadings may indicate the presence of a method effect (Marsh, 1986): a source of indicator (or item) variability that is unrelated to the substantive, underlying dimension; in this case, the source may be differential response patterns to positively and negatively worded items. In such a case, a unidimensional construct, as we hypothesize psychological flexibility/inflexibility to be, may spuriously appear as two separate ones. Furthermore, the intercorrelation between the two factors is

Table 1
Factor Loadings From Principal Axis Factoring, Means, Standard Deviations, and Alpha Levels ($N = 206$)

AAQ-II Item	Two-Factor Solution		One-Factor Solution
	Factor 1	Factor 2	
1. It's OK if I remember something unpleasant. [a]	.18	**.34**	-
2. My painful experiences and memories make it difficult for me to live a life that I would value.	**.61**	.14	**.70**
3. I'm afraid of my feelings.	**.75**	-.05	**.70**
4. I worry about not being able to control my worries and feelings.	**.72**	-.03	**.71**
5. My painful memories prevent me from having a fulfilling life.	**.77**	.09	**.82**
6. I am in control of my life. [a]	.25	**.40**	-
7. Emotions cause problems in my life.	**.89**	-.16	**.79**
8. It seems like most people are handling their lives better than I am.	**.55**	.17	**.65**
9. Worries get in the way of my success.	**.42**	.29	**.59**
10. My thoughts and feelings do not get in the way of how I want to live my life. [a]	-.07	**.69**	-
Percent explained variance	41.47	4.94	50.68
Scale mean	30.69		21.41
Scale *SD*	9.91		7.97
Coefficient α for scale	.87		.88

Note. Coefficients in bold load onto the corresponding factor.
[a] Item reversed for scoring purposes.

.5, which also suggests that the two factors represent the same construct (Clark & Watson, 1995).

Recent psychometric research has shown that items of opposite valence function poorly in the context of an overall scale (e.g., Credé, Chernyshenko, Bagraim, & Sully, 2009), and so we explored the psychometric functioning of the three positively worded items on the second factor. Their reliability was poor: Cronbach α was .55, as compared to .88 for the 7 negatively worded items that constituted the first factor, and .87 for all 10 items together. In addition, we examined category response curves for all 10 items. Category response options represent the probability of responding at each interval of an item, depending upon the person's overall level on the characteristic being examined: in this case, psychological flexibility. For a well-performing item, category response curves are narrow and peaked, indicating that the points on the scale clearly differentiate among overall levels of the characteristic. Poorly performing items have flatter curves and do not differentiate very well among levels of a characteristic. Results (available from the corresponding author) indicated that the three items with the poorest category response curves were the three positively worded items that comprised the second factor. Perhaps most importantly, though, and which we discuss in Study 3 below, the predictive validity of the 7 items on the first factor was not at all enhanced, for any criterion, when we formed a 10-item AAQ-II by adding the 3 positively worded items on the second factor (see Table 4); in addition, when we partialed the three positively worded items from the association between the seven negatively worded items and each criterion in Table 4, the correlation coefficient in every case was 1.00. Thus, for these reasons of internal and external validity, we did not retain the three items on the second factor.[1]

We reran the same extraction and rotation procedures, described above, on the seven negatively worded items; and as can be seen in the last column of Table 1, we identified one factor that accounted for 50.68% of the variance and had a Cronbach α coefficient of .88. The mean and standard deviation of this measure for Sample 1 can be seen in Table 1. Study 2 used confirmatory factor analysis to test the fit and parameters of this unidimensional measure.

[1] Non-English translations and empirical studies have already appeared using the 10-item version of the AAQ-II. While the data we report here suggests that the 7-item version is psychometrically stronger, the 10-item version is not significantly weaker predictively (see Table 4), and the 7- and 10-item versions correlate at $r=.96$; thus it should not be assumed that studies conducted with the 10-item version are invalid.

Study 2: Confirmatory Factor Analyses

Unlike exploratory factor analysis, confirmatory factor analysis allows for the a priori specification and then testing of different parameters of latent factor measurement models (Kline, 2005). These include the number of factors, the indicators that reflect those dimensions, and covariances among them. Confirmatory factor analysis also accounts for measurement error, or variance in indicator scores that is not explained by one or more factors; correlations among these errors can be specified. Doing so allows one to test hypotheses that certain indicators are jointly affected by a variable that is not associated with an underlying factor (Kline). These method effects can result from sources such as high content overlap, similar item phrasings, differential susceptibility to demand characteristics, and differential ease in reading or understanding positively and negatively worded items (Brown, 2003; Byrne, Shavelson, & Muthen, 1989). We tested the factor structure of the AAQ-II using three new samples. We also used these samples to test for measurement invariance, or whether the seven AAQ-II items operate equivalently across these groups.

