European Journal of

Psychological Assessment

Official Organ of the European Association of Psychological Assessment

hogrefe

Multistudy Report

Experiential Avoidance as a Common Psychological Process in European Cultures

Jean-Louis Monestès,[1] Maria Karekla,[2] Nele Jacobs,[3] Michalis P. Michaelides,[2] Nic Hooper,[4] Marco Kleen,[5] Francisco J. Ruiz,[6] Giovanni Miselli,[7] Giovambattista Presti,[8] Carmen Luciano,[9] Matthieu Villatte,[10] Frank W. Bond,[11] Naoko Kishita,[12] and Steven C. Hayes[13]

[1]LIP Lab, Univ. Grenoble Alpes and Epsylon Lab, Montpellier, France
[2]University of Cyprus, Cyprus
[3]Hasselt University and FARESA Evidence-Based Psychological Centre, Hasselt, Belgium
[4]University of the West of England, Bristol, UK
[5]GGZ Drenthe, The Netherlands
[6]Fundación Universitaria Konrad Lorenz, Bogotá, Colombia
[7]Fondazione Istituto Ospedaliero Sospiro, Cremona, Italy
[8]Kore University, Enna, Italy
[9]University Almería, Almería, Spain
[10]Evidence Based Practice Institute, Seattle, WA, USA
[11]University of London, UK
[12]University of East Anglia, Norwich, UK
[13]University of Nevada, Reno, NV, USA

Abstract: Experiential avoidance, the tendency to rigidly escape or avoid private psychological experiences, represents one of the most prominent transdiagnostic psychological processes with a known role in a wide variety of psychological disorders and practical contexts. Experiential avoidance is argued to be based on a fundamental verbal/cognitive process: an overextension of verbal problem solving into the world within. Although cultures apparently differ in their patterns of emotional expression, to the extent that experiential avoidance is based on a fundamental verbal/cognitive process, measures of this process should be comparable across countries, with similar relationships to health outcomes regardless of the language community. This research tests this view in European countries. The psychometric properties of the Acceptance and Action Questionnaire-II, a measure of experiential avoidance, are compared across six languages and seven European countries, for a total of 2,170 nonclinical participants. Multiple group analysis showed that the instrument can be considered invariant across the language samples. The questionnaire constitutes a unidimensional instrument with similar relationships to psychopathology, and has good and very similar psychometric properties in each assessed country. Experiential avoidance reveals not just as transdiagnostic, but also as a transcultural process independent of a specific language community.

Keywords: experiential avoidance, Acceptance and Action Questionnaire, AAQ-II, cross-cultural validation, multiple group analysis

Experiential avoidance represents one of the psychological processes that have received much interest in recent years (Boulanger, Hayes, & Pistorello, 2010; Hayes, Wilson, Gifford, Follette, & Strosahl, 1996; Luciano & Hayes, 2001). It is defined as the tendency to rigidly escape or avoid private psychological experiences (thoughts, emotions, sensations, memories, urges), even when doing so is futile or interferes with valued actions (Hayes et al., 1996). Research suggests that excessive emotion regulation and high experiential avoidance contribute to the development and maintenance of various forms of psychopathology, through the narrowing of one's behavioral repertoire (Hooper, Stewart, Duffy, Freegard, & McHugh, 2012). Indeed, individuals exhibiting inflexibility utilizing experiential avoidance are found to be more prone to develop a variety of different psychopathological conditions (Boulanger et al., 2010; Chawla & Ostafin, 2007; Hayes, Luoma, Bond, Masuda, & Lillis, 2006; Hayes et al., 2004; Ruiz, 2010).

DOI: 10.1027/1015-5759/a000327

Theoretical Account of Experiential Avoidance

Advocates of experiential avoidance as a central psychopathological process (e.g., Hayes et al., 1996) conceptualize it as a process tied to basic processes of human language and cognition (Hayes, Barnes-Holmes, & Roche, 2001). Experiential avoidance is a broad functional concept that includes many of the more specific and topographically defined processes of emotional and cognitive adjustment, such as thought suppression or emotional distancing.

Experiential avoidance is explained by the Relational Frame Theory (Hayes et al., 2001), which views the core of human language and cognition as the derivation of relationship among events under the control, at least in part, of arbitrary cues. Experimental studies in the field of Relational Frame Theory demonstrate that the functions of stimuli can transform the functions of other stimuli in a relational network based on the derived relation between them. For example, a normal adult trained to relate "serpent" to "snake" as equivalent, and actual snakes to the word snake, will transfer any fear acquired from direct experiences with snakes to the word "serpent" (for a review, see Sidman, 2009). As such, human beings often try to avoid thinking of fearful stimuli as they would avoid these stimuli themselves if confronted by them. For example, an individual may respond to thoughts about the possibility of seeing a snake, as if the snake was really present and make attempts to avoid not just the snakes but any thoughts related to them.

Experiential avoidance is argued to be an extension of verbal problem solving into the world within, resulting in the attempt to regulate unwanted internal stimuli by direct change efforts, or avoidance of reactions and the situations that give rise to them. In essence, experiential avoidance corresponds to attempts to solve the problem of difficult thoughts and feelings the same way we deal with external problems, that is, using verbal problem-solving abilities, even if this strategy proves to be unproductive in managing unwanted private events (Hooper, Saunders, & McHugh, 2010, Wegner, Schneider, Carter, & White, 1987).

Measuring Experiential Avoidance

So far, experiential avoidance was measured by the Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004) and the Acceptance and Action Questionnaire-II (AAQ-II; Bond et al., 2011), two brief self-report measures.

