Penn State Worry Questionnaire-11 validity in Colombia and factorial equivalence across gender and nonclinical and clinical samples

Francisco J. Ruiz[1*], Andrea Monroy-Cifuentes[1], and Juan C. Suárez-Falcón[2]

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

Título: Validez del Penn State Worry Questionnaire-11 en Colombia y equivalencia factorial a través de género y muestras clínicas y no clínicas.
Resumen: El Penn State Worry Questionnaire (PSWQ) es la medida principal de la preocupación inespecífica que caracteriza al trastorno de ansiedad generalizada (TAG). Este estudio explora la validez del PSWQ en Colombia y su invarianza de medida entre género y participantes clínicos y no clínicos. Se utilizó la versión abreviada de 11 ítems del PSWQ propuesta por Sandín, Chorot, Valiente y Lostao (2009) porque los ítems redactados negativamente han mostrado ser problemáticos para los hispanohablantes. Adicionalmente, la investigación ha sugerido que los ítems redactados en negativa carecen de utilidad práctica. El PSWQ-11 se administró a un total de 1045 participantes, incluyendo una muestra de participantes no clínicos (*N* = 710) y una muestra de participantes clínicos (*N* = 335). La consistencia interna del PSWQ-11 a través de muestras fue excelente. El modelo de un factor mostró un ajuste aceptable a los datos. Se observó invarianza métrica y escalar a través de género y muestras clínicas y no clínicas. En conclusión, el PSWQ-11 parece ser una medida válida de la preocupación tipo TAG, mientras que los datos sobre equivalencia factorial permiten la comparación de puntuaciones a través de género y participantes clínicos y no clínicos.
Palabras clave: Penn State Worry Questionnaire; Preocupación; Trastorno de Ansiedad Generalizada; Propiedades psicométricas; Invarianza de medida.

Abstract: The Penn State Worry Questionnaire (PSWQ) is the gold standard measure of the unspecific worry that characterizes generalized anxiety disorder. This study aims to explore the validity of the PSWQ in Colombia and its measurement invariance across gender and clinical and nonclinical participants. An abbreviated, 11-item version of the PSWQ by Sandín, Chorot, Valiente, and Lostao (2009) was used because the negatively worded items have shown to be problematic for Spanish speakers. Additionally, research has suggested that the negatively worded items of the PSWQ lack of practical utility. The PSWQ-11 was administered to a total of 1045 participants, including a sample of nonclinical participants (*N* = 710) and a sample of clinical participants (*N* = 335). The internal consistency of the PSWQ-11 across samples was excellent. The one-factor model showed an acceptable fit to the data. Metric and scalar invariance were observed across gender and clinical and nonclinical samples. In conclusion, the PSWQ-11 seems to be a valid measure of GAD-related worry in Colombia, whereas data on factorial equivalence data warrant the comparison of scores across gender and clinical and nonclinical samples.
Key words: Penn State Worry Questionnaire; Worry; Generalized anxiety disorder; Psychometric properties; Measurement invariance.

Introduction

Worry consists of repetitive thoughts that are experienced as unpleasant and that concern an uncertain future outcome that is considered undesirable (e.g., Berenbaum, 2010; Mennin, Heimberg, Turk, & Fresco, 2002). Worrying is one of the most evolved types of behavior because it allows individuals to anticipate future danger, planning, experiment with ideas before implementing them, and evaluate alternative options (Mathews, 1990). However, worry loses its adaptive functions when people engage in it chronically to the extent that it feels uncontrollable. When this occurs, worry is associated with deteriorated functioning and low quality of life (see review in Watkins, 2008). Indeed, excessive worry is considered as one of the main symptoms in several diagnostic categories such as generalized anxiety disorder (GAD), depression, eating disorders, and hypochondria (Olatunji, Wolitzky-Taylor, Sawchuk, & Ciesielski, 2010). Furthermore, worry is considered as a pervasive process involved in the onset and maintenance of emotional disorders (Ehring & Watkins, 2008; Harvey, Watkins, Mansell, & Shafran, 2004).