PARTICIPANTS AND PROCEDURES

Three new samples participated in this study. Sample 2 was comprised of 433 undergraduate students from the University of Kentucky; they had a mean age of 21 ($SD = 3.54$) years, 68% were female, and 90% identified as white. They completed the 49-item trial version of the AAQ-II as part of a larger packet of questionnaires for another project. In exchange, they received credit for a research participation requirement.

Sample 3 was comprised of 290 people who were seeking outpatient psychological treatment for substance misuse in a New York City university hospital; their mean age was 39 years ($SD = 10.20$), 43% were female, 37%, 29%, and 28% identified as Caucasian, Hispanic, and Black, respectively. They completed the 49-item trial version of the AAQ-II as part of larger intake packet that respondents knew would be used for research purposes.

Participants in Sample 4 were 583 employees of a UK retail bank; they were located in offices across England and Scotland and had a mean age of 34 years ($SD = 9.76$), 58% were female, and 97% identified as white. Sample 4 was reassessed 3- and 12-months later as well. These participants completed the AAQ-II as part of a larger packet of questionnaires for another project. Thirty-nine percent of these respondents left formal education at 16 with qualifications (O-levels/General Certificate of Secondary Education), 40% left formal education at 18 with qualifications (A-levels), 13% obtained an undergraduate degree, 0.4% had a postgraduate qualification, and 7% had some "other" type of qualification (e.g., a National Vocational Qualification). In comparison to UK norms, approximately 23% fewer people in this sample obtained a first university degree (i.e., a bachelor's degree; Office for National Statistics, 2001).

ANALYSES

For each sample, covariance matrices were used to analyze the measurement models, and maximum likelihood estimation was used to assess their fit. (These analyses were conducted with the structural equation modeling software program AMOS 5; Arbuckle, 2003). As the chi-square (χ^2) statistic is very sensitive to sample size and may overestimate the lack of model fit (Bollen, 1989), we selected four additional indicators, based upon Bollen, and Hu and Bentler (1998). The first was the normed chi-square (*NC*), which is the chi-square value divided by the degree of freedom (χ^2/df). Consistent with Bollen, we specified a value of 3 as indicating good model fit. The other three fit indicators were the root-mean-square error of approximation (*RMSEA*), the standardized root-mean-square residual (*SRMR*), and the comparative fit index (*CFI*). Hu and Bentler suggest that the values of .06, .08, and .95 are, respectively, indicative of good model fit.

As Brown (2003) notes, it is frequently necessary to specify correlated measurement errors among items that have similar content and, in particular, use the same key terms (e.g., the word "painful"). For the AAQ-II, we specified correlated measurement errors between Items 2 and 5 (i.e., respectively, "My painful experiences and memories make it difficult for me to live a life that I would value" and "My painful memories prevent me from having a fulfilling life"). We suspected that the content of this item pair was so similar (even involving the same key terms) that it could be influenced by method effects.

Results and Discussion

Before conducting the confirmatory factor analyses on each sample, we tested their data for univariate and multivariate normality. All items, and AAQ-II total scores, were in acceptable ranges (Muthen & Kaplan, 1985).

The one-factor model, which specified method effects between the two "painful" items (2 and 5), fit the data very well in all samples (see Table 2). For example, *RMSEA* was not above .06 in any sample (i.e., .06, .04, and .05, respectively), and the *CFI* was above .95 (i.e., .96, .99, and .98, respectively). As Table 2 indicates, all other indices also indicated good model fit for all samples.

All unstandardized factor loadings were significant at $p < .000$ and ranged from .75 to 1.61 (see Table 3). In each sample, the unstandardized covariance between the item pair with similar content (which ranged from .16–.88) was significant at $p < .000$.

Scale means and standard deviations can be seen in Table 3, as can their alpha coefficients, which were in the acceptable range (i.e., .78–.88). Findings from Sample 4 showed that the 3- and 12-month test–retest reliabilities for the AAQ-II were very acceptable at .81 and .79, respectively.

Consistent and encouraging results across three very different samples—U.S. university students, those receiving treatment for substance misuse in New York City, and UK financial services workers—suggest the same conclusion: Covariance among the seven AAQ-II items is due to a single latent dimension: psychological inflexibility, or experiential avoidance.

