MANUSCRIPT PUBLISHED IN THE PSYCHOLOGICAL RECORD

https://doi.org/10.1007/s40732-016-0183-2

Psychometric properties of the Acceptance and Action Questionnaire – II in Colombia

Francisco J. Ruiz[1] (franciscoj.ruizj@konradlorenz.edu.co)

Juan C. Suárez-Falcón[2]

Sebastián Cárdenas-Sierra[3]

Yesica A. Durán[3]

Karen Guerrero[3]

Diana Riaño-Hernández[1]

[1]Fundación Universitaria Konrad Lorenz, [2]Universidad Nacional de Educación a Distancia,

[3]Universidad Santo Tomás

Correspondence address: Fundación Universitaria Konrad Lorenz, Carrera 9 bis, Nº 62-43, Bogotá (Cundinamarca, Colombia), teléfono: (+57 1) 347 23 11 ext. 185

Abstract

The Acceptance and Action Questionnaire – II (AAQ-II) is a widely used measure of experiential avoidance and psychological inflexibility, two key constructs of acceptance and commitment therapy (ACT). The main aim of this study was to analyze the psychometric properties and factor structure of the Spanish version of the AAQ-II in Colombia. A secondary aim was to analyze measurement invariance of the AAQ-II across nonclinical and clinical samples in view that previous research has yielded mixed results. The AAQ-II was administered to a total of 1759 participants, including a sample of undergraduates, one of general population, and a clinical sample. Data were very similar to the ones obtained in the original AAQ-II version. The internal consistency across the different samples was good (Cronbach's alpha between .88 and .91). The one-factor model found in the original scale showed a good fit to the data. Measurement invariance was also found across sample (clinical and nonclinical) and gender. The mean score of the clinical sample on the AAQ-II was significantly higher than the scores of the nonclinical samples. The AAQ-II was sensitive to the effects of a 1-session ACT intervention and AAQ-II scores were significantly related to emotional symptoms, mindfulness, satisfaction with life, and dysfunctional attitudes. The Spanish version of the AAQ-II shows good psychometric properties in Colombia.

Key words: Acceptance and Action Questionnaire; Acceptance and commitment therapy; Experiential avoidance; Psychological inflexibility; Instrumental study.

Psychometric properties of the Acceptance and Action Questionnaire – II in Colombia

Acceptance and commitment therapy (ACT; Hayes, Strosahl, & Wilson, 1999) is a functional-contextual model of psychological intervention that has been successfully applied to a wide range of psychological disorders, health problems, and performance-related issues (e.g., A-Tjak et al., 2015; Ruiz, 2012). The cornerstone of ACT is that private events of any kind (such as negative thoughts, aversive memories, feelings, and bodily sensations) do not directly cause psychopathology and behavioral ineffectiveness. Instead, it is how the individual reacts to these private events what can cause problems or not.

The ACT model has evolved across the years, but maintaining the previous core assumption intact. In its beginnings, the ACT model emphasized the role of experiential avoidance as a behavioral process pervasive across psychological disorders. Experiential avoidance is a behavior regulation pattern aimed to avoid and/or escape from private experiences (i.e., thoughts, memories, feelings, etc.) that are experienced as aversive, even when doing so leads to behaving in ways incompatible with one's values and goals (Hayes et al., 1999; Hayes, Wilson, Gifford, Follette, & Strosahl, 1996; Luciano & Hayes, 2001).

In later years, psychological inflexibility has emerged in the ACT literature as a broader construct that contains experiential avoidance (Hayes & Strosahl, 2004). Psychological inflexibility can be defined as the dominance of the discriminative functions of ongoing private events (including thoughts, memories and sensations with appetitive and aversive functions) over other sources of stimulus control such as chosen values (Bond et al., 2011; Törneke, Luciano, Barnes-Holmes, & Bond, 2016).

Parallel to the development and expansion of ACT, an increasing interest has been posited on the design of self-report measures of experiential avoidance and psychological inflexibility (Hayes et al., 2004; Bond et al., 2011). The first attempt to measure

experiential avoidance was the Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004). A good part of the evidence showing the pernicious role of experiential avoidance in relation to mental health and behavioral effectiveness (e.g., Hayes, Luoma, Bond, Masuda, & Lillis, 2006; López et al., 2010; Ruiz, 2010) and the reduction of experiential avoidance levels as a process of change in ACT (e.g., Ruiz, 2012) has been obtained through the AAQ. The AAQ was designed to measure general levels of experiential avoidance, as averaged across different contexts, in clinical and community samples. Nevertheless, the ACT model suggests that experiential avoidance/psychological inflexibility is contextually controlled and, therefore, can vary across different contexts. Accordingly, a growing number of AAQ versions specifically tailored to particular areas has appeared during the last years (e.g., Bond, Lloyd, & Guenole, 2013; Jurascio, Forman, Timko, Butryn, & Goodwin, 2011; Ruiz & Odriozola-González, 2014).

Despite the advancement in the collection of empirical data, the AAQ showed some problems with regard to its internal consistency and factor structure (Bond et al., 2011). For this reason, an improved, second version of the AAQ (i.e., the AAQ-II) was developed to overcome such limitations. The AAQ-II (Bond et al., 2011) is dedicated to measure experiential avoidance and psychological inflexibility. The AAQ-II possesses good internal consistency and a one-factor structure in clinical and nonclinical participants (Bond et al., 2011; Fledderus, Oude, Klooster, & Bohlmeijer, 2012) while maintaining similar levels of external validity as the AAQ. It has been translated to multiple languages, including Spanish (e.g., Ruiz, Langer, Luciano, Cangas, & Beltrán, 2013), showing similar psychometric properties and factor structure among them (Monestès et al., in press).