The original AAQ (Hayes et al., 2004) actually had two versions, one with 9 and one with 16 items, answered on a 7-point Likert-type scale (from 1 = *never true* to 7 = *always true*). Although presenting as a useful tool, especially for its potential to predict a wide range of psychological conditions (Hayes et al., 2006), the AAQ presented with several shortcomings. Notably, its internal consistency (Cronbach's α = .70) and test-retest reliability (.64 over 4 months) were weak, falling just below accepted cut-offs (Hayes et al., 2004). In addition, AAQs' internal consistency was found to fall below accepted cut-offs for less well-educated participants and those with English as a second language (Hayes et al., 2004), suggesting that items may include an unnecessary complexity in their formulation, potentially leading to misunderstanding. Finally, the factorial structure of the AAQ proved to be unstable across studies, presenting with one or two factors depending on the evaluation (Bond & Bunce, 2003).

The AAQ-II (Bond et al., 2011) was designed to assess the same construct as the original AAQ and aimed to improve its psychometric properties by withdrawing the negatively formulated items, and simplifying the wording of several items. The AAQ-II consists of seven items and retains the same 7-point Likert-type answering scale (from 1 = *never true* to 7 = *always true*). Its scores range from 7 to 49, with higher scores indicating greater experiential avoidance. In a study examining 2,816 participants from different nonclinical groups, Bond et al. (2011) found good internal consistency (Cronbach's α = .88) – higher than that for the original AAQ – with exploratory and confirmatory factor analyses (CFAs) showing a unifactorial structure. Also, the AAQ-II scores were stable, with good test-retest reliability after 3 months (r = .81) and 1 year (r = .79), and found to be highly correlated with the original AAQ (r = .82, $p < .01$). It showed good predictive and concurrent validity, correlating with various measures of psychopathology, including: the Beck Depression Inventory (Beck, Steer, & Brown, 1996; r range across two samples = .69–.71, $p < .01$), the Symptom Checklist-90-Revised (DeRogatis, 1992; r = .65, $p < .01$), and the Beck Anxiety Inventory (Beck & Steer, 1993, r = .58, $p < .01$). Further, the AAQ-II was able to demonstrate convergent validity when compared to measures of similar constructs, such as that of thought suppression, assessed via the White Bear Suppression Inventory (Wegner & Zanakos, 1994; r range across three samples = .57–.60). Divergent validity was also established, given that no relationship was found between the AAQ-II and dissimilar constructs such as social desirability, assessed via the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960; $r = -.09$).

As experiential avoidance is considered to be a general possible pathogenic process potentially shared across different cultures, there have been a growing number of translations of the AAQ-II in different languages (26 so far to our knowledge). Such translations have produced the same adequate psychometric properties. Overall, the updated AAQ-II presents with good reliability and good predictive, convergent, concurrent, and discriminant validity, both in the original English version as well as in other languages.

Overview of the Present Study

Conceptualizing experiential avoidance as aforementioned means that it is a fundamental verbal/cognitive process built into derivation of relationships permitted by language. If this is correct, experiential avoidance as a process should be relatively independent of cultural or linguistic differences since mastering any language necessarily provides basic verbal problem-solving skills. Whether or not cultures are expressive, or more or less open to discussion of emotions, might alter the baseline level of experiential avoidance or its expression, but not the basic process or its impact. Indeed, previous research has displayed differences of experiential avoidance between various groups. Asian samples, for example, are known to present more experiential avoidance than Caucasian ones (Cook & Hayes, 2010). Differences of experiential avoidance are also observed in parents of children with anxiety disorders (Cheron, Ehrenreich, & Pincus, 2009), with autism (Blackledge & Hayes, 2006), or in dementia family caregivers (Spira et al., 2007). Despite these differences, if experiential avoidance relies on basic language processes, the structure and psychometric properties of its evaluation, and its basic relationships to psychopathology, should be common to any group, and independent of potential differences in experiential avoidance level.

Some limited evidence exists to support the prediction that experiential avoidance is a reliable construct across cultures. For example, Cook and Hayes (2010) found that experiential avoidance is related to health outcome regardless of acculturation, religion, or recency of immigration when comparing Asian American and Caucasian Americans. However, this study used the first version of the Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004) known to present with weak internal consistency. The study also only included English-speaking participants, limiting the ability to examine language community differences.

The present study tested the idea that experiential avoidance is a psychological process common to European countries by examining multiple translations of a measure of experiential avoidance. While comparing different translations of the AAQ-II (Bond et al., 2011) – a more recent, psychometrically improved version of the AAQ – within nonclinical samples and across seven countries and six different language communities, we hypothesized that these translations have comparable psychometric properties regardless of the language community, and despite potential differences in mean scores. Notably, the unifactorial structure predicted theoretically and observed in the original version of the AAQ-II (Bond et al., 2011) should be found with each language adapted version of the AAQ-II studied. In addition, AAQ-II scores in different language versions are expected to correlate with health outcomes and associated constructs such as depression, thought suppression, and mindfulness, as was found in the validation study of the AAQ-II (Bond et al., 2011) and in other studies using the English version of the AAQ-II (see Boulanger et al., 2010, for a description of the relationships between experiential avoidance and other constructs).