The measurement of worry has been typically conducted with self-report instruments such as the Worry Domains Questionnaire (WDQ; Tallis, Eysenck, & Mathews, 1992), the Student Worry Scale (SWS; Davey, Hampton, Farrell, & Davidson, 1992), and the Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990). While some of the instruments explore worry in relation to some domains (e.g., SWS and WDQ), the PSWQ was designed as a general self-report that measures the tendency to engage in worry and the difficulty to control it without focusing on the worry content (Meyer et al., 1990).

The PSWQ is broadly considered as the gold standard measure of GAD-related worry (e.g., Hanrahan, Field, Jones, & Davey, 2013). It consists of 16 items that are responded on a 5-point Likert-type scale (5 = *very typical of me*; 1 = *not at all typical of me*), with scores ranging from 16 to 80 points. Higher scores indicate higher worry degree. The internal consistency of the PSWQ is excellent, with Cronbach's alpha ranging from .91 to .94 (Meyer et al., 1990). The factor structure of the PSWQ has been a source of debate because some studies have found the better fit of a two-factor structure (Meloni & Gana, 2001; Olatunji, Schottenbauer, Rodríguez, Glass, & Arnkoff, 2007), one related with the direct measurement of worry (11 items) and the other one corresponding to the absence of worry (5 reverse-scored items). However, the second factor can be considered a statistical artifact rather than a meaningful construct, and some

* Correspondence address [Dirección para correspondencia]:
Francisco J. Ruiz. Fundación Universitaria Konrad Lorenz, Carrera 9 bis, Nº 62-43, Bogotá, Cundinamarca (Colombia).
E-mail: franciscoj.ruizj@konradlorenz.edu.co

authors have argued that the PSWQ is better represented by only one factor (Brown, 2003; Brown, Antony, & Barlow, 1992; Hazlett-Stevens, Ullman, & Craske, 2004; Korte, Allan, & Schmidt, 2016). Indeed, Olatunji et al. (2007) suggested that the five negatively worded items lack practical utility because they do not correlate adequately with other psychological variables. Therefore, the authors suggested using only the 11 positively worded items of the PSWQ.

Some studies have explored the factorial equivalence of the PSWQ across gender and clinical and nonclinical participants. Brown (2003) found factorial equivalence across male and female clinical participants. Nuevo, Mackintosh, Gatz, Montorio, and Wetherell (2007) analyzed measurement invariance of an abbreviated, 8-item version of the PSWQ in American and Spanish older adults. They found that factorial equivalence across countries could be assumed for women but not for men. Lastly, Păsărelu et al. (2017) found factorial equivalence across gender, age and clinical diagnosis using the version of the PSWQ for children (i.e., PSWQ-C). The PSWQ-C was the result of a grammar analysis and the items were reworded to be readable for children (Chorpita, Tracey, Brown, Collica, & Barlow, 1997). However, the measurement invariance of the PSWQ across gender and clinical and nonclinical adult participants remains largely unexplored. This is an important issue for a gold standard measure such as the PSWQ because, in the absence of evidence in this regard, it is not justified to compare the PSWQ scores across gender and clinical and nonclinical samples.

Several Spanish translations of the PSWQ have been conducted that have shown good psychometric properties (e.g., Nuevo, Montorio, & Ruiz, 2002; Rodríguez-Biglieri & Vetere, 2011; Sandín, Chorot, Valiente, & Lostao, 2009). However, a common concern has been raised regarding the negatively worded items because they are difficult to understand for Spanish speakers. Indeed, Sandín et al. (2009) have suggested the use of only the positively worded items (an abbreviated version of the PSWQ that was called PSWQ-11) in Spain because the instrument showed better psychometric properties than the complete PSWQ. This suggestion by Sandín et al. is in line with Olatunji et al. (2007) findings regarding the lack of practical utility of the negatively worded items. Sandín et al. found a Cronbach's alpha of .92 for the PSWQ-11 in a nonclinical sample.

