Validity Evidence of the Spanish Version of the Automatic Thoughts Questionnaire–8 in Colombia

Francisco J. Ruiz[1], Juan C. Suárez-Falcón[2] and Diana Riaño-Hernández[1]

[1] *Fundación Universitaria Konrad Lorenz (Colombia)*
[2] *Universidad Nacional de Educación a Distancia (Spain)*

Abstract. The Automatic Thoughts Questionnaire (ATQ) is a widely used, 30-item, 5-point Likert-type scale that measures the frequency of negative automatic thoughts as experienced by individuals suffering from depression. However, there is some controversy about the factor structure of the ATQ, and its application can be too time-consuming for survey research. Accordingly, an abbreviated, 8-item version of the ATQ has been proposed. The aim of this study was to analyze the validity evidence of the Spanish version of the ATQ-8 in Colombia. The ATQ-8 was administered to a total of 1587 participants, including a sample of undergraduates, one of general population, and a clinical sample. The internal consistency across the different samples was good ($\alpha = .89$). The one-factor model found in the original scale showed a good fit to the data (RMSEA = .083, 90% CI [.074, .092]; CFI = .96; NNFI = .95). The clinical sample's mean score on the ATQ-8 was significantly higher than the scores of the nonclinical samples. The ATQ-8 was sensitive to the effects of a 1-session acceptance and commitment therapy focused on disrupting negative repetitive thinking. ATQ-8 scores were significantly related to dysfunctional schemas, emotional symptoms, mindfulness, experiential avoidance, satisfaction with life, and dysfunctional attitudes. In conclusion, the Spanish version of the ATQ-8 showed good psychometric properties in Colombia.

Received 17 February 2016; Revised 19 January 2017; Accepted 20 January 2017

Keywords: Automatic Thoughts Questionnaire–8, confirmatory factor analysis, cognitive therapy, depression.

Psychological models of psychopathology have long employed a great amount of effort to analyze unipolar depression, as it is the most frequent psychiatric complaint and the first cause of disability worldwide (e.g., Murray & López, 1996). One of the most popular models of depression is the cognitive model advocated by Beck, Rush, Shaw, and Emery (1979), which suggests that the core characteristic of depression is a pattern of negative thinking about the self, the world, and the future, known as the cognitive triad. According to this model, negative automatic thoughts are the most proximal cause of depressive symptoms and they are produced when stressful events activate depressogenic schemas. These schemas are considered the distal cause of depression, as they are thought to be shaped by early negative life experiences, to be relatively stable, and to remain latent until the individual encounters negative events that activate them. When this occurs, depressogenic schemas skew the information processing system, leading to the production of negative automatic thoughts, which are more unstable and state-dependent.

Given the relevance of negative automatic thoughts, several measures of the frequency with which individuals experience them have been proposed (Nezu, Ronan, Meadows, & Clure, 2000). The most widely used instrument of negative thoughts is the Automatic Thoughts Questionnaire (ATQ; Hollon & Kendall, 1980). The ATQ is a 30-item, 5-point Likert-type scale (5 = *all the time*; 1 = *not at all*) that measures the frequency of negative automatic thoughts experienced during the past week. The ATQ was developed by asking undergraduates to recall dysphoric experiences and report associated thoughts. One hundred items were generated that were administered to an undergraduate sample and then reduced to the final 30 items. Hollon and Kendall (1980) found that the ATQ had excellent internal consistency, a four-factor structure, and it discriminated between depressive and nondepressive participants.

Numerous studies have confirmed that the ATQ possesses excellent internal consistency, temporal consistency, and convergent and discriminant validity (e.g., Chioqueta & Stiles, 2004; Hollon & Kendall, 1980; Hollon, Kendall, & Lumry, 1986; Kazdin, 1990). With regard to the factor structure of the ATQ, exploratory factor analyses have usually yielded more than one factor (e.g., Bryant & Baxter, 1997; Cano-García & Rodríguez-Franco, 2002; Chioqueta & Stiles, 2006; Deardorf, Hopkins, & Finch Jr, 1984; Ghassemzadeh, Mojtabai, Karamghadiri, & Ebrahimkhani, 2006; Joseph, 1994; Kazdin, 1990; Oei & Mukhtar, 2008; Sahin & Sahin, 1992; see reviews in Netemeyer et al., 2002; Zettle, Webster, Gird, Wagener, & Burdsal, 2013), but most

Correspondence concerning this article should be addressed to Dr. Francisco J. Ruiz. Facultad de Psicología. Director Doctorado en Psicología. Fundación Universitaria Konrad Lorenz. Bogotá (Colombia).