Method

Participants

Six different data sets from independent validation studies of the AAQ-II in seven different countries (Belgium, Cyprus, France, Italy, Netherlands, Spain, and United Kingdom) were included in this secondary analysis, for a total of 2,170 nonclinical volunteers who spoke one of the six languages studied (Dutch for both the Belgium and the Netherlands samples, Greek, French, Italian, Spanish, or English). Participants were students recruited through in-class announcements or on-campus flyers, except for Belgium/Netherlands (email). They completed the AAQ-II in classrooms using paper and pencil, or online (Belgium/Netherlands) as part of a larger packet of questionnaires (Spain, UK), or specifically for the validation of the AAQ-II. Participants completed the questionnaires voluntarily without any compensation. In case of retest, participants were asked to place a code of their choice on the questionnaire during the first administration, and to indicate that code during the second administration.

The total sample consisted of 487 (22.44%) males and 1,526 (70.32%) females (157 unknown, 7.24%). The mean age of the total sample was 24.84 years (*SD* = 9.35). Sample size, gender distribution, and mean age for each country are presented in Table 1. Regarding ethnicity, not all countries provided this data. From those who reported this information, 100% of the sample classified themselves as Spanish in the Spanish sample, French in the French sample, Greek-Cypriot in the Greek sample, 88% as Italian in the Italian sample (4.6% stated that they are foreigners and 7.4% did not specify), and the majority (86.8%) from the Dutch sample identified themselves as Belgians, 12.1% as Dutch (Netherlands), 0.2% as Italian, 0.2% as Flamish, and 0.2% as Georgian. Given that all participants were recruited from universities, it was assumed that all completed at least high school level of education.

Table 1. Sample characteristics by country and total sample

	Belgium/Netherlands	UK	France	Cyprus	Italy	Spain	Total sample
N	370	423	207	177	365	628	2,170
Gender: male	96 (25.90%)	57 (13.50%)	27 (13.00%)	35 (19.80%)	82 (22.50%)	190 (30.30%)	487 (22.40%)
Gender unknown	0	83 (19.60%)	14 (6.8%)	4 (2.30%)	19 (5.20%)	37 (5.90%)	157 (6.26%)
Mean age (*SD*)	25.80 (11.23)	25.88 (11.29)	22.95 (6.27)	25.83 (7.78)	23.29 (5.47)	24.74 (9.45)	24.84 (9.35)

Measures

Four measures were used in the present study: the AAQ-II, the Beck Depression Inventory II (BDI-II; Beck et al., 1996), the White Bear Suppression Inventory (WBSI; Wegner & Zanakos, 1994), and the Mindfulness Attention and Awareness Scale (MAAS; Brown & Ryan, 2003). Each scale was used in the corresponding language version.

The AAQ-II was translated into each language by a team of clinicians from the corresponding country, except for Belgium/Netherlands (translation completed by a nonpsychologist native English speaker). Each version was back-translated in English by a native English speaker unaware of the purpose of the scale, except for Spain (back-translation completed by a professional translator). Finally, the original and back-translated versions were compared by a local team to check for accuracy. Minor discrepancies were detected and adjusted[1].

The BDI-II is a 21-item self-report measure designed to assess depressive symptoms in adolescents and adults. Participants are asked to rate each of the symptoms, ranging from 0 (not present) to 3 (severe), in terms of how they have been feeling during the past 2 weeks. Beck et al. (1996) reported a good convergent validity and excellent internal consistency of scores (Cronbach's α = .91; α in the present overall sample = .93).

The WBSI measures the tendency of individuals to suppress unwanted thoughts. Participants are asked to rate each of the 15 statements (e.g., "I wish I could stop thinking of certain things") on a 5-point Likert-type scale, ranging from *strongly disagree* to *strongly agree*. The WBSI scores have been shown to have good predictive and convergent validity (Wegner & Zanakos, 1994) and internal consistency (Cronbach's $\alpha \geq .87$; α in the present overall sample = .92).

The MAAS is a unidimensional measure of mindfulness, defined as "paying attention in a particular way: on purpose, in the present moment, and nonjudgmentally" (Kabat-Zinn, 1994, p. 4). It consists of 15 items (e.g., "I rush through activities without being really attentive to them") focused on a lack of attention, and a lack of awareness of what is occurring in the present. The MAAS scores have been shown to have good convergent validity (Brown & Ryan, 2003) and internal consistency (Cronbach's α = .81; α in the present overall sample = .89).

Procedure

Institutional Review Board (IRB) approval was obtained for each independent study in their respective country, and for subsequent analysis. Each participant gave informed consent after the study was described to them; participants completed the questionnaires anonymously. In case of retest, participants were given a number to note during the next evaluation a month later (France, UK) or two weeks later (Italy, Cyprus).

Statistical Analyses Performed

To examine whether the scores obtained from the AAQ-II are comparable across the subgroups, the measurement model should be invariant. Measurement invariance was carried out under a confirmatory factor analysis framework (CFA) in IBM SPSS AMOS 20 (Arbuckle, 2011), by setting increasingly more stringent criteria for invariant parameters.

After examining the internal consistency of the AAQ-II in each country through Cronbach's alpha coefficients, separate CFAs were run on each language sample to examine the fit of the data on a single experiential avoidance factor, the structure found in the scale validation study (Bond et al., 2011). In that study, a correlation between two error variances was specified: items "My painful experiences and memories make it difficult for me to live a life that I would value" and "My painful memories prevent me from having a fulfilling life" (items 1 and 4 in AAQ-II) were considered very similar. A second correlation between measurement errors was reported by Gloster, Klotsche, Chaker, Hummel, and Hoyer (2011) between items 2 and 3 which involved common wording and meaning: "I am afraid of my feelings" and "I worry about not being able to control my worries and feelings." The chi-square statistic was used as a measure of the overall model fit but since it is sensitive to sample size and may overestimate the lack of model fit (Bollen, 1990), we additionally examined the following goodness-of-fit indices: the Root Mean Square Error of

[1] The final version of the AAQ-II in each language is available upon request from the authors.