Table 2
Confirmatory Factor Analysis Results for the AAQ-II in Three Samples

Model	χ^2	*df*	χ^2_{diff}	Δdf	*NC*	*RMSEA*	*SRMR*	*CFI*
					(≤3)	(≤.06)	(≤.08)	(≥.95)
Sample 2 (*N*=433)	38.70**	13			2.97	.06	.04	.96
Sample 3 (*N*=290)	19.41	13			1.49	.04	.03	.99
Sample 4 (*N*=583)	25.79*	13			1.98	.05	.04	.98
Measurement Invariance Across Samples 2–4								
Baseline	104.22**	39			2.52	.04	.05	.95
Equality constraints (Item 4 fixed to 1.0)	116.06**	53	11.84 (*ns*)	14	2.25	.05	.06	.96
Equality constraints (Item 2 fixed to 1.0)	125.38**	53	21.16 (*ns*)	14	2.31	.04	.05	.95

Note. *NC* = normed chi-square; *RMSEA* = root–mean–square error of approximation; *SRMR* = standardized root–mean–square residual; *CFI* = comparative fit index; values in parentheses define good model fit for the respective fit index; *ns* = nonsignificant.
$^*p < .05$, $^{**}p < .001$.

Table 3
Unstandardized Factor Loadings From Confirmatory Factor Analyses, Means, Standard Deviations, and Alpha Coefficients in Three Samples

AAQ-II Item	Sample 2		Sample 3		Sample 4	
	Factor loading	*SE*	Factor loading	*SE*	Factor loading	*SE*
2	1.000		1.000		1.000	
3	1.54	.18	1.35	.13	1.61	.14
4	1.46	.18	1.17	.12	1.54	.13
5	1.33	.14	1.22	.10	1.32	.14
7	1.16	.15	1.10	.11	1.18	.11
8	0.89	.13	0.75	.10	.88	.12
9	1.06	.14	0.87	.10	1.06	.12
Scale mean	17.34		28.34		18.53	
Scale *SD*	4.37		9.92		7.52	
α coefficient	.78		.87		.81	

Note. All factor loadings are significant at $p < .001$.

MEASUREMENT INVARIANCE

Finally, in order to determine further whether the seven items of the AAQ-II assess psychological inflexibility in a similar manner across different samples, we compared the relative fit of two models in Samples 2–4. The first allowed the seven unstandardized factor loadings and one error covariance to vary across the three samples, and the second placed equality, or invariance, constraints on those loadings and error covariance. If the constrained model does not generate a significantly worse fit than the unconstrained model, the items are likely to be assessing the same construct in a comparable way (Byrne, 2001). Equality constraints were not placed on estimates of the factor variances, since these are known to vary across groups even when the indicators are measuring the same construct in a similar manner (Kline, 2005; MacCallum & Tucker, 1991). As can be seen in Table 2, the baseline model fit the data well; for example, *NC*, *RMSEA*, *SRMR*, and *CFI* values all suggested a good-fitting solution. When we placed equality constraints on the factor loadings and error covariance, there was not a significant decrement in goodness of fit (as assessed by the χ^2_{diff} test), suggesting that the measures were invariant across the three samples.

In order to statistically identify the measurement models just tested for invariance, the same indicator had to be set to 1.0 in each sample: Item 4 was randomly selected. As a result, that item could not be tested for group differences because it was a constant. This leaves open the possibility that Item 4 does not load in a similar pattern across the three samples. To test for this possibility, we reanalyzed the model by freeing Item 4 to be estimated and fixing the loading of another indicator to 1.0 (Kline, 2005): Item 2 was randomly selected. Once again, the model with and without equality constraints fit the data equally well (Table 2), suggesting that the seven AAQ-II items measure psychological inflexibility in a comparable manner in these three very different samples.

Study 3: Concurrent, Predictive, Discriminant, Convergent, and Incremental Validities

The two previous studies provided support, across four samples, for the factorial validity and internal consistency of the AAQ-II; in addition, findings from Sample 4 showed good test–retest reliability for this measure. The next step in assessing the construct validity of the AAQ-II was to ensure that it correlated with theoretically expected outcomes (concurrent and predictive validity); that it was significantly related to similar constructs (convergent validity); that it was not strongly related to theoretically distinct constructs (i.e., discriminant validity); and that it demonstrated theoretical or practical utility above and beyond established measures of related constructs (incremental validity; Cronbach & Gleser, 1957).

Regarding concurrent and predictive validities, psychological inflexibility, or experiential avoidance, is purported to be an important determinant of psychological distress and behavioral ineffectiveness (Hayes et al., 1996, 2006); thus greater levels of inflexibility, indicated by higher AAQ-II scores, should be related to greater emotional distress (e.g., worse general mental health, as well as higher levels of depression, anxiety, and stress), and poorer life functioning (e.g., show more absences from work).[2]

Concerning convergent validity, higher scores on the AAQ-II should correlate with greater levels of thought suppression (e.g., the White Bear Suppression Inventory; Wegner & Zanakos, 1994), as thought suppression is one indicator of psychological inflexibility or experiential avoidance. Unlike thought suppression, however, psychological inflexibility can involve avoidance of all internal events, for example, emotions (e.g., “I'm afraid of my feelings”); thus the relationship between the two measures should not be so strong as to suggest that they are assessing the same construct.