To our knowledge, the validity of the PSWQ has not been explored in Colombia, which makes it difficult to conduct studies on GAD-related worry in this country. Additionally, testing measures in culturally diverse samples enhances both our confidence in the measure and the cross-cultural relevance of the underlying theory being measured (Elosua, Mujika, Almeida, & Hermosilla, 2014). Due to the concern raised with regard to the comprehensibility of the negatively worded items of the PSWQ in Spanish, we selected the version by Sandín et al. (2009) to explore the validity of the PSWQ in Colombian samples. Additionally, a secondary aim of this study was to explore the measurement invariance of the PSWQ-11 across gender and clinical and non-

clinical participants. The PSWQ-11 was administered to two samples: a nonclinical sample (*n* = 710) and a clinical sample (*n* = 335).

Method

Participants

Sample 1. The sample consisted of 710 participants (71.4% females) with age ranging between 18 and 89 years (*M* = 27.51, *SD* = 10.18). The relative educational level of the participants was: 44.8% primary studies (i.e., compulsory education) or mid-level study graduates (i.e., high school or vocational training), 34.9% were undergraduates or college graduates, and 20.3% were currently studying or had a postgraduate degree. They responded to an anonymous internet survey distributed through social media. All of them were Colombian. Only 7.4% of participants in this sample were receiving psychological/psychiatric treatment. Also, 3.7% of participants reported consumption of some psychotropic medication.

Sample 2. It consisted of 335 patients (74% of them were women), with an age range of 18 to 63 years (*M* = 27.40, *SD* = 9.93). All participants were being evaluated in the institutional psychological consultation center, in which inexpensive psychological therapy is offered to general population in Bogotá. Most of the participants stated that the reason for consultation was suffering from emotional symptoms (91%), whereas the remaining participants consulted for family problems or social skills deficits. Only 7.2% of the participants reported that they were consuming some psychotropic medication.

Instruments

Penn State Worry Questionnaire - 11 (PSWQ-11; Meyer et al., 1990; Spanish version by Sandín et al., 2009). The PSWQ was designed to evaluate the permanent and unspecific degree of worry that characterizes GAD. The Spanish version of the PSWQ showed excellent psychometric properties although the authors recommended eliminating the negatively worded items because they were difficult to understand for Spanish speakers (Sandín et al., 2009). This recommended version was named PSWQ-11.

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). Strong positive correlations

were expected between the PSWQ-11 and the DASS-21 subscales.

Generalized Anxiety Disorder – 7 (GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006). The GAD-7 is a 7-item, 4-point Likert-type scale (3 = *nearly every day*; 0 = *not at all*), self-report instrument that was designed as a diagnostic and severity measure of GAD. We used the Spanish translation of the GAD-7 for Colombia distributed by Pfizer. The GAD-7 showed good psychometric properties with α = .90 in Sample 1 and α = .87 in Sample 2. Strong positive correlations were expected between the PSWQ-11 and the GAD-7.

Ruminative Responses Scale – Short Form (RRS-SF; Treynor, Gonzalez, & Nolen-Hoeksema, 2003; Spanish version by Hervás, 2008). The RRS-SF is a 10-item, 4-point Likert scale (4 = *almost always*; 1 = *almost never*) self-report instrument that was designed to measure the tendency to ruminate in response to feelings of sadness and depression. It contains two subscales called Brooding and Reflection. According to Treynor et al., brooding is the most maladaptive form of rumination, whereas reflection could have both maladaptive and adaptive aspects. The psychometric properties of the RRS-SF in Colombia are adequate (Ruiz, Suárez-Falcón, et al., 2017). Strong and medium positive correlations were expected between the PSWQ-11 and the Brooding and Reflection subscales, respectively.