Table 2. Total AAQ-II scores and Cronbach's alpha coefficients for each country and total sample

	Total sample	Belgium/Netherlands	UK	France	Cyprus	Italy	Spain
Cronbach's alpha	.86	.89	.87	.88	.89	.84	.85
Mean AAQ-II total score (*SD*)	20.07 (7.71)	18.80 (7.17)	19.00 (7.08)	21.77 (8.96)	19.00 (8.39)	20.11 (7.30)	21.23 (7.76)

Approximation (RMSEA), the Comparative Fit Index (CFI), and the standardized root mean square residual (SRMR). Following suggested criteria by Cheung and Rensvold (2002) and Fan and Sivo (2007), models with CFI > .90 and RMSEA and SRMR < .08 were indicative of acceptable model fit (see also Browne & Cudeck, 1993; MacCallum, Browne, & Sugawara, 1996). If the fit of the model was inadequate, residual covariances and modification indices were inspected to address any theoretically justifiable adjustments.

Configural invariance was the first form of measurement invariance examined to establish a baseline model without any parameter constraints. As a general principle, if a test of invariance is met, then subsequent more strict constraints were imposed. At the second step, fixed factor loadings were added to test for metric invariance. Next, factor variances were set to be invariant in all subgroups (since the model is unifactorial, no factor covariances were present). Finally, as a test of strict invariance, error variances and covariances were set to be invariant across language samples. The nonsignificant difference in chi-square ($\Delta\chi^2$) is considered as evidence of invariance, but since it is very sensitive to sample size, we based our decisions on the difference in CFI (ΔCFI; Cheung & Rensvold, 2002). If ΔCFI < .01, then the model is considered invariant.

In the presence of invariance across language samples, tests for comparing latent group means were carried out. We conducted *Z*-test pairwise comparisons between each group's latent mean and all the remaining groups' means, by fixing the former to zero and allowing the latter to be freely estimated. Since multiple comparisons were made, a Bonferroni-type correction was applied.

The covariance matrix for each subsample was entered in IBM SPSS AMOS 20 (Arbuckle, 2011), and maximum likelihood estimation (MLE) was used to assess the fit with the Multiple Groups Analysis command. Although the 7-point response scale employed in the AAQ-II is ordinal, MLE was found to provide similar results to categorical least square methodology when response scales had more than 5 points (Rhemtulla, Brosseau-Liard, & Savalei, 2012). With regard to the normality of the data, kurtosis is of particular concern in analyses of covariance structures (Byrne, 2010). Examination of univariate kurtosis indices by sample revealed that the largest kurtosis appeared on item 4 in the English and Dutch samples (2.14 and 1.62, respectively), estimates well below significant kurtosis values (Bentler, 2006).

Finally, concurrent validity of scores was examined in the total sample and for all country samples by means of Pearson's correlation coefficients that estimated the association of the AAQ-II and other measures utilized in the study, namely the BDI-II, the WBSI, and the MAAS. Additionally, temporal reliability was estimated by comparison of test and retest AAQ-II scores when available (Pearson's correlation coefficients). All correlational analyses were performed using SPSS version 16.0 for Windows.

Results

Descriptive Statistics

Data screening of AAQ-II responses revealed two missing values, both in the English sample for item 6. Pairwise deletion was used to handle these missing values.[2] The total AAQ-II mean was 20.07 (*SD* = 7.71) when the scores from each country were merged, with scores ranging from 7 to 49. Mean scores on the AAQ-II ranged from 18.80 to 21.77 in the six samples (see Table 2 for details). Internal consistency (α coefficient) was high in each sample, ranging from .84 to .89. While there were significant differences in mean age across the six samples ($F_{(5,\ 1,840)} = 5.28$; $p < .01$), and gender and language distributions were not independent ($\chi^2_{(5)} = 44.27$; $p < .01$), Pearson's correlation coefficients were very close to zero for the merged group: between age and AAQ-II score ($r = -.02$; $p > .01$), and between gender and AAQ-II score ($r = -.06$; $p < .01$).

Factorial Structure of the AAQ-II in Each Language Sample

A CFA model with the seven items-indicators loading on a single latent factor was first fitted on the six subsamples separately; scaling of the latent factor was determined by a fixed loading of 1 on item 7 in all cases. Fit indices were not acceptable. Inspection of standardized residual covariances (Kline, 2010) presented with large residuals for items

[2] When the missing values were replaced using an EM algorithm approach, inter-item correlations, means, and standard deviations for the English sample were identical up to the second decimal point. Item 6's mean differed by < .005, correlations and standard deviation differed < .002 compared to estimates after pairwise deletion.

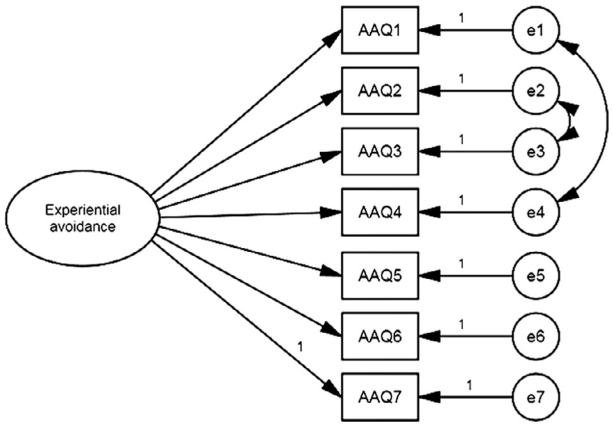

Figure 1. The baseline model for AAQ-II.