Some criticisms have emerged during the last years, however, with respect to the AAQ-II. Specifically, some authors have suggested that the AAQ-II content may lead to

criterion contamination because the items include descriptions of negative affectivity (Gámez, Chmielewski, Kotov, Ruggero, & Watson, 2011; Gámez et al., 2014). In this sense, Wolgast (2014) suggested that the AAQ-II is more a measure of psychological distress than a measure of experiential avoidance. However, other authors have shown that the AAQ-II scores explains additional variance above and beyond traditional measures of affect (Gloster, Klotsche, Chaker, Hummel, & Hoyer, 2011; Ruiz, 2010). Lastly, other authors have shown that scores on experiential avoidance, as measured by the AAQ and the AAQ-II, are stable across time in clinical and nonclinical samples notwithstanding the fluctuation of emotional symptoms (Ruiz & Odriozola-González, 2015; Spinhoven, Drost, de Rooij, van Hemert, & Penninx, 2014).

Although the psychometric properties of the AAQ-II have been analyzed in numerous studies (e.g., Monestès et al., in press), the measurement invariance (i.e., whether the AAQ-II items operate equivalently across samples) across clinical and nonclinical samples has not been systematically explored. This is important because violations of measurement invariance might impede a meaningful comparison between the scores of clinical and nonclinical samples. In the original study, Bond et al. (2011) obtained measurement invariance across different samples of undergraduates, employees, and a people seeking outpatient psychological treatment for substance misuse. However, Costa, Marôco, Pinto-Gouveia, and Galhardo (2014) did not find conclusive evidence of measurement invariance across a general normative group and a clinical sample in Portugal. In conclusion, additional studies are needed to explore the measurement invariance of the AAQ-II across clinical and nonclinical samples.

To our knowledge, there is no measure of experiential avoidance and psychological inflexibility validated in Colombian samples, which makes difficult the conduction of

ACT-related studies in this country. Additionally, testing measures in culturally diverse samples enhances both our confidence in the measure and the cross-cultural relevance of the underlying theory being measured (Elosua, Mujika, Almeida, & Hermosilla, 2014). Since the Spanish translation of the AAQ (Barraca, 2004) has similar limitations to the ones mentioned for the original version, the primary aim of this study was to analyze the factor structure and psychometric properties of the Spanish version of the AAQ-II by Ruiz et al. (2013) in Colombia. Additionally, the secondary aim of the study was to explore the measurement invariance of the AAQ-II across clinical and nonclinical samples. Importantly, although the studies of Barraca (2004) with the AAQ and Ruiz et al. (2013) with the AAQ-II included both clinical and nonclinical samples, measurement invariance was not analyzed.

A small pilot study was conducted first to explore whether Colombian people experienced difficulties in understanding the items of the Spanish versions of the AAQ-II. After confirming that the AAQ-II items were understandable by Colombian participants, the AAQ-II was administered in conjunction with other related measures to three samples with a total of 1759 participants: a sample of 762 undergraduates, a sample of 724 Colombian people recruited through internet, and a clinical sample of 277 participants.

Method

Participants

Sample 1. This sample consisted of 762 undergraduates (age range 18-63, M = 21.16, SD = 3.76) from seven universities of Bogotá. Forty-six percent of the sample was studying Psychology. The other studies included Law, Engineering, Philosophy, Communication, Business, Medicine, and Theology. Sixty-two percent were women. Of the overall sample, 26% of participants had received psychological or psychiatric treatment

at some time, but only 4.3% were currently in treatment. Also, 2.9% of participants were taking some psychotropic medication.

Sample 2. The sample consisted of 724 participants (74.4% females) with age ranging between 18 and 88 years (M = 26.11, SD = 8.93). The relative educational level of the participants was: 17.8% primary studies (i.e., compulsory education) or mid-level study graduates (i.e., high school or vocational training), 63.8% were undergraduates or college graduates, and 18.4% were currently studying or had a postgraduate degree. They responded to an anonymous internet survey distributed through social media. All of them were Colombian. Forty-five percent reported having received psychological or psychiatric treatment at some time, but only 8.4% were currently in treatment. Also, 5.4% of participants reported consumption of some psychotropic medication.

Sample 3. It consisted of 277 patients (64% of them were women) with an age range of 18 to 67 years (M = 28.4, SD = 11.33), suffering from emotional (88.4%) or sexual disorders (11.6%). All participants were being evaluated in some private psychological consultation center of Bogotá or were assessed to participate in a research about the efficacy of brief ACT protocols to treat emotional disorders. Only 6.1% of the participants reported that they were consuming some psychotropic medication.

Sample 4. This sample consisted of 11 participants (2 men, mean age = 22.18, SD = 4.40, age range: 18 to 32) who participated in a multiple baseline design study that analyzed the effect of a 1-session ACT intervention to disrupt problematic worry and rumination. The relative educational level of the participants was as follows: 9% mid-level study graduates, 55% undergraduate students, and 36% were college graduates. Participants were recruited through advertisements in social media and had spent at least 6 months

entangled in thoughts, memories, and/or worries that provoked significant interference in at least two life areas.

