Further Evaluation of the Psychometric Properties of the Acceptance and Action Questionnaire–II

Martine Fledderus, Martijn A. H. Oude Voshaar, Peter M. ten Klooster, and Ernst T. Bohlmeijer
University of Twente

The Acceptance and Action Questionnaire–II (AAQ-II) is a self-report measure designed to assess experiential avoidance as conceptualized in acceptance and commitment therapy (ACT). The current study is the first to evaluate the psychometric properties of the AAQ-II in a large sample of adults ($N = 376$) with mild to moderate levels of depression and anxiety who participated in a study on the effects of an ACT intervention. The internal construct validity and local measurement precision were investigated by fitting the data to a unidimensional item response theory (IRT) model, and the incremental validity of the AAQ-II beyond mindfulness, as measured by the Five Facet Mindfulness Questionnaire, was assessed. Results of the IRT analyses suggest that the AAQ-II is a unidimensional measure of experiential avoidance and has satisfactory reliability for group comparisons in mild to moderately depressed and anxious populations. Item functioning was found to be independent of gender and slightly dependent on age in this sample. Furthermore, the AAQ-II showed incremental validity beyond 5 mindfulness facets in explaining depression, anxiety, and positive mental health. This study suggests the AAQ-II shows promise as a useful tool for the measurement of experiential avoidance in mild to moderately depressed and anxious populations.

Keywords: item response theory, Acceptance and Action Questionnaire–II, experiential avoidance, differential item functioning, depression

There is growing interest in experiential avoidance (EA) as a risk factor for psychopathology (Biglan, Hayes, & Pistorello, 2008). EA has been defined as the unwillingness to remain in contact with experiences such as feelings, thoughts, and bodily sensations, as an attempted means of behavioral regulation (Hayes et al., 2004). As a consequence, a person will try to use methods that alter, control, predict, or avoid the form, the frequency, or the contexts in which these experiences arise, even when these methods lead to behaviors that cause harm to physical, emotional, or psychological well-being (Hayes, Luoma, Bond, Masuda, & Lillis, 2006). A behavior therapy that is focused on decreasing EA is acceptance and commitment therapy (ACT; Hayes et al., 2006). In ACT, clients are encouraged to accept their private experiences when these experiences help them in engaging in value-based behavior. Studies have shown that ACT is effective in reducing depression and anxiety (e.g., Bohlmeijer, Fledderus, Rokx, & Pieterse, 2011; Fledderus, Bohlmeijer, Pieterse, & Schreurs, 2012; Forman, Herbert, Moitra, Yeomans, & Geller, 2007) and chronic pain (e.g., Vowles & McCracken, 2008) and in increasing positive mental health (Fledderus, Bohlmeijer, Smit, & Westerhof, 2010). Meta-analyses have shown medium to large effect sizes of ACT

interventions on different symptoms of psychological distress (Hayes et al., 2006; Powers, Zum Vörde Sive Vörding, & Emmelkamp, 2009). Recently, more studies have shown that reducing EA (or enhancing acceptance) is an important process of change through which ACT leads to observed improvements in mental health (Ciarrochi, Bilich, & Godsell, 2010). For example, two studies provided preliminary evidence that changes in EA mediated the effect of an ACT intervention on social anxiety (Dalrymple & Herbert, 2007; Kocovski, Fleming, & Rector, 2009).

For examining the effects and the mediating role of EA, it is important that the assessment of EA is carried out with a proper and well-validated general questionnaire. The Acceptance and Action Questionnaire (AAQ) is the most frequently used measure of EA and is available in versions of nine or 16 items (Bond & Bunce, 2003; Hayes et al., 2004). The AAQ measures various theoretically linked aspects of EA, including the need for emotional and cognitive control, avoidance of negative private events, inability to take needed action in the face of private events, and forms of cognitive entanglement (Hayes et al., 2004). Both versions of the AAQ have shown their usefulness in assessing EA in psychopathology (for reviews, see Chawla & Ostafin, 2007; Hayes et al., 2006). Moreover, several versions have been developed that are tailored to populations with specific problems, such as chronic pain (McCracken, Vowles, & Eccleston, 2004), smoking (Gifford et al., 2004), and weight-related difficulties (Lillis & Hayes, 2008). Although these questionnaires have shown their usefulness in predicting relevant outcomes of ACT interventions in these areas or populations (e.g., Gifford et al., 2004; Lillis & Hayes, 2008), a general measure of EA that can be used in any context is important

for studying the processes underlying ACT interventions (Bond et al., 2011).

Although the AAQ is widely used, it has demonstrated two important limitations (Chawla & Ostafin, 2007). First, the AAQ has shown problems with its factor structure and internal consistency in various settings. Due to the broad item content of the different related constructs, it is unclear whether the AAQ measures one overarching construct or a multidimensional construct (Chawla & Ostafin, 2007). To illustrate this problem, the nine-item AAQ showed a one-factor solution (Hayes et al., 2004), while the 16-item AAQ showed a two-factor solution of EA, consisting of willingness and overt action (Bond & Bunce, 2003). Furthermore, internal consistency of the scale is often low, which is probably a result of the complex items (Bond et al., 2011). In the study of Hayes et al. (2004), Cronbach's alpha barely reached an acceptable level ($\alpha = .70$), while other studies found even lower internal consistency (e.g., Boelen & Reijntjes, 2008).

Second, there is uncertainty about the incremental validity of the AAQ because it is unclear what the AAQ adds to other theoretically related measures that also address motivation to accept or avoid aversive private experiences, such as mindfulness and thought-suppression scales (Chawla & Ostafin, 2007). To overcome these limitations, the AAQ-II was developed from an item pool generated by ACT researchers and therapists (Bond et al., 2011). It is the current form for assessing acceptance (e.g., Costa & Pinto-Gouveia, 2011; Wheaton, Berman, & Abramowith, 2010) and measures the ability to accept aversive internal experiences and to pursue goals and values in the presence of these experiences.

To date, only a few studies have assessed the psychometric properties of the AAQ-II. In Bond et al. (2011), confirmatory factor analyses (CFAs) in three different samples, including university students ($N = 433$), financial services workers ($N = 583$), and people seeking treatment for substance misuse ($N = 290$), indicated that the 10-item AAQ-II had a one-factor solution, after allowing a residual correlation between Item 2 ("My painful experiences and memories make it difficult for me to live a life that I would value") and Item 5 ("My painful memories prevent me from having a fulfilling life"). In all samples tested by Bond et al. (2011), the root-mean-square error of approximation (RMSEA) was ≤ 0.06, the standardized root-mean-square residual (SRMR) was ≤ 0.04, and the comparative fit index (CFI) was ≥ 0.95. Furthermore, the AAQ-II showed good internal consistency (Cronbach's $\alpha = .78–.88$) and was related to theoretically linked constructs such as depression, anxiety, and thought suppression, showing adequate construct validity (Bond et al., 2011). The psychometric properties of the AAQ-II were further examined in a sample of people seeking treatment for chronic pain ($N = 144$; McCracken & Zhao-O'Brien, 2010). Exploratory factor analysis (EFA) demonstrated that the AAQ-II had a unitary factor structure in this population as well. The AAQ-II showed good internal consistency (Cronbach's $\alpha = .89$) and construct validity as it was associated with pain-related anxiety, depression, and mindfulness. The Dutch translation of the AAQ-II was tested in a general sample ($N = 374$) and in a sample of patients in psychiatric hospitals ($N = 124$). In both samples, a one-factor structure was found using principal component analyses, and the scale demonstrated good internal consistency (Cronbach's $\alpha = .89$ in both samples) and satisfactory construct validity (Jacobs, Kleen, de

Groot, & A-Tjak, 2008). Finally, CFAs in patients with panic disorder with agoraphobia ($N = 368$), patients with clinically relevant symptoms of social anxiety ($N = 209$), students ($N = 495$), and employment office visitors ($N = 95$) again showed that the one-factor model adequately fitted only after allowing the error terms between the previously mentioned items to correlate. This finding was observed in all of the studied samples individually and also in a CFA analysis of the total data. The fit indices for the latter analysis were CFI = 0.98, Tucker-Lewis index (TLI) = 0.97, SRMR = 0.24, and RMSEA = 0.59 (Gloster, Klotsche, Chaker, Hummel, & Hoyer, 2011). The incremental validity of the AAQ-II has been investigated in two studies. McCracken and Zhao-O'Brien (2010) found that the AAQ-II added significant variance to the prediction of the quality of daily patient functioning above and beyond acceptance of pain and general mindfulness. Karekla and Panayiotoua (2011) showed that the AAQ-II explained unique variance in psychological distress and quality of life above and beyond various coping styles (e.g., active coping, emotional support). Taken together, these studies provide promising support for the psychometric qualities of the AAQ-II. However, previous studies used approaches based on classical test theory only for examining the internal construct validity of the scale. Moreover, the incremental validity of the AAQ-II over closely related aspects of mindfulness has yet to be established.

