EVALUAR

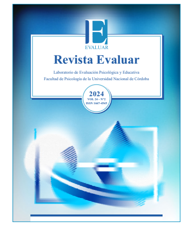
Revista Evaluar

Laboratorio de Evaluación Psicológica y Educativa
Facultad de Psicología de la Universidad Nacional de Córdoba

2024
VOL 24 - N°2
ISSN 1667-4545

2024, Vol. 24, No. 2
ISSN 1667-4545
Recuperado de https://revistas.unc.edu.ar/index.php/revaluar

Laboratorio de Evaluación Psicológica y Educativa
Facultad de Psicología - Universidad Nacional de Córdoba



Short Versions of the Penn State Worry Questionnaire: Psychometric Performance in a Sample of People Seeking Help

Versiones breves del cuestionario Penn State Worry Questionnaire: Rendimiento psicométrico en una muestra de personas que buscan ayuda

Pablo D. Valencia * [1] , Anabel De la Rosa-Gómez [1] , Ana Belén González-Hernández [2] , Juan Carlos Claros-Guzman [1] , Aylin Torres-Guffanti [3] , Alejandrina Hernández-Posadas [1]

1- Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Tlalnepantla, México.
2 - Universidad Virtual del Estado de Guanajuato, Purísima del Rincón, México.
3 - Instituto Peruano de Altas Capacidades, Lima, Peru.

Recibido: 15/10/2023 **Revisado:** 03/01/2024 **Aceptado:** 04/01/2024

Abstract

Worry is a cognitive transdiagnostic variable that is relevant for psychopathology research. The Penn State Worry Questionnaire (PSWQ) was developed to measure it. We aimed to examine the psychometric properties of three short versions of the PSWQ (11, 8, and 7 items) in a sample of Mexicans seeking help. A sample of 1391 individuals (82.2% women) seeking online psychological help completed the 11-item PSWQ, as well as measures of depression and anxiety. Single and multi-group confirmatory factor analyses were conducted. Good fit was achieved in the three versions only after adding correlated residuals to the models. Internal consistency reliability was excellent for the PSWQ-11 (ω = .93) and the PSWQ-A (ω = .90); it was acceptable for the PSWQ-5 (ω = .81). Furthermore, evidence of approximate invariance between sexes and age groups was found. Finally, the three versions were similarly associated with depression and anxiety.

Keywords: *worry, anxiety, psychopathology, validation studies, factor analysis, Mexico*

Resumen

La preocupación es una variable cognitiva transdiagnóstica relevante para la investigación en psicopatología. El Penn State Worry Questionnaire (PSWQ) fue desarrollado para medirla. Nuestro objetivo fue examinar las propiedades psicométricas de tres versiones cortas del PSWQ (11, 8 y 7 ítems) en una muestra de personas de origen mexicano que buscaban ayuda. Una muestra de 1391 individuos (82.2% mujeres) que buscaban ayuda psicológica en línea completaron el PSWQ-11, así como medidas de depresión y ansiedad. Se realizaron análisis factoriales confirmatorios de grupo único y multigrupo. Se alcanzó un buen ajuste en las tres versiones solo después de añadir residuos correlacionados a los modelos. La fiabilidad de la consistencia interna para el PSWQ-11 (ω = .93) y el PSWQ-A fue excelente (ω = .90), mientras que para el PSWQ-5 fue aceptable (ω = .81). Asimismo, se encontraron evidencias de invarianza aproximada entre sexos y grupos de edad. Por último, las tres versiones se asociaron de forma similar con la depresión y la ansiedad.

Palabras clave: *preocupación, ansiedad, psicopatología, estudios de validación, análisis factorial, México*

***Correspondence to:** Pablo D. Valencia. E-mail: pablo.valencia@ired.unam.mx. Address: Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Av. De los Barrios 1, Col. Los Reyes Iztacala, Edo. de Mexico, 54090, Tlalnepantla, México.

How to cite: Valencia, D., P., De la Rosa-Gómez, A., González-Hernández, B., A., Claros-Guzman, C., J., Torres-Guffanti, A., & Hernández-Posadas, A., (2024): Short Versions of the Penn State Worry Questionnaire: Pyschometric Performance in a Sample of People Seeking Help. *Revista Evaluar 24*(2), 1-12. Retrieved from https://revistas.unc.edu.ar/index.php/revaluar

Author's Note: We confirm that there are no known conflicts of interest associated with this publication. This work was supported by the UNAM-PAPIIT Project (IT300721). The funding institution had no role in the design of the study or in the collection, analysis, and interpretation of the data, and had no role in the writing of the manuscript.

Participaron en la edición de este artículo: Vanesa Toledo, Gloria Nieve, Abigail Pérez, Florencia Ruiz, Jorge Bruera, Benjamín Casanova.

Introduction

Worry is a cognitive phenomenon that involves focusing on adverse consequences related to future events, characterized by catastrophic anticipations, and correlated with various mental health disorders (Borkovec et al., 1998). Indeed, worry can have a negative impact on individuals and it is strongly associated with anxiety, depression, and other diagnoses (Wu et al., 2013); in this context, it can be considered a transdiagnostic feature (Ehring & Behar, 2020). Therefore, studies show the necessity to use questionnaires or reliable instruments to measure this variable, mainly because of the relevance of interventions supported by measurements and monitoring of therapeutic progress (Puccinelli et al., 2023; Wuthrich et al., 2014).

The Penn State Worry Questionnaire (Meyer et al., 1990), known as PSWQ, measures worry intensity. Originally, it was developed based on a single general dimension of the trait worry, which has been considered in its application. However, this unidimensionality has been questioned in some studies, identifying two factors: the first factor relates to the general tendency to worry which comprises 11 positively worded items and the second factor relates to the absence of worry, which includes 5 negatively worded items. This last factor, however, seems to be a methodological artifact and lacks a substantial interpretation (Brown, 2003). Due to the above, the elimination of the five negatively worded items has been suggested, resulting in the 11-item PSWQ-11 (Nuevo Benítez et al., 2002; Sandín et al., 2009). Likewise, there are antecedents on other abbreviated versions of the PSWQ, which seek to shorten the administration time in clinical practice without compromising reliability and efficiency (Carbonell-Bártoli & Tume-Zapata, 2022), such as the 8-item PSWQ-A (Hopko et al., 2003) and the 5-item Brief PSWQ (Topper et al., 2014).

Research has consistently shown that shorter versions of the PSWQ perform better than the original 16-item version (Padros-Blazquez et al., 2018; Valencia & Paredes-Angeles, 2021). For example, the PSWQ-11 has demonstrated adequate factorial fit and evidence of gender invariance (Ruiz et al., 2018). Likewise, the PSWQ-A has exhibited a much better fit than the original PSWQ-16, along with evidence of measurement invariance between racial groups in the United States (Cares et al., 2022; DeLapp et al., 2016). One possible exception is a study conducted with Chinese adolescents, where it was necessary to include two pairs of residual correlations (items 4–5 and 7–8 as numbered in the PSWQ-11) for the PSWQ-A to achieve a good fit (Xie et al., 2023). However, after adding these modifications, the PSWQ-A showed evidence of invariance by sex and age within this adolescent sample (Xie et al., 2023). Regarding the PSWQ-5, there have been few studies conducted in the literature. In a study conducted with Peruvian university students, a good model fit was found for this version, but no data were reported regarding measurement invariance (Valencia & Paredes-Angeles, 2021). In a previous Mexican study, the PSWQ-11 and PSWQ-A performed adequately in the adult population (Padros-Blazquez et al., 2018). However, this research had two main limitations: (a) all study samples were non-clinical, and (b) measurement invariance was not examined.

Therefore, the present study aimed to examine the psychometric properties of the PSWQ-11, PSWQ-A, and PSWQ-5 in a Mexican sample of individuals seeking psychological care online. Specifically, we examined the factor structure, reliability, invariance by sex and age, and validity evidence based on relations to other variables (depression and anxiety, akin to the examination conducted by Becerra Herrera et al., 2023).

Method

Design

The present study follows an instrumental design since its objective is to examine the psychometric properties of a test (Ato et al., 2013).

Participants

The sample consisted of 1391 individuals (82.2% women) aged 18 to 76 (M = 31.67, SD = 9.92). Most participants were single (52.9%), followed by those married or cohabiting (34.0%). The great majority (69.0%) reported having a university education. Regarding their place of residence, all except 49 individuals lived in Mexico; the states with the highest representation were the State of Mexico (33.8%) and Mexico City (30.8%). Of the participants, 15.5% indicated that they were undergoing psychological or psychiatric treatment, while 10.6% reported being under psychiatric pharmacological treatment. Regarding their occupation, 27.0% were students, 26.2% were employed, 14.5% were professionals, 9.8% were unemployed, 9.1% were homemakers, 8.3% were self-employed, 4.5% were employed in another job, and 0.6% were retired. The only inclusion criterion was to be 18 years old or older and to have completed the PSWQ-11. For the present study, no exclusion criteria were considered.

Measures

Penn State Worry Questionnaire (PSWQ-11). The Penn State Worry Questionnaire (Meyer et al., 1990), known as PSWQ-11 in the version applied in this study, is an instrument for measuring trait worry. It consists of 11 Likert-type items (1 = *not at all*, 5 = *very much*) and is recommended to measure the intensity of worry, as well as to help in the diagnosis of generalized anxiety disorder (González et al., 2007; Nuevo Benítez et al., 2002). For this study, a translated version with psychometric properties analyzed with Mexican samples was used; both the PSWQ-11 and the PSWQ-A showed good reliability (α = .88 and α = .85, respectively; Padros-Blazquez et al., 2018). The detailed psychometric properties of this measure in our data are presented in the Results section.

Beck Depression Inventory-II (BDI-II). The Beck Depression Inventory (Second Version) is an instrument seeking to assess the severity of depressive symptoms during the last two weeks (Beck et al., 2006). It comprises 21 items (e.g., *Loss of interest*) and a response scale from 0 to 3, resulting in total scores ranging from 0 to 63. The present study employed the Mexican adaptation of the BDI-II by González et al. (2015), who found high internal consistency in students (α = .92) and a community sample (α = .87). Finally, the instrument presented a Fernandez-Huerta index of 80, evidencing that the Mexican adaptation is very readable/accessible. In the present study, the reliability of this instrument was optimal (α = .91).

Beck Anxiety Inventory (BAI). The Beck Anxiety Inventory (BAI) represents an empirically validated psychometric assessment instrument designed to quantify the severity of anxious symptomatology in adolescent and adult populations (Robles et al., 2001). This self-report inventory comprises 21 items covering a wide range of anxiety-related symptoms, including both physical and cognitive manifestations. Each BAI item is assessed using a four-point Likert scale ranging from 0 (indicating the absence of the symptom) to 3 (indicating the severe presence of symptoms). Participants are instructed to rate each item based on their experience during the week prior to the time of the assessment. Subsequently, the item scores are summed to obtain a total score ranging from 0 to 63. In the present study, the reliability of the scores was good (α = .92).

Procedure

Data collection was part of a larger project, which consisted of a clinical trial that tested two online psychotherapy interventions (de la Rosa-Gómez et al., 2023). Dissemination was carried out in social networks and institutional channels, inviting individuals interested in applying for a free online psychotherapeutic intervention, which required them to answer a series of questionnaires as an initial screening. For the present study, only data from this initial screening (baseline) were used. These data were collected via a SurveyMonkey form, and the instruments were administered in randomized order to control for participant fatigue.

Ethical Considerations

At the beginning of the SurveyMonkey form, participants were provided with information regarding confidentiality, data handling, potential risks, and benefits. Individuals were required to provide consent to participate in the study. Throughout the project, the complete baseline data were exclusively managed by two research assistants, who created anonymized versions of the databases for use by other team members. The intervention project was approved by the Ethics Committee of the Facultad de Estudios Superiores Iztacala of the Universidad Nacional Autónoma de México (CE/FESI/082020/1363).