Finally, we examined whether psychological inflexibility was associated with social desirability.

[2] In order to minimize the length of the manuscript, other validity measures were eliminated from this report, including bankers' sales figures, learning achievement, job satisfaction, turnover intention, openness to experience, conscientiousness, extraversion, and agreeableness. Effects all fit well with the underlying theory, and findings are available from the corresponding author.

As we predicted that no such significant association would exist, this test served to examine the discriminant validity of the AAQ-II.

PARTICIPANTS

All four samples described above comprised the participants of this study to which data from two additional samples were added: Sample 5 consisted of 872 employees of a financial services organization in the United Kingdom; 58% of the respondents were female, their mean age was 35 ($SD = 9.47$), and 95% identified as white. Sample 6 consisted of 432 students from Northern Illinois University; they had a mean age of 19 years ($SD = 2.63$), 100% were female, and 72% identified as white. All samples completed the AAQ-II, and with the exception of Sample 2, they completed at least one additional measure; however, Sample 2 did provide data on the association between the AAQ-II and the demographic variables, age, gender, and race.

MEASURES

Each of the following, well-validated, and widely used measures was used to assess the various forms of validity, just discussed: Beck Depression Inventory-Second Edition (BDI-II; Beck, Steer, & Brown, 1996), Beck Anxiety Inventory (BAI; Beck & Steer, 1990), Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1995), General Health Questionnaire-12 (GHQ; Goldberg, 1978), Global Severity Index of the Symptom Checklist-90-Revised (SCL-90-R-GSI; DeRogatis, 1992), White Bear Suppression Inventory (WBSI; Wegner & Zanakos, 1994), General Job Satisfaction scale (Hackman & Oldham, 1975), the number of occasions of full-day absences from work for each participant in a retail bank was obtained from that banks' human resources department), and the Marlowe-Crowne Social Desirability scale (Crowne & Marlowe, 1960).

Results and Discussion

As can be seen in Table 4, the results were consistent with our predictions. Higher levels of psychological inflexibility, or experiential avoidance, were concurrently associated with greater depressive symptoms (on both the BDI-II and the DASS), more anxiety-related symptoms (on both the BAI and the DASS), more stress (on the DASS), and greater overall psychological ill health (on both the GHQ and SCL-90-GSI).[3] It is possible that these correlations are due to common method variance: the propensity to answer similarly to multiple scales even when there is no true correlation among their constructs (Campbell & Fiske, 1959). This likelihood was first examined by considering the degree to which the AAQ-II was associated not just with other self-report

Table 4
Correlations Between the AAQ-II and Other Measures

Measure	Sample	*N*	*r* With AAQ-II	*r* With AAQ-II + Three Omitted Items
AAQ-I	1	206	.97**	.82*
BDI-II	1	206	.71**	.71*
	3	281	.70**	.69*
BAI	1	206	.61**	.58*
DASS: Depression	6	432	.61**	.61*
Anxiety	6	432	.49**	.51*
Stress	6	432	.57**	.54*
GHQ-12	1	206	.30**	.30*
	4	583	.34* [a]	.32* [a]
	5	872	.53*	.51*
SCL-90-R–GSI	1	206	.70**	.65*
WBSI	1	206	.63**	.60*
	4	583	.59*	.58*
	5	872	.59*	.57*
Absence occasions from work	4	583	.25* [a]	.25* [a]
MCSD	5	872	-.09	-.09
Age	1	206	.13	.14
	2	427	.10	.10
	3	281	-.09	-.10
	4	583	-.07	-.09
	5	872	-.08	-.08
Gender	1	206	-.19	-.20*
	2	429	-.04	-.03
	3	289	-.09	-.10
	4	583	-.04	-.04
	5	872	-.07	-.07
Race	1	206	.03	.01
	2	427	-.03	-.03
	3	290	.10	.02
	4	583	-.02	-.03
	5	872	-.02	-.02
	6	432	-.04	-.03

Note. The final column shows the correlation coefficient of each measure with the 10-item AAQ-II scale seen in Table 1. Three of those items were omitted from the final version of the AAQ-II. BDI-II = Beck Depression Inventory-II; BAI = Beck Anxiety Inventory; DASS = Depression Anxiety Stress Scales; GHQ-12 = General Health Questionnaire, 12-item version; SCL-90-R–GSI = Symptom Checklist-90 items-Revised–Global Severity Index; WBSI = White Bear Suppression Inventory; the number of occasions of full-day absences from work for each participant in a retail bank was obtained from that bank's human resources department; MCSD = Marlowe-Crowne Social Desirability Scale; gender was coded so that 1 = female, 2 = male; race was coded so that 1 = white/Caucasian, 2 = not white/not Caucasian. To minimize family-wise Type I error, we set the alpha level significant at .01.
$^{*}p < .01$, $^{**}p < .001$.
[a] AAQ-II predicts the criterion 1 year later.