This study aims to provide further empirical support for the internal construct validity of the AAQ-II by showing that the responses to the AAQ-II fit a unidimensional item response theory (IRT) model. Fitting an IRT model can validate the scoring rule of the AAQ-II by verifying that the variance in observed responses can be attributed to both item and person parameters that are related to a single underlying trait of EA (Glas, 1998). Construct validity further implies that expected scores on items should not differ between subpopulations (e.g., gender, age) when their overall level of EA is the same (Chang & Mazzeo, 1994). This dependence of item response on background variables is known as differential item functioning (DIF). IRT provides the possibility to thoroughly investigate if DIF is present, and if so, it can be investigated if the same latent trait of EA still applies to all groups, despite observed differences in response behavior (Gebhardt & Adams, 2007; Glas, 1998). Although CFA and IRT models are closely related (Reise, Widaman, & Pugh, 1993), IRT is a stronger model than CFA, with more parameters (location parameters for the items in addition to factor loadings, i.e., item discrimination parameters) allowing stronger conclusions regarding DIF (Fischer & Molenaar, 1995). Furthermore, the test information curve (TIC) can be evaluated in an IRT model. This is a more advanced method for assessing the reliability of the AAQ-II than classical approaches that summarize the average measurement precision of a scale in a single index score (such as Cronbach's alpha). This feature of IRT is especially relevant for the analysis of self-report measures because it is a common feature of the items of such instruments to differentiate best between respondents at a specific level of the latent trait (Embretson & Reise, 2000). If, for example, relative item difficulties would cluster together at a narrow range in the middle of the latent trait scale, the measure would perform poorly in differentiating between persons at the extreme ends of the latent trait.

Further support for the incremental validity would be obtained by demonstrating that the AAQ-II contributes to information be-

yond that which is attained by a comprehensive measure of mindfulness in a sample eligible for ACT especially because mindfulness is incorporated in ACT and acceptance is included in most definitions of mindfulness (Fletcher & Hayes, 2005). For instance, Bishop et al. (2004) defined mindfulness as "an orientation that is characterized by curiosity, openness and acceptance" (p. 232). They described acceptance as being in the present moment and open to experiences (Bishop et al., 2004). This is in accordance with the ACT theory on the definition of acceptance (Fletcher & Hayes, 2005). The only previous study that assessed the incremental validity of the AAQ-II over mindfulness used a unidimensional measure of mindfulness (McCracken & Zhao-O'Brien, 2010). A comprehensive multifaceted and often-used measure of mindfulness is the Five Facet Mindfulness Questionnaire (FFMQ; Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006). The FFMQ consists of five facets of mindfulness: (a) observing, defined in terms of noticing or attending to internal and external experiences; (b) describing, defined in terms of labeling internal experiences with words; (c) acting with awareness, defined in terms of attending to one's activities of the moment (the opposite of acting on automatic pilot); (d) nonjudging of inner experience, defined in terms of taking a nonevaluative stance toward thoughts and feelings; and (e) nonreactivity to inner experience, defined in terms of allowing thoughts and feelings to come and go without getting caught up in or carried away by them. Baer et al. (2006) stated that nonreactivity and nonjudging may be seen as ways of operationalizing acceptance. They found a high correlation ($r = .49$) between the AAQ-II and the nonjudging facet. Although the AAQ-II and FFMQ are not meant to measure to same construct, it is important to examine whether the AAQ-II adds additional variance in explaining relevant outcomes such as depression, anxiety, and positive mental health given the possible overlap between acceptance and several aspects of mindfulness.

Finally, this is the first study aimed at assessing these psychometric properties of the AAQ-II in a sample with mild to moderate depression and anxiety. As many people suffer from mild to moderate depression and anxiety (World Health Organization, 2008), there is a growing implementation of ACT and mindfulness-based interventions in this population. Although the efficacy of these treatments has been established (e.g., Forman et al., 2007; Segal, Williams, & Teasdale, 2002), it is increasingly important to study the underlying processes of change for understanding how and why these treatments work, to allow further optimization. In ACT and mindfulness-based treatments for depression and anxiety, acceptance is considered an important process of change (Ciarrochi et al., 2010). Therefore, it is important that this process is assessed with a reliable and valid measure for this population.

Therefore, the current study had two aims. The first aim was to use IRT-based methods to further assess the internal construct validity of the AAQ-II and to provide insight into its local measurement precision using IRT-based methods in sample of adults with mild to moderate depression and anxiety. The second aim was to further examine whether the AAQ-II has additional variance in explaining depression, anxiety, and positive mental health over the mindfulness facets as measured by the FFMQ.

Method

Participants

Baseline data were used from a randomized controlled trial of the effects of a guided self-help ACT intervention on psychological distress and positive mental health (Fledderus et al., 2012). In September 2009, participants were recruited through advertisements in Dutch newspapers for a study on the effects of guided self-help based on ACT. In the advertisement, the target group of the intervention was described as people who wanted to get more out of their life but who were hindered by depressive or anxiety symptoms.

Inclusion criteria were an age of 18 years or older and mild to moderate depressive symptoms (>10 and <39 on the Center of Epidemiologic Studies Depression Scale [CES-D]; Radloff, 1977) and anxiety symptoms (>3 and <15 on the Hospital Anxiety and Depression Scale–Anxiety [HADS-A]; Zigmond & Snaith, 1983). People with severe depressive symptomatology and/or anxiety (more than one standard deviation above the population mean on the CES-D [cutoff score ≥ 39; Bouma, Ranchor, Sanderman, & van Sonderen, 1995] and/or HADS-A [cutoff score ≥ 15; Olssøn, Mykletun, & Dahl, 2005]) were excluded because severe distress would require more intensive individual diagnostics and treatment. For the remaining participants, it was checked who was still responding positively to a screener for a depressive disorder (Web Screening Questionnaire [WSQ] Q1 ≥ 6 and Q2 = 1; Donker, van Straten, Marks, & Cuijpers, 2009). As the WSQ yields a high number of false positives (Donker et al., 2009), those who were screened as having a depressive disorder underwent a telephone interview that employed the depressive episode module of the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). People whom the MINI diagnosed as having a severe depressive episode were excluded.

Other exclusion criteria were (a) few depressive (≤ 10 on the CES-D) and/or anxiety symptoms (≤ 3 on the HADS-A), (b) receiving psychological or psychopharmacological treatment within the last 3 months, and (c) high suicide risk (Q15 = 3 on the WSQ).

Procedure

A total of 625 people responded to the advertisements and received an information sheet explaining the study and an informed-consent form. This was signed by 507 people who then received an e-mail with a screening questionnaire comprising the CES-D, HADS-A, WSQ, and demographic items. First, 54 respondents were excluded because they had severe depression and/or anxiety according to the scores on the CES-D and HADS-A. They were advised to contact their general practitioner. Second, 44 respondents were diagnosed by the WSQ as having a depressive disorder and subsequently underwent a telephone interview using the MINI. These interviews were conducted by master's degree students of psychology who were trained and supervised by a clinical psychologist. Of the 43 respondents (one respondent could not be contacted), two were diagnosed with a severe depressive episode and were excluded and advised to contact their general practitioner. In all, 56 respondents were excluded because they had severe depression or anxiety. A further 75 respondents were ex-

cluded because they had few depression and/or anxiety symptoms ($N = 58$), did not complete the screening questionnaire ($N = 15$), could not be contacted for the interview ($N = 1$), or currently received psychological treatment ($N = 1$). In total, 376 participants were included in the study and were randomly assigned to the ACT intervention with minimal e-mail support ($N = 125$), to the same intervention with extensive e-mail support ($N = 125$), or to a waiting list ($N = 126$). The waiting-list group received the intervention after the intervention period of 9 weeks. More detailed information about the study can be found in Fledderus et al. (2012). Table 1 shows an overview of the participants' characteristics. Their mean age was 42 years (range = 18–73). The majority was female (70%) and of Dutch origin (93%). Most of the participants had a high level of education (86%), held a paid job (76%), and were not married (47%).