2 and 3, and for items 1 and 4. The corresponding error correlations for these pairs of items also had high modification indices with substantial expected parameter changes, if added to the model. An error covariance was added between items 1 and 4, similar to Bond et al. (2011) and Gloster et al.'s (2011) analyses. Fit indices were still not acceptable. Items 2 and 3, "I'm afraid of my feelings" and "I worry about not being able to control my worries and feelings," respectively, are not only consecutive in order, but have also similar content and share the key term *my feelings*. Hence, we specified a second error covariance, similar to Gloster et al. (2011). The model fitted can be seen in Figure 1. Model fit statistics were notably improved in all samples. Fit indices, standardized estimates, factor variances, and error correlations by sample are shown in Table 3.

For all six samples, the chi-square statistic was significant, but CFI was well above and SRMR well below the suggested cut-off points. RMSEA was also acceptable in all cases, except in the French sample. Overall, the results indicated acceptable model fit. Standardized factor loadings were relatively consistent across samples ranging between .543 and .853. Comparing each factor loading to the average loading across samples (see Table 3, rightmost column), none of the standardized estimates had a deviation larger than 0.2 in absolute value. Latent factor variances were significant and about 1 or higher. Both error covariances were positive and significant in all six subsamples: the correlation between items 1 and 4 ranged from .38 to .59, and that between items 2 and 3 from .19 to .40; both specifications result in improved model fit (Schweizer, 2012) and were previously encountered in the literature (Bond et al., 2011; Gloster et al., 2011). The size and significance of the estimated correlation across six linguistically diverse samples may be interpreted as a replicable finding, a systematic source of shared method variance in the current version of the scale that should be included in the model specification (cf. Cole, Ciesla, & Steiger, 2007), and not as a sample-specific, random finding.

Measurement Invariance Across Languages

The Multiple Groups Analysis to examine measurement invariance began by establishing a baseline model (Figure 1) across all six samples simultaneously, without any equality constraints. The configural invariance model had an excellent fit, $\chi^2(72) = 226.077$, $p < .001$, CFI = .976, RMSEA = .031, SRMR = .028 (Table 4; Model 1), indicating that the same latent structure fits the six samples. Constraining the factor loadings to be equal did not result in

Table 3. Fit indices and parameter estimates of the baseline model

	Sample						
	Belgium/Netherlands	UK	France	Cyprus	Italy	Spain	Average
Fit indices							
χ^2 (df = 12)	35.801***	28.593**	49.725***	24.215*	33.383**	54.362***	
CFI	0.982	0.988	0.948	0.980	0.975	0.975	
RMSEA	0.073	0.057	0.124	0.076	0.070	0.075	
SRMR	0.028	0.027	0.050	0.037	0.036	0.036	
Standardized factor loadings							
AAQ1	0.711***	0.557***	0.558***	0.660***	0.564***	0.636***	0.614
AAQ2	0.647***	0.599***	0.724***	0.705***	0.647***	0.549***	0.645
AAQ3	0.708***	0.769***	0.730***	0.720***	0.731***	0.638***	0.716
AAQ4	0.744***	0.632***	0.686***	0.734***	0.656***	0.694***	0.691
AAQ5	0.705***	0.765***	0.853***	0.814***	0.543***	0.624***	0.717
AAQ6	0.719***	0.765***	0.624***	0.680***	0.626***	0.654***	0.678
AAQ7	0.772***	0.730***	0.674***	0.671***	0.697***	0.740***	0.714
Latent factor variance	1.040***	1.022***	1.029***	1.231***	0.990***	1.253***	1.094
Error Correlation AAQ2-AAQ3	0.248***	0.194**	0.290**	0.404***	0.249**	0.351***	0.289
Error Correlation AAQ1-AAQ4	0.565***	0.564***	0.590***	0.382***	0.423***	0.502***	0.504

Note. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 4. Goodness-of-fit statistics for the multigroup invariance tests

Model	Model comparison	χ^2	*df*	$\Delta\chi^2$	Δdf	CFI	ΔCFI
1. No equality constraints imposed; configural invariance	–	226.077	72	–	–	.976	–
2. Invariant factor loadings; metric invariance	1	312.802	102	86.725	30	.968	.008
3. Invariant factor loadings and variances	2	327.854	107	15.052	5	.966	.002
4. Invariant factor loadings, variances and error variances and covariances	3	662.925	152	335.071	45	.922	.044
→4a. Invariant factor loadings, variances and error variances except on items 1–4 for DU and UK and on item 5 for IT and SP	3	405.627	132	77.773	25	.958	.008
→4b. Model 4a and invariant error covariances	3	454.946	142	127.092	35	.952	.014
→4c. Model 4b except error covariance for items 2 and 3 for DU and UK	3	422.700	140	94.846	33	.957	.009

Note. CFI = Comparative Fit Index; Δ = change in statistic.