[3] As can be seen in Table 4, all of these correlations are nearly identical to those that result from an AAQ that includes all of the 10 items listed in Table 1. This indicates that the three positively worded items are redundant in terms of criterion-related and predictive validities; this further supports the decision to omit them from the AAQ-II.

Table 5
Confirmatory Factor Analyses That Test for the Distinction of the AAQ-II and BDI-II

Model	χ^2	*df*	χ^2_{diff}	Δdf	*NC* (≤3)	*RMSEA* (≤.06)	*SRMR* (≤.08)	CFI (≥.95)
Sample 1								
AAQ-II and BDI-II as two factors	719.94*	349			2.06	.07	.07	.86
AAQ-II and BDI-II as one factor	1051.49*	350	331.55*	1	2.73	.10	.32	.74
Sample 3								
AAQ-II and BDI-II as two factors	683.49*	349			1.96	.06	.07	.89
AAQ-II and BDI-II as one factor	1197.69*	350	514.20*	1	3.42	.09	.31	.75

Note. *NC* = normed chi-square; *RMSEA* = root–mean–square error of approximation; *SRMR* = standardized root–mean–square residual; *CFI* = comparative fit index; values in parentheses define good model fit for the respective fit index; BDI-II = Beck Depression Inventory-II.
*$p<.001$.

measures but also with different behavior sets. In particular, we found that the population seeking treatment for substance misuse was more psychologically inflexible (Sample 3: $M=28.34$, $SD=9.92$) than the samples that did not have that behavioral status ($M=18.51$, $SD=7.05$). This difference was statistically significant, $F(1,\ 3282)=362.21$, $p<.000$), with a very large effect size ($d=1.12$).[4] This finding indicates that the AAQ-II yields substantially different scores in groups of people who are engaged in very different behaviors, which undermines common method variance as an explanation for its predictive validity.

Further challenging such an explanation are findings (seen in Table 4) that greater levels of inflexibility were associated both with greater psychological distress (GHQ) 1 year later, and more occasions of full-day work absence over the following year (taken from absentee records, not participant self-report). Because the AAQ-II can longitudinally predict both a self-report measure, and an objectively measured variable, it is once again unlikely that common method variance is the primary reason why correlations are seen between it and other measures.

As can be seen in Table 4, the associations between the AAQ-II and the BDI-II were the strongest convergent and predictive relationships ($r=.71$ and .70 in Samples 1 and 3, respectively). (This is similar to findings for the AAQ-I; Hayes et al., 2006). These correlation coefficients are not so high as to suggest that these measures are assessing the same construct (Nunnally & Bernstein, 1994), but in order to directly test the hypothesis that the AAQ-II and BDI-II are measuring separate constructs, we conducted a confirmatory factor analysis (using AMOS 5.0; Arbuckle, 2003) following procedures outlined by Kline (2005). As can be seen in Table 5, for both Samples 1 and 3, the model specifying the AAQ-II and the BDI-II as representing different latent variables had a significantly better fit than the one specifying both latent variables as the same construct. For both samples, treating the constructs as distinct provided an adequate fit to the data (e.g., Sample 1: $RMSEA=.07$, $SRMR=.07$; Sample 3: $RMSEA=.06$, $SRMR=.07$), whereas treating them as a single construct did not produce a good fit (e.g., Sample 1: $RMSEA=.10$, $SRMR=.32$; Sample 3: $RMSEA=.10$, $SRMR=.31$).

Regarding convergent validity, the AAQ-II was, as hypothesized, positively correlated with the WBSI in three samples (Table 4), thus providing evidence for this type of validity. The AAQ-II was not significantly associated with social desirability. This provides evidence for discriminant validity, as it suggests that participants' responses to the AAQ-II were not influenced by any need that they had to react in a culturally appropriate and acceptable manner (Nunnally & Bernstein, 1994).