Measures

All participants completed online measures at baseline and directly after the intervention (9 weeks). Those assigned to the experimental conditions completed a third assessment at 5 months after baseline. For this study, the baseline data were used. The internal consistency of the used measures was examined by Cronbach's alpha coefficients, where values above .70 were considered

Table 1

Respondents' Characteristics and Scores on AAQ-II, CES-D, HADS-A, MHC-SF, and FFMQ

Characteristics	Scores
Age, years ($N = 376$): *M, SD*	42.49 (11.09)
Age groups ($N = 376$): *N (%)*	
18–36 years	116 (30.9)
37–48 years	132 (35.1)
49 and older	128 (34.0)
Gender ($N = 376$): % female	69.7
Marital status ($N = 375$): *N (%)*	
Married	164 (43.7)
Divorced	32 (8.5)
Widowed	4 (1.1)
Never married	175 (46.7)
Race ($N = 376$): *N (%)*	
Dutch	349 (97.8)
Other	27 (2.2)
Educational level ($N = 376$): *N (%)*	
Low (primary school, lower vocational education)	19 (5.1)
Intermediate (secondary school, vocational education)	62 (16.5)
High (higher vocational education, university)	295 (78.5)
Acceptance (AAQ-II) ($N = 372$): *M, SD*	40.72 (8.59)
Depression (CES-D) ($N = 364$): *M, SD*	22.70 (6.63)
Anxiety (HADS-A) ($N = 373$): *M, SD*	9.47 (2.50)
Positive mental health (MHC-SF) ($N = 362$): *M, SD*	3.13 (.76)
Mindfulness (FFMQ): *M, SD*	
Observe ($N = 372$)	25.09 (5.17)
Describe ($N = 373$)	25.69 (6.23)
Act With Awareness ($N = 375$)	20.94 (4.96)
Nonjudging ($N = 374$)	22.98 (5.38)
Nonreactive ($N = 372$)	19.18 (3.78)

Note. AAQ-II = Acceptance and Action Questionnaire–II; CES-D = Center for Epidemiologic Studies Depression Scale; HADS-A = Hospital Anxiety and Depression Scale–Anxiety; MHC-SF = Mental Health Continuum–Short Form; FFMQ = Five Facet Mindfulness Questionnaire.

acceptable and values of .80 or higher as good (Nunnally, 1978). Only fully completed questionnaires were used in the analysis.

The AAQ-II (Bond et al., 2011) is a 10-item questionnaire. Participants were asked to rate on a 7-point Likert-type scale the degree to which each statement was true for them. A total score, ranging from 10 to 70, was computed by summing the scores on the individual items. Higher scores indicate higher levels of general acceptance and less experiential avoidance. The Dutch AAQ-II (Jacobs et al., 2008) showed good internal consistency in the current study ($\alpha = .85$).

The FFMQ (Baer et al., 2006) is a 39-item questionnaire that measures five facets of mindfulness: observing (eight items), describing (eight items), acting with awareness (eight items), nonjudging (eight items), and nonreactivity (seven items). Participants were asked to rate the degree to which each statement was true for them on a 5-point Likert-type scale ranging from 1 (*never or very rarely true*) to 5 (*very often or always true*). Facet scores were computed by summing the scores on the individual items. Facet scores range from 8 to 40 (except for the nonreactivity facet, which ranges from 7 to 35), with higher scores indicating more mindfulness. The Dutch FFMQ was developed by translation and back-translation of the original FFMQ and has shown adequate construct validity and test–retest reliability in patients with fibromyalgia (Veehof, ten Klooster, Taal, Westerhof, & Bohlmeijer, 2011) and factorial validity in people with depressive symptomatology (Bohlmeijer, ten Klooster, Fledderus, Veehof, & Baer, 2011). All five facets showed acceptable to good internal consistency in this study, ranging from .70 for observing to .91 for describing.

The CES-D (Radloff, 1977) is a 20-item questionnaire that measures depressive symptoms in the general population. Respondents rated on a 4-point scale ranging from hardly ever (less than 1 day) to predominantly (5–7 days) to what extent they had experienced depressive symptoms in the previous week. Summation of the scores results in a total score ranging from 0 to 60. A score of 16 or higher is considered to indicate the presence of clinically relevant depressive symptoms. The CES-D has shown good psychometric properties in a general sample (Radloff, 1977). The Dutch translation demonstrated similar psychometric properties in a group of elderly people in the Netherlands (Haringsma, Engels, Beekman, & Spinhoven, 2004). In this study, the scale showed acceptable internal consistency ($\alpha = .78$).

The HADS-A (Zigmond & Snaith, 1983) was used to measure the presence and severity of anxiety symptoms. Participants were asked to rate the degree to which they experienced several emotions in the past week. All items were rated on a 4-point scale. Scale scores were computed by summing the scores on the individual items. Scale scores range from 0 to 21, with higher scores indicating more anxiety. The Dutch HADS has shown good psychometric properties across diverse general and clinical populations (Spinhoven et al., 1997). In this study, the scale showed low internal consistency at baseline ($\alpha = .56$).

The Mental Health Continuum–Short Form (MHC-SF) was used to measure positive mental health (Keyes, 2002). The MHC-SF is a 14-item questionnaire that measures three dimensions of mental health: (a) emotional well-being (three items), defined in terms of positive feelings and satisfaction with life; (b) psychological well-being (six items), defined in terms of positive functioning in individual life (self-realization); and (c) social well-

being (five items), defined in terms of positive functioning in community life (being of social value). Participants were asked to rate the frequency of feelings they experienced in the past month. Items were scored on a 6-point scale ranging from 1 (*never*) to 6 (*every day*). A total score was computed by summing the scores on the individual items and dividing these by the number of items. Higher scores indicate better positive mental health. The Dutch MHC-SF has shown good construct validity and test–retest reliability in the general adult population (Lamers, Westerhof, Bohlmeijer, ten Klooster, & Keyes, 2011) and good internal consistency in this study ($\alpha = .88$).

Statistical Analysis

CFA was used to test whether the AAQ-II items were sufficiently unidimensional for IRT-based analyses. A one-factor model was tested with maximum likelihood parameter estimates with standard errors and a mean-adjusted chi square statistic that are robust to nonnormality (i.e., Satorra-Bentler scaling), using Mplus 5.2 (Muthén & Muthén, 2008). Because all previous CFAs of the AAQ-II noted a pronounced method effect in responses to Items 2 and 5 due to their highly similar content (Bond et al., 2011; Gloster et al., 2011), we compared the fit of a model where the error terms between these items were allowed to correlate versus a model with no error correlations. The overall fit of the models was evaluated using commonly accepted criteria for the fit indices provided by Mplus (Hu & Bentler, 1999).

For assessing the internal construct validity and local measurement precision of the AAQ-II, the two-parameter generalized partial credit model (GPCM; Muraki, 1992) was estimated using the MIRT software (Glas, 2010). The GPCM pertains to polytomously scored items, such as the items of the AAQ-II. In this model, each item is described by a number of category intersection parameters equal to the number of response options minus 1 and one discrimination parameter. Category intersection parameters indicate the location on the latent EA continuum where two consecutive response options are equally likely to be endorsed by respondents. Discrimination parameters represent the degree to which an item discriminates between persons with different levels of the latent trait and can be interpreted like factor loadings in factor analysis, that is, they represent the strength of the association of the item with the latent trait. The item parameters were estimated using the marginal maximum likelihood (MML) procedure. MML is the most commonly used estimation procedure in IRT. In contrast to other methods (e.g., joint maximum likelihood), it produces consistent estimates of the structural model parameters (Bock & Aitkin, 1981). Response Options 1 and 2 and Response Options 6 and 7 of the AAQ-II were collapsed to obtain stable category intersection parameters.