Table 5. Pairwise comparisons of latent means

	Comparison group									
	English		French		Cypriot		Italian		Spanish	
Group used as baseline (latent mean fixed to 0)	*Z*	*p*	*Z*	*p*	*Z*	*p*	*Z*	*p*	*Z*	*p*
Dutch	−0.109	.160	−0.301	.006	−0.016	.888	−0.234	.004	−0.401	< .001
English			−0.192	.073	0.093	.393	−0.125	.112	−0.293	< .001
French					0.285	.032	0.067	.542	−0.101	.334
Cypriot							−0.219	.049	−0.386	< .001
Italian									−0.167	.025

a worse model fit judging by the small reduction in CFI (ΔCFI = .008; Model 2). Given the evidence for metric invariance, constraining the latent factor variances to be equal as well, did not lead to a large decrease in CFI (ΔCFI = .002; Model 3). Finally, a strict form of invariance suggests that error variances and covariances could be constrained to be equal. This test led to a significant reduction in CFI (ΔCFI = .044; Model 4). Gradual inspection of each error variance at a time revealed some patterns: error variances for items 1–4 for the Dutch and the English samples were smaller than the rest of the samples; for item 5 error variances for the Italian and the Spanish samples were larger. Freeing the above constraints led to a non-substantial decrease in CFI (ΔCFI = .008; Model 4a). Since the baseline model included additional error covariances, we tested the equivalence of these covariances in the six samples (Model 4b), but the difference in CFI was larger than the cut-off (ΔCFI = .014). Freeing the error covariances in the Dutch and the English samples which were larger than in the other four samples, the reduction in CFI was non-substantial (ΔCFI = .009; Model 4c). Equivalence of error variances and covariances "represents an overly restrictive test of the data" (Byrne, 2010, p. 199), but it was found that with few exceptions (especially in the Dutch and English samples) even this form of strict invariance holds to a large extent. Overall, the instrument can be considered invariant across the six language samples, allowing further mean comparisons.

Returning to the metric invariance model (Model 2) for the six samples, that is, assuming invariant factor loadings, we proceeded with the estimation of means and intercepts in AMOS and the additional constraint of equivalent measurement intercepts – and all factor means constrained to zero. This model had an acceptable, but significantly worse fit compared to Model 2: $\chi^2(137) = 748.460$, $p < .001$, CFI = .906 (ΔCFI = .062), RMSEA = .045, SRMR = .036, indicating significant differences in the group mean structures. Subsequently, we fitted a series of models where one group's latent mean was fixed to zero and all other latent means were freely estimated; for this process, IBM SPSS AMOS provides a Z statistic for pairwise latent mean differences. The results appear in Table 5. The ordering of the groups with respect to the mean level of experiential avoidance starting with the group with the highest estimate was: Belgium/Netherlands, Cyprus, UK, Italy, France, and Spain. Since 15 pairwise comparisons were carried out, a Bonferroni-type correction to the significance level was used: .05/15 = .003. The only significant differences were between the Spanish and the three groups with high experiential avoidance: English, Cypriot, and Dutch. The mean differences between Dutch and Italian, and between Dutch and French were marginally nonsignificant.

Concurrent Validity of the AAQ-II Scores Across Languages

We estimated AAQ-II scores concurrent validity by calculating Pearson's correlation coefficients between AAQ-II and other measures for all samples merged. The number

Table 6. Number (percentage) of participants who filled the different measures in each country, and correlations of each measure with the AAQ-II

	BDI-II		WBSI		MAAS	
	N (%)	Correlation with AAQ-II (*r*)	*N* (%)	Correlation with AAQ-II (*r*)	*N* (%)	Correlation with AAQ-II (*r*)
Belgium/Netherlands	14 (3.78%)	.75*	16 (4.32%)	.80*		
UK	354 (83.69%)	.63*	354 (83.69%)	.62*		
France	207 (100%)	.68*	207 (100%)	.70*	207 (100%)	−.47*
Cyprus	164 (92.66%)	.65*	169 (95.48%)	.47*		
Italy	299 (81.92%)	.55*				
Spain	123 (19.59%)	.45*				
Total	1161	.57	746	.57	207	−.47

Note. *$p < .01$.

and percentages of participants who filled the different measures are presented in Table 6 for each country.

For the merged samples, the AAQ-II correlated positively with the BDI-II ($r = .57$, $p < .01$; range across countries: .45–.75), the WBSI ($r = .57$, $p < .01$, range across countries: .47–.80), and negatively with the MAAS ($r = -.47$, $p < .01$), showing that higher experiential avoidance is related to higher depression and thought control, and lower mindfulness, with the same pattern of correlations in each country (see Table 6 for details).

The high correlations between the AAQ-II, WBSI, and BDI-II may suggest that the three scales are measuring similar constructs. Nevertheless, comparing a three-factor and a single-factor model by means of confirmatory factor analysis showed a very bad fit for the single-factor model ($\chi^2 = 4{,}294.193$, $df = 856$, $p < .001$, CFI = 0.72, RMSEA = 0.07, SRMR = 0.08), and a significantly improved fit for the three-factor model ($\chi^2 = 2{,}269.013$, $df = 853$ $p < .001$, CFI = 0.89, RMSEA = 0.05, SRMR = 0.05), lending credence to separate constructs measured by the three scales.

Temporal Stability of the AAQ-II Scores

Temporal stability was estimated by comparing the AAQ-II scores for test and retest for countries that collected these data, for a total of 336 participants (67 English, 126 French, 98 Cypriot, and 45 Italian). A large and significant correlation (Cohen, 1988) between test and retest for this merged group was observed ($r = .81$, $p < .01$) indicating good temporal reliability of scores. The same analysis for each country for which retest was available showed similarly large correlations ($r = .83$, $p < .01$ for UK; $r = .77$, $p < .01$ for France; $r = .82$, $p < .01$ for Cyprus; and $r = .87$, $p < .01$ for Italy).

Discussion

In this study we examined, in a secondary analysis, the psychometric properties of the Acceptance and Action Questionnaire-II, a measure of experiential avoidance, across six languages and seven different European countries, to test for comparable properties across multiple translations, regardless of the language community. The theory underlying the concept of experiential avoidance (e.g., Hayes et al., 1996) emphasizes that its toxic effects are based on an overextension of human problem solving, and thus to a degree is built into human language and cognition. If that is correct, its basic structure and impact should not vary widely across language communities, even in case of different levels of experiential avoidance. In the present study, despite few differences in mean AAQ-II scores between countries, the psychometric properties found across the multiple samples were similar, which lends support to this view.