The AAQ-II was not designed as a tool for diagnosing mental disorders. Measures that are intended for such use (e.g., the BDI-II and GHQ) are necessarily based upon the symptoms that define that disorder, whereas the AAQ-II was designed to assess a specific model of psychopathology that emphasizes psychological inflexibility. Accordingly, this measure was not devised to establish a cutoff point at which people are likely to meet the criteria for a diagnosable psychiatric disorder. Nevertheless, we were able to identify a range of AAQ-II scores that is associated with the cutoff points of three measures of psychopathology. To do this, we inserted the relevant values into a regression equation to find the AAQ-II score (Y′) that is predicted by the cutoff score (X) of the BDI-II (using Sample 1), the GHQ-12 (using Samples 1, 4, and 5), and the GSI scale of the SCL-90-R (using Sample 1). As can be seen in Table 6, these cutoff values predicted AAQ-II scores that ranged from 24 to 28. Thus, scores in this range or above are associated with GHQ-12, BDI-II, and GSI values

[4] Consistent with the findings shown in Table 4, this effect size is very similar to the one that results when these two groups are compared against an AAQ that includes the 10 items that are listed in Table 1 (i.e., $d=1.18$).

Table 6
Predicting AAQ-II Scores From Values (or Cutoff Scores) That Indicate Psychological Distress

Predictor	Sample	AAQ-II Score
BDI-II cutoff score of 14	1	28
GHQ cutoff score of 4	1	24
	4	25
	5	25
GSI cutoff score of .6	1	28

Note. BDI-II = Beck Depression Inventory-Second Edition; GHQ = General Health Questionnaire-12; GSI = Global Severity Index of the Symptom Checklist-90-Revised. Cutoff score is the value at which a psychological disorder is indicated on the given measure. These values were obtained from Beck, Steer, and Brown (1996) for the BDI-II; Goldberg (1978) and Schmitz, Kruse, Heckrath, Alberti, and Tress (1999) for the GHQ-12; and Schmitz et al. (1999) for the GSI.

that indicate psychological distress. This range of 24–28 falls between the mean AAQ-II scores of the sample that is seeking treatment for substance misuse, 28.34, and the samples that are not, 18.51. Taken together, these findings present a consistent, albeit limited and preliminary, indication of which scores on the AAQ-II may indicate a clinically relevant level of distress.

The AAQ-II was not associated with age, gender, or race (categorized as white or not white) across the diverse samples that assessed this demographic (Table 4). These concurrent, predictive, and discriminant validity findings for the AAQ-II closely reflect those that were found for the AAQ-I (e.g., Bond & Flaxman, 2006; Hayes et al., 2004). This would be expected, as the AAQ-II is intended to assess the same construct, only more reliably. That it does so is more directly indicated by results from Sample 1, which show that the correlation between the AAQ-II and the AAQ-I was .97 (Table 4).

General Discussion

The overall aim of this research was to begin examining the psychometric properties of a second version of the AAQ, and the results from these three studies, across six samples with a total of 2,816 participants, provide promising evidence as to the adequate structure, reliability, and validity of this measure. To elaborate, after ACT experts generated an AAQ-II item pool with good content validity, corrected item-total correlations and an exploratory factor analysis suggested a two-factor solution for a 10-item scale. Notably, however, the second factor consisted of only the three positively worded items on the scale, thus suggesting that the second factor resulted from a method effect and did not represent a second substantive dimension. Various tests comparing the internal and external validities of a 7- and 10-item scale also led us to reject a two-factor solution and so we did not retain the three items on the second factor. Thus, as hypothesized, and consistent with the AAQ-I, the AAQ-II appears to be a unidimensional measure that assesses the construct of psychological inflexibility, and results indicate that it does so in a comparable manner across very different samples.

The primary, immediate need for the AAQ-II is that the AAQ-I shows insufficient levels of reliability in various populations. Findings from these studies indicate that the reliability of the AAQ-II is consistently above the AAQ-I, with a mean alpha coefficient across the six samples of .84 (.78–.88), and the 3- and 12-month test–retest reliability is .81 and .79, respectively.

In addition to its sound factor structure and good reliability, findings from these studies indicate that the AAQ-II is associated with variables to which it is theoretically tied. Specifically, higher levels of psychological inflexibility, as measured by the AAQ-II, are related to greater levels of depression, anxiety, stress, and overall psychological distress. Beyond mere association, however, results indicate that higher levels of psychological inflexibility may serve as a risk factor for mental ill health, as higher scores on the AAQ-II predicted, 1 year later, greater psychological distress. The AAQ-II is not just associated with other self-report measures, however. Findings show that it also longitudinally predicts occasions of workplace absence over a 1-year period. As predicted, psychological inflexibility, or experiential avoidance, was not significantly associated with social desirability.