The first step of the IRT analysis was to evaluate the presence of DIF across gender and age in the items of the AAQ-II. To optimize the stability of the resulting parameters, three equally large age groups were created for this analysis, resulting in the following groups: 18–36 years, 37–48 years, and 49 years and older. Age and gender are important background variables that might confound the outcomes of effectiveness studies if the items of the instrument used are biased against subgroups, especially in smaller trials. Analysis of DIF is therefore an important step in ascertaining the external construct validity of the AAQ-II for

depressed and anxious populations. Items show DIF if the probability of choosing a given response option differs between groups with the same level of EA. The presence of DIF and the unidimensionality of the AAQ-II were evaluated with Lagrange multiplier (LM) tests (Glas, 1998). Although relatively new in the personality assessment literature, LM tests are more widespread in other areas of research (e.g., Glas, Geerlings, van de Laar, & Taal, 2009; van Groen, ten Klooster, Taal, van de Laar, & Glas, 2010; Weisscher, Glas, Vermeulen, & de Haan, 2010). The LM test is asymptotically equivalent to the likelihood ratio test but has some computational advantages (Glas, 1998, 1999). From a practical point of view, the LM statistics are useful item-oriented diagnostic tools, which give an indication of the source of model violations. They are based on a difference between observed and expected frequencies, so the importance of a significant DIF finding can be assessed in a framework that is directly related to observed data. Another advantageous aspect of the LM statistics is that they offer the possibility of directed model relaxation to obtain sufficient model fit.

The LM statistics are accompanied by effect-size statistics that show the seriousness of model violation. These statistics are the absolute difference between observed average item scores minus average item scores expected by the model. Because response options were collapsed, the effect-size statistics theoretically ranged from 0 to 4. We evaluated whether the DIF effect sizes were significant, correcting for multiple comparisons ($p < .01$), and considered significant effects sizes > 0.10 indicative of substantial DIF. If DIF is not present, unambiguous support for the construct validity of the AAQ-II is obtained. If DIF is found, this indicates that response behavior is inconsistent across groups. This could indicate that an item is systematically more difficult for one group. It may also indicate that the same latent dimension of EA does not apply to one of the groups at all, which is a more serious violation of construct validity (Glas, 1998). When a limited number of items show substantial DIF, construct validity may still be defendable if it can be explicitly shown that the same underlying latent variable EA pertains to both groups. That is, the same IRT model should hold for the entire set of response data after assigning separate item parameters to the items that show substantial DIF (Fischer & Molenaar, 1995; Gebhardt & Adams, 2007; Glas, 1998). The second step of the IRT analysis was therefore to investigate whether the same latent scale, with gender- or age-specific parameters where necessary, still applied. To this end, LM statistics pertaining to the form of the item response curves were calculated after assigning separate parameters to items with substantial DIF (Glas, 1999). Because, in this analysis, LM statistics are computed within groups for each of the items, corrected p values $< .01$ were, again, considered to indicate significant misfit.

After the data were adequately modeled, the TIC was calculated to provide insight into the local measurement precision of the AAQ-II. The TIC is calculated from the item parameters and provides information about the range of latent scores where the AAQ-II is best at discriminating among individuals. The height of the curve (denoting the amount of information at a given point of the latent scale) is a function of the discrimination parameters and the threshold parameters of the items making up the scale. To better interpret the outcomes of this analysis, the amount of information provided by the AAQ-II for relevant ranges along the latent

metric of EA was also converted to reliability coefficients ($r = 1 - 1/information$; Reeve & Fayers, 2005).

For examining the incremental validity of the AAQ-II, Pearson's correlation coefficients were first calculated between acceptance (AAQ-II) and depression (CES-D), positive mental health (MHC-SF), and mindfulness facets (FFMQ). In line with previous studies (e.g., Fledderus, Bohlmeijer, & Pieterse, 2010; Hayes et al., 2006) negative correlations between the AAQ-II and anxiety and depression and a positive correlation with positive mental health were expected. On the basis of previous studies (Baer et al., 2006, 2008; Veehof et al., 2011), positive correlations were expected between the mindfulness facets (except for observing) and the AAQ-II. Nonjudging was predicted to show the strongest correlation with the AAQ-II. Furthermore, at least moderate negative correlations were predicted between the mindfulness facets (except observing) and anxiety and depression. Also, moderate positive correlations between the mindfulness facets (except observing) with positive mental health were predicted. Hierarchical multiple regression analyses were then conducted to examine the incremental validity of AAQ-II beyond the facets of the FFMQ in explaining depression, anxiety, and positive mental health. In the first block, the facets of the FFMQ were entered that were univariately significantly related to the dependent variables depression (CES-D), anxiety (HADS-A), and positive mental health (MHC-SF). In the second block, acceptance (AAQ-II) was included. The change in variance accounted for from Block 2 served as a test for the incremental validity of the AAQ-II ($p < .05$). We also performed reversed hierarchical multiple regression analyses in which the AAQ-II was entered in the first block and the mindfulness facets were added in the second block.

Results

Internal Construct Validity

The one-factor model for the AAQ-II showed poor overall fit (CFI = 0.74, TLI = 0.67, RMSEA = 0.12, SRMR = 0.06), although all factor loadings were > 0.5. The second model that allowed the error terms between Items 2 and 5 to correlate showed a marked improvement in overall fit (CFI = 0.95, TLI = 0.93, RMSEA = 0.05, SRMR = 0.04). Given the high observed factor loadings, the overlapping content in Items 2 and 5 appears to be the source of misspecification in the first model. All the fit indices for the second model were in acceptable ranges according to recommended criteria (Hu & Bentler, 1999). The overall conclusion of the CFA was that the one-factor model fitted the data very well after accounting for the previously found method effect produced by the overlapping content in Items 2 and 5, and therefore, the analysis supports the notion that the AAQ-II data were essentially unidimensional.

In the IRT analysis, LM tests indicated no substantial DIF for gender for any of the items of the AAQ-II (i.e., all $ps > .01$). In Table 2, the results for DIF for age are shown. Substantial age DIF was found in Item 9 ("Worries get in the way of my success") of the AAQ-II. Because the presence of DIF may bias the parameter estimates of the other items, the presence of DIF was reevaluated after assigning age-specific parameters to Item 9 (Fischer & Molenaar, 1995; Glas, 1998). In the respecified model with age-specific parameters for Item 9, the DIF in Item 4 ("I worry about

Table 2

Differential Item Functioning Across Age Groups

Item	LM[a]	p	ES
1	4.48	.11	0.10
2	7.04	.03	0.10
3	0.18	.92	0.01
4	9.76	.01	0.12
5	4.82	.09	0.08
6	1.31	.52	0.04
7	5.97	.05	0.08
8	5.94	.05	0.09
9	16.39	.00	0.14
10	2.32	.31	0.07

Note. LM = Lagrange multiplier; ES = effect size.
[a] $df = 2$.

not being able to control my worries and feelings") became more pronounced (LM = 12.39, $p < .01$, effect size = 0.13). Therefore, Item 4 was assigned age-specific parameters as well. After these respecifications, no more significant LM tests were found (i.e., all $ps > .01$).

The three panels in Table 3 contain the DIF statistics of the items with substantial DIF separately for each age group. In each panel, the observed average score and the expected average score of the age group are compared to the observed and expected averages of the other age groups together. Table 3 illustrates the nature of DIF present in Item 4. The first panel shows that the observed score on Item 4 in the age group of 18 to 36 years was 2.00 compared to an average observed score in the other groups of 2.30. Because people aged 18 to 36 years were expected to score 2.20 according to the model (see Table 2), this results in a DIF effect size of -0.20, reflecting the fact that these respondents scored lower than expected. Likewise, it can be seen that the average observed score of the respondents in the age group of 49+ years was 0.11 points higher than expected under the model. The same pattern was noticed for Item 9. In the next step of the analysis, it was examined whether the respecified model, with age-group specific item parameters for Items 4 and 9, fitted the data. The results for the age group of 18–36 years are shown in Table 4. The results for the two other age groups were analogous (i.e., range of $ps = .08–.98$). Note that none of the outcomes of the LM tests were below the significance level of 1%, indicating that all items adequately fitted to the respecified model. So, although younger respondents scored systematically worse on Items 4 and 9 than is to be expected based on their total score, this analysis showed that both items still relate to the same latent trait of EA and that the observed bias is relatively minor.