Data Analysis

First, the descriptive statistics of mean, standard deviation, skewness, and kurtosis were examined for each item and instrument (PSWQ-11, PSWQ-A and PSWQ-5). Skewness and kurtosis values within the range [-1, +1] were considered as evidence that the item follows an approximately normal distribution (Ferrando et al., 2022). In addition, the response percentages for each option were analyzed to identify potential floor or ceiling effects. The corrected item-test correlations for each dimension were also examined to determine whether any should be eliminated for having a value of less than .30; values greater than .30 were considered acceptable.

Subsequently, a confirmatory factor analysis (CFA) based on Pearson correlations was performed. The method used was a robust variant of maximum likelihood (MLR) considered appropriate when the items have five or more response options (Rhemtulla et al., 2012). Model fit was assessed with the following approximate indices (the good fit criterion is mentioned in parentheses): CFI (> .95), TLI (> .95), RMSEA (< .06), and SRMR (< .08). Reliability was estimated from the results of the factor analysis through the omega coefficient. In a complementary manner, Cronbach's alpha coefficient was also calculated.

Next, measurement invariance was examined regarding sex (male vs. female) and age (< 30 vs. ≥ 30). Models of increasing invariance were tested sequentially: factor loadings (metric invariance), intercepts (scalar invariance) and residuals (strict invariance). To assess whether invariance was met, we examined the change in CFI (ΔCFI). Compared with the previous model, if the CFI of the new model decreased by more than .01, invariance was considered not to be met at that level (Cheung & Rensvold, 2002). Specifically, the robust CFI proposed by Brosseau-Liard and

Savalei (2014) was used for such comparisons.

Finally, Pearson correlation coefficients were estimated as evidence of associative validity. All analyses were performed in the *R* 4.3.0 program, using the following packages: lavaan 0.6-16, semPlot 1.1.6 and psych 2.3.3.

Results

Preliminary Item Analysis

When analyzing the questionnaire items across the PSWQ-11, PSWQ-A, and PSWQ-5 versions, most of the skewness and kurtosis values were within the range [-1, +1] and no evidence of a floor or ceiling effect was observed. Furthermore, all item-test correlations were examined, and all of them were greater than .30 (Table 1).

Confirmatory Factor Analysis and Internal Consistency Reliability

To empirically test the proposed dimensionality, a CFA was performed for each PSWQ questionnaire (Table 2). The one-factor model of the PSWQ-11 presented a good fit only regarding SRMR. Consequently, the modification indices were examined, which suggested allowing for covariation between the errors of items 1 and 2. However, the fit was still suboptimal even after allowing this pair of correlated errors. Subsequently, we proceeded to test a model that incorporated the correlations between the errors of items 1–2 and items 7–8, and this model achieved an acceptable fit across most of the indices (Table 2). Similarly, the PSWQ-A version required the same pair of correlated errors to achieve a good fit (Table 2). Finally, the PSWQ-5 achieved an acceptable fit only after including the correlation between the residuals of items 7 and 8 (Table 2). The standardized factor loadings of the final models are shown in Figure 1.

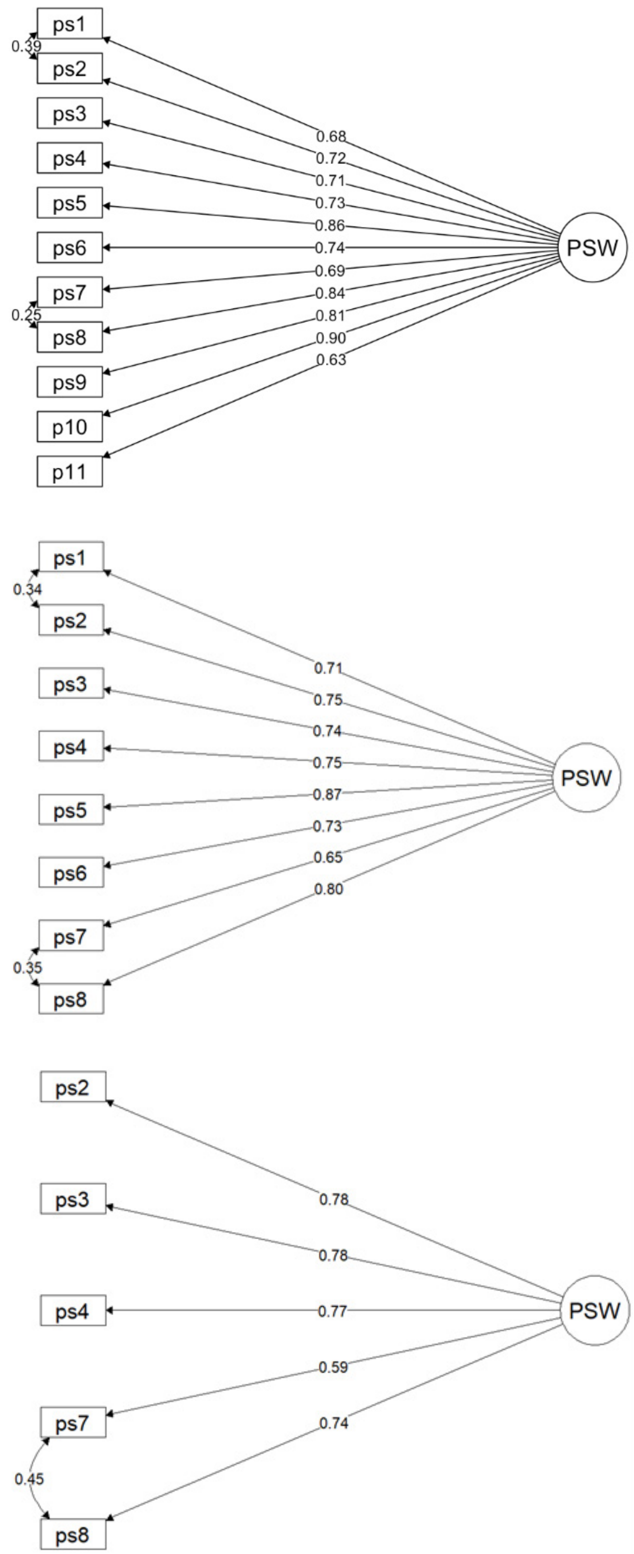


Figure 1
Standardized Coefficients of the Confirmatory Factor Analyses of PSWQ (Short Versions).

Table 1
Descriptive statistics and item-total correlations of the Penn State Worry Questionnaire's items.

Item	*M*	*DE*	g_1	g_2	% of responses per option					r_{it}		
					1	2	3	4	5	PSWQ-11	PSWQ-A	PSWQ-5
1. Sus preocupaciones le agobian [Your worries overwhelm you].	4.01	0.94	-0.83	0.24	1	7	17	41	34	.68	.69	
2. Hay muchas circunstancias que hacen que se preocupe [Many situations make you worry.].	3.88	0.97	-0.63	-0.28	1	10	19	41	29	.72	.73	.69
3. Sabe que no debería preocuparse por las cosas, pero no puede evitarlo [You know you should not worry about things, but you just cannot help it.].	3.98	1.00	-0.82	-0.03	1	9	17	37	36	.69	.70	.67
4. Cuando está bajo tensión tiende a preocuparse mucho [When you are under pressure, you worry a lot].	4.13	0.95	-1.10	0.82	2	6	13	37	43	.71	.71	.68
5. Siempre está preocupándose por algo [You are always worrying about something].	3.77	1.10	-0.60	-0.49	3	12	21	33	31	.82	.81	
6. Tan pronto como termina una tarea, en seguida empieza a preocuparse por alguna otra cosa que debe hacer [As soon as you finish one task, you start to worry about everything else you have to do].	3.68	1.22	-0.62	-0.61	6	12	21	29	32	.72	.69	
7. Ha estado preocupado toda su vida [You have been a worrier all your life].	3.39	1.25	-0.28	-1.01	8	20	22	27	23	.67	.65	.62
8. Se da cuenta de que siempre está preocupándose por las cosas [You notice that you have been worrying about things.].	3.72	1.16	-0.59	-0.68	4	15	18	33	31	.81	.79	.76
9. Una vez que comienza a preocuparse por algo, ya no puede parar [Once you start worrying, you cannot stop].	3.67	1.16	-0.57	-0.58	5	13	21	32	29	.77		
10. Está todo el tiempo preocupándose por algo [You worry all the time].	3.60	1.19	-0.52	-0.73	5	16	19	33	27	.85		
11. Se preocupa por un proyecto hasta que está acabado [You worry about projects until they are all done].	3.72	1.18	-0.61	-0.60	5	13	20	30	32	.61		

Note. *N* = 1731. g_1 = skewness; g_2 = kurtosis (zero-centered); r_{it} = corrected item-total correlation.

Table 2
Fit indices of the confirmatory factor analyses for PSWQ-11, PSWQ-A, and PSWQ-5.

Model	Correlated residuals	χ^2	*gl*	*p*	CFI	TLI	RMSEA	SRMR	α	ω
PSWQ-11	—	572.29	44	<.001	.93	.91	.09	.04	.94	.94
	1 & 2	413.36	43	<.001	.95	.94	.08	.04	.94	.93
	1 & 2, 7 & 8	357.05	42	<.001	.96	.95	.07	.04	.94	.93
PSWQ-A	—	355.25	20	<.001	.93	.90	.11	.04	.91	.91
	7 & 8	235.91	19	<.001	.95	.93	.09	.04	.91	.90
	1 & 2, 7 & 8	139.33	18	<.001	.97	.96	.07	.03	.91	.90
PSWQ-5	—	179.57	5	<.001	.92	.84	.16	.04	.86	.86
	7 & 8	11.22	4	.024	1	.99	.04	.01	.86	.81

Note. *N* = 1731. The estimation method used was robust maximum likelihood (MLR).

Table 2 also illustrates the internal consistency reliability estimates for each version. In the final models, the coefficients of the PSWQ-11 (ω = .93) and the PSWQ-A (ω = .90) were very similar. On the other hand, the reliability of the PSWQ-5 was relatively lower (ω = .81), although still acceptable.

Measurement Invariance

Table 3 shows the results of the invariance analysis. Notably, strict invariance was met in all the brief versions concerning sex. On the other hand, regarding age, strict invariance was met in the PSWQ-11 and PSWQ-A, but only scalar invariance in the case of the PSWQ-5.

Associative Evidence of Validity

A subset of individuals also reacted to measures of depressive (*n* = 1323) and anxious (*n* = 1327) symptomatology. As shown in Table 4, correlations were very similar across the three versions. Indeed, correlations of the PSWQ-11 and PSWQ-A with both measures were virtually identical.

Discussion

In the present study, the psychometric properties of three brief versions of the PSWQ (PSWQ-11, PSWQ-A, and PSWQ-5) were examined in a sample of people seeking psychotherapeutic help. The scale was found to function almost unidimensionally, but it was necessary to consider the correlation between residuals. Likewise, reliability in all three versions was adequate and evidence of invariance about sex and age was found. Finally, the three versions offered similar correlations with measures of anxiety and depression.

Previous studies have also found that the PSWQ, in its different brief versions, functions adequately as a unidimensional measure and, in addition, shows good reliability (Cares et al., 2022; Padros-Blazquez et al., 2018; Ruiz et al., 2018; Sandín et al., 2009; Valencia & Paredes-Angeles, 2021). On the other hand, in the present study, the final models of the three versions included correlated errors, which is considered an undesirable psychometric characteristic (Dominguez-Lara, 2019). However, it is important to identify and monitor this inter-item dependence, otherwise, the internal consistency reliability estimates will be

Table 3
Measurement invariance of the brief versions of the PSWQ by sex and age.