Overall, the common psychometric properties of the AAQ-II across countries showed that experiential avoidance constitutes a psychological process consistent across European countries. First, the questionnaire exhibited good internal consistency in all language versions and all subsamples. Confirmatory factor analysis carried out across all samples found that, consistent with its theoretical basis and with the original English version (Bond et al., 2011), the AAQ-II is a unidimensional instrument with two residual covariances for item-pairs with parallel wording (Gloster et al., 2011). These modifications are theoretically justified and cross-validate across languages; although a reformulation of the items involved may be considered in future revisions of the scale, currently the evidence suggests that they emerge as necessary model specifications (cf. Cole et al., 2007). Also, multiple group analysis showed that the instrument can be considered invariant across the six language samples. In addition, AAQ-II scores were temporally stable, with a large correlation between test and retest.

Finally, positive correlations with the BDI-II and WBSI and a negative correlation with the MAAS were observed in all countries, which support an association between higher levels of experiential avoidance with greater levels of depression and thought suppression, and lower levels of mindfulness, as theoretically predicted. Although these correlations may suggest that the three scales are measuring similar or overlapping constructs, our analyses show that experiential avoidance constitutes a distinct psychological process. Notably, regarding the rather high correlation of the AAQ-II and the BDI-II, the present results are in line with several studies that showed experiential avoidance and depression to constitute different constructs. Indeed, experiential avoidance has been shown to mediate the effects of various factors on depression (Kashdan, Barrios, Forsyth, & Steger, 2006; Kashdan & Breen, 2007), to evolve independently of depression (Berking, Neacsiu, Comtois, & Linehan, 2009), to predict future variance of depression (Williams, Ciarrochi, & Patrick Deane, 2010), and to reliably distinguish between participants with and without clinical levels of depression (Tull & Gratz, 2008).

The study itself has limitations emerging from the fact that it constitutes a secondary analysis of multiple versions of the AAQ-II. First, the measures used to examine concurrent validity of AAQ-II scores (i.e., BDI, WBSI, MAAS) were not available in all countries, limiting the analysis to the countries for which data were available. Second, test-retest delays in these validation studies were not the same, limiting the reliability of results for temporal validity of the AAQ-II scores. In addition, while AAQ-II scores were not the central interest of this study, and neither gender nor age was correlated with mean scores, the restricted age range and the overrepresentation of females in the different samples may decrease the generalization of the present results to broader populations. Also, some demographic information was lacking for some of the samples, limiting the generalizability of the findings to the representative populations of all countries assessed. A final limitation stands on the high RMSEA value found for the French sample. While this fit index could have been improved by imposing additional modifications on the model for this sample, we chose to keep the same baseline model across samples to examine invariance for the sake of parsimony. Despite these limitations, this study gives support to the overall reliability of experiential avoidance as a pervasive psychological function across European countries.

Specific measures of experiential avoidance have been created by modifying AAQ items to focus on disorder-specific content. These specific measures are available in the areas of psychosis (Shawyer et al., 2007), body image (Sandoz, 2010), smoking dependence (Gifford et al., 2004), weight-related difficulties (Lillis & Hayes, 2008), chronic pain (Vowles, McCracken, McLeod, & Eccleston, 2008), epilepsy (Lundgren, Dahl, & Hayes, 2008), diabetes (Gregg, Callaghan, Hayes, & Glenn-Lawson, 2007), and tinnitus (Westin, Hayes, & Andersson, 2008), and seem to be particularly useful as measures of treatment impact (Hayes et al., 2006). Although the present results are not determinative, they lend credence to the possibility that these measures will also be reliable across cultural and linguistic contexts.

The field has begun to focus on measures of transdiagnostic processes because they hold out hope for a simplification of case analysis, conceptualization, and treatment (Harvey, Watkins, Mansell, & Shafran, 2004). That makes good intellectual and practical sense, but some of this good would be undermined if these transdiagnostic processes were strongly sensitive to cultural and linguistic factors. The common psychometric properties of the AAQ-II across different European countries represent an argument in favor of experiential avoidance as one transdiagnostic factor for which that concern is less likely (see also Cook & Hayes, 2010). As recent data point to different representation of emotion regulation strategies such as suppression and reappraisal across cultures (Matsumoto et al., 2008; Soto, Perez, Kim, Lee, & Minnick, 2011), further studies are needed to explore the cultural differences responsible for different expressions of experiential avoidance, notably by including data from non-European samples. Nonetheless, as the properties of its evaluation and its basic relationships to psychopathology suggest, experiential avoidance stands as a psychological function potentially common to any group.

References

Arbuckle, J. L. (2011). AMOS (Version 20) [Computer software]. Chicago, IL: IBM SPSS.

Beck, A. T., & Steer, R. A. (1993). *Beck Anxiety Inventory Manual*. San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.

Bentler, P. M. (2006). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software.

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

Blackledge, J. T., & Hayes, S. (2006). Using acceptance and commitment training in the support of parents of children diagnosed with autism. *Child and Family Behavior Therapy, 28*, 1–18.

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin, 107*, 256–259.

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

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, M., 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.

Boulanger, J. L., Hayes, S. C., & Pistorello, J. (2010). Experiential avoidance as a functional contextual concept. In A. A. Kring & D. M. Sloan (Eds.), *Emotion regulation and psychopathology: A transdiagnostic approach to etiology and treatment* (pp. 107–136). New York, NY: Guilford Press.