In the final step of the IRT analysis, the TIC was calculated from the resulting item parameters. Because age-group-specific item parameters were assigned to two items of the AAQ-II, the TIC was plotted for the three age groups separately, with age-group-specific parameters for Items 4 and 9. In Figure 1, it can be seen that measurement precision is somewhat lower across the range of latent scores for Age Group 2. This is because the discrimination parameters of Items 4 and 9 in Age Group 2 were lower than for the other age groups, indicating a less strong association with EA for both items within Age Group 2. Overall, between -2 and 1.6, the measurement precision of the AAQ-II exceeded 4, which

Table 3

Age-Based Differential Item Functioning (DIF) for Items With Substantial DIF

Age group	Focal group		Reference group				
	Obs	Exp	Obs	Exp	ES	LM[a]	p
18–36							
Item 4	2.00	2.20	2.30	2.30	−0.20	10.59	.03
Item 9	1.40	1.58	1.65	1.65	−0.18	14.32	.01
37–48							
Item 4	2.27	2.21	2.17	2.20	0.06	4.10	.39
Item 9	1.53	1.57	1.59	1.57	−0.03	1.91	.75
49+							
Item 4	2.32	2.21	2.15	2.21	0.11	4.41	.35
Item 9	1.77	1.57	1.47	1.57	0.20	16.21	.00

Note. Obs = observed average score; Exp = expected average score; ES = effect size; LM = Lagrange multiplier.
[a] df = 4.

corresponds to $r = .75$. This indicates that with collapsed response options, the AAQ-II is reliable for inferences at the group level within this range. In our sample, 93.5% of respondents fell within this range.

Correlations Between the AAQ-II and Other Measures

The correlations between AAQ-II and the other constructs can be found in Table 5. As predicted, the AAQ-II correlated moderately negatively with depression and anxiety. The AAQ-II correlated positively with positive mental health. As expected, no significant relation was found between the AAQ-II and the observing facet of the FFMQ. The highest significant relation was found between the AAQ-II and nonjudging. For the other facets of the FFMQ, moderate correlations were found. Depression and anxiety were, as predicted, not related to observing and significantly related to acting with awareness, nonjudging, and nonreacting. Depression was, as expected, also related to the describing facet, but no relation was found between anxiety and this facet. As predicted, positive mental health was significantly related to four of the mindfulness facets, but also to observing.

Incremental Validity of the AAQ-II

In Table 6, the hierarchical regressions for examining the incremental validity of the AAQ-II can be found. In the first step, the mindfulness facets that correlated significantly with depression were included. Table 6 shows that the mindfulness facets acting with awareness and nonjudging significantly explained variance in depression. In the second step, when the AAQ-II was entered as well, only the AAQ-II was significantly related to depressive symptoms. The AAQ-II explained a significant proportion of the variance in depressive symptoms beyond the contribution of the mindfulness facets (adjusted R^2 change = .07, $p < .001$). The same procedure was repeated with anxiety as the outcome measure. As with depression, the facets acting with awareness and nonjudging significantly explained variance in anxiety. After including the AAQ-II, acting with awareness and AAQ-II were significantly related to anxiety. The AAQ-II again explained a

significant proportion of variance beyond the contribution of the mindfulness facets (adjusted R^2 change = .02, $p < .001$). Finally, the same procedure was repeated with positive mental health as outcome. In the first step, the facets observing, describing, and nonjudging were significantly related to positive mental health. After adding the AAQ-II, the mindfulness facets observing and describing and the AAQ-II were significantly related to positive mental health. Again, the AAQ-II made a significant contribution to the explained variance (adjusted R^2 change = .08, $p < .001$). The reversed hierarchical analyses revealed mainly the same results in that the FFMQ adds unique explained variance beyond the AAQ-II in explaining anxiety and positive mental health. The results of these analyses seem to imply that the instruments assess differing constructs that are both, to some extent, uniquely related to anxiety and positive mental health. However, for depression, the FFMQ did not explain additional variance beyond the AAQ-II in the reversed analysis, suggesting that the FFMQ scales have no usefulness in predicting depression beyond the AAQ-II.

Discussion

In this study, the psychometric properties of the AAQ-II were investigated in a sample of adults with mild to moderate depression and anxiety ($N = 376$). The first aim of the study was to use IRT-based methods to examine the internal construct validity of the AAQ-II in a sample of adults with mild to moderate depression and anxiety. The CFA indicated that the AAQ-II was sufficiently unidimensional for IRT analysis. Although the fit of a one-dimensional structure was adequate only after allowing the error terms between Items 2 and 5 to correlate, it is generally accepted that empirical data are never strictly unidimensional because the presence of minor factors, such as method effects, also influences item response behavior (Hambleton, Swaminathan, & Rogers, 1991). Moreover, the finding that the observed data, with age-group-specific parameters for items exhibiting significant DIF, adequately conformed to the GPCM model expectations as tested with the LM statistics also supports the notion of essential unidimensionality.

Table 4

Outcomes of Tests for Model Fit in Score Level Groups for Respondents in Age Group 18–36 Years

Item	LM[a]	p	ES	18–36 years		Other age groups			
				Obs	Exp	Obs	Exp	Obs	Exp
1	3.52	.17	0.14	1.94	1.69	2.21	2.22	2.92	2.76
2	3.69	.16	0.10	1.49	1.46	2.58	2.36	3.21	3.28
3	0.09	.96	0.05	1.41	1.39	2.17	2.20	3.08	3.16
4	5.73	.06	0.10	1.41	1.26	1.81	1.93	2.77	2.80
5	0.85	.65	0.05	1.67	1.66	2.76	2.66	3.49	3.52
6	3.50	.17	0.11	2.05	1.90	2.25	2.41	2.95	2.92
7	1.11	.57	0.07	1.18	1.08	1.65	1.62	2.55	2.48
8	1.26	.53	0.10	0.82	0.92	1.62	1.67	2.54	2.67
9	7.28	.03	0.07	0.73	0.62	1.18	1.25	2.26	2.28
10	0.21	.90	0.09	0.74	0.78	1.29	1.28	1.82	2.02

Note. LM = Lagrange multiplier; ES = effect size; Obs = observed average score; Exp = expected average score.
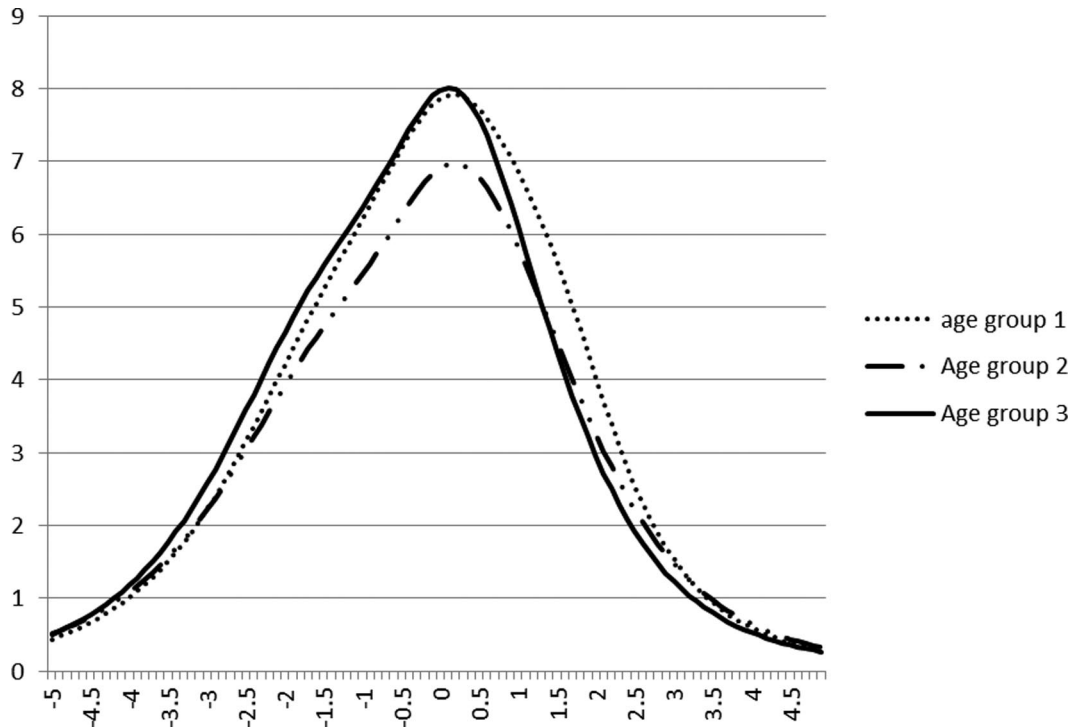[a] df = 2.