E-mail: Franciscoj.ruizj@konradlorenz.edu.co

studies showed different factorial solutions to the four factors identified by Hollon and Kendall (1980). As noted by Netemeyer et al. (2002), all studies found that the first factor accounted for a large proportion of the variance, and the remaining factors for only small percentages. Thus, data suggest that a single factor might underlie the 30 items of the ATQ. Furthermore, most studies only use the overall score of the ATQ obtained by summing the scores of the 30 items, which treats the scale as represented by only one factor.

In spite of its excellent psychometric properties, the length of the ATQ can be an obstacle for administering it in some kinds of studies where participants are asked to complete several measures. In other words, measurement efficiency is an issue in this type of research because nonresponse and respondent fatigue need to be considered (Cortina, 1993; Netemeyer et al., 2002). In this sense, longer measures usually have better psychometric properties but might also contain redundant items. Indeed, in narrowly defined constructs, fewer items can result in reliable and valid measures (Clark & Watson, 1995).

Taking into account the results obtained in factor-analytic research and that measurement efficiency can be an issue for using the ATQ, Netemeyer et al. (2002) offered 15– and 8-item shortened versions of the ATQ. The authors used two samples of 439 and 419 nonclinical participants, respectively, to derive the reduced one-factor versions. Subsequently, two additional cross-validation samples showed that the reduced versions had good psychometric properties, with the ATQ-8 showing coefficient alphas between .85 and .92 across the samples. Additionally, a one-factor structure was obtained in the exploratory and confirmatory factor analyses conducted on the ATQ-8, while less clear results were obtained for the ATQ-15. Correlations between the ATQ, ATQ-15, and ATQ-8 were almost perfect, whereas correlations with other related instruments were basically the same. In conclusion, the ATQ-8 seems to be an efficient measure of the frequency of negative automatic thoughts.

To our knowledge, there is no measure of the frequency of negative automatic thoughts validated in Colombian samples, which makes it difficult to conduct studies in this country related to the cognitive model of depression. 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). Accordingly, the aim of this study was to analyze the factor structure and psychometric properties of the ATQ-8 in Colombia. We used the 8 items of the ATQ-8 from the Spanish version of the ATQ (full scale) by Cano-García and Rodríguez-Franco (2002). These authors showed that the ATQ had excellent psychometric properties in patients suffering from chronic pain in Spain. A four-factor structure was found with the first one accounting for 53.4% of the variance, which indicated that a single factor might underlie the 30 items of the ATQ.

A secondary aim of this study was to explore metric and scalar invariances across clinical and nonclinical samples. This is important because violations of measurement invariance might impede a meaningful comparison between the scores of clinical and nonclinical samples. To our knowledge, no study has explored the measurement invariance of the ATQ-8 between clinical and nonclinical samples.

After confirming that the ATQ-8 items were understandable by Colombian participants, the test was administered in conjunction with other related measures to three samples with a total of 1587 participants: a sample of 762 undergraduates, a sample of 724 Colombian people recruited through internet, and a clinical sample of 101 participants. We expected the ATQ-8 to show similar psychometric characteristics in Colombia to the original scale and to be sensitive to the effect of the 1-session acceptance and commitment therapy intervention.

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 using psychotropic medication.

Sample 3

It consisted of 101 patients (52 of them were women) with an age range of 18 to 67 years (*M* = 32.22, *SD* = 12.09), suffering from emotional (67.3%) or sexual disorders (32.7%). All participants were being evaluated in some private psychological consultation centers in Bogotá. Only 5% of the participants reported that they were consuming some psychotropic medication.