Instruments

Acceptance and Action Questionnaire – II (AAQ-II; Bond et al., 2011; Spanish translation by Ruiz et al., 2013). The AAQ-II is a 7-item, 7-point Likert-type scale (7 = *always*; 1 = *never true*) that measures general experiential avoidance or psychological inflexibility. The items reflect: (a) unwillingness to experience unwanted emotions and thoughts (e.g., "I am afraid of my feelings," "I worry about not being able to control my worries and feelings"), and (b) the inability to be in the present moment and behave according to value-directed actions when experiencing unwanted psychological events (e.g., "My painful experiences and memories make it difficult for me to live a life that I would value," "My painful memories prevent me from having a fulfilling life," "Worries get in the way of my success"). The Spanish version by Ruiz et al. (2013) showed good psychometric properties (mean alpha of .88) and a one-factor structure.

Depression, Anxiety, and Stress Scales – 21 (DASS-21; Antony, Bieling, Cox, Enns, & Swinson, 1998; Spanish version by Daza, Novy, Stanley, & Averill, 2002). The DASS-21 is a 21-item, 4-point Likert-type scale (3 = *applied to me very much, or most of the time*; 0 = *did not apply to me at all*) consisting of sentences describing negative emotional states. It contains three subscales (Depression, Anxiety, and Stress) and has shown good internal consistency and convergent and discriminant validity. Alpha values in this study were good for all subscales (for the Depression, Anxiety, and Stress subscales, respectively, Sample 1: .86, .80, and .80; Sample 2: .92, .85, and .86; Sample 3: .92, .85, and .90). Medium to strong positive correlations were expected between the AAQ-II and the DASS-21 subscales.

General Health Questionnaire – 12 (Goldberg & Williams, 1988; Spanish version by Rocha, Pérez, Rodríguez-Sanz, Borrell, & Obiols, 2011). The GHQ-12 is a 12-item, 4-point Likert-type scale that is frequently used as screening for psychological disorders. Respondents are asked to indicate the degree to which they have recently experienced a range of common symptoms of distress, with higher scores reflecting greater levels of psychological distress. The Likert scoring method was used in this study, with scores ranging from 0 to 3 assigned to each of the four response options. Alpha values for the GHQ-12 in this study were good (Sample 1: .88; Sample 3: .93). Medium to strong positive correlations were expected between the AAQ-II and the GHQ-12.

Dysfunctional Attitude Scale - Revised (DAS-R; de Graaf, Roelofs, & Huibers, 2009; Spanish version by Ruiz et al., 2015, in press). The DAS is a classic measure of dysfunctional schemas. The revised version of the DAS is a 17-item, 7-point Likert-type scale (7 = *fully agree*; 1 = *fully disagree*) grouped in two factors: Perfectionism/Performance evaluation and Dependency. In a Colombian sample, the DAS-R showed excellent psychometric properties and a factor structure with two-correlated factors and a second-order factor (Ruiz et al., in press). Medium to strong positive correlations were expected between the AAQ-II and the DAS-R.

Satisfaction with Life Survey (SWLS; Diener, Emmons, Larsen, & Griffin, 1985; Spanish version by Atienza, Pons, Balaguer, & García-Merita, 2000). The SWLS is a 5-item, 7-point Likert-type scale (7 = *strongly agree*; 1 = *strongly disagree*) that measures self-perceived well-being. Example of items are “I am satisfied with my life” and “In most ways, my life is close to my ideal.” The SWLS has good psychometric properties and convergent validity. Alpha values in this study for the SWLS were good (Sample 1: .85;

Sample 2: .89; Sample 3: .84). Medium to strong negative correlations were expected between the AAQ-II and SWLS.

Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003; Spanish version by Soler et al., 2012). The MAAS is a 15-item, 6-point Likert-type scale (6 = *almost never*; 1 = almost never) designed to measure the extent to which individuals pay attention during several tasks or, in contrast, behave on "autopilot," without paying enough attention to them. The MAAS does not require familiarity with meditation. Higher scores indicate greater mindfulness level. The MAAS has shown good psychometric properties and a one-factor structure in a Colombian sample (Ruiz, Suárez-Falcón, & Riaño-Hernández, in press). Medium to strong negative correlations were expected between the AAQ-II and MAAS.

Procedure

The procedure of the study was approved by the Center for Psychological Research of the Fundación Universitaria Konrad Lorenz. Following the suggestions by Elosua et al. (2014), a small pilot study was conducted first to explore whether Colombian people experienced difficulties in understanding the items of the Spanish versions of the AAQ-II, DASS-21, GHQ-12, and SWLS. Ten Colombian undergraduates found no difficulties to understand the AAQ-II items; therefore, we decided to apply the original scale without changes.

In Sample 1, the administration of the questionnaire package was collective and conducted in the participants' classrooms during the beginning of a regular class. Participants in Sample 2 responded to an anonymous internet survey distributed through social media. Lastly, participants in Sample 3 responded to the questionnaires during one of

the clinical assessment interviews at the beginning of the treatment in the presence of their therapist.

In Samples 1 to 3, the study was presented, and individuals who signed an informed consent were provided with a questionnaire packet. Specifically, participants in Sample 1 responded to the AAQ-II, DASS-21, GHQ-12, DAS-R, MAAS, and SWLS. Participants in Sample 2 responded to the AAQ-II, SWLS, and DASS-21. Lastly, participants in Sample 3 responded to the AAQ-II, DASS-21, GHQ-12, and SWLS. Upon completion of the study, participants were debriefed about the aims of the study and thanked for their participation.