Figure 1. Test information curves of the Acceptance and Action Questionnaire–II. Age Group 1 = 18–36 years; Age Group 2 = 37–48 years; Age Group 3 = 49 years and older.

In the IRT analysis, DIF across age and gender and local measurement precision were assessed first. The results showed no DIF for gender, and only two items of the AAQ-II showed DIF across age. These results indicate that most items function equivalently across the variables gender and age and that the items have the same meaning and difficulty for men and women and across different age groups. However, Items 4 and 9 both showed that the youngest age group scored lower and the highest age group higher than expected on acceptance. Since there are only age differences in these two items, the DIF may be related to differences in interpreting these particular items. In both items, *worry* is the subject: worrying about not having worries under control and that

worries get in the way of a successful life. This indicates that older people might worry less compared with younger people. Earlier research has shown that older and younger people differ in their worry content and that worries decline with age (Diefenbach et al., 2001; Lindesay et al., 2006). For example, Lindesay et al. (2006) found that older people worried less about relationship/family, finances/housing, and work compared with younger people. So, maybe, different ideas of what worrying means and the frequency of worries could explain these differences. On the other hand, although these two items were found to exhibit statistically significant DIF, subsequent analysis indicated that a model with age-group-specific parameters for Items 4 and 9 fitted the unidimen-

Table 5
Correlations Between AAQ-II and Depression (CES-D), Anxiety (HADS-A), Positive Mental Health (MHC-SF), and Five Facets of Mindfulness (FFMQ)

Variables	CES-D	HADS-A	MHC-SF	FFMQ–Observe	FFMQ–Describe	FFMQ–Act With Awareness	FFMQ–Nonjudging	FFMQ–Nonreactive
AAQ-II	−.40**	−.31**	.45**	.10	.31**	.30**	.54**	.37**
N	360	371	359	368	369	371	370	368
CES-D		−.47**	−.34**	−.03	−.11*	−.20**	−.25**	−.16**
N		361	350	360	361	363	362	360
HADS-A			.06	−.03	−.02	−.22**	−.24**	−.20**
N			360	369	370	372	371	369
MHC-SF				.30**	.38**	.20**	.20**	.22**
N				358	359	361	360	358

Note. AAQ-II = Acceptance and Action Questionnaire–II; CES-D = Center for Epidemiologic Studies Depression Scale; HADS-A = Hospital Anxiety and Depression Scale–Anxiety; MHC-SF = Mental Health Continuum–Short Form; FFMQ = Five Facet Mindfulness Questionnaire.
* $p < .05$.   ** $p < .01$.

Table 6

Hierarchical Regression Analyses for Depression (CES-D), Anxiety (HADS-A), and Positive Mental Health (MHC-SF) With the Facets of Mindfulness (FFMQ) and Acceptance (AAQ-II)

Dependent variables and predictors	Step 1		Step 2	
	Beta	Adjusted R^2	Beta	Adjusted R^2
CES-D				
FFMQ–Describe	−.04		.04	
FFMQ–Act With Awareness	−.13*		−.10	
FFMQ–Nonjudging	−.20***		−.03	
FFMQ–Nonreactive	−.04		.02	
AAQ-II			−.37***	
		.08***		.15***
HADS-A				
FFMQ–Act With Awareness	−.16**		−.13*	
FFMQ–Nonjudging	−.17**		−.08	
FFMQ–Nonreactive	−.10		−.07	
AAQ-II			−.20**	
		.09***		.11**
MHC-SF				
FFMQ–Observe	.17**		.20***	
FFMQ–Describe	.28***		.18**	
FFMQ–Act With Awareness	.00		−.03	
FFMQ–Nonjudging	.15**		−.02	
FFMQ–Nonreactive	.10		.03	
AAQ-II			.38***	
		.20***		.28***

Note. AAQ-II = Acceptance and Action Questionnaire–II; FFMQ = Five Facet Mindfulness Questionnaire; CES-D = Center for Epidemiologic Studies Depression Scale; HADS-A = Hospital Anxiety and Depression Scale–Anxiety; MHC-SF = Mental Health Continuum–Short Form.
* $p < .05$. ** $p < .01$. *** $p < .001$.

sional GPCM. These findings indicate that the same underlying latent variable of EA applies to all age groups but that the overall level of EA required for a respondent to endorse a specific response option differs systematically. Considering that the absolute magnitude of observed DIF was relatively minor, the results suggest that the measurement model of the AAQ-II is valid for people with mild to moderate anxious and depressive symptoms and that the AAQ-II can reasonably be considered a unidimensional scale in this population.

The analysis of the TIC suggests that the AAQ-II is a reliable instrument for measuring EA in adults with mild to moderate depression and anxiety. However, it should be noted that the response options were collapsed to five instead of seven response options. So, the results of our analysis pertain to the AAQ-II with five response options. Collapsing response options leads to loss of variability and consequent loss of measurement precision. The average reliability of the AAQ-II with seven response options will therefore exceed the reliability of the AAQ-II with five response options. However, as very few respondents in our sample elected Response Options 1 and 2 and Response Options 6 and 7, it was not possible to estimate stable threshold parameters for these response options. Although having many response options is an appealing feature of a scale for reasons mentioned above, it is vital for the validity of inferences drawn from the total score that respondents can consis-

tently distinguish between response options, especially since the sum score of AAQ-II is used. Previous research has shown that respondents find it difficult to discriminate between more than six response options (Lopez, 1996). Therefore, it would be worthwhile to further investigate the utility of the current response format in different settings with larger samples.

Another point of interest is the three positively framed items. Bond et al. (2011) found in an EFA that these items loaded on a second factor. They concluded that these three items should be deleted and that a seven-item version of the AAQ-II should be used. However, using EFA, it is difficult to distinguish between true multidimensionality and method effects stemming from negatively and positively framed items. The IRT analysis carried out in this study is more sensitive in assessing multidimensionality and revealed no problems with multidimensionality after the DIF for age was taken into account. Furthermore, in McCracken and Zhao-O'Brien (2010), a unitary factor structure was found as well. Future research could examine whether it is necessary to delete these three items.

The second aim of the study was to examine whether the AAQ-II adds additional variance to mindfulness facets as measured with the FFMQ (Baer et al., 2006) in explaining depression, anxiety, and positive mental health. Results showed that the AAQ-II was negatively related to depression and anxiety and positively related to mindfulness facets (except for the facet observing) and positive mental health. We found no relation with the AAQ-II and the mindfulness facet observing. The other facets correlated significantly with the AAQ-II, with the strongest relation between AAQ-II and nonjudging. These findings correspond to earlier research (e.g., Baer et al., 2006; Veehof et al., 2011). The facet nonjudging is also theoretically the most closely related to the AAQ-II. Both measure the willingness to accept private experiences in the present moment without trying to avoid or change these experiences. Furthermore, the AAQ-II accounts for a higher proportion of variance in depression, anxiety, and positive mental health when added to the mindfulness facets. This finding corroborates earlier research that showed that the AAQ-II had incremental validity beyond mindfulness in predicting chronic pain (McCracken & Zhao-O'Brien, 2010). This earlier study used a unidimensional measure of mindfulness, while the current study used a multifaceted measure of mindfulness and looked in more detail to the separate facets of mindfulness. When the AAQ-II was added, the mindfulness facets acting with awareness and nonjudging were no longer related to depression, and only the facet acting with awareness was still related to anxiety. When the AAQ-II was included, observing and describing were still related to positive mental health, but nonjudging was no longer significant. This implies that the AAQ-II has a unique role in predicting these outcomes above and beyond these mindfulness facets, which justifies its use in ACT and mindfulness-based intervention studies. An explanation of this finding is that although both questionnaires assess the ability to contact the present moment and to accept private experiences, the AAQ-II also assesses taking value-based actions even in the face of unwanted thoughts, feelings, and other private events that might occur. Besides acceptance of experiences, value-based behavior is important for the enhancement of psychological flexibility, which is a core process of ACT (Hayes et al., 2006). Value-based behavior might be particularly important in individuals experiencing significant levels of depression and anxiety. This relevance is already

underscored by earlier research that has shown that improved value-based actions at the end of an ACT treatment for chronic pain patients were associated with less pain and depression after the treatment (Vowles & McCracken, 2008). More research is needed to investigate whether value-based living is an important predictor of depression, anxiety, and positive mental health beyond acceptance and mindfulness. Furthermore, future research could further examine the incremental validity of the AAQ-II with other closely related constructs, such as thought suppression.