Sample 4

This sample consisted of 11 participants (2 men, age range: 18 to 32 years, *M* = 22.18, *SD* = 4.40,) who participated in a multiple baseline design study that analyzed the effect of a 1-session acceptance and commitment therapy (ACT; Hayes, Strosahl, & Wilson, 1999) intervention to disrupt repetitive negative thinking. The relative educational level of the participants was as follows: 1 with mid-level studies, 6 undergraduate students, and 4 college graduates.

Instruments

Automatic Thoughts Questionnaire – 8 (ATQ-8; Netemeyer et al., 2002; Spanish version by Cano-García & Rodríguez-Franco, 2002). The ATQ-8 is the result of reducing the ATQ to only 8 items while maintaining the same instructions and Likert-type scale (Netemeyer et al., 2002). Examples of items are "I'm no good," "Nothing feels good anymore," "What's wrong with me?" and "I'm worthless." To administer the ATQ-8 in this study, we used only the 8 items of the Spanish version by Cano-García and Rodríguez-Franco (2002).

Dysfunctional Attitude Scale - Revised (DAS-R; de Graaf, Roelofs, & Huibers, 2009; Spanish version by Ruiz et al., 2015). 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, Suárez-Falcón, Barón-Rincón et al., 2016). In this study, the DAS-R showed a Cronbach's alpha of .91. According to the cognitive model of depression, medium to strong correlations between the ATQ-8 and the DAS-R were expected.

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. The DASS-21 has shown good psychometric properties in Colombia (Ruiz, García-Martín, Suárez-Falcón, & Odriozola-González, 2017) and hierarchical factor structure with three first-order factors (Depression, Anxiety, and Stress) and a second-order factor (Emotional Symptoms). The DASS-21 was administered because the frequency of negative thoughts strongly correlated with emotional symptoms in previous studies, not only with depression. Therefore, strong correlations between the ATQ-8 and the DASS-21 subscales were expected.

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). The GHQ-12 was administered because the frequency of negative thoughts is positively correlated with measures of psychological distress. Medium to strong positive correlations were, therefore, expected between the ATQ-8 and the GHQ-12.

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. Examples 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). As previous research has shown that the frequency of negative thoughts negatively correlates with life satisfaction, medium to strong negative correlations were expected between the ATQ-8 and the SWLS.

Acceptance and Action Questionnaire – II (AAQ-II; Bond et al., 2011; Spanish translation by Ruiz, Langer, Luciano, Cangas, & Beltrán, 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 unwillingness to experience unwanted emotions and thoughts and the inability to be in the present moment and behave

according to value-directed actions when experiencing unwanted psychological events. The Spanish version by Ruiz, Suárez-Falcón, Cárdenas-Sierra, et al. (2016) showed good psychometric properties (mean alpha of .90) and a one-factor structure in Colombian samples. The AAQ-II was administered because previous research has shown strong positive correlations between the AAQ-II and ATQ scores (e.g., Ruiz & Odriozola-González, 2016).

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 the tasks. 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, 2016). The MAAS was administered in view of the established negative correlation between present moment awareness and frequency of negative thoughts. Medium to strong negative correlations were expected between the ATQ-8 and the MAAS.

Procedure

A small pilot study was conducted to explore whether the items of the ATQ-8 were understandable by Colombian participants. Specifically, ten Colombian undergraduates were asked to rate item clarity. The undergraduates rated the items as highly understandable, so we did not change any of them.

In Sample 1, the administration of the questionnaire package was 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 (e.g., institutional webpages, Facebook profiles, etc). 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, participants provided informed consent and were given a questionnaire packet. Participants in Sample 1 responded to the ATQ-8, AAQ-II, DASS-21, GHQ-12, MAAS, DAS-R, and SWLS. Participants in Sample 2 responded to the ATQ-8, AAQ-II, SWLS, and DASS-21. Participants in Sample 3 responded to the GHQ-12 in addition to the questionnaires administered in Sample 2. The number of questionnaires applied varied across samples according to the available time for conducting the assessment. Upon completion of the study, participants were debriefed about the aims of the study and

thanked for their participation. No incentives were provided for participation.