Participants in Sample 4 completed a baseline period ranging between 2 to 10 weeks and then received an ACT intervention specifically oriented to disrupt problematic worry and rumination. After that, participants completed follow-up measures for 6 weeks. The ACT protocol consisted of an approximately 75-minute, individual session. The main objectives of the protocol were: (a) to identify triggers for worrying/ruminating and experiential avoidance strategies related to them, (b) to promote creative hopelessness regarding the counterproductive effect of engaging in worry/rumination and the other experiential avoidance strategies, (c) to promote values clarification and the commitment to valued actions, and (d) to introduce defusion training.

Data analysis

Prior to conducting factor analyses, data from Samples 1 to 3 were examined, searching for missing values. Only one value of the AAQ-II was missing for Item 6. This datum was imputed using the imput missing values option of LISREL© (version 8.71, Jöreskog & Sörbom, 1999), which was the software used to conduct the confirmatory factor analyses (CFA).

Because the AAQ-II uses a Likert-type scale measured on an ordinal scale, a weighted least squares (WLS) estimation method using polychoric correlations was used in conducting CFA. The WLS method is recommended in large samples with fewer than 20 items (Holgado-Tello, Chacón-Moscoso, Barbero-García, & Villa-Abad, 2010; Jöreskog & Sörbom, 1996) as in the current study. In order to use the matrix of polychoric correlations, the assumption of bivariate normal distribution was analyzed by means of the chi squared test and the percentage of tests that rejected the null hypothesis of bivariate normality for each pair of correlations. Due to the sensitivity of the chi-square test, the Root Mean Square Error of Approximation (RMSEA) was also analyzed for each pair of correlations. Hooper, Coughlan, and Mullen (2008) point out that the parameter estimation is not very affected when RMSEA values are not higher than .1 (Cano-García et al., 2015). Since previous CFAs on the AAQ-II found method effect in responses to Items 1 and 4 due to their highly similar content (Bond et al., 2011; Monèstes et al., in press), we compared the fit of a one-factor model where error terms between these items were allowed to correlate versus a one-factor model with no error correlations.

Goodness of fit was examined computing the following fit indexes: (a) RMSEA; (b) the Comparative Fit Index (CFI); (c) the Non-Normed Fit Index (NNFI); and (d) the Expected Cross-Validation Index (ECVI). According to Kelloway (1998) and Hu and Bentler (1999), RMSEA values of .10 represent a good fit, and values below .05 represent a very good fit to the data. With respect to the CFI and NNFI, values above .90 indicate well-fitting models, and above .95 represent a very good fit to the data. The ECVI was computed to compare the goodness of fit of the one-factor model to the one-factor model with error terms for Items 1 and 4 allowed to correlate (lower values indicate better fit to the model). Lastly, the difference between the chi-square-values for the two models was calculated

following a likelihood ratio test under the null hypothesis that the one-factor model fits as well as the one-factor model with error terms for Items 1 and 4 allowed to correlate. This chi-square difference is also chi-square distributed with degrees of freedom equal to the difference between the degrees of freedom of the two compared models.

As in Bond et al. (2011), additional CFA were performed to test for measurement invariance across samples and gender. In other words, we analyzed whether the item factor loadings are invariant across the three samples and between men and women. In so doing, the relative fit of two models was compared. The first model (the multiple-group baseline model) allowed the seven unstandardized factor loadings and one error covariance between Items 1 and 4 to vary across the three samples, whereas the second model (constrained model) placed equality constraints (i.e., invariance) on those loadings and error covariance. Equality constraints were not placed on estimates of the factor variances because these are known to vary across groups even when the indicators are measuring the same construct in a similar manner (Kline, 2005). Given the hierarchy of these nested and increasingly restrictive models, they were compared to each other. The parsimonious model (constrained model) was selected if the following four criteria suggested by Cheung and Rensvold (2002) and Chen (2007) were met: (a) the constrained model did not generate a significantly worse fit than the unconstrained model (the multiple-group baseline model) according to the chi-square test; (b) the difference in RMSEA (ΔRMSEA) was lower than .01; (c) the difference in CFI (ΔCFI) was greater than -.01; and (d) the difference in NNFI (ΔNNFI) was greater than -.01.

The remaining statistical analyses were performed on SPSS 19©. Cronbach's alphas were computed providing 95% confidence intervals (CI) to explore the internal consistency of the AAQ-II in Samples 1 to 3 and the overall sample. Corrected item-total correlations

were obtained to identify items that should be removed because of low discrimination item index (i.e., values below .20). Descriptive data were also calculated, and gender differences in AAQ-II scores were explored by computing Student's *t*. To examine discriminant construct validity, scores on the AAQ-II were compared between (a) participants with scores above and below the cutoff on the GHQ-12 (overall score higher than 11) in Sample 1, and (b) participants in Samples 1 and 2 with respect to participants in Sample 3. Pearson correlations between the AAQ-II and other scales were calculated to assess convergent construct validity. Lastly, to explore whether the AAQ-II scores were sensitive to the effects of a 1-session ACT intervention, Student's *t*-tests for dependent data were conducted between the last AAQ-II score of participants' baseline and the 6-week follow-up. Cohen's *d* for within-participant studies was also computed.