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.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newsbury Park, CA: Sage.

Byrne, B. M. (2010). *Structural Equation Modeling with AMOS* (2nd ed.). New York, NY: Routledge.

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

Cheron, D. M., Ehrenreich, J. T., & Pincus, D. B. (2009). Assessment of parental experiential avoidance in a clinical sample of children with anxiety disorders. *Child Psychiatry and Human Development, 40*, 383–403.

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.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods, 12*, 381–398.

Cook, D., & Hayes, S. C. (2010). Acceptance-based coping and the psychological adjustment of Asian and Caucasian Americans. *International Journal of Behavioral Consultation and Therapy, 6*, 186–197.

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

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

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research, 42*, 509–529.

Gifford, E. V., Kohlenberg, B. S., Hayes, S. C., Antonuccio, D. O., Piasecki, M. M., Rasmussen-Hall, M. L., & Palm, K. M. (2004). Acceptance theory-based treatment for smoking cessation: An initial trial of Acceptance and Commitment Therapy. *Behavior Therapy, 35*, 689–705.

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.

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

Harvey, A., Watkins, E., Mansell, W., & Shafran, R. (2004). *Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment*. Oxford, UK: Oxford University Press.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York, NY: Plenum Press.

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

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

Hooper, N., Saunders, J., & McHugh, L. (2010). The derived generalization of thought suppression. *Learning & Behavior, 38*, 160–168.

Hooper, N., Stewart, I., Duffy, C., Freegard, G., & McHugh, L. (2012). Modelling the direct and indirect effects of thought suppression on behavioral preference. *Journal of Contextual Behavioral Science, 1*, 73–82. doi: 10.1016/j.jcbs.2012.06.001

Kabat-Zinn, J. (1994). *Wherever you go, there you are: Mindfulness meditation in everyday life*. New York, NY: Hyperion.

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

Kashdan, T. B., & Breen, W. E. (2007). Materialism and diminished well-being: Experiential avoidance as a mediating mechanism. *Journal of Social and Clinical Psychology, 26*, 521–539.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Lillis, J., & Hayes, S. C. (2008). Measuring avoidance and inflexibility in weight related problems. *International Journal of Behavioral Consultation and Therapy, 4*, 348–354.

Luciano, M. C., & Hayes, S. C. (2001). Experiential avoidance disorder. *International Journal of Clinical and Health Psychology, 1*, 109–157.

Lundgren, T., Dahl, J., & Hayes, S. C. (2008). Evaluation of mediators of change in the treatment of epilepsy with Acceptance and Commitment Therapy. *Journal of Behavioral Medicine, 31*, 225–235.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149.

Matsumoto, D., Yoo, S. H., Nakagawa, S., Alexandre, J., Altarriba, J., Anguas-Wong, A. M., & Zengeya, A. (2008). Culture, emotion regulation, and adjustment. *Journal of Personality and Social Psychology, 94*, 925–937.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354–373.

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.

Sandoz, E. K. (2010). *Assessment of body image flexibility: An evaluation the Body Image – Acceptance and Action Questionnaire.* Unpublished dissertation. University of Mississippi.

Schweizer, K. (2012). On correlated errors. *European Journal of Psychological Assessment, 28*, 1–2.

Shawyer, F., Ratcliff, K., Mackinnon, A., Farhall, J., Hayes, S. C., & Copolov, D. (2007). The voices acceptance and action scale (VAAS): Pilot data. *Journal of Clinical Psychology, 63*, 593–606.

Sidman, M. (2009). Equivalence relations and behavior: An introductory tutorial. *The Analysis of Verbal Behavior, 25*, 5–17.

Soto, J. A., Perez, C. R., Kim, Y. H., Lee, E. A., & Minnick, M. R. (2011). Is expressive suppression always associated with poorer psychological functioning? A cross-cultural comparison between European Americans and Hong Kong Chinese. *Emotion, 11*, 1450–1455.

Spira, A. P., Beaudreau, S. A., Jimenez, D., Kierod, K., Cusing, M. M., Gray, H. L., & Gallagher-Thompson, D. (2007). Experiential avoidance, acceptance, and depression in dementia family caregivers. *Clinical Gerontologist, 30*, 55–64.

Tull, M. T., & Gratz, K. L. (2008). Further examination of the relationship between anxiety sensitivity and depression: The mediating role of experiential avoidance and difficulties engaging in goal-directed behavior when distressed. *Journal of Anxiety Disorders, 22*, 199–210.

Vowles, K. E., McCracken, L. M., McLeod, C., & Eccleston, C. (2008). The Chronic Pain Acceptance Questionnaire: Confirmatory factor analysis and identification of patient subgroups. *Pain, 140*, 284–291.

Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology, 53*, 5–13.

Wegner, D. M., & Zanakos, S. (1994). Chronic thought suppression. *Journal of Personality, 62*, 616–640.

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

Williams, V., Ciarrochi, J., & Patrick Deane, F. (2010). On being mindful, emotionally aware, and more resilient: Longitudinal pilot study of police recruits. *Australian Psychologist, 45*, 274–282.

Received September 15, 2014
Revision received August 25, 2015
Accepted August 25, 2015
Published online April 29, 2016

Jean-Louis Monestès
LIP Lab
Univ. Grenoble Alpes
UFR SHS, 1251 avenue Centrale
38040 Grenoble Cedex 9
France
Tel. +33 670 127-567
E-mail jlmonestes@yahoo.fr