Participants in Sample 4 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. Additional information of the inclusion and exclusion criteria can be seen in Ruiz, Riaño-Hernández, Suárez-Falcón, and Luciano (2016). Participants completed a baseline period ranging between 2 to 10 weeks during which the ATQ-8 was administered every 2 weeks in addition to other measures. Subsequently, participants received a one-session ACT intervention specifically oriented to disrupt repetitive negative thinking in the form of worry and rumination. Participants then completed follow-up measures for 6 weeks (one measurement each 2 weeks). The ACT protocol consisted of an approximately 75-minute, individual session. The main objectives of the protocol were: (a) to identify triggers for negative repetitive thinking and experiential avoidance strategies related to them, (b) to promote creative hopelessness regarding the counterproductive effect of engaging in repetitive negative thinking and the other experiential avoidance strategies, (c) to promote values clarification and the commitment to valued actions, and (d) to introduce defusion training. It was expected that participants would show significant changes on the ATQ-8 to the extent that disrupting worry/rumination would decrease the frequency of negative thoughts.

Data analysis

Prior to conducting factor analyses, data from Samples 1 to 3 (Sample 4 was too small to conduct factor analyses and was only used to analyze the treatment sensitivity of the ATQ-8) were examined searching for missing values, which were imputed using the matching response pattern of LISREL© (version 8.71, Jöreskog & Sörbom, 1999), which was the software used to conduct the confirmatory factor analyses (CFA). Fifty-nine values were missing (from 7 for Item 6 to 11 for Items 2 and 8), which represents only 0.60% of the data. In this imputation method, the value to be substituted for the missing value of a single case is obtained from another case (or cases) having a similar response pattern over the remaining items of the ATQ-8.

Because the ATQ-8 uses a Likert-type scale measured on an ordinal scale, a weighted least squares (WLS) estimation method using polychoric correlations was used in conducting the CFA. The WLS method is recommended in large samples with fewer than 20 items, as in the current study (Jöreskog & Sörbom, 1996). We computed the chi-square test and the following

goodness of fit indexes for the one-factor model: (a) the root mean square error of approximation (RMSEA); (b) the comparative fit index (CFI); and (c) the non-normed fit index (NNFI). According to Kelloway (1998) and Hu and Bentler (1999), RMSEA values below .10 represent an acceptable fit, and values below .05 represent a very good fit to the data. With respect to the CFI and NNFI, values above .90 indicate acceptable-fitting models, and above .95 represent a good fit to the data.

Additional CFA were performed to test for metric and scalar invariances across samples and gender following Jöreskog (2005), and Millsap and Yun-Tein (2004). In other words, we analyzed whether the item factor loadings and items intercepts are invariant across the three samples and between men and women. In so doing, the relative fits of three increasingly restrictive models were compared: the multiple-group baseline model, the metric invariance model, and the scalar invariance model. The multiple-group baseline model allowed the eight unstandardized factor loadings to vary across the three samples and across men and women. The metric invariance model, which was nested within the multiple-group baseline model, placed equality constraints (i.e., invariance) on those loadings across groups. Lastly, the scalar invariance model, which was nested within the metric invariances model, is tested by constraining the factor loadings and the items intercepts to be the same across groups. 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). For the model comparison, the RMSEA, CFI, and NNFI indices between nested models were compared. The more constrained model was selected (i.e., second model versus first model, and third model versus second model) if the following criteria suggested by Cheung and Rensvold (2002) and Chen (2007) were met: (a) the difference in RMSEA (ΔRMSEA) was lower than .01; (b) the differences in CFI (ΔCFI) and NNFI (ΔNNFI) were equal to or greater than –.01.

The remaining statistical analyses were performed on SPSS 19©. Coefficient alphas were computed providing 95% confidence intervals (CI) to explore the internal consistency of the ATQ-8 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 ATQ-8 scores were explored by computing independent sample *t*-test. To examine criterion validity, scores on the ATQ-8 were compared between participants in Sample 1 and 2 (nonclinical participants) and participants in Sample 3 (clinical participants).