Results

Factor structure

The results of the chi-square test to explore bivariate normality showed that this assumption was accepted in 19% of the correlations. However, the RMSEA values were lower than .1 in all correlations, which supports the use of the matrix of polychoric correlations to conduct the CFA. Table 1 presents the results of the CFA conducted on the overall sample comparing the one-factor model to the one-factor model with error terms for Items 1 and 4 allowed to correlate. As in Bond et al. (2011), results showed that the latter model obtained the best fit. Specifically, the chi-square difference between the two competing models was 124.05 ($df = 1$, $p < .01$), indicating that the one-factor model with error terms allowed to correlate showed a significantly better fit to the data. This factor model also had the lowest ECVI value (.087, 90% CI [.069, .11]). Scores on the goodness-of-fit indexes for this model were good for the RMSEA (RMSEA = .069, 90% CI [.059,

.081]), and very good for the CFI and NNFI (.98 and 0.97, respectively). Figure 1 depicts the results of the standardized solution of the one-factor model with correlated measurement errors between Items 1 and 4.

INSERT TABLE 1 ABOUT HERE

INSERT FIGURE 1 ABOUT HERE

Measurement invariance

Table 2 shows that the multiple-group baseline models fit the data well, with all values of the goodness-of-fit indexes suggesting good-fitting solutions. When equality constraints were placed on the factor loadings and error covariance between Items 1 and 4, there was no significant decrement in goodness of fit, suggesting that the measures were invariant across clinical and nonclinical samples and gender. The four criteria recommended by Cheung and Rensvold (2002) and Chen (2007) were met. Specifically, the χ^2 diff test were not statistically significant (across samples: $\chi^2(14) = 23.29$, $p > .01$; across gender: $\chi^2(7) = 6.19$, $p > .01$), the differences in RMSEA were lower than .01, and the differences in CFI and NNFI were higher than -.01.

INSERT TABLE 2 ABOUT HERE

Internal Consistency, Descriptive Data and Criterion Validity

Table 3 shows that Cronbach's alpha of the AAQ-II ranged from .88 (Sample 1) to .91 (Sample 2), with an overall alpha of .91. Corrected item-total correlations of the AAQ-II ranged from .63 to .69 in Sample 1, from .67 to .78 in Sample 2, and from .66 to .77 in Sample 3. This indicates that all items presents a high level of discrimination index.

INSERT TABLE 3 ABOUT HERE

The mean score of men (M = 18.92, SD = 8.04) in Sample 1 was slightly lower than that of women (M = 20.64, SD = 8.50), with a statistically significant difference (t = -2.76, p = .006). No statistically significant differences (t = 1.22, p = .22) were found in Sample 2 between men (M = 23.59, SD = 9.46) and women (M = 22.57, SD = 9.47). Likewise, no statistically significant differences were found in Sample 3 with relation to sex (men: M = 28.25, SD = 12.23; women: M = 30.40, SD = 9.12; t = -1.52, p = .13).

In Sample 1, participants with scores above the cutoff on the GHQ-12 scored statistically significantly higher on the AAQ-II (M = 24.88, SD = 8.25) than those with scores below the cutoff (M = 16.83, SD = 6.79; t = 14.00, p < .001). Mean score of participants in the clinical sample (Sample 3) was higher than those of participants in Sample 1 (t = -12.33, p < .001) and Sample 2 (t = -9.86, p < .001).

Pearson Correlations with other Related Constructs

The AAQ-II showed correlations with all the other assessed constructs in theoretically coherent ways (see Table 4). Specifically, the AAQ-II showed positive correlations with psychological distress (GHQ-12), depression, anxiety, and stress symptoms (DASS-21), and dysfunctional schemas (DAS-R); and negative correlations with mindful awareness (MAAS), and satisfaction with life (SWLS).

INSERT TABLE 4 ABOUT HERE

Sensitivity to treatment

In Sample 4, participants' mean score in the last baseline assessment was 29.09 (SD = 6.14), whereas the mean score at the 6-week follow-up was 18.82 (SD = 6.57). The difference was statistically significant and with a very large effect size (t = 7.13, p < .001, d = 2.16).

Discussion

In order to advance in the research of the ACT model, several attempts to measure experiential avoidance and psychological inflexibility have been proposed during the last years. By far, the most widely used measures of these constructs have been the two versions of the AAQ. While the psychometric properties and factor structure of the AAQ-I were questionable, the AAQ-II is widely considered as a psychometrically sound instrument (Bond et al., 2011). Indeed, the AAQ-II has yielded the same one-factor structure in a wide range of European countries (Monestès et al., in press). A less explored topic has been the measurement invariance of the AAQ-II across clinical and nonclinical samples, which has obtained mixed results. There was no published data of the AAQ-II in Colombia, which made more difficult the research in ACT in this country. Accordingly, the current study aimed to explore the psychometric properties and factor structure of the Spanish version of the AAQ-II by Ruiz et al. (2013), which was previously validated in Spain. Additionally, we explored the measurement invariance of the AAQ-II across large clinical and nonclinical samples.