The present studies provide preliminary evidence that the AAQ-II has psychometric properties that, in comparison to the AAQ-I, are stronger and more stable across different groups. Despite its variable internal consistency, the AAQ-I has performed very well in relation to its underlying theory (Chawla & Ostafin, 2007; Hayes et al., 2006); further research will be needed to more fully examine the psychometric properties and usefulness of this new version of the AAQ. The present data provide a strong beginning of that process.

References

Arbuckle, J. L. (2003). *Amos 5.0 update to the Amos user's guide*. Chicago: Small Waters.

Baer, R. A. (2003). Mindfulness training as a clinical intervention: A conceptual and empirical review. *Clinical Psychology: Science and Practice*, *10*, 125–143.

Beck, A. T., & Steer, R. A. (1990). *Manual for the Beck Depression Inventory*. San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Brown, G. (1996). *Beck Depression Inventory II manual*. San Antonio, TX: Psychological Corporation.

Berking, M., Neacsiu, A., Comtois, K. A., & Linehan, M. M. (2009). The impact of experiential avoidance on the reduction of depression in treatment for borderline personality disorder. *Behaviour Research and Therapy*, *47*, 663–670.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bond, F. W., & Bunce, D. (2000). Mediators of change in emotion-focused and problem-focused worksite stress management interventions. *Journal of Occupational Health Psychology*, *5*, 156–163.

Bond, F. W., & Bunce, D. (2003). The role of acceptance and job control in mental health, job satisfaction, and work performance. *Journal of Applied Psychology*, *88*, 1057–1067.

Bond, F. W., & Flaxman, P. E. (2006). The ability of psychological flexibility and job control to predict learning, job performance, and mental health. *Journal of Organizational Behavior Management*, *26*, 113–130.

Bond, F. W., Flaxman, P. E., & Bunce, D. (2008). The influence of psychological flexibility on work redesign: Mediated moderation of a work reorganization intervention. *Journal of Applied Psychology*, *93*, 645–654.

Bond, F. W., Flaxman, P. E., van Veldhoven, M. J. P. M., & Biron, M. (2010). The impact of psychological flexibility and acceptance and commitment therapy (ACT) on health and productivity at work. In J. Houdmont & S. Leka (Eds.). *Contemporary occupational health psychology: Global perspectives on research, education, and practice* Chichester, UK: Wiley-Blackwell.

Bond, F. W., Hayes, S. C., & Barnes-Holmes, D. (2006). Psychological flexibility, ACT and organizational behavior. In S. C. Hayes, F. W. Bond, D. Barnes-Holmes, & J. Austin (Eds.). *Acceptance and mindfulness at work* (pp. 25–54). Binghamton, NY: Haworth Press.

Brown, R. A., Lejuez, C. W., Kahler, C. W., & Strong, D. R. (2002). Distress tolerance and duration of past smoking cessation attempts. *Journal of Abnormal Psychology*, *111*(1), 180–185.

Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects. *Behavior Research and Therapy*, *41*, 1411–1426.

Byrne, B. M. (2001). *Structural equation modeling with Amos: Basic concepts, applications and programming*. Mahwah, NJ: Erlbaum.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial invariance. *Psychological Bulletin*, *105*, 456–466.

Campbell, D. T., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Chawla, N., & Ostafin, B. (2007). Experiential avoidance as a functional dimensional approach to psychopathology: An empirical review. *Journal of Clinical Psychology*, *63*, 871–890.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319.

Credé, M., Chernyshenko, O. S., Bagraim, J., & Sully, M. (2009). Contextual performance and the job satisfaction–dissatisfaction distinction: Examining artifacts and utility. *Human Performance*, *22*, 246–272.

Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions*. Urbana: University of Illinois Press.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349–354.

DeRogatis, L. R. (1992). *The SCL-90-R: Administration scoring and procedure manual II*. Baltimore: Clinical Psychometric Research.

Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International Journal of Selection and Assessment*, *1*, 84–94.

Flaxman, P. E., & Bond, F. W. (2010). A randomised worksite comparison of acceptance commitment therapy and stress inoculation training. *Behaviour Research and Therapy*, *48*, 816–820.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286–299.

Gifford, E. V., Kohlenberg, B. S., Hayes, S. C., Antonuccio, D. O., Piasecki, M. M., Rasmussen-Hall, M. L., & Palm, K. M. (2004). Applying a functional acceptance based model to smoking cessation: An initial trial of acceptance and commitment therapy. *Behavior Therapy*, *35*, 689–705.

Goldberg, D. (1978). *Manual of the General Health Questionnaire*. Windsor, UK: National Foundation for Educational Research.

Gregg, J. A., Callaghan, G. M., Hayes, S. C., & Glenn-Lawson, J. L. (2007). Improving diabetes self-management through acceptance, mindfulness, and values: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *75*(2), 336–343.