Pearson correlations between the ATQ-8 and other scales were calculated to assess convergent construct validity. Lastly, to explore whether the ATQ-8 scores were sensitive to the effects of a 1-session ACT intervention, a paired sample *t*-test was conducted between the last ATQ-8 score of participants' baseline and the 6-week follow-up. Cohen's *d* for within-participant studies was also computed.

Results

Descriptive data and psychometric quality of the items

Table 1 shows the original items of the ATQ-8, their translation into Spanish, the descriptive data and corrected item-total correlations for each sample. All items showed good discrimination, with corrected item-total correlations ranging from .56 to .64 in Sample 1, from .63 to .78 in Sample 2, and from .46 to .73 in Sample 3.

Table 2 shows that Coefficient alpha of the ATQ-8 ranged from .85 (Sample 1) to .91 (Sample 2), with an overall alpha of .89. There were no differences on ATQ-8 scores between men and women across Samples 1 to 3.

Validity evidence based on internal structure

Dimensionality

The overall fit of the one-factor model was adequate, and scores on the goodness-of-fit indexes were good: $\chi^2(20) = 237.62$, $p < .01$; RMSEA = .083, 90% CI [.074, .092], CFI = .96, NNFI = .95. Figure 1 depicts the results of the standardized solution of the one-factor model.

Measurement invariance

Table 3 shows the results of the metric and scalar invariance analyses. Parameter invariance was supported at both the metric and scalar levels across samples and gender because changes in RMSEA, CFI, and NNFI were lower than .01.

Validity evidence based on relationships with other variables

The ATQ-8 showed correlations with all the other assessed constructs in theoretically coherent ways (see Table 4). Specifically, the ATQ-8 showed positive correlations with psychological inflexibility (AAQ-II), depression, anxiety, and stress symptoms (DASS-21), emotional distress (GHQ-12), and dysfunctional attitudes (DAS-R); and negative correlations with mindful awareness (MAAS), and satisfaction with life (SWLS).

Means and standard deviations of the ATQ-8 score for each sample can be seen in Table 2. Participants' mean

Table 1. *Item Description, Corrected Item-Total Correlations, Descriptive Data*

Item number and description	Corrected item-total correlation			M (SD)		
	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
1. No soy bueno [I'm no good].	.59	.68	.70	1.86 (1.02)	2.14 (1.09)	2.14 (1.11)
2. ¡Soy tan decepcionante hasta para mí mismo! [I'm so disappointed in myself].	.64	.78	.73	1.48 (.91)	1.78 (1.15)	1.91 (1.23)
3. ¿Qué es lo que funciona mal en mí? [What's wrong with me?].	.61	.76	.71	2.28 (1.18)	2.37 (1.21)	2.97 (1.24)
4. Soy un inútil, no valgo para nada [I'm worthless].	.61	.75	.70	1.38 (.80)	1.65 (1.09)	1.80 (1.09)
5. Me siento tan impotente, tan desamparado [I feel so helpless].	.61	.74	.69	1.74 (1.07)	2.05 (1.23)	2.43 (1.34)
6. Algo tiene que cambiar [Something has to change].	.58	.64	.46	2.96 (1.32)	3.22 (1.23)	3.61 (1.30)
7. Mi futuro es un desierto [My future is bleak].	.58	.68	.60	1.78 (1.11)	2.07 (1.29)	2.36 (1.31)
8. No consigo terminar nada de lo que empiezo [I can't finish anything].	.56	.63	.61	1.81 (1.10)	2.13 (1.24)	2.53 (1.33)