Groups	Version	Model	χ^2	*df*	CFI	$\Delta\chi^2$	Δ*df*	*p*	ΔCFI
Females vs. Males	PSWQ-11	Configural	390.02	84	.96				
		Metric	409.73	94	.96	9.97	10	.443	0
		Scalar	444.49	104	.96	32.40	10	<.001	-.002
		Strict	479.48	115	.96	36.54	11	<.001	-.003
	PSWQ-A	Configural	157.14	36	.98				
		Metric	168.23	43	.98	5.17	7	.639	0
		Scalar	198.38	50	.97	30.73	7	<.001	-.003
		Strict	229.75	58	.97	31.43	8	<.001	-.005
	PSWQ-5	Configural	16.47	8	1.00				
		Metric	22.92	12	1.00	5.72	4	.221	0
		Scalar	25.33	16	1.00	1.88	4	.758	.001
		Strict	45.32	21	.99	18.22	5	.003	-.006
Age < 30 vs. Age ≥30	PSWQ-11	Configural	398.59	84	.96				
		Metric	422.63	94	.96	17.3	10	.069	-.001
		Scalar	450.33	104	.96	23.47	10	.009	-.001
		Strict	523.05	115	.95	69.38	11	<.001	-.008
	PSWQ-A	Configural	159.33	36	.98				
		Metric	173.49	43	.98	10.09	7	.183	0
		Scalar	191.15	50	.98	15.55	7	.030	-.001
		Strict	236.85	58	.97	43.61	8	<.001	-.008
	PSWQ-5	Configural	21.41	8	1.00				
		Metric	26.10	12	1.00	3.57	4	.467	0
		Scalar	32.86	16	.99	6.51	4	.164	-.001
		Strict	72.66	21	.98	36.52	5	<.001	-.013

Note. The CFI values correspond to the robust coefficient proposed by Brosseau-Liard and Savalei (2014).

Table 4
Correlations between the three brief versions of the PSWQ and two measures of symptomatology.

Brief version	Depression (BDI-II)	Anxiety (BAI)
PSWQ-11	.54 [.50, .58]***	.49 [.45, 53]***
PSWQ-A	.54 [.50, .58]***	.48 [.44, .52]***
PSWQ-5	.53 [.49, .56]***	.46 [.42, .50]***

Note. Sample sizes were 1323 for depression and 1327 for anxiety.
****p* < .001.

biased (Viladrich et al., 2017). It should be noted that, in a previous study, covariation was also observed between the residuals of items 7 (*You have been a worrier all your life*) and 8 (*You notice that you have been worrying about things*) (Xie et al., 2023). This result seems to be explained by the fact that both statements refer to a temporal aspect of worry (i.e., chronicity). Future studies should examine whether this result replicates in similar samples to the one used in this study, and if it is confirmed, consider potential modifications to the instrument for this population.

Given the existence of three short versions that function similarly, the question arises as to which of them is preferable. The answer to this question, however, depends on each research project. When dealing with a substantial number of measures, and worry is a secondary variable in the study, researchers may opt for the shortest possible version that maintains good psychometric properties (Schetsche et al., 2022). Indeed, there has even been a proposal for a single-item version of the PSWQ (Schroder et al., 2019). On the other hand, if the number of items is not an issue or worry is the principal outcome variable, a version with more items will almost always be preferable (Petersen et al., 2023). When examining the three short versions in this study, it is important to consider that the 11- and 8-item versions demonstrated similar performance, while the 5-item version exhibited slightly lower performance in terms of internal consistency and the attenuation of its correlation with other variables. This result coincides with the findings of another study that also compared these three versions (Valencia & Paredes-Angeles, 2021).

Limitations

The present study has several limitations. First, although the objective was to examine three brief versions, these were not administered independently. In fact, only the PSWQ-11 was administered in the study, and the analyses of the PSWQ-A and PSWQ-5 were conducted by selecting the corresponding items in the database. Second, the PSWQ-16, which could be an interesting point of comparison, was not considered. Third, it is worth noting that the majority of participants (69%) had higher education, which is probably not representative of the Mexican population requiring psychological help. Fourth, all the data in this study were cross-sectional, so it was not possible to assess properties such as longitudinal invariance or test-retest reliability. Despite these limitations, this study reveals several strengths, including using a large sample of people seeking professional help (as opposed to other studies that were limited to university samples; Valencia & Paredes-Angeles, 2021).

Conclusion

The current findings demonstrated that the three short versions of the PSWQ (11-item, 8-item, and 5-item) function adequately within a sample of individuals seeking psychological help. This performance is similar in males and females and between adults younger and older than 30. Future studies should examine whether the presence of error correlations replicates in other populations. Researchers are encouraged to use the brief version of the PSWQ that best suits the needs of their projects.

References

Ato, M., López-García, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología [A classification system for research designs in psychology]. *Anales de Psicología / Annals of Psychology, 29*(3), 1038-1059. https://doi.org/10.6018/analesps.29.3.178511

Becerra Herrera, W. A., Quispe Escobar, A. R., Cárcamo-Zepeda, E., Torres Guffanti, A., & Yupanqui Lorenzo, D. E. (2023). Cuestionario de Pensamiento Automático (ATQ-8): Nueva evidencia basada en AFC, TRI y SEM [Automatic Thinking Questionnaire (ATQ-8): New evidence based on CFA, IRT, and SEM]. *Revista Evaluar, 23(*1), 51-60. https://doi.org/10.35670/1667-4545.v23.n1.41010

Beck, A. T., Steer, R. A., & Brown, G. K. (2006). *Inventario de Depresión de Beck: BDI-II*. Ediciones Paidós Ibérica S. A.

Borkovec, T. D., Ray, W. J., & Stober, J. (1998). Worry: A cognitive phenomenon intimately linked to affective, physiological, and interpersonal behavioral processes. *Cognitive Therapy and Research, 22*(6), 561-576. https://doi.org/10.1023/A:1018790003416

Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting incremental fit indices for nonnormality. Multivariate *Behavioral Research, 49*(5), 460-470. https://doi.org/10.1080/00273171.2014.933697

Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy, 41*(12), 1411-1426. https://doi.org/10.1016/S0005-7967(03)00059-7

Carbonell-Bártoli, E. L., & Tume-Zapata, V. N. (2022). Cuestionario de Preocupación de Pensilvania (PSWQ 11): Evidencias psicométricas en una muestra de adultos de Lima Metropolitana [Penn State Worry Questionnaire (PSWQ 11): Psychometric evidence in a sample of adults in Metropolitan Lima]. *PsiqueMag, 11*(2), 31-41. https://doi.org/10.18050/psiquemag.v11i2.2108

Cares, S. R., Younce, J. A., Mangen, K. H., Winder, J. R., Fergus, T. A., & Wu, K. D. (2022). I am worried: Do the PSWQ and PSWQ-A display measurement invariance across four racial groups? *Assessment, 29*(6), 1320-1330. https://doi.org/10.1177/10731911211013911

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

de la Rosa-Gómez, A., Hernández Posadas, A., Valencia, P. D., Flores-Plata, L. A., Serrano Zárate, B., Flores Elvira, A. I., Dominguez-Rodriguez, A., Vázquez Sánchez, M. F., & González Santiago, E. (2023). Online transdiagnostic intervention for emotional, trauma-and stressor-related disorders in the Mexican population: A randomized pilot and feasibility trial. *Contemporary Clinical Trials Communications, 35*, 101204. https://doi.org/10.1016/j.conctc.2023.101204

DeLapp, R. C. T., Chapman, L. K., & Williams, M. T. (2016). Psychometric properties of a brief version of the Penn State Worry Questionnaire in African Americans and European *Americans. Psychological Assessment, 28*(5), 499-508. https://doi.org/10.1037/

pas0000208

Dominguez-Lara, S. (2019). Correlación entre residuales en análisis factorial confirmatorio: Una breve guía para su uso e interpretación [Correlation between residuals in confirmatory factor analysis: A brief guide to their use and interpretation]. *Interacciones, 5*(3), e207. https://doi.org/10.24016/2019.v5n3.207

Ehring, T., & Behar, E. (2020). Transdiagnostic view on worrying and other negative mental content. In A. L. Gerlach & A. T. Gloster (Eds.), *Generalized Anxiety Disorder and Worrying: A comprenhensive handbook for clinicians and researchers* (1st ed., pp. 43-68). Wiley. https://doi.org/10.1002/9781119189909.ch4

Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A., & Muñiz, J. (2022). Decálogo para el análisis factorial de los ítems de un test [Decalogue for the factor analysis of test items]. *Psicothema, 34*(1), 7-17. https://doi.org/10.7334/psicothema2021.456

González, A., Mata, S., Lavie, R., & Resler, G. (2007). Validación del cuestionario de preocupaciones de la Universidad estatal de Pensilvania (PSWQ). *Revista de La Facultad de Medicina, 30*(1), 38-42.

Hopko, D. R., Reas, D. L., Beck, J. G., Stanley, M. A., Wetherell, J. L., Novy, D. M., & Averill, P. M. (2003). Assessing worry in older adults: Confirmatory factor analysis of the Penn State Worry Questionnaire and psychometric properties of an abbreviated model. *Psychological Assessment, 15(*2), 173-183. https://doi.org/10.1037/1040-3590.15.2.173

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*(6), 487-495. https://doi.org/10.1016/0005-7967(90)90135-6

Nuevo Benítez, R., Montorio Cerrato, I., & Ruiz Díaz, M. Á. (2002). Aplicabilidad del Inventario de Preocupación de Pensilvania (PSWQ) a población de edad avanzada [Application of the Penn State Worry Questionnaire (PSWQ) to elderly population]. *Ansiedad y Estrés, 8*(2-3), 157-172.

Padros-Blazquez, F., Gonzalez-Betanzos, F., Martinez-Medina, M. P., & Wagner, F. (2018). Psychometric characteristics of the original and brief version of the Penn State Worry Questionnaire (PSWQ) in Mexican samples. *Actas Españolas de Psiquiatría, 46*(4), 117-124.

Petersen, M. Aa., Vachon, H., Groenvold, M., & the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group. (2023). Development of a diverse set of standard short forms based on the EORTC CAT Core item banks. *Quality of Life Research, 32*(7), 2037-2045. https://doi.org/10.1007/s11136-023-03373-6

Puccinelli, C., Cameron, D. H., Ouellette, M. J., McCabe, R. E., & Rowa, K. (2023). Psychometric properties of the Penn State Worry Questionnaire-Past Week (PSWQ-PW) in an anxiety and related disorders sample. *Journal of Psychopathology and Behavioral Assessment, 45*(2), 549-557. https://doi.org/10.1007/s10862-023-10029-9

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*(3), 354-373. https://doi.org/10.1037/a0029315

Robles, R., Varela, R., Jurado, S., & Páez, F. (2001). Versión mexicana del Inventario de Ansiedad de Beck: Propiedades psicométricas . *Revista Mexicana de Psicología, 18*(2), 211-218.