The data obtained showed that the AAQ-II had good psychometric properties in Colombia and very similar to the ones found in Spain. Specifically, the AAQ-II showed construct validity to the extent that factor analyses showed the same one-factor solution as in Bond et al. (2011). The internal consistency of the AAQ-II was very good with an overall alpha of .91. The AAQ-II also showed discriminant validity to the extent that its scores discriminated between clinical and nonclinical samples. The instrument showed convergent validity in view of the positive correlations found with emotional symptoms and dysfunctional attitudes and divergent validity according to the negative correlations with mindfulness and life satisfaction. Lastly, the AAQ-II was shown to be sensitive to the

effect of a one-session ACT intervention with people suffering from problematic worry and rumination.

The current study also provides evidence of the measurement invariance of the AAQ-II in Colombia across large clinical and nonclinical samples. This is especially relevant because previous studies with the AAQ and AAQ-II in Spain (Barraca, 2004; Ruiz et al., 2013) did not addressed this issue. Additionally, the empirical evidence on this topic is scarce yet, with studies showing mixed results: while Bond et al. (2011) found measurement invariance across clinical and nonclinical samples, Costa et al. (2014) did not find conclusive evidence.

Some limitations of this study are worth mentioning. Firstly, no systematic information was obtained concerning the diagnosis in clinical participants. Secondly, some of the instruments used to explore the convergent and divergent validity of the AAQ-II lacked formal validation in Colombian samples (DASS-21, GHQ-12, and SWLS); however, their internal consistencies were adequate and similar to the ones obtained in the validation studies. Thirdly, the samples used in this study were mostly composed by relatively young and well-educated participants so that further analyses are needed to warrant the psychometric properties of the AAQ-II in older and less educated participants.

In conclusion, the Spanish translation of the AAQ-II by Ruiz et al. (2013) can be used to measure experiential avoidance and psychological inflexibility in Colombia according to the reliability and validity data provided in this study. Further studies might analyze the measurement invariance of the AAQ-II in Spanish-speaking countries and across clinical and nonclinical samples.

Compliance with Ethical Standards:

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales (DASS) in clinical groups and a community sample. *Psychological Assessment, 10*, 176-181. doi:10.1037/1040-3590.10.2.176

Atienza, F. L., Pons, D., Balaguer, I., & García-Merita, M. (2000). Propiedades psicométricas de la Escala de Satisfacción con la Vida en adolescentes [Psychometric properties of the Satisfaction with Life Scale in adolescents]. *Psicothema, 12*, 314-319.

A-Tjak, J. G. L., Davis, M. L., Morina, N., Powers, M. B., Smits, J. A. J., & Emmelkamp, P. M. G. (2015). A meta-analysis of the efficacy of acceptance and commitment therapy for clinically relevant mental and physical health problems. *Psychotherapy and Psychosomatics, 84*, 30-36.

Barraca, J. (2004). Spanish adaptation of the Acceptance and Action Questionnaire (AAQ). *International Journal of Psychology and Psychological Therapy, 4*, 505-515.

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

Bond, F. W., Lloyd, J., & Guenole, N. (2013). The Work-related Acceptance and Action Questionnaire (WAAQ): Initial psychometric findings and their implications for measuring psychological flexibility in specific contexts. *Journal of Occupational and Organizational Psychology, 86*, 331-347. doi:10.1111/joop.12001

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*, 822-848. doi:10.1037/0022-3514.84.4.822

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 464-504. doi:10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 233-255. doi:10.1207/S15328007SEM0902_5

Costa, J., Marôco, J., Pinto-Gouveia, J., & Galhardo, A. (2014). Validation of the psychometric properties of Acceptance and Action Questionnaire-II in clinical and nonclinical groups of Portuguese population. *International Journal of Psychology and Psychological Therapy, 14*, 353-364.

Daza, P., Novy, D. M., Stanley, M., & Averill, P. (2002). The Depression Anxiety Stress Scale-21: Spanish translation and validation with a Hispanic sample. *Journal of Psychopathology and Behavioral Assessment, 24*, 195-205.

de Graaf, L. E., Roelofs, J., & Huibers, M. J. H. (2009). Measuring dysfunctional attitudes in the general population: The Dysfunctional Attitude Scale (form A) Revised. *Cognitive Therapy and Research, 33*, 345-355. doi:10.1007/s10608-009-9229-y

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment, 49*, 71-75.

Elosua, P., Mujika, J., Almeida, L. S., & Hermosilla, D. (2014). Procedimientos analítico-racionales en la adaptación de tests. Adaptación al español de la Batería de Pruebas de Razonamiento [Judgmental-analytical procedures for adapting tests: Adaptation

to Spanish of the Reasoning Tests Battery]. *Revista Latinoamericana de Psicología, 46*, 117-126. doi:10.1016/S0120-0534(14)70015-9

Fledderus, M., Oude Voshaar, M. A., ten Klooster, P. M., & Bohlmeijer, E. T. (2012). Further evaluation of the psychometric properties of the Acceptance and Action Questionnaire–II. *Psychological assessment*, *24*, 925-936.

Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., Suzuki, N., & Watson, D. (2014). The Brief Experiential Avoidance Questionnaire: Development and initial validation. *Psychological Assessment, 26*, 35-45. doi:10.1037/a0034473

Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., & Watson, D. (2011). Development of a measure of experiential avoidance: The Multidimensional Experiential Avoidance Questionnaire. *Psychological Assessment, 23*, 692-713. Doi:10.1037/a0023242

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

Goldberg, D., & Williams, P. (1988). *A user's guide to the General Health Questionnaire*. Windsor, UK: NFER-Nelson.

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. (Eds.) (2004). *A practical guide to Acceptance and Commitment Therapy*. NY: Springer.