Table 2. *Coefficient alphas and Descriptive Data across Samples*

	Sample 1: Undergraduates (N = 762)	Sample 2: General population online (N = 724)	Sample 3: Clinical (N = 101)	Overall Sample (N = 1587)
Alpha 95% CI	.85 [.84, .87]	.91 [.90, .92]	.88 [.84, .91]	.89 [.88 .90]
Mean score (SD)	15.28 (6.01)	17.41 (7.46)	19.75 (7.35)	16.53 (6.92)
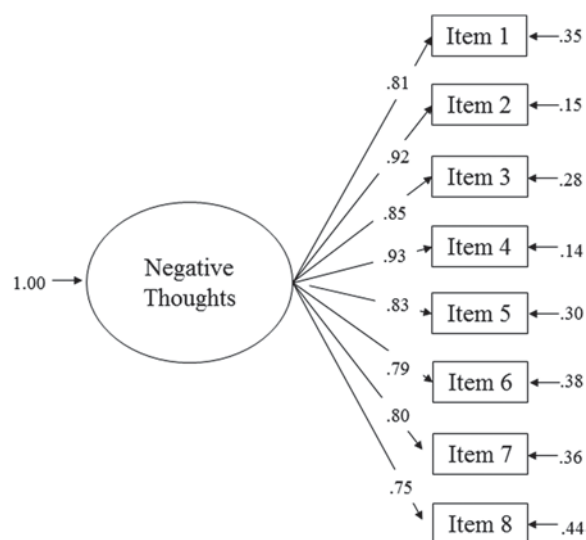
Figure 1. Results of the confirmatory factor analysis conducted with the overall sample to analyze the fit of a one-factor model.

score in the clinical sample (Sample 3) was higher than those of participants in Sample 1 ($t = -5.75$, $p < .001$) and Sample 2 ($t = -2.95$, $p = .004$).

Sensitivity to treatment

In Sample 4, participants' mean score during the baseline assessment was 21.45 (SD = 5.58), whereas the mean score at the 6-week follow-up was 12.45 (SD = 4.78). The difference was statistically significant and with a very large effect size ($t = 6.05$, $p < .001$, $d = 1.61$).

Discussion

While some Spanish versions of the ATQ already existed, to our knowledge, no study has explored the psychometric properties and factor structure of the ATQ-8 in Spanish, which was the main aim of this study. The data obtained showed that this Spanish version of the ATQ-8 has good psychometric properties in Colombia. Specifically, the ATQ-8 showed construct validity to the extent that factor analysis yielded a one-factor solution, as in Netemeyer et al. (2002). The internal consistency of the ATQ-8 was good, with an overall alpha of .89, and it showed criterion validity to the extent that its scores discriminated between clinical and nonclinical samples. The instrument also showed convergent validity in view of the positive correlations found with dysfunctional schemas, emotional symptoms, and

Table 3. *Metric and Scalar Invariance across Samples and Gender*

Model	RMSEA	ΔRMSEA	CFI	ΔCFI	NNFI	ΔNNFI
Measurement invariance across samples						
MG Baseline model	.107		.972		.960	
Metric invariance	.095	.012	.972	.000	.969	.009
Scalar invariance	.101	−.006	.963	−.009	.965	−.004
Measurement invariance across gender						
MG Baseline model	.105		.978		.969	
Metric invariance	.103	.002	.975	−.003	.970	.001
Scalar invariance	.101	.002	.973	−.002	.972	.002

Table 4. *Pearson Correlations between the ATQ-8 Scores and Other Relevant Self-report Measures*

Measure	S	N	r with ATQ-8
DAS-R	1	762	.43*
DASS-21 – Depression	1	762	.59*
	2	724	.80*
	3	101	.65*
DASS-21 – Anxiety	1	762	.46*
	2	724	.65*
	3	101	.66*
DASS-21 – Stress	1	762	.47*
	2	724	.64*
	3	101	.57*
GHQ-12	1	762	.60*
	3	101	.50*
AAQ-II	1	762	.56*
	2	724	.71*
	3	101	.59*
MAAS	1	762	−.29*
SWLS	1	762	−.44*
	2	724	−.60*
	3	101	−.54*

Note: AAQ-II: Acceptance and Action Questionnaire – II; ATQ-8: Automatic Thoughts Questionnaire – 8; 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. *$p < .001$.

psychological inflexibility, and the negative correlations with mindfulness and life satisfaction. Lastly, the ATQ-8 was shown to be sensitive to the effect of a one-session ACT intervention focused on disrupting repetitive negative thinking.