Although this study has provided robust IRT analyses and a more detailed insight into the incremental validity of AAQ-II, there are several limitations as well. First, the psychometric properties were analyzed cross-sectionally, and no longitudinal analyses were done. Future research could examine DIF for the items of the AAQ-II over a longer time period, to investigate the stability of the parameters over time. Second, the DIF analyses were performed with age and gender. Future research could examine other demographic variables, such as education or ethnicity. In our sample, our participants were mainly Dutch and highly educated, so generalization of the present results should be made with prudence. Third, although it was the first time that the AAQ-II was assessed in a sample with mild to moderate depressive and anxiety symptoms, generalizations to other samples (e.g., major depression) have to be made with care. Finally, future latent trait studies of the AAQ-II would probably benefit from larger sample sizes. Various simulation studies have shown that parameter recovery is influenced by sample size. Recommendations for adequate sample sizes of IRT studies in the peer-reviewed literature range from 250 to 500 depending on the software used and the models estimated (Choi, Cook, & Dodd, 1997; DeMars, 2003; Reise & Yu, 1990). Lord (1983) also noted that small sample sizes (e.g., $N < 200$) generally argue for simple models (e.g., the Rasch model); however, he also argued that this is mainly the case if the discrimination parameters are difficult to estimate. In our sample, the discrimination parameters fell within reasonable limits (i.e., small standard deviations), indicating that this was not the case in our sample. Considering this result, we believe that our sample size of 376 is sufficiently larger than 200 to justify the use of the GPCM.

To conclude, this study suggests that the AAQ-II is a valid and reliable measure to assess experiential avoidance or its reverse acceptance in people with mild to moderate depression and anxiety. This study expands previous evaluations of the psychometric properties of the AAQ-II by using more advanced and robust IRT methods. Furthermore, the AAQ-II showed incremental validity in explaining depression, anxiety, and positive mental health over different mindfulness facets. An important research question for future studies is to examine whether changes in acceptance mediate the effects of an ACT intervention on depression and anxiety.

References

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facts of mindfulness. *Assessment, 13,* 27–45. doi:10.1177/1073191105283504

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

Biglan, A., Hayes, S. C., & Pistorello, J. (2008). Acceptance and commit-

ment: Implications for prevention science. *Prevention Science, 9,* 139–152. doi:10.1007/s11121-008-0099-4

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., . . . Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice, 11,* 230–241. doi:10.1093/clipsy.bph077

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443–459. doi:0.1007/BF02293801

Boelen, P. A., & Reijntjes, A. (2008). Measuring experiential avoidance: Reliability and validity of the Dutch 9-item Acceptance and Action Questionnaire (AAQ). *Journal of Psychopathology and Behavioral Assessment, 30,* 241–251. doi:10.1007/s10862-008-9082-4

Bohlmeijer, E. T., Fledderus, M., Rokx, T. A. J. J., & Pieterse, M. E. (2011). Efficacy of an early intervention based on acceptance and commitment therapy for adults with depressive symptomatology: Evaluation in a randomized controlled trial. *Behaviour Research and Therapy, 49,* 62–67. doi:10.1016/j.brat.2010.10.003

Bohlmeijer, E. T., ten Klooster, P. M., Fledderus, M., Veehof, M., & Baer, R. (2011). Psychometric properties of the Five Facet Mindfulness Questionnaire in depressed adults and development of a short form. *Assessment, 18,* 308–320. doi:10.1177/1073191111408231

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. doi:10.1037/0021-9010.88.6.1057

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. C., Guenole, N., Orcutt, H. K., . . . Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological flexibility and acceptance. *Behavior Therapy, 42,* 676–688. doi:10.1016/j.beth.2011.03.007

Bouma, J., Ranchor, A. V., Sanderman, R., & van Sonderen, E. (1995). *Het meten van symptomen van depressie met de CES-D, een handleiding* [Measuring symptoms of depression with the CES-D, a guide]. Groningen, the Netherlands: Noordelijk Centrum voor Gezondheidsvraagstukken.

Chang, H. H., & Mazzeo, J. (1994). The unique correspondence of the item response function and item category response functions in polytomously scored item response models. *Psychometrika, 59,* 391–404. doi:10.1007/BF02296132

Chawla, N., & Ostafin, B. (2007). Experiential avoidance as a functional dimensional approach to psychopathology: An empirical review. *Journal of Clinical Psychology, 63,* 871–890. doi:10.1002/jclp.20400

Choi, S. W., Cook, K. F., & Dodd, B. G. (1997). Parameter recovery for the partial credit model using MULTILOG. *Journal of Outcome Measurement, 1,* 114–142.

Ciarrochi, J., Bilich, L., & Godsell, C. (2010). Psychological flexibility as a mechanism of change in acceptance and commitment therapy. In R. Baer (Ed.), *Assessing mindfulness and acceptance processes in clients: Illuminating the theory and practice of change* (pp. 51–75). Oakland, CA: Context Press/New Harbinger Publications.

Costa, J., & Pinto-Gouveia, J. (2011). The mediation effect of experiential avoidance between coping and psychopathology in chronic pain. *Clinical Psychology & Psychotherapy, 18,* 34–47. doi:10.1002/cpp.699

Dalrymple, K. L., & Herbert, J. D. (2007). Acceptance and commitment therapy for generalized social anxiety disorder: A pilot study. *Behavior Modification, 31,* 543–568. doi:10.1177/0145445507302037

DeMars, C. E. (2003, April). *Recovery of graded response and partial credit parameters in MULTILOG and PARSCALE.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Diefenbach, G., McCarthy-Larzelere, M., Williamson, D., Mathews, A., Manguno-Mire, G., & Bentz, B. (2001). Anxiety, depression and the content of worries. *Depression and Anxiety, 14,* 247–250. doi:10.1002/da.1075

Donker, T., van Straten, A., Marks, I., & Cuijpers, P. (2009). A brief web-based screening questionnaire for common mental disorders: Development and validation. *Journal of Medical Internet Research, 11*(3), Article 19. doi:10.2196/jmir.1134

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Fischer, G. H., & Molenaar, I. W. (1995). *Rasch models: Foundations, recent developments and applications.* New York, NY: Springer Verlag.

Fledderus, M., Bohlmeijer, E. T., & Pieterse, M. E. (2010). Does experiential avoidance mediate the effects of maladaptive coping styles on psychopathology and mental health? *Behavior Modification, 34,* 503–519. doi:10.1177/0145445510378379

Fledderus, M., Bohlmeijer, E. T., Pieterse, M. E., & Schreurs, K. M. G. (2012). Acceptance and commitment therapy as guided self-help for psychological distress and positive mental health: A randomized controlled trial. *Psychological Medicine, 42,* 485–495. doi:10.1017/S0033291711001206

Fledderus, M., Bohlmeijer, E. T., Smit, F., & Westerhof, G. J. (2010). Mental health promotion as a new goal in public mental health care: A randomized controlled trial of an intervention enhancing psychological flexibility. *American Journal of Public Health, 100,* 2372–2378. doi:10.2105/AJPH.2010.196196

Fletcher, L., & Hayes, S. C. (2005). Relational frame theory, acceptance and commitment therapy, and a functional analytic definition of mindfulness. *Journal of Rational-Emotive and Cognitive-Behavioral Therapy, 23,* 315–336. doi:10.1007/s10942-005-0017-7

Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification, 31,* 772–799. doi:10.1177/0145445507302202

Gebhardt, E., & Adams, R. J. (2007). The influence of equating methodology on reported trends in PISA. *Journal of Applied Measurement, 8,* 305–322.