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

Hayes, S. C., Strosahl, K. D., Wilson, K. G., Bissett, R., Pistorello, J., Toarmino, D., Polusny, M. A., Dykstra, T. A., Batten, S. V., Bergan, J., Stewart, S. H., Zvolensky, M. J., Eifert, G. H., Bond, F. W., Forsyth, J. P., Karekla, M., & McCurry, S. M. (2004). Measuring experiential avoidance: A preliminary test of a working model. *The Psychological Record, 54*, 553-578.

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

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity: International Journal of Methodology, 44*, 153-166. doi:10.1007/s11135-008-9190-y

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*, 53-60.

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. doi:10.1080/10705519909540118

Jöreskog, K. G., & Sörbom, D. (1996). PRELIS 2 User's Reference Guide: A Program for Multivariate Data Screening and Data Summarization: A Preprocessor for LISREL. Chicago, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1999). *LISREL 8.30*. Chicago, IL: Scientific Software International.

Jurascio, A., Forman, E., Timko, C. A., Butryn, M., & Goodwin, C. (2011). The development and validation of the Food Craving Acceptance and Action Questionnaire (FAAQ). *Eating Behaviors, 12,* 182-187. doi:10.1016/j.eatbeh.2011.04.008

Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide*. Thousand Oaks, CA: Sage.

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

López, J. C., Ruiz, F. J., Feder, J., Barbero, A., Suárez, J. J., Rodríguez, J. A., & Luciano, C. (2010). The role of experiential avoidance in the performance on a high cognitive demand task. *International Journal of Psychology and Psychological Therapy, 10*, 475-488.

Luciano, M. C., & Hayes, S. C. (2001). Trastorno de evitación experiencial [Experiential avoidance disorder]. *International Journal of Clinical and Health Psychology, 1*, 109-157.

Monestès. J. L., Karekla, M., Jacobs, N., Michaelides, M., Hooper, N., Kleen, M., Ruiz, F. J., Miselli, G., Presti, G., Luciano, C., Villatte, M., Bond, F. W., Kishita, N., & Hayes, S. C. (in press). Toward a transdiagnostic approach to psychological disorders: Measuring experiential avoidance across six cultures with the Acceptance and Action Questionnaire – II. *European Journal of Psychological Assessment*.

Rocha, K., Pérez, K., Rodríguez-Sanz, M., Borrell, C., & Obiols, J. E. (2011). Propiedades psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en

población española [Psychometric properties and normative scores of the General Health Questionnaire (GHQ-12) in general Spanish population]. *International Journal of Clinical and Health Psychology, 11*, 125-139.

Ruiz, F. J. (2010). A review of Acceptance and Commitment Therapy (ACT) empirical evidence: Correlational, experimental psychopathology, component and outcome studies. *International Journal of Psychology and Psychological Therapy, 10*, 125-162.

Ruiz, F. J. (2012). Acceptance and commitment therapy versus traditional cognitive behavioral therapy: A systematic review and meta-analysis of current empirical evidence. *International Journal of Psychology and Psychological Therapy, 12*, 333-357.

Ruiz, F. J., Langer, A. I., Luciano, C., Cangas, A. J., & Beltrán, I. (2013). Measuring experiential avoidance and psychological inflexibility: The Spanish translation of the Acceptance and Action Questionnaire. *Psicothema, 25,* 123-129. doi:10.7334/psicothema2011.239

Ruiz, F. J., & Odriozola-González, P. (2014). The Spanish version of the Work-related Acceptance and Action Questionnaire (WAAQ). *Psicothema, 26*, 63-68. doi:10.7334/psicothema2013.110

Ruiz, F. J., & Odriozola-González, P. (2015). Comparing cognitive, metacognitive, and acceptance and commitment therapy models of depression: A longitudinal study survey. *Spanish Journal of Psychology, 18*, e39, 1-9. doi:10.1017/sjp.2015.31

Ruiz, F. J., Suárez-Falcón, J. C., Barón-Rincón, D., Barrera, A., Martínez, A., & Peña, A. (2016). Factor structure and psychometric properties of the Dysfunctional Attitude

Scale Revised in Colombian undergraduates. *Revista Latinoamericana de Psicología*, *48*, 81-87.

Ruiz, F. J., Suárez-Falcón, J. C., Odriozola-González, P., Barbero-Rubio, A., López-López, J. C., Eisenbeck, N., Budziszewska, L., & Gil, E. (2015). Factor structure and psychometric properties of the Spanish version of the Dysfunctional Attitude Scale Revised. *Behavioral Psychology*, *23*, 287-303.

Ruiz, F. J., Suárez-Falcón, J. C., & Riaño-Hernández, D. (2016, in press). Psychometric properties of the Mindful Attention Awareness Scale in Colombian undergraduates. *Suma Psicológica*.

Soler, J., Tejedor, R., Feliu-Soler, A., Pascual, J. C., Cebolla, A., Soriano, J., Alvarez, E., & Pérez, V. (2012). Propiedades psicométricas de la versión española de la escala Mindful Attention Awareness Scale (MAAS) [Psychometric propoerties of the Spanish versión of the Mindful Attention Awareness Scale (MAAS)]. *Actas Españolas de Psiquiatría*, *40*, 18-25.