Procedure

Participants in Sample 1 responded to an anonymous internet survey distributed through social media (e.g., institutional web-pages, Facebook profiles, etc.). The survey was called "Survey of Emotional Health in Colombia" and was responded on the platform www.typeform.com. After finishing data collection, a general inform was sent to the participants who provided an email address for that purpose. Afterwards, personal scores and options for receiving inexpensive psychological treatment were provided when requested by the person. Participants in Sample 2 responded to the questionnaires during one of the clinical assessment interviews at the beginning of treatment in the presence of their therapist. All participants provided informed consent and were given a questionnaire packet.

All participants in Sample 1 responded to the PSWQ-11 and DASS-21. One part of Sample 1 also responded to the RRS-SF (*N* = 370), whereas the other part responded to the GAD-7 (*N* = 340). With regard to Sample 2, all participants responded to the PSWQ-11, 242 responded to the DASS-21 and RRS-SF, and 94 to the GAD-7.

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.

Statistical and Psychometric Analysis

Prior to conducting factor analysis, data from all samples were examined searching for missing values that were im-

puted 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). 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 PSWQ-11. Only two values were missing.

Because the PSWQ uses a Likert-type scale measured on an ordinal scale, a robust diagonally weighted least squares (Robust DWLS) estimation method using polychoric correlations was used to conduct the CFA. The WLS method is recommended in large samples with fewer than 20 items (Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2010; Jöreskog & Sörbom, 1996) as in the current study. In order to use the matrix of polychoric correlations, the assumption of bivariate normal distribution was analyzed by means of the chi-squared test and the percentage of tests that rejected the null hypothesis of bivariate normality for each pair of correlations. Due to the sensitivity of the chi-square test, the root mean square error of approximation (RMSEA) was also analyzed for each pair of correlations. Hooper, Coughlan, and Mullen (2008) point out that the parameter estimation is not very affected when RMSEA values are not higher than 0.1.

We computed the Satorra-Bentler chi-square test and the following goodness-of-fit indexes for the one-factor model: (a) RMSEA; (b) the comparative fit index (CFI); and (c) the non-normed fit index (NNFI). According to Kelloway (1998), RMSEA values of .10 represent an acceptable fit although for Hu and Bentler (1999) RMSEA values to consider acceptable fit should be .08. With respect to the CFI and NNFI, values above .90 indicate well-fitting models, and above .95 represent a very good fit to the data.

Additional CFA were performed to test for metric and scalar invariance across gender and clinical and nonclinical participants following Jöreskog (2005) and Millsap and Yun-Tein (2004). In other words, we analyzed whether the item factor loadings and item intercepts were invariant across 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 eleven unstandardized factor loadings to vary across the samples and across gender. 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 invariance model, was tested by constraining the factor loadings and the item 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.

Coefficients alpha and McDonald's omega were computed providing percentile bootstrap 95% confidence intervals (CI) to explore the internal consistency of the PSWQ-11 in Samples 1 and 2 and the overall sample (Viladrich, Angulo-Brunet, & Doval, 2017). In order to calculate these coefficients, the MBESS package in R was used (Kelley & Lai, 2012; Kelley & Pornprasertmanit, 2016). The remaining statistical analyses were performed on SPSS 20©. 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 PSWQ-11 scores were explored by computing independent sample *t*-test. To examine criterion validity, scores on the PSWQ-11 were compared between participants in Sample 1 (nonclinical participants) and participants in Sample 2 (clinical participants). Pearson correlations between the PSWQ-11 and other scales were calculated to assess validity evidence based on relationships with other variables.

Results

Descriptive data and psychometric quality of the items

Table 1 shows the descriptive data and corrected item-total correlations for Samples 1 and 2. All items showed

good discrimination, with corrected item-total correlations ranging from .67 to .88 in Sample 1, and from .65 to .85 in Sample 2.

Table 1. Corrected Item-Total Correlations and Descriptive Data.