Gifford, E. V., Kohlenberg, B. S., Hayes, S. C., Antonuccio, D. O., Piasecki, M. M., Rasmussen-Hall, M. L., & Palm, K. M. (2004). Acceptance-based treatment for smoking cessation. *Behavior Therapy, 35,* 689–705. doi:10.1016/S0005-7894(04)80015-7

Glas, C. A. W. (1998). Detection of differential item functioning using Lagrange multiplier tests. *Statistica Sinica, 8,* 647–668.

Glas, C. A. W. (1999). Modification indices for the 2-PL and the nominal response model. *Psychometrika, 64,* 273–294. doi:10.1007/BF02294296

Glas, C. A. W. (2010). *Preliminary manual of the software program Multidimensional Item Response Theory (MIRT).* Retrieved from http://www.utwente.nl/gw/omd/afdeling/Glas/

Glas, C. A. W., Geerlings, H., van de Laar, M. A. F. J., & Taal, E. (2009). Analysis of longitudinal randomized clinical trials using item response models. *Contemporary Clinical Trials, 30,* 158–170. doi:10.1016/j.cct.2008.12.003

Gloster, A. T., Klotsche, J., Chaker, S., Hummel, K. V., & Hoyer, J. (2011). Assessing psychological flexibility: What does it add above and beyond existing constructs? *Psychological Assessment, 23,* 970–982. doi:10.1037/a0024135

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Haringsma, R., Engels, G. I., Beekman, A. T. F., & Spinhoven, P. (2004). The criterion validity of the Center for Epidemiological Studies Depression Scale (CES-D) in a sample of self-referred elders with depressive symptomatology. *International Journal of Geriatric Psychiatry, 19,* 558–563. doi:10.1002/gps.1130

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,* 1–25. doi:10.1016/j.brat.2005.06.006

Hayes, S. C., Strosahl, K. D., Wilson, K. G., Bissett, R. T., Pistorello, J.,

Toarmino, D., . . . McCurry, S. M. (2004). Measuring experiential avoidance: A preliminary test of a working model. *Psychological Record, 54,* 553–578.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Jacobs, N., Kleen, M., de Groot, F., & A-Tjak, J. (2008). Het meten van experiëntiële vermijding: De Nederlandstalige versie van de Acceptance and Action Questionnaire-II (AAQ-II) [Measuring experiential avoidance: Dutch translation of the Acceptance and Action Questionnaire-II (AAQ-II)]. *Gedragstherapie, 41,* 349–361.

Karekla, M., & Panayiotoua, G. (2011). Coping and experiential avoidance: Unique or overlapping constructs? *Journal of Behavior Therapy and Experimental Psychiatry, 42,* 163–170. doi:10.1016/j.jbtep.2010.10.002

Keyes, C. L. M. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior, 43,* 207–222. doi:10.2307/3090197

Kocovski, N. L., Fleming, J., & Rector, N. A. (2009). Mindfulness and acceptance-based group therapy for social anxiety disorder: An open trial. *Cognitive and Behavioral Practice, 16,* 276–289. doi:10.1016/j.cbpra.2008.12.004

Lamers, S. M. A., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., & Keyes, C. L. M. (2011). Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF). *Journal of Clinical Psychology, 67,* 99–110. doi:10.1002/jclp.20741

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

Lindesay, J., Baillon, S., Brugha, T., Dennis, M., Stewart, R., Araya, R., & Meltzer, H. (2006). Worry content across the lifespan: An analysis of 16- to 74-year-old participants in the British National Survey of Psychiatric Morbidity 2000. *Psychological Medicine, 36,* 1625–1633. doi:10.1017/S0033291706008439

Lopez, W. (1996). Communication validity and rating scales. *Rasch Measurement Transactions, 10,* 482–483.

Lord, F. M. (1983). Small *n* justifies the Rasch model. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 51–61). New York, NY: Academic Press.

McCracken, L. M., Vowles, K. E., & Eccleston, C. (2004). Acceptance of chronic pain: Component analysis and a revised assessment method. *Pain, 107,* 159–166. doi:10.1016/j.pain.2003.10.012

McCracken, L. M., & Zhao-O'Brien, J. (2010). General psychological acceptance and chronic pain: There is more to accept than the pain itself. *European Journal of Pain, 14,* 170–175. doi:10.1016/j.ejpain.2009.03.004

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159–176. doi:10.1177/014662169201600206

Muthén, L. K., & Muthén, B. (2008). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.

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

Olssøn, I., Mykletun, A., & Dahl, A. A. (2005). The Hospital Anxiety and Depression Rating Scale: A cross-sectional study of psychometrics and case finding abilities in general practice. *BMC Psychiatry, 5,* Article 46. doi:10.1186/1471-244X-5-46

Powers, M. B., Zum Vörde Sive Vörding, M. B., & Emmelkamp, P. M. G. (2009). Acceptance and commitment therapy: A meta-analytic review. *Psychotherapy and Psychosomatics, 78,* 73–80. doi:10.1159/000190790

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401. doi:10.1177/014662167700100306

Reeve, B. B., & Fayers, P. (2005). Applying item response theory model-

ling for evaluating questionnaire item and scale properties. In P. Fayers & R. D. Hays (Eds.), *Assessing quality of life in clinical trials: Methods of practice* (2nd ed., pp. 55–73). Oxford, England: Oxford University Press.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552–566. doi:10.1037/0033-2909.114.3.552

Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement, 27,* 133–144.

Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. New York, NY: Guilford Press.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janav, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuro-psychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59,* 22–33.

Spinhoven, P. H., Ormel, J., Sloekers, P. P. A., Kempen, G. I. J. M., Speckens, A. E. M., & van Hemert, A. M. (1997). A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychological Medicine, 27,* 363–370. doi:10.1017/S0033291796004382

van Groen, M. M., ten Klooster, P. M., Taal, E., van de Laar, M. A., & Glas, C. A. (2010). Application of the Health Assessment Questionnaire

Disability Index to various rheumatic diseases. *Quality of Life Research, 19,* 1255–1263. doi:10.1007/s11136-010-9690-9

Veehof, M. M., ten Klooster, P. M., Taal, E., Westerhof, G. J., & Bohlmeijer, E. T. (2011). Psychometric properties of the Dutch Five Facet Mindfulness Questionnaire (FFMQ) in patients with fibromyalgia. *Clinical Rheumatology, 30,* 1045–1054. doi:10.1007/s10067-011-1690-9

Vowles, K. E., & McCracken, L. M. (2008). Acceptance and values-based action in chronic pain: A study of treatment effectiveness and process. *Journal of Consulting and Clinical Psychology, 76,* 397–407. doi:10.1037/0022-006X.76.3.397

Weisscher, N., Glas, C. A., Vermeulen, M., & de Haan, R. J. (2010). The use of an item response theory based disability item bank across diseases: Accounting for differential item functioning. *Journal of Clinical Epidemiology, 63,* 543–549. doi:10.1016/j.jclinepi.2009.07.016

Wheaton, M. G., Berman, N. C., & Abramowitz, J. S. (2010). The contribution of experiential avoidance and anxiety sensitivity in the prediction of health anxiety. *Journal of Cognitive Psychotherapy, 24,* 229–239. doi:10.1891/0889-8391.24.3.229

World Health Organization. (2008). *The global burden of disease: 2004 update*. Retrieved from http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67,* 361–370.

Call for Papers: Special Issue
Ethical, Regulatory, and Practical Issues in Telepractice

Professional Psychology: Research and Practice will publish a special issue on recent ethical, regulatory and practical issues related to telepractice. In its broadest definition the term telepractice refers to any contact with a client/patient other than face-to-face in person contact. Thus, telepractice may refer to contact on a single event or instance such as via the telephone or by means of electronic mail, social media (e.g., Facebook) or through the use of various forms of distance visual technology. We would especially welcome manuscripts ranging from the empirical examination of the broad topic related to telepractice to those manuscripts that focus on a particular subset of issues associated with telepractice. Although manuscripts that place an emphasis on empirical research are especially encouraged, we also would welcome articles on these topics that place an emphasis on theoretical approaches as well as an examination of the extant literature in the field. Finally, descriptions of innovative approaches are also welcome. Regardless of the type of article, all articles for the special issue will be expected to have practice implications to the clinical setting. Manuscripts may be sent electronically to the journal at http://www.apa.org/pubs/journals/pro/index.aspx to the attention of Associate Editor, Janet R. Matthews, Ph.D.