Ruiz, F. J., Monroy-Cifuentes, A., & Suárez-Falcón, J. C. (2018). Penn State Worry Questionnaire-11 validity in Colombia and factorial equivalence across gender and nonclinical and clinical samples. *Anales de Psicología / Annals of Psychology, 34*(3), 451-457. https://doi.org/10.6018/analesps.34.3.300281

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 Penn State Worry Questionnaire: Factor structure and psychometric properties]. *Revista de Psicopatología y Psicología Clínica, 14(*2), 107-122. https://doi.org/10.5944/rppc.vol.14.num.2.2009.4070

Schetsche, C., Jaume, L. C., & Azzollini, S. (2022). Desarrollo de una versión breve del Coping Strategies Inventory [Development of a short version of the Coping Strategies Inventory]. *Revista Evaluar, 22(*1), 01-16. https://doi.org/10.35670/1667-4545.v22.n1

Schroder, H. S., Clark, D. A., & Moser, J. S. (2019). Screening for problematic worry in adults with a single item from the Penn State Worry Questionnaire. *Assessment, 26*(2), 336-346. https://doi.org/10.1177/1073191117694453

Topper, M., Emmelkamp, P. M. G., Watkins, E. R., & Ehring, T. (2014). Development and assessment of brief versions of the Penn State Worry Questionnaire and the Ruminative Response Scale. *British Journal of Clinical Psychology, 53*(4), 402-421. https://doi.org/10.1111/bjc.12052

Valencia, P. D., & Paredes-Angeles, R. (2021). Dimensionality of the Penn State Worry Questionnaire and its brief versions. *Revista Argentina de Ciencias Del Comportamiento, 13*(2), 38-51. https://doi.org/10.32348/1852.4206.v13.n2.27395

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna [A journey around alpha and omega to estimate internal consistency reliability]. *Anales de Psicología, 33*(3), 755-782. https://doi.org/10.6018/analesps.33.3.268401

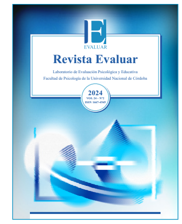
Wu, S. M., Schuler, T. A., Edwards, M. C., Yang, H.-C., & Brothers, B. M. (2013). Factor analytic and item response theory evaluation of the Penn State Worry Questionnaire in women with cancer. *Quality of Life Research, 22*(6), 1441-1449. https://doi.org/10.1007/s11136-012-0253-0

Wuthrich, V. M., Johnco, C., & Knight, A. (2014). Comparison of the Penn State Worry Questionnaire (PSWQ) and abbreviated version (PSWQ-A) in a clinical and non-clinical population of older adults. *Journal of Anxiety Disorders, 28(*7), 657-663. https://doi.org/10.1016/J.JANXDIS.2014.07.005

Xie, S.-S., Xiao, H.-W., & Lin, R.-M. (2023). Abbreviated version of Penn State Worry Questionnaire for Chinese adolescents: Age, gender and longitudinal invariance. Frontiers in Psychiatry, 14, 1086592. https://doi.org/10.3389/fpsyt.2023.1086592

2024, Vol. 24, No. 2
ISSN 1667-4545
Recuperado de https://revistas.unc.edu.ar/index.php/revaluar

Laboratorio de Evaluación Psicológica y Educativa
Facultad de Psicología - Universidad Nacional de Córdoba



Propiedades Psicométricas del Inventario de Pensamiento Dicotómico (DTI) en español

Psychometric Properties of the Spanish version of the Dichotomous Thinking Inventory (DTI)

Juan Ignacio, López * [1], Santiago Resett [2],
Pablo Christian González-Caino [3], Atsushi Oshio [4]

1 - Graduate School of Letters, Arts and Sciences, Waseda University, Tokyo, Japan.
2 - Universidad Austral, CONICET.
3 - Universidad Argentina de la Empresa.
4 - Faculty of Letters, Arts and Sciences, Waseda University, Tokyo, Japan.

Recibido: 04/01/2024 **Revisado:** 14/02/2024 **Aceptado:** 21/03/2024

Resumen

El pensamiento dicotómico es una distorsión cognitiva que se caracteriza por emplear categorías extremas para clasificar situaciones o personas. El presente estudio tuvo como objetivo adaptar el Inventario de Pensamiento Dicotómico (DTI) en una muestra de adultos argentinos. Se constituyó una muestra intencional (n = 470) residentes en el Gran Buenos Aires y la Ciudad Autónoma de Buenos Aires, de entre 18 y 63 años (M = 30.71, DE = 10.76), de los cuales 367 eran mujeres. Los resultados del análisis factorial exploratorio y confirmatorio indicaron una estructura de dos dimensiones con adecuada confiabilidad (α = .78 para creencias dicotómicas, α = .72 para preferencia por la dicotomía y α = .82 para el puntaje total). Se encontraron evidencias de validez convergente y discriminante y se identificaron diferencias de género, lo que mostró niveles más altos de pensamiento dicotómico en hombres. La adaptación al español sugiere adecuadas propiedades psicométricas.

Palabras clave: *pensamiento dicotómico, propiedades, psicometría, DTI, estructura factorial*

Abstract

Dichotomous thinking is a cognitive distortion characterized by the use of extreme categories to classify situations or people. This study aimed to adapt the Dichotomous Thinking Inventory (DTI) in a sample of Argentine adults. An intentional sample was constituted (n = 470) residing in Greater Buenos Aires and the Autonomous City of Buenos Aires, between 18 to 63 years (M = 30.71, SD = 10.76), of which 367 were women. Results from exploratory and confirmatory factor analyses indicated a two-dimensional structure with adequate reliability (α = .78 for dichotomous beliefs, α = .72 for preference for dichotomy, and α = .82 for the total score). Evidence of convergent and discriminant validity was found, and gender differences were identified, showing higher levels of dichotomous thinking in men. The adaptation to Spanish suggests adequate psychometric properties.

Keywords: *dichotomous thinking, properties, psychometrics, DTI, factor structure*

***Correspondencia a:** Juan Ignacio López, Graduate School of Letters, Arts and Sciences, Waseda University, 1-24-1 Toyama, Shinjuku, Tokyo, 162-8644, Japan. Email: juanii.lopez@gmail.com.

How to cite: López, I., J., Resett, S., Caino-González, C., & Oshio, A., (2024): Propiedades Psicométricas del Inventario de Pensamiento Dicotómico (DTI) en español. *Revista Evaluar, 24*(2), 13-25. Retrieved from https://revistas.unc.edu.ar/index.php/revaluar

Participaron en la edición de este artículo: Marianela López, Stefano Macri, Eugenia Barrionuevo, Florencia Ruiz, Jorge Bruera.

Introducción

Las distorsiones cognitivas refieren a patrones de pensamiento que predisponen al procesamiento de la información de acuerdo a esquemas cognitivos preexistentes (Beck, 1963). Estos esquemas cognitivos son estructuras relativamente permanentes que se desarrollan a través de la experiencia previa y guían la percepción, codificación y evaluación de estímulos. Asimismo, influyen en la toma de decisiones subsiguientes (Guglielmo, 2015). Además, Beck (1967) propone seis distorsiones cognitivas, que incluyen la inferencia arbitraria, abstracción selectiva, generalización, maximización-minimización, personalización y el pensamiento absolutista o dicotómico. Este último se caracteriza por la tendencia a clasificar las experiencias, situaciones o personas, mediante solo dos categorías extremas, tales como "todo o nada", "blanco o negro", "correcto o incorrecto", "bueno o malo", evitando la utilización de matices.

Si bien este estilo de pensamiento facilita la comprensión de información y simplifica el proceso de toma de decisiones, el acto de clasificar la información de manera binaria puede dar lugar a malentendidos entre personas con puntos de vista y opiniones divergentes (Oshio, 2009). De igual modo, dado su papel en la resolución rápida de situaciones, el pensamiento dicotómico se ha considerado útil para la supervivencia (Bodenhausen et al., 2012). Además, se ha observado que las tendencias de pensamiento dicotómico suelen manifestarse con mayor frecuencia en ambientes hostiles, a la vez que se relacionan con ciertas habilidades cognitivas (Mieda et al., 2021). Dada esta combinación de factores, no es de extrañar que las creencias dicotómicas se hayan relacionado, en particular, con comportamientos impulsivos y agresivos (Oshio et al., 2016). A su vez, se considera un sesgo cognitivo necesario para aquellos que siguen estilos de vida parasitarios y explotadores, orientados hacia la satisfacción inmediata de metas a expensas de los resultados a largo plazo (Jonason et al., 2018).

Al clasificar repetidamente la realidad en términos binarios, las decisiones que se toman pueden derivar de una percepción distorsionada de la realidad (Bonfá-Araujo et al., 2022). Por lo tanto, el pensamiento dicotómico ha sido identificado como un factor de riesgo en los trastornos de personalidad, y se constituye como una característica presente en los estilos de pensamiento de dichos trastornos a lo largo de todos los grupos (Beck et al., 2004; Oshio, 2012; Veen & Arntz, 2000).

Asimismo, existen dos conceptos estrechamente relacionados con el pensamiento dicotómico, pero que deben diferenciarse: la intolerancia a la ambigüedad (Budner, 1962) y la intolerancia a la incertidumbre (Dugas et al., 2001). Ambos se definen como la tendencia a percibir e interpretar una situación como una amenaza o una fuente de malestar o ansiedad, lo que resulta en una serie de respuestas cognitivas, emocionales y conductuales (Grenier et al., 2005). La principal diferencia entre ellos radica en su orientación temporal. La intolerancia a la ambigüedad se manifiesta como la incapacidad de tolerar estímulos o situaciones en el tiempo presente, mientras que la intolerancia a la incertidumbre se refiere a la aprehensión de eventos futuros negativos (Dugas et al., 2001). Por consiguiente, tanto la intolerancia a la ambigüedad como la intolerancia a la incertidumbre no solo están relacionadas con un estilo de pensamiento dicotómico, sino que también involucran respuestas emocionales y conductuales frente a estímulos o situaciones inciertas o ambiguas (Oshio, 2009).

En la actualidad, existe solo un instrumento que evalúa exclusivamente el estilo de pensamiento dicotómico: el *Dichotomous Thinking*

Inventory (DTI; Oshio, 2009). Este instrumento consta de tres dimensiones: preferencia por la dicotomía, creencias dicotómicas y pensamiento ganancia-pérdida. La primera dimensión se refiere a la tendencia a mostrar una preferencia por la distinción, la claridad y la precisión en la clasificación de información o personas, en lugar de tolerar la ambigüedad y la vaguedad en estas situaciones. Por otro lado, los individuos con altos niveles de creencias dicotómicas piensan que al juzgar, categorizar o dividir la información del entorno siempre debe hacerse en dos categorías (por ejemplo, bueno o malo, correcto o incorrecto). Finalmente, el pensamiento ganancia-pérdida se caracteriza por evaluar la ventaja o desventaja de una situación y la tendencia a evitar situaciones que puedan traer resultados negativos. Este instrumento ha demostrado tener buenas propiedades psicométricas y ha sido utilizado para evaluar el pensamiento dicotómico en diversos estudios a nivel mundial. Sin embargo, hasta la fecha, no existe una versión en español que permita estudiar este constructo en países hispanohablantes. Por ende, la fortaleza del presente estudio es ser el primero en evaluar sus propiedades psicométricas en una muestra de habla española.

Con respecto a los estudios sobre sus propiedades psicométricas, a pesar de que el inventario de pensamiento dicotómico ha sido utilizado en diversos idiomas y poblaciones, únicamente las versiones originales en inglés y japonés de la escala se han sometido a un análisis de su estructura factorial (Oshio, 2009; Oshio, 2010). Al analizar los estudios que utilizaron este instrumento relevados por Bonfá-Araujo et al. (2022), la confiabilidad total del instrumento osciló entre α = .80 y α = .84, y la dimensión creencias dicotómicas presentó los valores más altos de alpha de Cronbach, entre .77 y .82. Preferencia por la dicotomía mostró valores de .72 a .76, mientras que para el factor pensamiento ganancia-pérdida, se registraron valores entre α = .71 y α = .79 (Jonason et al., 2018; Mieda et al., 2021; Namatame et al., 2015; Palascha et al., 2015; Ueno et al., 2017). No obstante, la versión rusa del instrumento presentó los valores más bajos, de .79, .51 y .54 para las tres dimensiones respectivamente (Oshio & Meshkova, 2012).

Por lo tanto, considerando lo mencionado anteriormente y la importancia psicosocial del pensamiento dicotómico, esta investigación tuvo los siguientes objetivos:

- Explorar la estructura factorial del Inventario de Pensamiento Dicotómico (DTI; Oshio, 2009) y su consistencia interna en una muestra de adultos argentinos.
- Examinar su validez concurrente y discriminante con el perfeccionismo y la tolerancia a la ambigüedad.
- Determinar si existen diferencias en el estilo de pensamiento dicotómico según la edad y el género.

Método

Participantes

La muestra intencional no probabilística utilizada para el presente estudio estuvo compuesta por un total de 470 participantes, residentes en el Gran Buenos Aires y la Ciudad Autónoma de Buenos Aires, Argentina, de los cuales 367 fueron mujeres (73.5%) y 103, hombres (26.5%). La edad de los participantes osciló entre 18 y 63 años, con una media de M = 30.71 y un desvío estándar de DE = 10.76. Los criterios de inclusión para el presente estudio fueron que los participantes tuvieran entre 18 y 65 años, fueran residentes del Gran Buenos Aires o Ciudad Autónoma de Buenos Aires y tuvieran el nivel educativo secundario completo.