Hackman, J. R., & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. *Journal of Applied Psychology*, *60*, 159–170.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behavior Research and Therapy*, *44*, 1–25.

Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York: Guilford Press.

Hayes, S. C., Strosahl, K. D., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., Polusny, M. A., et al. (2004). Measuring experiential avoidance: A preliminary test of a working model. *Psychological Record*, *54*, 553–578.

Hayes, S. C., Wilson, K. G., Gifford, E. V., Follette, V. M., & Strosahl, K. (1996). Experiential avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology*, *64*(6), 1152–1168.

Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometricka*, *30*, 179–185.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453.

Kashdan, T. B., Barrios, V., Forsyth, J. P., & Steger, M. F. (2006). Experiential avoidance as a generalized psychological vulnerability: Comparisons with coping and emotion regulation strategies. *Behavior Research and Therapy*, *9*, 1301–1320.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Lillis, J., & Hayes, S. C. (2008). Measuring avoidance and inflexibility in weight related problems. *International Journal of Behavioral Consultation and Therapy*, *4*, 372–378.

Linehan, M. M. (1993). *Cognitive behavioral treatment of borderline personality disorder*. New York: Guilford Press.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behavior Research and Therapy*, *33*, 335–343.

Lundgren, T., Dahl, J., & Hayes, S. C. (2008). Evaluation of mediators of change in the treatment of epilepsy with acceptance and commitment therapy. *Journal of Behavior Medicine*, *31*, 225–235.

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common factor model: Implications for theory and practice. *Psychological Bulletin*, *109*, 502–511.

Marsh, H. W. (1986). The bias of negatively worded items in rating scales for young children: A cognitive-developmental phenomenon. *Developmental Psychology*, *22*, 37–49.

Masuda, A., Hayes, S. C., Fletcher, L. B., Seignourel, P. J., Bunting, K., Herbst, S. A., Twohig, M. P., et al. (2007). The impact of acceptance and commitment therapy versus education on stigma toward people with psychological disorders. *Behavior Research and Therapy*, *45*, 2764–2772.

McCracken, L. M. (1998). Learning to live with the pain: Acceptance of pain predicts adjustment in persons with chronic pain. *Pain*, *74*, 21–27.

McCracken, L. M., Vowles, K. E., & Eccleston, C. (2004). Acceptance of chronic pain: Component analysis and a revised assessment method. *Pain*, *107*, 159–166.

Muthen, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171–189.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Office for National Statistics. (2001). *Graduation rates from first university degrees: EU comparison*. London: Her Majesty's Stationery Office.

Reise, S. P., Morizot, J., & Hays, R. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*, 19–31.

Robins, R. W., Fraley, R. C., & Kruger, R. F. (Eds.). (2007). *Handbook of research methods in personality psychology* New York: Guilford Press.

Rüsch, N., Schiel, S., Corrigan, P. W., Leihener, F., Jacob, G. A., Olschewski, M., Lieb, K., et al. (2008). Predictors of dropout from inpatient dialectical behavior therapy among women with borderline personality disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*, 497–503.

Schmidt, N. B., Richey, J. A., Cromer, K. R., & Buckner, J. D. (2007). Discomfort intolerance: Evaluation of a potential risk factor for anxiety psychopathology. *Behavior Therapy*, *38*, 247–255.

Schmitz, N., Kruse, J., Heckrath, C., Alberti, L., & Tress, W. (1999). Diagnosing mental disorders in primary care: The General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Social Psychiatry and Psychiatric Epidemiology*, *34*, 360–366.

Segal, Z. V., Williams, J. M. G., & Teasdale, J. T. (2001). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. New York: Guilford Press.

Shawyer, F., Ratcliff, K., Mackinnon, A., Farhall, J., Hayes, S. C., & Copolov, D. (2007). The Voices Acceptance and Action Scale (VASS). Pilot data. *Journal of Clinical Psychology*, *63*, 593–606.

Wegner, D. M., & Zanakos, S. (1994). Chronic thought suppression. *Journal of Personality*, *62*, 615–640.

Wells, A. (2000). *Emotional disorders and metacognition: Innovative cognitive therapy*. Chichester, UK: Wiley.

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, *51*, 59–91.

Westin, V., Andersson, G., & Hayes, S. C. (2008). Is it the sound or your relationship to it? The role of acceptance in predicting tinnitus impact. *Behaviour Research and Therapy*, *46*, 1259–1265.

Zwick, W., & Velicer, W. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432–442.

Received: April 9, 2010
Accepted: March 5, 2011
Available online 25 May 2011