	Corrected item-total correlations		M (SD)	
	Sample 1 nonclinical	Sample 2 clinical	Sample 1 nonclinical	Sample 2 clinical
Item 1	.73	.70	2.67 (1.12)	3.68 (1.05)
Item 2	.76	.75	2.60 (1.06)	3.38 (1.11)
Item 3	.79	.69	2.45 (1.19)	3.42 (1.21)
Item 4	.76	.73	2.97 (1.20)	3.79 (1.14)
Item 5	.82	.77	2.27 (1.11)	2.95 (1.18)
Item 6	.74	.71	2.39 (1.20)	3.11 (1.25)
Item 7	.76	.65	2.14 (1.12)	2.71 (1.28)
Item 8	.75	.70	2.63 (1.35)	3.51 (1.30)
Item 9	.82	.78	2.24 (1.16)	3.12 (1.17)
Item 10	.88	.85	2.22 (1.14)	3.07 (1.20)
Item 11	.67	.65	2.88 (1.18)	3.51 (1.16)

Table 2 shows that the alpha and omega coefficients of the PSWQ-11 were almost identical and excellent in all cases.

Table 2. Alpha and Omega Coefficients and Descriptive Data across Samples.

	Sample 1: Nonclinical (n = 710)	Sample 2: Clinical (n = 335)	Overall Sample (n = 1045)
Alpha 95% CI	.949 [.942, .954]	.933 [.920, .942]	.951 [.946, .955]
Omega 95% CI	.949 [.943, .954]	.934 [.921, .944]	.952 [.947, .956]
Mean score (SD)	27.47 (10.44)	36.26 (10.13)	30.28 (11.12)

Validity evidence based on internal structure

Dimensionality

The results of the chi-square test to explore bivariate normality showed that this assumption was accepted in 22 % of the correlations. However, the RMSEA values were lower than 0.1 in all correlations, which supports the use of the matrix of polychoric correlations to conduct the CFA.

The goodness-of-fit values of the one-factor model were: $S\text{-}B\chi^2(44) = 430.73$, $p < .01$; CFI = .99; NNFI = .98; and RMSEA = .092, 90% CI [.084, .099]. The CFI and NNFI values indicated a very good fir to the data, and the RMSEA showed an acceptable fit according to the guidelines provided by Kelloway (1998), but poorer according to Hu and Bentler (1999). Overall, the fit of the one-factor model seemed to be acceptable. Figure 1 depicts the results of the standardized solution of the one-factor model.
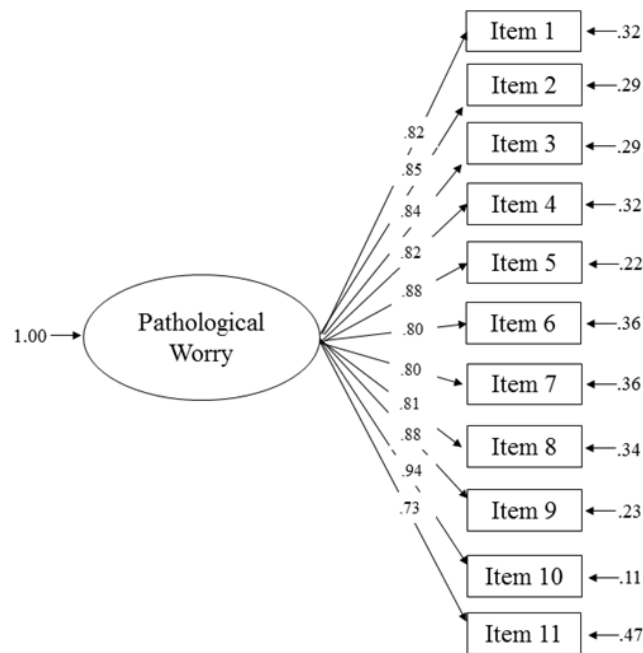
Figure 1. Results of the confirmatory factor analysis conducted with the overall sample to analyze the fit of a one-factor model.

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 gender and clinical and nonclinical participants because changes in RMSEA, CFI, and NNFI were lower than .01.

Table 3. Metric and Scalar Invariance across Clinical and Nonclinical Samples and Gender.