Instrumentos

Cuestionario socio-demográfico. El mismo incluía preguntas que indagaban sobre variables sociodemográficas como edad, género y lugar de residencia.

Inventario de Pensamiento Dicotómico (DTI; Oshio, 2009). Este instrumento está compuesto por 15 afirmaciones que deben responderse por medio de una escala Likert de seis puntos (1 = *fuertemente en desacuerdo*, 6 = *fuertemente de acuerdo*). Evalúa los niveles de pensamiento dicotómico en tres subescalas: preferencia por la dicotomía (ítems 1, 4, 7, 10, 13), creencias dicotómicas (ítems 2, 5, 8, 11, 14) y pensamiento ganancia-pérdida (ítems 3, 6, 9, 12, 15). Los puntajes totales de cada factor se obtienen sumando los ítems correspondientes a cada uno de ellos, a la vez que el puntaje total del inventario se obtiene sumando los puntajes obtenidos en las 3 dimensiones. La confiabilidad total de la escala original fue de α = .84, con α = .81 para preferencia por la dicotomía α = .74 para creencias dicotómicas y α = .75 para pensamiento ganancia-pérdida. La validez del constructo en su publicación original se evidenció frente a la intolerancia a la ambigüedad y al perfeccionismo. La adaptación del instrumento al español argentino se realizó utilizando un método de traducción inversa y siguiendo recomendaciones internacionales (International Test Commission, 2017; Muñiz et al., 2013). Dos traductores profesionales en inglés llevaron a cabo la traducción del documento original al español. Posteriormente, en colaboración con los autores del presente estudio, se evaluó la equivalencia de ambas versiones y se realizó una revisión sistemática de cada ítem hasta lograr un consenso. Luego, esta versión fue traducida nuevamente al inglés por otro traductor (un hablante nativo de inglés con habilidades en español) y se llevó a cabo una comparación entre la versión original en inglés y la traducida para asegurar la equivalencia, tanto a nivel semántico como conceptual.

Medida Multidimensional de Perfeccionismo (APS-R; Slaney et al., 2001). Este instrumento está integrado por 23 afirmaciones que se responden a través de una escala Likert de 7 puntos (1 = *fuertemente en desacuerdo*, 7 = *fuertemente de acuerdo*). La escala evalúa el nivel de perfeccionismo por medio de un modelo de tres factores compuesto por: altos estándares (ítems 1, 5, 8, 12, 14, 18, 22), orden (ítems 2, 4, 7, 10) y discrepancia (ítems 3, 6, 9, 11, 13, 15, 16, 17, 19, 20, 21, 23). Este instrumento demostró buenas propiedades psicométricas en su adaptación Argentina (Arana et al., 2009), tanto de confiabilidad, como de validez. En la muestra estudiada, la escala presentó una confiabilidad de α = .68 para la dimensión altos estándares, de α = .74 para orden, α = .92 en su factor discrepancia y una confiabilidad total de α = .86.

Escala de Tolerancia a la Ambigüedad Multiestímulo II (MSTAT-II; Arquero & McLain, 2010). Este instrumento tiene como objetivo evaluar los niveles de tolerancia a la ambigüedad del sujeto por medio de 13 afirmaciones que se responden con una escala Likert de 5 puntos, en la que los valores son 1= *fuertemente en desacuerdo* y 5 = *fuertemente de acuerdo*. El puntaje total obtenido indica que a una mayor puntuación se corresponde una mayor tolerancia a la ambigüedad. En el presente estudio la confiabilidad de la prueba fue de .82.

Procedimiento de recolección de datos

Los participantes se reclutaron fueron reclutados por medio a través de distintos canales de comunicación virtual y redes sociales como *Instagram* y *Facebook*. La recolección de datos se realizó a travéspor medio de la utilización de un formulario de *Google form*, en el cual los sujetos aceptaban un consentimiento informado, previo a la presentación de los formularios, que garantizaba el anonimato y la confidencialidad de las respuestas, a la vez que informaba de los requisitos necesarios para que puedan formar parte de la presente investigación. Una vez aceptado, se presentaron a los encuestados las escalas que componían este estudio. La universidad donde se realizó la investigación aprobó el estudio.

Procedimiento de análisis de datos

Los datos del presente estudio se analizaron utilizando el programa estadístico *R* Studio (versión 2023.03.0 *Build* 386) y el paquete estadístico psych. Para indagar en la estructura factorial subyacente de los ítems y luego confirmar en un análisis factorial confirmatorio, el total de la muestra se dividió de manera aleatoria en dos partes (Brown, 2006). Para determinar la cantidad de factores a extraer, se llevó a cabo un análisis paralelo con el programa Factor 8.10, mientras que para señalar la retención de factores, se empleó el método de implementación clásico de Horn (1965). Para este fin, se compararon los autovalores empíricos con los autovalores (medias) aleatorios; por consiguiente, se escogieron los que se encontraban por encima de la media aleatoria (O'Connor, 2000). Se usó un número de replicaciones igual a 450 y un percentil de representación de simulaciones igual a .95. Una vez obtenidos los factores a retener, para el análisis factorial exploratorio (AFE) se usó rotación Promin, ya que dicha rotación permite que los factores sean oblicuos, lo que facilita la simplificación de la estructura factorial en caso de que emerja más de un factor (Lorenzo-Seva, 2013). Se empleó el método *Unweighted Least Squares* (ULS), debido a la naturaleza ordinal de los datos. La adecuación de los datos se evaluó mediante la prueba de esfericidad de Bartlett y la medida de adecuación de muestreo *Kaiser-Meyer-Olkin* (KMO). El análisis factorial confirmatorio se realizó sobre una muestra compuesta de 300 participantes utilizando el paquete *lavaan*; la composición de las dos muestras se dividió aleatoriamente. Los valores de asimetría y curtosis se apartaban ligeramente de la normalidad, con valores de entre -1.69 a 1.34 y -1.19 a 2.27, respectivamente. Aunque no resulten lo suficientemente extremos, se llevó a cabo un análisis factorial confirmatorio (AFC) con métodos robustos de máxima verosimilitud para datos continuos que se alejan de la normalidad (Byrne, 2008), tanto en el modelo propuesto por el análisis factorial exploratorio como en la estructura factorial sugerida por el autor del instrumento. El ajuste de los modelos se evaluó utilizando los criterios de bondad de ajuste como el índice de ajuste comparativo (CFI), el índice Tucker-Lewis (TLI) y el error cuadrático medio de aproximación de la raíz (RMSEA). Se consideraron como adecuados valores de CFI y TLI superiores a .90 y un RMSEA inferior a .10 (Bentler, 1992). En el proceso de ajuste del modelo, no se consideró la significancia del X^2, debido a que este criterio puede ser exigente y está influenciado por el tamaño de la muestra (Byrne, 2012). Sin embargo, es posible dividir el X^2 por los grados de libertad, considerando valores entre 2 y 3, o incluso hasta 5, como aceptables (Cupani, 2012). Para este análisis, se puso a prueba el modelo surgido en el análisis factorial exploratorio y el modelo de tres factores de los autores del instrumento. Por último, para analizar las diferencias según género y edad, se utilizaron las pruebas *t* de Student de muestras independientes y ANOVA de un factor, respectivamente. Se evaluó la validez concurrente y discriminante del instrumento mediante el coeficiente de correlación de Pearson, calculando correlaciones positivas o negativas entre el Inventario de Pensamiento Dicotómico, la Medida Multidimensional de Perfeccionismo (APS-R) y la Escala de Tolerancia a la Ambigüedad Multiestímulo II (MSTAT-II), respectivamente. Para este estudio, se estableció un nivel de significancia de .05, como es habitual en investigaciones en ciencias sociales (Labovitz, 1968).

Resultados

Resultados del Análisis Factorial Exploratorio del DTI en una muestra de adultos argentinos

En primer lugar, se realizó un análisis paralelo con el objetivo de detectar el número adecuado de factores a retener del Inventario de Pensamiento Dicotómico (DTI) sobre la muestra de $n = 170$.

En la Tabla 1, se muestran los autovalores empíricos emergidos del análisis. Al compararlos, solamente los dos primeros autovalores empíricos se hallaban por encima de la media aleatoria, por lo cual, se decidió retener solamente dos factores, como lo sugiere O'Connor (2000).

A partir de lo emergido en el análisis paralelo, se aplicó un criterio de solicitar dos factores para el AFE (Tabla 2), y se obtuvo/el resultado fue una estructura factorial que explicaba el 37% de la varianza. El factor "creencias dicotómicas" explicó un 19%, mientras que el factor "preferencia por la dicotomía" explicó un 18%. La prueba de esfericidad de Bartlett indicó resultados significativos con un X^2 (105) = 617.50, $p < .001$. A su vez, el indicador de adecuación del tamaño de la muestra *Kaiser-Meyer-Olkin* demostró ser adecuado (.82). Como se muestra en la Tabla 2, no existían cargas cruzadas mayores a .35 y todos las preguntas cargaban por encima de .36 en su respectivo factor. Solamente los ítems 1 y 6 no cargaban por encima de .35 en ninguno de los dos factores. Posteriormente, se realizó un segundo AFE para reevaluar la estructura factorial del instrumento y se eliminaron los ítems 1 y 6 por su baja carga (Tabla 3). Los resultados mostraron una adecuación significativa del modelo X^2 (78) = 493.56, $p < .001$ y un KMO total de .82.

Resultados del análisis factorial confirmatorio del DTI en una muestra de adultos argentinos

Al poner a prueba el modelo sugerido por el AFE, se excluyeron los ítems 1 y 6 debido a que no alcanzaron el punto de corte establecido para su carga factorial de .35 (Kline, 2015; Matsunaga, 2010). En la Tabla 4 se pueden observar los resultados del análisis factorial confirmatorio. El ajuste del modelo propuesto por el AFE (Modelo 1) resultó adecuado con un CFI de .923, TLI .906 y RMSEA = .055. Los resultados del Modelo 2 (Oshio, 2012) indicaron un ajuste levemente menos aceptable de RMSEA = .052, CFI = .901 y TLI = 881. Por dicho motivo, la validez se analizó con el modelo 1.

Tabla 1
Análisis paralelo del DTI en una muestra de adultos argentinos.

	Datos reales	Media	Percentil
1	3.29	.57	.68
2	1.12	.46	.53
3	0.43	.37	.45
4	0.24	.29	.36
5	0.19	.22	.28
6	0.17	.16	.21
7	0.03	.09	.15
8	-0.02	.05	.09
9	-0.05	-.01	.03
10	-0.08	-.06	-.03
11	-0.15	-.11	-.08
12	-0.17	-.16	-.13
13	-0.21	-.23	-.18
14	-0.26	-.27	-.23
15	-0.29	-.33	-.26

Tabla 2
Análisis factorial exploratorio del DTI en una muestra de adultos argentinos.

	Factor1	Factor2
Item 1		
Item 2	.61	
Item 3		.43
Item 4		.51
Item 5	.80	
Item 6		
Item 7		.51
Item 8	.55	
Item 9		.69
Item 10		.43
Item 11	.56	
Item 12	.36	
Item 13		.60
Item 14	.78	
Item 15		.59

Nota. Se omiten las cargas factoriales menores a .35.

Tabla 3
Segundo análisis factorial exploratorio del DTI sin preguntas 1 y 6 en una muestra de adultos argentinos.

	Factor1	Factor2
Item 2	.59	
Item 3		.43
Item 4		.52
Item 5	.80	
Item 7		.50
Item 8	.53	
Item 9		.67
Item 10		.44
Item 11	.53	
Item 12	.36	
Item 13		.60
Item 14	.73	
Item 15		.58

Nota. Se omiten las cargas factoriales menores a .35.

Consistencia interna del DTI en una muestra de adultos argentinos

En el estudio actual, el Inventario de pensamiento dicotómico presentó una confiabilidad total de $\alpha = .82$, $\alpha = .72$ y $\alpha = .78$, para las dimensiones de preferencia por la dicotomía, creencias dicotómicas, respectivamente.