Spinhoven, P., Drost, J., de Rooij, M., van Hermert, A. M., Pennix, B. W. (2014). A longitudinal study of experiential avoidance in emotional disorders. *Behavior Therapy*, *45*, 840-850. doi:10.1016/j.beth.2014.07.001

Törneke, N., Luciano, C., Barnes-Holmes, Y., & Bond, F. W. (2016). Relational Frame Theory and three core strategies in understanding and treating human suffering. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 254-272). Wiley-Blackwell.

Wolgast, M. (2014). What does the Acceptance and Action Questionnaire (AAQ-II) really measure? *Behavior Therapy*, *45*, 831-839. doi:10.1016/j.beth.2014.07.002

Table 1

Goodness-of-Fit Indexes of the One-Factor Model and One-Factor Model with Items 1 and 4 with Error Terms Allowed to Correlate (Overall Sample: N = 1759)

Goodness-of-fit indicators	One-factor model	One-factor model with error terms allowed to correlate
RMSEA [90% CI]	.097 [.087, .11]	.069 [.059, .081]
CFI	.97	.98
NNFI	.95	.97
ECVI [90% CI]	.16 [.13, .19]	.087 [.069, .11]
χ^2 (*df*)	247.23 (14)	123.18 (13)

Table 2

Measurement Invariance Across Samples and Gender

Model	χ^2	df	$\Delta\chi^2$	Δdf	RMSEA	ΔRMSEA	CFI	ΔCFI	NNFI	ΔNNFI
Measurement invariance across samples										
MG Baseline model	161.59	39			.073		.98		.97	
Constrained model	184.88	53	23.29	14	.065	.008	.98	.00	.98	.01
Measurement invariance across gender										
MG Baseline model	134.04	26			.069		.99		.98	
Constrained model	140.23	33	6.19	7	.061	.008	.99	.00	.98	.00

Table 3

Cronbach's Alphas and Descriptive Data across Samples

	Sample 1: Undergraduates (*N* = 762)	Sample 2: Online (*N* = 724)	Sample 3: Clinical (*N* = 277)	Overall Sample (*N* =1759)
Alpha	.88	.91	.90	.91
95% CI	[.87, .89]	[.90, .92]	[.88, .92]	[.90, .92]
Mean score	19.99	22.86	29.67	22.69
(*SD*)	(8.37)	(9.51)	(10.27)	(9.74)

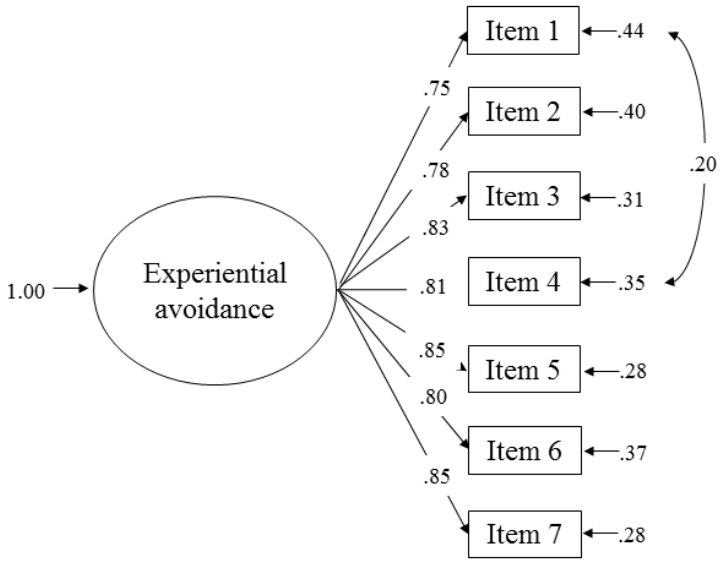
Table 4

Pearson Correlations between the AAQ-II Scores and Other Relevant Self-report Measures

Measure	S	*N*	*r* with AAQ-II
DASS-21 – Depression	1	762	.58*
	2	724	.73*
	3	277	.70*
DASS-21 – Anxiety	1	762	.49*
	2	724	.61*
	3	277	.65*
DASS-21 – Stress	1	762	.50*
	2	724	.62*
	3	277	.68*
SWLS (life satisfaction)	1	762	-.42*
	2	724	-.57*
GHQ-12 (mental health)	1	762	.55*
	3	277	.60*
MAAS (mindfulness)	1	762	-.31*
DAS-R (dysfunctional schemas)	1	762	.42*

Note. AAQ-II: Acceptance and Action Questionnaire – II; DAS-R: Dysfunctional Attitude Scale – Revised; DASS: Depression, Anxiety, and Stress Scales – 21; GHQ-12: General Health Questionnaire – 12; MAAS: Mindful Attention Awareness Scale; SWLS: Satisfaction with Life Scale. *$p < .001$

Figure 1. Standardized solution for the one-factor model with correlated measurement errors between items 1 and 4 in the overall sample.



Appendix 1. Items of the Acceptance and Action Questionnaire – II.

Item number and description
1. Mis experiencias y recuerdos dolorosos hacen que me sea difícil vivir la vida que querría.
2. Tengo miedo de mis sentimientos.
3. Me preocupa no ser capaz de controlar mis preocupaciones y sentimientos.
4. Mis recuerdos dolorosos me impiden llevar una vida plena.
5. Mis emociones interfieren en cómo me gustaría que fuera mi vida.
6. Parece que la mayoría de la gente lleva su vida mejor que yo.
7. Mis preocupaciones interfieren en el camino de lo que quiero conseguir.