Model	RMSEA	ΔRMSEA	CFI	ΔCFI	NNFI	ΔNNFI
Measurement invariance across clinical and nonclinical samples						
MG Baseline model	.0955		.984		.981	
Metric invariance	.0952	.0003	.983	-.001	.981	.000
Scalar invariance	.1010	-.0058	.979	-.004	.978	-.003
Measurement invariance across gender						
MG Baseline model	.0893		.985		.981	
Metric invariance	.0884	.0009	.984	-.001	.982	.001
Scalar invariance	.0861	.0023	.983	-.001	.983	.001

Validity evidence based on relationships with other variables

The PSWQ-11 showed correlations with all the other assessed constructs in theoretically coherent ways (see Table 4). Specifically, the PSWQ-11 showed strong positive correlations with emotional symptoms as measured by the DASS-21, symptoms of GAD, and brooding. Lower positive correlations were found with reflection.

Means and standard deviations of the PSWQ-11 scores for nonclinical and clinical samples can be seen in Table 2. Participants' mean score in the clinical sample (Sample 2) was higher than that of the nonclinical participants (Sample 1) ($t = -12.80$, $p < .001$). There were no statistically significant differences across gender in the PSWQ-11 in any of the samples.

Table 4. Pearson Correlations between the PSWQ-11 Scores and Other Relevant Self-report Measures.

Measure	S	N	r with PSWQ-11
DASS-21 – Depression	1	710	.58*
	2	242	.46*
DASS-21 – Anxiety	1	710	.60*
	2	242	.58*
DASS-21 – Stress	1	710	.67*
	2	242	.63*
RRS-SF Brooding	1	370	.65*
	2	242	.49*
RRS-SF Reflection	1	370	.33*
	2	242	.30*
GAD-7	1	340	.73*
	2	94	.73*

Note. DASS-21 = Depression Anxiety and Stress Scale – 21; GAD-7 = Generalized Anxiety Disorder – 7; PSWQ-11 = Penn State Worry Questionnaire – 11; RRS-SF = Ruminative Response Style – Short Form; S = Sample; 1 = Nonclinical sample; 2 = Clinical sample; *$p < .001$

Discussion

Although unconstructive worry is pervasive among emotional disorders (Watkins, 2008), GAD represents the prototype disorder in which worry plays a central role. Individuals suffering from GAD experience unspecific and permanent worry that is usually felt as uncontrollable. To assess this type of worry, Meyer et al. (1990) designed the PSWQ-11, which soon became the gold standard measure of GAD-related worry (Hanrahan et al., 2013).

Several Spanish versions of the PSWQ exist that have shown excellent psychometric properties although significant concerns have been raised with regard to the practical utility and comprehensibility of the negatively worded items (e.g., Olatunji et al., 2007; Sandín et al., 2009). Accordingly, Sandín et al. (2009) recommended using the PSWQ-11 in Spain, which is the result of eliminating the five negatively worded items. This version showed very good psychometric properties and a one-factor structure. We selected this version to explore the validity of the PSWQ in Colombia.

The PSWQ-11 was administered to a nonclinical sample (*N* = 710) and a clinical sample (*N* = 335) showing an excellent internal consistency (overall alpha of .95). The one-factor model showed an acceptable fit to the data, and measurement invariance at both metric and scalar levels was obtained across samples and gender. This indicates that the PSWQ-11 is measuring the same construct across nonclinical and clinical samples, and across gender. The PSWQ-11 also showed validity evidence based on relationships with other variables in view of the strong correlations found with measures of GAD, depression, anxiety, stress, and rumination. Lastly, the PSWQ-11 scores discriminated between clinical and nonclinical samples.

One important finding of this study is the factorial equivalence across gender, and nonclinical and clinical participants. These proofs of measurement invariance are important because the studies that use the PSWQ usually compare scores from these types of samples. In the absence of data supporting the factorial equivalence of the PSWQ, the comparison of the scores across these samples is not justified. With regard to factorial equivalence across gender, the current study extends the findings by Brown (2003) who found measurement invariance across male and female clinical participants. However, the factorial equivalence across gender in nonclinical adult participants had not been analyzed. On the other hand, the analysis of factorial equivalence across nonclinical and clinical adult participants had not been explored. Overall, the findings of this study, in conjunction with those of Păsărelu et al. (2017) with the children version of the PSWQ, point to considering that the PSWQ is invariant across gender and clinical and nonclinical participants.