Validez concurrente y discriminante del DTI en una muestra de adultos argentinos

Para analizar la validez de la escala, se llevaron a cabo correlaciones de Pearson entre las dimensiones del Pensamiento Dicotómico, la tolerancia a la ambigüedad y el perfeccionismo. Como se puede observar en la Tabla 6, se encontraron correlaciones positivas estadísticamente significativas entre el perfeccionismo y el pensamiento dicotómico en todas sus dimensiones. En relación con la validez discriminante, se obtuvieron correlaciones negativas y estadísticamente significativas entre el pensamiento dicotómico y la tolerancia a la ambigüedad.

Diferencias según género y edad en los puntajes del DTI en una muestra de adultos argentinos

Se realizaron pruebas de *t* de Student para evaluar diferencias por género y análisis de varianza (ANOVAs) e investigar diferencias asociadas a grupos de edad en los niveles de pensamiento dicotómico. Los resultados revelaron diferencias estadísticamente significativas a favor de los hombres, tanto en el puntaje total de la escala ($t = 2.12$, $df = 468$, $p < .05$) como en la dimensión de creencias dicotómicas ($t = 2.19$, $df = 468$, $p < .05$). Sin embargo, no se identificaron diferencias en la dimensión "preferencia por la

Tabla 4
Índices de ajuste de los modelos del Inventario de Pensamiento Dicotómico (DTI) en una muestra de adultos argentinos.

	RMSEA	CFI	TLI	X^2	*df*	X^2/df	*p*
Modelo 1 (2 factores)	.055	.932	.900	136.695	64	2.135	.001
Modelo 2 (3 factores) (Oshio, 2012)	.052	.901	.881	156.743	87	1.801	.001

Tabla 5
Análisis factorial confirmatorio del Inventario de Pensamiento Dicotómico (DTI) en una muestra de adultos argentinos.

	Creencias dicotómicas (α=.78; ω=.76)	Preferencia por la dicotomía (α=.72; ω=.70)
Ítem 2	.87	_
Ítem 5	.88	_
Ítem 8	.75	_
Ítem 11	.86	_
Ítem 12	.74	_
Ítem 14	.72	_
Ítem 3	_	.52
Ítem 4	_	.58
Ítem 7	_	.73
Ítem 9	_	.48
Ítem 10	_	.76
Ítem 13	_	.57
Ítem 15	_	.51
Covarianza entre factores		
Creencias dicotómicascicotómicas	1	
Preferencias por la dicotomía	.49**	1

Nota. ** $p < .01$.

Tabla 6
Estadística descriptiva y coeficientes de correlación entre perfeccionismo, tolerancia a la ambigüedad, y pensamiento dicotómico.

	M (DE)	1	2	3	4
1. Perfeccionismo.	101.18 (16.77)				
2. Tolerancia a la ambigüedad.	41.24 (7.22)	-.31**			
3. Preferencia por la Dicotomía.	31.95 (4.81)	.33**	-.37**		
4. Creencias Dicotómicas.	16.30 (5.80)	.37**	-.32**	.48**	
5. Total Pensamiento Dicotómico.	48.26 (9.14)	.41**	-.40**	.83**	.89**

Nota. ** $p < .01$.

dicotomía". En cuanto a las diferencias según la edad, la muestra se dividió en tres grupos: adultos jóvenes (18 a 29 años), adultos de mediana edad (30 a 59 años) y adultos mayores (mayores de 60 años) y se realizó una serie de ANOVAs. De acuerdo a los resultados obtenidos en las pruebas de Tukey, no se evidenciaron diferencias significativas entre las medias de las puntuaciones ($p >$ .05).

Discusión

El presente estudio tuvo como objetivo principal abordar la estructura factorial, la consistencia interna y la validez del Inventario de Pensamiento Dicotómico (DTI; Oshio, 2009), en nuestra muestra de adultos argentinos.

Con relación a su estructura factorial, los análisis exploratorios y confirmatorios realizados revelaron que un modelo de dos factores presentó un índice satisfactorio en comparación con la estructura de tres factores de la versión japonesa de la escala. En base a estos resultados, la versión argentina del Inventario de Pensamiento Dicotómico se compone de dos dimensiones: creencias dicotómicas (ítems 2, 5, 8, 11, 12, 14) y preferencia por la dicotomía (ítems 3, 4, 7, 9, 10, 13, 15). Excluyendo los ítems cuyas cargas factoriales no resultaron adecuadas, la composición de la dimensión creencias dicotómicas se mantuvo similar a la escala original (con la diferencia del ítem 12: *Prefiero clasificar toda información como útil o inútil para mí*, que en su versión japonesa pertenece a la dimensión pensamiento ganancia-pérdida). A su vez, la dimensión "preferencia por la dicotomía" derivada de este estudio cuenta con tres ítems que integran la dimensión pensamiento ganancia-pérdida en su versión japonesa (ítems 3: *Quiero poder distinguir claramente lo que es seguro de lo que es peligroso*, 9: *Me gusta tener en claro cuáles cosas son beneficiosas para mí y cuáles no*, 15: *Es mejor cuando las competencias tienen resultados claros*). Las diferencias culturales en la interpretación de conceptos psicológicos pueden haber implicado diferencias en la estructura factorial de la escala (Orcan, 2018; van de Vijver & Tanzer, 2004), principalmente en una cultura oriental como la japonesa. Además, las diferencias también podrían explicarse por la muestra intencional utilizada, lo que sugiere la necesidad de realizar más investigaciones al respecto.

Con relación a la consistencia interna del instrumento, la dimensión creencias dicotómicas presentó un $\alpha = .78$, la dimensión preferencia por la dicotomía demostró un $\alpha = .72$ y la escala en su conjunto mantuvo un valor de $\alpha = .82$. Si se considera que un índice entre .70 y .80 es una estimación adecuada de la consistencia interna (DeVellis, 2012; Kaplan & Saccuzzo, 2006), la escala presentó valores adecuados de consistencia interna en la población estudiada. Los resultados obtenidos en este estudio muestran similitudes con los hallazgos informados en otras investigaciones que aplicaron la misma escala en distintas poblaciones de interés (Jonason et al., 2018; Mieda et al., 2021; Oshio, 2009).

El segundo objetivo del presente estudio consistió en evaluar la validez de la adaptación del Inventario de Pensamiento Dicotómico para su aplicación en la población argentina. En relación a la validez concurrente, se encontraron correlaciones positivas y estadísticamente significativas entre todas las dimensiones del Inventario de Pensamiento Dicotómico y el perfeccionismo. Estos hallazgos concuerdan con los resultados previamente reportados por Oshio (2009) durante el desarrollo del instrumento en su versión japonesa. Al interpretar el entorno en términos de categorías extremas, como *correcto* o *incorrecto*, el pensamiento dicotómico puede propiciar la rigidez cognitiva y fomentar la búsqueda constante de estándares, ya sean inalcanzables o realistas. De este modo, los resultados obtenidos en el presente estudio apoyan la relación entre el pensamiento dicotómico y la configuración de actitudes perfeccionistas (Stoeber, 2018). A su vez, las correlaciones negativas significativas entre la versión argentina del Inventario de Pensamiento Dicotómico y la tolerancia a la ambigüedad apoyan la validez discriminante del instrumento. En consecuencia, individuos con tendencias de pensamiento absolutista tienden a rechazar estímulos ambiguos en diversas situaciones, lo cual se posiciona como un elemento cognitivo fundamental en la intolerancia a la ambigüedad. Este fenómeno refleja la preferencia por respuestas definitivas a preguntas, en contraposición a tolerar la incertidumbre, la confusión o la ambigüedad (Lauriola et al., 2016).

En relación a las diferencias de género, los resultados revelaron diferencias a favor de los hombres, quienes exhibieron niveles más altos de pensamiento dicotómico en comparación con las mujeres. Estos hallazgos se alinean con investigaciones previas, como los obtenidos por Nguyen (2020), donde las tendencias a creencias dicotómicas se asociaron positivamente con creencias de inflexibilidad, más altas en la población masculina. Contrariamente a los hallazgos de investigaciones anteriores, que señalaron que la edad modera los efectos del pensamiento dicotómico o sugirieron que un constructo estrechamente vinculado con dicho tipo de pensamiento, como el perfeccionismo, tiende a disminuir con el paso de los años, en el presente estudio no se observaron diferencias significativas en los niveles de pensamiento dicotómico en función de la edad de los participantes (Dibb-Smith et al., 2019; Oshio et al., 2016).

En conclusión, los resultados del presente estudio sugieren que el Inventario de Pensamiento Dicotómico mantiene propiedades psicométricas adecuadas en una muestra de adultos argentinos, aunque destaca una estructura factorial diferente, compuesta por dos factores. Se reconocen ciertas limitaciones, a pesar del significativo aporte que la adaptación del Inventario de Pensamiento Dicotómico al español brinda al estudio de este constructo en la población argentina. La elección de un método de recolección de datos intencional, llevado a cabo de manera *online* y limitado a residentes del Gran Buenos Aires y Ciudad Autónoma de Buenos Aires, Argentina, no asegura la validez intercultural de la escala. La estructura factorial final del instrumento podría haberse visto afectada por la composición demográfica de la muestra (Tabachnick & Fidell, 2013). Aunque se examinaron diferencias de género y edad, no se examinó la invarianza métrica a este respecto. Otra limitación es que la ausencia de adaptaciones culturales y lingüísticas previas en español impidió la comparación de nuestra versión de la escala con otras poblaciones hispanohablantes. Por otra parte, todos los datos se recogieron con el autoinforme, el cual tiene conocidas limitaciones. Finalmente, los datos se recogieron de forma *online*, lo cual puede introducir sesgos en las respuestas al no estar presente un responsable de la investigación para evacuar dudas de los participantes. Para abordar estas limitaciones, se sugiere que los futuros estudios expandan sus muestras a diversos países hispanohablantes recogidas en forma aleatoria y empleando otras técnicas de recolección de datos, además del autoinforme. También se debería controlar mediante grupos focales aquellas preguntas que tuvieron baja carga factorial para explorar el significado que los sujetos le dan a las mismas. Asimismo, se alienta a explorar el uso de otras técnicas de análisis de datos, como el modelo factorial dinámico y comparaciones transculturales, como una posible vía para profundizar en la comprensión de la estructura subyacente del instrumento, así como para evaluar su consistencia y la estabilidad temporal del constructo estudiado.

Finalmente, a pesar de las limitaciones mencionadas, la versión propuesta por esta investigación del Inventario de Pensamiento Dicotómico presenta evidencia de niveles adecuados de confiabilidad y validez. Ya que se lo considera una herramienta válida, confiable y breve, y dada la relevancia psicosocial de este constructo, futuras investigaciones podrían explorar su aplicación en diversos contextos, tanto en el ámbito clínico como en la investigación académica.

Referencias

Arana, F. G., Keegan, E., & Rutsztein, G. (2009). Adaptación de una medida multidimensional de perfeccionismo: La Almost Perfect Scale-Revised (APS-R). Un estudio preliminar sobre sus propiedades psicométricas en una muestra de estudiantes universitarios argentinos. *Revista Evaluar, 9*(1), 35-53. https://doi.org/10.35670/1667-4545.v9.n1.463

Arquero, J. L., & McLain, D. L. (2010). Preliminary validation of the Spanish version of the Multiple Stimulus Types Ambiguity Tolerance Scale (MSTAT-II). *The Spanish Journal of Psychology, 13*(1), 476-484. https://doi.org/10.1017/s1138741600004029

Beck, A. T. (1963). Thinking and depression: I. Idiosyncratic content and cognitive distortions. *Archives of General Psychiatry, 9*(4), 324-333. https://doi.org/10.1001/archpsyc.1963.01720160014002

Beck, A. T. (1967). *Depression: Clinical, experimental and theoretical aspects*. Harper and Row.