This study also provides evidence of the measurement invariance (metric and scalar) of the ATQ-8 in Colombia across clinical and nonclinical samples. This indicates that the ATQ-8 scores have a similar origin (metric invariance) and that they can be compared since equality of factor loadings and intercepts has been

found (scalar invariance). These findings are especially relevant because no study to date had explored the measurement invariance of the ATQ-8. Further studies might try to replicate this finding in other countries including additional Spanish-speaking countries.

Some limitations of this study are worth mentioning. Firstly, the functioning of the ATQ-8 was tested only in one relatively small clinical sample so that further research is necessary in larger clinical samples to confirm the results obtained in this study. Secondly, no systematic information was obtained concerning the specific diagnosis in clinical participants because they were categorized in broad categories such as emotional and sexual disorders. Thirdly, some of the instruments used to explore the convergent and divergent validity of the ATQ-8 lacked formal validation in Colombian samples (e.g., SWLS); however, their internal consistencies in the current study were adequate and similar to the ones obtained in the original language validation studies.

In conclusion, according to the reliability and validity data provided in this study, this Spanish version of the ATQ-8 can be used to measure the frequency of negative thoughts as experienced by individuals suffering from depression in Colombia.

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. https://doi.org/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.

Beck A. T., Rush A. J., Shaw B. F., & Emery G. (1979). *Cognitive therapy of depression*. New York, NY: Guilford Press.

Bond F. W., Hayes S. C., Baer R. A., Carpenter K. M., Guenole N., Orcutt H. K., … Zettle R. D. (2011). Preliminary psychometric properties of the Acceptance

and Action Questionnaire – II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, 42, 676–688. https://doi.org/10.1016/j.beth.2011.03.007

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. https://doi.org/10.1037/0022-3514.84.4.822

Bryant F. B., & Baxter W. J. (1997). The structure of positive and negative automatic cognition. *Cognition & Emotion*, 11, 225–258. https://doi.org/10.1080/026999397379908

Cano-García F. J., & Rodríguez-Franco L. (2002). Evaluación del lenguaje interno ansiógeno y depresógeno en la experiencia de dolor crónico [Assessment of anxious and depressive self-talk in chronic pain experience]. *Apuntes de Psicología*, 20, 329–346.

Chen F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 464–504. https://doi.org/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. https://doi.org/10.1207/S15328007SEM0902_5

Chioqueta A. P., & Stiles T. C. (2004). Norwegian version of the Automatic Thoughts Questionnaire: A reliability and validity study. *Cognitive Behaviour Therapy*, 33, 79–82. https://doi.org/10.1080/16506070310016031

Chioqueta A. P., & Stiles T. C. (2006). Factor structure of the Dysfunctional Attitude Scale (Form A) and the Automatic Thoughts Questionnaire: An exploratory study. *Psychological Reports*, 99, 239–247. https://doi.org/10.2466/pr0.99.1.239-247

Clark L. A., Watson D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Cortina J. (1993). What is coefficient alpha: An examination of theory and applications. *Journal of Applied Psychology*, 78, 98–104. https://doi.org/10.1037/0021-9010.78.1.98

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.

Deardorf P. A., Hopkins L. R., & Finch A. J. Jr. (1984). Automatic Thoughts Questionnaire: A reliability and validity study. *Psychological Reports*, 55, 708–710. https://doi.org/10.2466/pr0.1984.55.3.708

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. https://doi.org/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. https://doi.org/10.1207/s15327752jpa4901_13

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. https://doi.org/10.1016/S0120-0534(14)70015-9

Ghassemzadeh H., Mojtabai R., Karamghadiri N., & Ebrahimkhani N. (2006). Psychometric properties of a Persian-language version of the Automatic Thoughts Questionnaire: ATQ-Persian. *International Journal of Social Psychiatry*, 52, 127–137. https://doi.org/10.1177/0020764006062095

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

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

Hollon S. D., & Kendall P. C. (1980). Cognitive self-statements in depression: Development of an Automatic Thoughts Questionnaire. *Cognitive Therapy and Research*, 4, 383–395. https://doi.org/10.1007/BF01178214

Hollon S. D., Kendall P. C., & Lumry A. (1986). Specificity of depressotypic cognitions in clinical depression. *Journal of Abnormal Psychology*, 95, 52–59. https://doi.org/10.1037/0021-843X.95.1.52

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. https://doi.org/10.1080/10705519909540118

Jöreskog K. G. (2005). *Structural equation modeling with ordinal variables using LISREL*. Lincolnwood, IL: Technical report, Scientific Software International, Inc.