Some limitations of this study are worth mentioning. Firstly, no systematic information was obtained concerning the diagnosis in clinical participants. Secondly, some validity aspects of the PSWQ-11 have not been analyzed in the current study (e.g., divergent validity, sensitivity to treatment effects, etc.). However, there is already evidence that the PSWQ-11 was sensitive to the treatment effect of brief acceptance and commitment therapy protocols focused on reducing repetitive negative thinking (Ruiz, Riaño-Hernández, Suárez-Falcón, & Luciano, 2016; Ruiz et al., in press). Thirdly, the percentage of women was significantly higher than the percentage of men in the composition of the samples. Also, the nonclinical sample was higher than the clinical sample. However, the number of male and clinical participants (approximately 300 for each category) was enough to conduct the measurement invariance analyses. Lastly, due to time constraints, all participants could not respond to all the questionnaires so that we decided to administer some of them a part of the sample and the others to the remaining part.

In conclusion, the abbreviated Spanish version of the PSWQ (i.e., PSWQ-11) suggested by Sandín et al. (2009) can be used to measure GAD-related worry in Colombia. The factorial equivalence found across gender and clinical and nonclinical samples justified the comparison of scores between male and women and between clinical and nonclinical participants. Further studies are needed to confirm the measurement invariance data found in this study across in other contexts. Also, additional studies might explore the psychometric properties of this version of the PSWQ-11 in other Spanish speaking countries and test for measurement invariance across countries.

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.

Berenbaum, H. (2010). An initiation-termination two-phase model of worrying. *Clinical Psychology Review, 30*, 962-975.

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

Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behaviour Research and Therapy, 30*, 33–37.

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

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.

Chorpita, B. F., Tracey, S. A., Brown, T. A., Collica, T. J., & Barlow, D. H. (1997). Assessment of worry in children and adolescents: An adaptation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 35*, 569-581.

Davey, G. C. L., Hampton, J., Farrell, J., & Davidson, S. (1992). Some characteristics of worrying: Evidence for worrying and anxiety as separate constructs. *Personality and Individual Differences, 13*, 133–147.

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.

Ehring, T., & Watkins, E. R. (2008). Repetitive negative thinking as a transdiagnostic process. *International Journal of Cognitive Therapy, 1*, 192-205.

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.

Hanrahan, F., Field, A. P., Jones, F. W., & Davey, G. C. (2013). A meta-analysis of cognitive therapy for worry in generalized anxiety disorder. *Clinical Psychology Review, 33*, 120-132.

Harvey, A. G., Watkins, E. R., Mansell, W., & Shafran, R. (2004). *Cognitive behavioural processes across disorders.* Oxford, UK: Oxford University Press.

Hazlett-Stevens, H., Ullman, J. B., & Craske, M. G. (2004). Factor structure of the Penn State Worry Questionnaire: Examination of a method factor. *Assessment, 11*, 361-370.

Hervás, G. (2008). Adaptación al castellano de un instrumento para evaluar el estilo rumiativo: la Escala de Respuestas Rumiativas [Spanish adaptation of an instrument to assess ruminative style: Ruminative Response Scale]. *Revista de Psicopatología y Psicología Clínica, 13,* 111-121.

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity, 44*, 153-166.

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

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

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

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.

Kelley, K., & Lai, K. (2012). MBESS: MBESS. R package version 3.3.2. Retrieved from http://CRAN.R-project.org/package=MBESS.

Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods, 21*, 69–92.

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

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

Korte, K. J., Allan, N. P., & Schmidt, N. B. (2016). Factor mixture modeling of the Penn State Worry Questionnaire: Evidence for distinct classes of worry. *Journal of Anxiety Disorders, 37,* 40-47.