Beck, A. T., Freeman, A., & Davis, D. D. (2004). *Cognitive therapy of personality disorders* (2nd ed.).The Guilford Press.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin.. *Psychological Bulletin, 112*(3), 400-404. https://doi.org/10.1037/0033-2909.112.3.400

Bodenhausen, G. V., Kang, S. K., & Peery, D. (2012). Social categorization and the perception of social groups. In S. T. Fiske & C. N. Macrae (Eds.), *The SAGE Handbook of Social Cognition* (pp. 311-329). Sage. https://doi.org/10.4135/9781446247631

Bonfá-Araujo, B., Oshio, A., & Hauck-Filho, N. (2022). Seeing things in black and white: A scoping review on dichotomous thinking style. *Japanese Psychological Research, 64*(4), 461-472. https://doi.org/10.1111/jpr.12328

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.

Budner, S. (1962). Intolerance of ambiguity as a personality variable. *Journal of Personality, 30*(1), 29-50. https://doi.org/10.1111/j.1467-6494.1962.tb02303.x

Byrne, B. M. (2008). Factor analysis: Confirmatory. *Wiley Encyclopedia of Clinical Trials*, 1-7. https://doi.org/10.1002/9780471462422.eoct946

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge. https://doi.org/10.4324/9780203807644

Cupani, M. (2012). Análisis de ecuaciones estructurales: Conceptos, etapas de desarrollo y un ejemplo de aplicación. *Revista Tesis, 2*, 186-199. https://rdu.unc.edu.ar/handle/11086/22039

DeVellis, R. (2012). *Scale development: Theory and applications*. Sage.

Dibb-Smith, A. E., Brindal, E., Chapman, J., & Noakes, M. (2019). A mixed-methods investigation of psychological factors relevant to weight maintenance. *Journal of Health Psychology, 24*(4), 440-452. https://doi.org/10.1177/1359105316678053

Dugas, M. J., Gosselin, P., & Ladouceur, R. (2001). Intolerance of uncertainty and worry: Investigating specificity in a nonclinical sample. *Cognitive Therapy and Research, 25*(5), 551-558. https://doi.org/10.1023/A:1005553414688

Grenier, S., Barrette, A. -M., & Ladouceur, R. (2005). Intolerance of uncertainty and intolerance of ambiguity: Similarities and differences. *Personality and Individual Differences, 39*(3), 593-600. https://doi.org/10.1016/j.paid.2005.02.014

Guglielmo, S. S. (2015). Cognitive distortion: Propositions and possible worlds. *Journal of Rational-Emotive & Cognitive-Behavior Therapy, 33*(1), 53-77. https://doi.org/10.1007/s10942-014-0202-7

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185. https://doi.org/10.1007/bf02289447

International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (2nd ed.). https://www.intestcom.org

Jonason, P. K., Oshio, A., Shimotsukasa, T., Mieda, T., Csathó, A., & Sitnikova, M. (2018). Seeing the world in black or white: The Dark Triad traits and dichotomous thinking. *Personality and Individual Differences, 120*, 102-106. https://doi.org/10.1016/j.paid.2017.08.030

Kaplan, R. M., & Saccuzzo, D. P. (2006). *Pruebas psicológicas. Principios, aplicaciones y temas* (6ta ed.). Thomson.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford.

Lauriola, M., Foschi, R., Mosca, O., & Weller, J. (2016). Attitude toward ambiguity: Empirically robust factors in self-report personality scales. *Assessment, 23*(3), 353-373. https://doi.org/10.1177/1073191115577188

Labovitz, S. (1968). Criteria for selecting a significance level: A note on the sacredness of .05. *The American Sociologist, 3*(3), 220-222. http://www.jstor.org

Lorenzo-Seva, U. (2013). *Why rotate my data using Promin? (Technical Report)*. Department of Psychology, Universitat Rovira i Virgili, Tarragona. https://www.psicologia.urv.cat/ca

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. International *Journal of Psychological Research, 3*(1), 97-110. https://doi.org/10.21500/20112084.854

Mieda, T., Taku, K., & Oshio, A. (2021). Dichotomous thinking and cognitive ability. *Personality and Individual Differences, 169*, 110008. https://doi.org/10.1016/j.paid.2020.110008

Muñiz, J., Elosua, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: Segunda edición [International Test Commission Guidelines for test translation and adaptation: Second edition]. *Psicothema, 25*(2), 151-157. https://doi.org/10.7334/psicothema2013.24

Namatame, H., Ueda, H., & Sawamiya, Y. (2015). Development of emotional intelligence through stress experiences: The role of dichotomous thinking. *Journal of Health Science, 5*(2), 42.

Nguyen, N. (2020). Relationships between dichotomous thinking and other cognitive distortions. *Undergraduate Psychology Research Methods Journal, 1*(22), Art. 5. https://digitalcommons.lindenwood.edu/psych_journals

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP Test. *Behavior Research Methods, Instrumentation, and Computers, 32*(3), 396-402. https://doi.org/10.3758/bf03200807

Orcan, F. (2018). Exploratory and confirmatory factor analysis: Which one to use first? *Journal of Measurement and Evaluation in Education and Psychology, 9*(4), 414-421. https://doi.org/10.21031/epod.394323

Oshio, A. (2009). Development and validation of the Dichotomous Thinking Inventory. *Social Behavior and Personality: An International Journal, 37*(6), 729-741. https://doi.org/10.2224/sbp.2009.37.6.729

Oshio, A. (2010). Characteristics of the Dichotomous Thinking Inventory Japanese version: Review of scale development process and relations with aspects of everyday life. *Journal of the College of Humanities (Chubu University), 23*, 45-57.

Oshio, A. (2012). The relationship between dichotomous thinking and music preferences among Japanese undergraduates. *Social Behavior and Personality: An International Journal, 40*(4), 567-574. https://doi.org/10.2224/sbp.2012.40.4.567

Oshio, A., & Meshkova, T. (2012). Eating disorders, body image, and dichotomous thinking among Japanese and Russian college women. *Health, 4*(7), 392-399. https://doi.org/10.4236/health.2012.47062

Oshio, A., Mieda, T., & Taku, K. (2016). Younger people, and stronger effects of all-or-nothing thoughts on aggression: Moderating effects of age on the relationships between dichotomous thinking and aggression. *Cogent Psychology, 3*(1), 1244874. https://doi.org/10.1080/23311908.2016.1244874

Palascha, A., van Kleef, E., & van Trijp, H. C. M. (2015). How does thinking in black and white terms relate to eating behavior and weight regain? *Journal of Health Psychology, 20*(5), 638-648. https://doi.org/10.1177/1359105315573440

Slaney, R. B., Rice, K. G., Mobley, M., Trippi, J., & Ashby, J. S. (2001). The Revised Almost Perfect Scale. *Measurement and Evaluation in Counseling and Development, 34*(3), 130–145. https://doi.org/10.1080/07481756.2002.12069030

Stoeber, J. (2018). The psychology of perfectionism: An introduction. In J. Stoeber (Ed.), *The psychology of perfectionism: Theory, research, applications* (pp. 3–16). Routledge/Taylor & Francis Group.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (*6th ed.*). Pearson

Ueno, Y., Mieda, T., & Oshio, A. (2017). Correlations between factors characteristic of competitive sports and dichotomous thinking. *Journal of Health Psychology Research, 30*(1). https://doi.org/10.11560/jhpr.161124063

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology / Revue Européenne de Psychologie Appliquée, 54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Veen, G., & Arntz, A. (2000). Multidimensional dichotomous thinking characterizes borderline personality disorder. *Cognitive Therapy and Research, 24*(1), 23–45. https://doi.org/10.1023/A:1005498824175

2024, Vol. 24, No. 2
ISSN 1667-4545
Recuperado de https://revistas.unc.edu.ar/index.php/revaluar

Laboratorio de Evaluación Psicológica y Educativa
Facultad de Psicología - Universidad Nacional de Córdoba

Medida de estrés minoritario LGBT: Análisis psicométrico en una muestra venezolana

LGBT Minority Stress Measure: Psychometric Analysis in a Venezuelan Sample

Daniel Chaustre * [1] , Nickolas González [1] , Camila Khatcherian [2]

1 - Centro de Investigación y Evaluación Institucional, Universidad Católica Andrés Bello, Caracas, Venezuela.
2 - Caracas, Venezuela.

Recibido: 14/02/2024 **Revisado:** 20/04/2024 **Aceptado:** 01/05/2024

Resumen

El estrés minoritario es experimentado por personas pertenecientes a minorías estigmatizadas y marginalizadas. Aunque existen instrumentos para personas con orientaciones sexuales e identidades de género minoritarias, hay pocos en español. Se propuso realizar un análisis psicométrico del instrumento de estrés minoritario LGBT de Outland en venezolanos. La muestra no probabilística constó de 223 personas LGBTI, quienes respondieron un formulario en línea. El análisis factorial confirmatorio permitió evidenciar un ajuste adecuado entre los datos y el modelo de siete dimensiones (χ^2/gl = 2.17; *GFI* = .97; *CFI* = .98; *RMSEA* = .07). Se halló evidencia de confiabilidad (ω = .88) y validez del instrumento: para validez convergente se hallaron asociaciones significativas y positivas con depresión (r = .53) e ideación suicida (r = .50); y para validez divergente una asociación negativa con bienestar recordado (r = -.44). Esta versión en español presenta adecuadas propiedades psicométricas para ser empleada en muestras venezolanas.

Palabras clave: *traducción, estrés minoritario, personas LGBTI, confiabilidad, validez*

Abstract

Minority stress is experienced by people belonging to stigmatized and marginalized minorities. Although instruments have been developed for people with minority sexual orientations and gender identities, few are available in Spanish. It was proposed to conduct a psychometric analysis of the Outland LGBT Minority Stress Measure in Venezuelans. Through a non-probabilistic sampling, 223 LGBTI people were surveyed via an online form. Confirmatory factor analysis provided evidence of an adequate fit between the data and the seven-dimensional model (χ^2/gl = 2.17; GFI = 097; CFI = .98; RMSEA = .07). Evidence of reliability (ω = .88) and validity of the instrument was found: for convergent validity, positive associations were found with depression (r = .53) and suicidal ideation (r = .50), and a negative association with remembered well-being (r = -.44), as evidence of divergent validity. This Spanish version presents adequate psychometric properties for being used in Venezuelan samples.

Keywords: *translation, minority stress, LGBTI people, reliability, validity*

***Correspondencia a:** Daniel Chaustre, Caracas, Distrito Capital, 1020. E-mail: dchaustr@ucab.edu.ve. Teléfono: + 58 212 4074561.

Cómo citar este artículo: Chaustre, D., González, N., & Khatcherian, C., (2024): Medida de estrés minoritario LGBT: Análisis psicométrico una muestra venezolana. *Revista Evaluar 24*(2), 26-38. Recuperado de https://revistas.unc.edu.ar/index.php/revaluar

Participaron en la edición de este artículo: Débora Camponetti, Andrea Suárez, Florencia Ruiz, Jorge Bruera.

Introducción

El estrés minoritario es definido como estrés crónico experimentado por las personas que forman parte de una categoría social estigmatizada o marginalizada (Meyer, 2003; Outland, 2016). Es considerado como un estrés cualitativamente distinto a aquel originado por estresores cotidianos (Vázquez-López et al., 2020); en cambio, este surge en diversos contextos sociales como producto de los conflictos entre los valores dominantes y aquellos compartidos por grupos minoritarios (Dentato, 2012).

El estrés minoritario ha sido identificado como un factor de riesgo para la salud mental (Mongelli et al., 2019; Tan et al., 2019), también se ha vinculado con conductas perjudiciales para la salud, especialmente en grupos con identidades de género y orientaciones sexuales minoritarias (Feinstein & Dyar, 2017; Goldbach et al., 2014; Pineda-Roa, 2019). En este sentido, el estrés minoritario se ha asociado con depresión e ideación suicida (Baams et al., 2015), ansiedad (Chodzen et al., 2019; Poetar & Crit, 2024) y otros efectos negativos relacionados con el bienestar subjetivo (Conlin et al., 2019; Meyer & Frost, 2013).