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.

Joseph S. (1994). Subscales of the Automatic Thoughts Questionnaire. *Journal of Genetic Psychology*, 155, 367–368. https://doi.org/10.1080/00221325.1994.9914786

Kazdin A. E. (1990). Evaluation of the Automatic Thoughts Questionnaire: Negative processes and depression among children. *Psychological Assessment*, 2, 73–79. https://doi.org/10.1037/1040-3590.2.1.73

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, NY: Guilford Press.

Millsap R. E., & Yun-Tein J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, 39, 479–515. https://doi.org/10.1207/S15327906MBR3903_4

Murray C. J. L., & López A. D. (1996). *The global burden of disease and injury series: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge, MA: Harvard University Press.

Netemeyer R. G., Williamson D. A., Burton S., Biswas D., Jindal S., Landreth S., … Primeaux S. (2002). Psychometric properties of shortened versions of the Automatic Thoughts Questionnaire. *Educational and Psychological Measurement*, 62, 111–129. https://doi.org/10.1177/0013164402062001008

Nezu A. M., Ronan G. F., Meadows E. A., & Clure K. S. (2000). *Practitioner's guide to empirically-based measures of depression. Clinical assessment series.* (Vol. 1). New York, NY: Kluwer/Plenum.

Oei T. P. S., & Mukhtar F. (2008). Exploratory and confirmatory factor validation and psychometric properties of the Automatic Thoughts Questionnaire (ATQ-Malay) in Malaysia. *Hong Kong Journal of Psychiatry*, *18*, 92–100.

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., García-Martín M. B., Suárez-Falcón J. C., & Odriozola-González P. (2017). The hierarchical factor structure of the Spanish version of Depression Anxiety and Stress Scale-21. *International Journal of Psychology and Psychological Therapy*, *17*, 93–101.

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. https://doi.org/10.7334/psicothema2011.239

Ruiz F. J., & Odriozola-González P. (2016). The role of psychological inflexibility in Beck's cognitive model of depression. *Anales de Psicología*, *32*, 441–447.

Ruiz F. J., Riaño-Hernández D., Suárez-Falcón J. C., & Luciano C. (2016). Effect of a one-session ACT protocol in disrupting repetitive negative thinking. *International Journal of Psychology and Psychological Therapy*, *16*, 213–233.

Ruiz F. J., Suárez-Falcón J. C., Barón-Rincón D., Barrera-Acevedo A., Martínez-Sánchez 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. https://doi.org/10.1016/j.rlp.2015.10.002

Ruiz F. J., Suárez-Falcón J. C., Cárdenas-Sierra S., Durán Y. A., Guerrero K., & Riaño-Hernández D. (2016). Psychometric properties of the Acceptance and Action Questionnaire – II in Colombia. *The Psychological Record*, *66*, 429–437. https://doi.org/10.1007/s40732-016-0183-2

Ruiz F. J., Suárez-Falcón J. C., Odriozola-González P., Barbero-Rubio A., López-López J. C., Eisenbeck N., … 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). Psychometric properties of the Mindful Attention Awareness Scale in Colombian undergraduates. *Suma Psicológica*, *23*, 18–24. https://doi.org/10.1016/j.sumpsi.2016.02.003

Sahin N. H., & Sahin N. (1992). Reliability and validity of the Turkish version of the Automatic Thoughts Questionnaire. *Journal of Clinical Psychology*, *48*, 334–340.

Soler J., Tejedor R., Feliu-Soler A., Pascual J. C., Cebolla A., Soriano J., … 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.

Zettle R. D., Webster B. K., Gird S. R., Wagener A. L., & Burdsal C. A. (2013). Factor structure of the Automatic Thoughts Questionnaire in a clinical sample. *International Journal of Cognitive Therapy*, *6*, 280–291. https://doi.org/10.1521/ijct.2013.6.3.280