Mathews, A. (1990). Why worry? The cognitive function of anxiety. *Behaviour Research and Therapy, 28*, 455-468.

Meloni, F., & Gana, K. (2001). Wording effects in the Italian version of the Penn State Worry Questionnaire. *Clinical Psychology & Psychotherapy, 8*, 282-287.

Mennin, D. S., Heimberg, R. G., Turk, C. L., & Fresco, D. M. (2002). Applying an emotion regulation framework to integrative approaches to generalized anxiety disorder. *Clinical Psychological: Science and Practice, 9*, 85-90.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28*, 487-495.

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

Nuevo, R., Mackintosh, M. A., Gatz, M., Montorio, I., & Wetherell, J. L. (2007). A test of the measurement invariance of a brief version of the Penn State Worry Questionnaire between American and Spanish older adults. *International Psychogeriatrics, 19*, 89-104.

Nuevo, R., Montorio, I., & Ruiz, M. A. (2002). Aplicabilidad del Inventario de Preocupación de Pensilvania (PSWQ) a población de edad avanzada [Applicability of the Penn State Worry Questionnaire in an older adult population]. *Ansiedad y Estrés, 8*, 157-172.

Olatunji, B. O., Schottenbauer, M. A., Rodríguez, B. F., Glass, C. R., & Arnkoff, D. B. (2007). The structure of worry: Relations between positive/negative personality characteristics and the Penn State Worry Questionnaire. *Journal of Anxiety Disorders, 21*, 540-553.

Olatunji, B. O., Wolitzky-Taylor, K. B., Sawchuk, C. N., & Ciesielski, B. G. (2010). Worry and the anxiety disorders: A meta-analytic synthesis of specificity to GAD. *Applied and Preventive Psychology, 14*, 1-24.

Păsărelu, C. R., Dobrean, A., Balazsi, R., Predescu, E., Șipos, R., & Lupu, V. (2017). The Penn State Worry Questionnaire for Children: Age, gender and clinical Invariance. *Child Psychiatry & Human Development, 48*, 359-369.

Rodríguez-Biglieri, R., & Vetere, G. L. (2011). Psychometric characteristics of the Penn State Worry Questionnaire in an Argentinean sample: A cross-cultural contribution. *The Spanish Journal of Psychology, 14*, 452-463.

Ruiz, F. J., Flórez, C. L., García-Martín, M. B., Monroy-Cifuentes, A., Barreto-Montero, K., García-Beltrán, D. M.,… & Gil-Luciano, B. (2018, in press). A multiple-baseline evaluation of a brief acceptance and commitment therapy protocol focused on repetitive negative thinking for moderate emotional disorders. *Journal of Contextual Behavioral Science.*

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 Scales-21. *International Journal of Psychology and Psychological Therapy, 17*, 97-105.

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: A randomized multiple-baseline design. *International Journal of Psychology and Psychological Therapy, 16*, 213-233.

Ruiz, F. J., Suárez-Falcón, J. C., Sierra, M. A., Barreto-Montero, K., García-Martín, M. B., Bernal, P. A., & Ramírez, E. S. (2017). Psychometric properties and factor structure of the Ruminative Response Scale-Shor Form in Colombia. *International Journal of Psychology and Psychological Therapy, 17*, 199-208.

Sandín, B., Chorot, P., Valiente, R. M., & Lostao, L. (2009). Validación española del cuestionario de preocupación PSWQ: estructura factorial y propiedades psicométricas [Spanish validation of the PSWQ Worry Questionnaire: Facture structure and psychometric properties]. *Revista de Psicopatología y Psicología Clínica, 14*, 107-122.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*, 1092-1097.

Tallis, F., Eysenck, M. W., & Mathews, A. (1992). A questionnaire for the measurement of nonpathological worry. *Personality and Individual Differences*, 13, 161-168.

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research, 27*, 247-259.

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Anales de Psicología, 33*, 755-782.

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin, 134*, 163-206.