A raíz de esto, se ha planteado la necesidad de elaborar instrumentos que permitan cuantificar este tipo de estrés para facilitar su estudio (Meyer & Frost, 2013). Entre las investigaciones orientadas a desarrollar instrumentos estandarizados para medir estrés minoritario se encuentra el de Outland (2016), quien elaboró un instrumento dirigido a personas lesbianas, gais, bisexuales y transexuales, usualmente denominadas como *personas LGBT*. Esta autora concibe que el estrés minoritario LGBT está conformado por siete dimensiones que se ven reflejadas en su instrumento, las cuales son: ocultamiento de la identidad, microagresiones o discriminación diaria, anticipación de rechazo, eventos de discriminación, estigma internalizado, eventos de victimización y conexión con la comunidad; estas dimensiones constituyen factores característicos del estrés minoritario.

Una de las ventajas que presenta este instrumento es que para su construcción se empleó el modelo de estrés LGB de Meyer (2003), el cual plantea que el estrés minoritario puede afectar a estas personas por dos vías: a) a través de los estresores distales o eventos de prejuicio externos y b) mediante estresores proximales o eventos internos y subjetivos. De acuerdo a este autor, los principales estresores proximales consisten en hipervigilancia, ocultamiento de la identidad y desprecio hacia sí mismo.

Dado que muchos de los instrumentos empleados en esta población tienden a centrarse principalmente en las microagresiones (Fisher et al., 2019), la propuesta de Outland (2016) constituye una herramienta útil para abordar el estudio del estrés minoritario en personas LGBT, al tomar en consideración un espectro más amplio de las dimensiones que lo conforman.

Aunque se han realizado estudios de validez con el instrumento original en otras poblaciones (Ogunbajo et al., 2020), actualmente no se dispone de una versión en español, por lo que son necesarios estudios para comprobar su adecuación en personas hispanoparlantes.

En el caso latinoamericano, durante las últimas décadas numerosos países han expandido las garantías de igualdad y no discriminación. A pesar de esto, países como Venezuela han mostrado un reducido avance en este sentido (Molina, 2022), por lo que el impacto que puede generar el estrés minoritario en personas LGBT aún requiere de investigaciones adicionales.

Dadas estas condiciones, se propuso como objetivo de investigación realizar una traducción al español y evaluar las propiedades psicométricas del instrumento de estrés minoritario LGBT de Outland (2016) en una muestra venezolana. Para determinar la validez de las mediciones hechas con este instrumento, se evaluó en conjunto variables como depresión e ideación suicida, con las cuales se esperaba encontrar una asociación positiva con estrés minoritario, así como bienestar recordado. Cabe destacar que la asociación propuesta, en este caso, es negativa.

Metodología

La investigación se puede catalogar como instrumental (Ato et al., 2013), de tipo cuantitativa y correlacional. Específicamente se considera no experimental ya que no se realizaron manipulaciones en las variables de estudio, sino que estas fueron medidas a través de cuestionarios, mientras que los datos se analizaron mediante técnicas basadas en la asociación con el fin de verificar la adecuación del instrumento *medida de estrés minoritario LGBT* en una muestra venezolana.

Participantes

La población objetivo consistió en personas LGBTI, es decir, lesbianas, gais, bisexuales, transgénero e intersexuales residenciadas en cualquier estado de Venezuela durante la recolección de datos. La muestra definitiva estuvo conformada por 223 personas, con una edad promedio de 26.9 años (S = 7.7; CV = 28.6%). Entre las variables sociodemográficas consideradas se incluyeron la orientación sexual, la identidad de género, la expresión de género, la intersexualidad y el estado de residencia. En relación con la orientación sexual, la mayor porción de las personas eran gais (51.6%), seguidas por bisexuales (27.4%), lesbianas (9.9%) y heterosexuales (5.8%), mientras que el resto reportó alguna otra orientación. Por su parte, en cuanto a la identidad de género, la mayoría de las personas reporta una identidad cisgénero masculino (52.9%), cisgénero femenino (24.2%), no binario (7.2%), *queer* (4.5%), entre otras.

De forma similar, la expresión de género más común fue la masculina (59,6%), seguida por la femenina (25.6%) y la neutra en tercer lugar (9%), mientras que otros se identificaban a sí mismos como andróginos (3.6%) o alguna expresión de género distinta. Por su parte, solo aproximadamente un 3% de los participantes reportó ser intersexual. Finalmente, la mayoría de los participantes residía en Distrito Capital (57.4%) y en el estado Miranda (33.2%), mientras que el resto habitaba en otros estados.

Dadas las características del proceso de muestreo, se puede catalogar como un muestreo no probabilístico y propositivo (Peña, 2017). Esto se debe a que los participantes no fueron seleccionados aleatoriamente para formar parte del estudio, sino que fueron elegidos en función de que cumplieran con ciertas características de interés: ser personas LGBTI y residir en Venezuela; así como que tuvieran la disposición para participar. Además de esto, la participación fue voluntaria y no se ofreció ningún tipo de compensación por la misma.

Instrumentos

Medida de estrés minoritario LGBT - Versión corta. Es un instrumento diseñado originalmente en inglés por Outland (2016) para evaluar estrés minoritario en personas lesbianas, gais, bisexuales y transexuales. La versión corta, presentada por esta misma autora, está conformada por 25 ítems tipo Likert, con cinco opciones de respuesta, que van desde 1 (*Nunca sucede*) hasta 5 (*Sucede todo el tiempo*). Estos ítems se agrupan en siete dimensiones: ocultamiento de la identidad, microagresiones o discriminación diaria, anticipación de rechazo, eventos de discriminación, estigma internalizado, eventos de victimización y, por último, *conexión con la comunidad*. Una vez que se inviertan las puntuaciones en conexión con la comunidad, los valores altos en el instrumento reflejan un mayor estrés minoritario. La versión en español del instrumento fue realizada para esta investigación. El instrumento presenta una consistencia interna adecuada evidenciada a través del coeficiente de confiabilidad omega de McDonald (ω = .88).

Subescala de bienestar recordado. Esta subescala mide al mismo tiempo los componentes hedónicos, eudaimónicos y sociales del bienestar; además, forma parte de las dos subescalas que constituyen el índice de felicidad de Pemberton (Hervás & Vázquez, 2013; Vázquez & Hervás, 2013). La subescala de bienestar recordado consta de 11 ítems tipo Likert, con opciones de respuesta que van desde 0 (*Totalmente en desacuerdo*) hasta 10 (*Totalmente de acuerdo*), más bien las puntuaciones altas indican un mayor bienestar recordado. La versión empleada en español fue presentada originalmente por Hervás y Vázquez (2013). El coeficiente de confiabilidad omega de McDonald fue de .92; por lo que se considera una medida confiable de bienestar recordado.

Inventario de depresión mayor. Propuesto originalmente en inglés por Bech y colaboradores (Bech et al., 2001; Bech et al., 2015), es un instrumento diseñado para medir la depresión. La versión en español (Herrera-Lopéz et al., 2022) cuenta con 10 ítems, tipo Likert, con puntajes entre 0 (*Nunca*) y 5 (*Todo el tiempo*, donde mayores puntuaciones indican una mayor depresión. Para los últimos tres ítems del inventario se presentan dos versiones para cada uno: versión a y versión b. Para obtener el puntaje global, se promedian 10 ítems de los cuales se seleccionan solo aquellos que puntúen más alto entre aquellos con dos opciones posibles. La confiabilidad del inventario se considera adecuada para la muestra estudiada (ω = .94).

Subescala de comportamientos de riesgo suicida. Es un instrumento compuesto por tres ítems extraídos de la *encuesta de riesgo juvenil* (Johns et al., 2019), que indaga acerca de conductas autolesivas, ideación suicida e intentos de suicidio durante los últimos 12 meses. Cada pregunta cuenta con tres opciones de respuesta; para el primero, estas consisten en: Nunca, De 1 a 10 *veces* y De 10 a 19 *veces o más*; para los dos ítems restantes las opciones incluían: No, *Sí, hace más de un año* y Sí, *durante este año*. En ambos casos, se asignaron respectivamente valores de 0 a 2 puntos; los mayores puntajes en esta subescala indican un mayor riesgo suicida. La versión en español de los ítems fue realizada durante esta investigación.

Cuestionario sociodemográfico. Es un conjunto de ítems destinados a evaluar distintas características sociodemográficas de los participantes. Estos ítems consisten en preguntas con opciones cerradas, a las cuales se les añade una alternativa final con respuesta abierta. En este caso, este cuestionario indaga acerca de: orientación sexual, identidad de género, expresión de género, intersexualidad y estado de residencia.

Procedimiento

Para la traducción inicial del instrumento medida de *estrés minoritario LGBT*, en su versión corta, se recurrió al método de comité (Erkut, 2010). Inicialmente, se contó con el apoyo de una traductora bilingüe para realizar la primera versión en español. Seguidamente, los investigadores compararon los ítems traducidos con su versión original para asegurar la equivalencia en el significado de los ítems y su pertinencia para el constructo y su dimensión respectiva. En los casos donde hubo discrepancias sobre la traducción de los ítems, estos fueron modificados una vez alcanzado el consenso sobre la versión más adecuada.

Finalmente, esta versión corregida fue presentada a un grupo reducido de personas pertenecientes a la población objetivo para asegurar la adecuada comprensión de los ítems, el patrón de respuesta, así como las instrucciones generales (Borsa et al., 2012).

Tanto la traducción en español del instrumento como el resto de cuestionarios se modificaron para evitar el uso de géneros específicos en la redacción de los ítems. Se optó por cambiar las palabras donde se presentará el género masculino o femenino asociado al individuo por la letra *e* para los casos en singular y *es* para los casos en plural.

Seguidamente, se construyeron las versiones digitales de los instrumentos a través de la plataforma *Google Forms*. Esta misma permite elaborar cuestionarios en línea, compartirlos mediante enlaces y se almacenan los registros de respuestas en la plataforma durante el proceso. Adicionalmente, se incluyó un mensaje inicial que explicara los objetivos de la investigación, indicará la confidencialidad de las respuestas y que la participación era voluntaria. La encuesta fue distribuida posteriormente a través de distintos medios (p. ej., correo electrónico, redes sociales y servicios de mensajería instantánea). El contactó inicial con los participantes se realizó a través de los registros de identificación de una organización venezolana centrada en derechos humanos de personas LGBTI+.

Con respecto a las consideraciones éticas presentes en la investigación, esta se llevó a cabo siguiendo los lineamientos propuestos por la Asociación Americana de Psicología (APA, por sus siglas en inglés, 2017). Entre estos se incluyen el consentimiento informado de los participantes al plantearles el objetivo de la investigación y el carácter voluntario de su participación; asimismo, se mantuvo la confidencialidad de los participantes ya que los resultados se presentan únicamente como estadísticos generales y no pueden ser asociados con una persona en particular.

Por otra parte, los investigadores no incurrieron en discriminaciones, ya fueran basadas en edad, identidad de género u orientación sexual. Por último, se procuró evitar posibles perjuicios a los participantes ya que algunas de las variables de interés podrían movilizar psicológicamente a los encuestados, especialmente al indagar sobre ideación suicida. En este caso, se incluyeron números de contacto de distintas organizaciones que ofrecían atención psicológica para las personas que necesitaran el servicio.

Análisis de Datos

Para todos los cálculos se hizo uso del programa estadístico *Jamovi* en su versión 2.3 (Proyecto Jamovi, 2023). Inicialmente, con el fin de comprobar la estructura factorial del instrumento de estrés LGBTI, se llevó a cabo un análisis factorial confirmatorio a partir de la estructura hallada por la autora original (Outland, 2016). Se empleó el método de mínimos cuadrados ponderados diagonalmente para estimar los factores ya que para este método no se asume una distribución normal de los datos (Li, 2016). Para cada reactivo se calcularon además las cargas factoriales estandarizadas, así como su significancia estadística y se tomó como punto de corte una significancia de 5%. Los 25 ítems se agruparon en los siete factores propuestos como se presenta a continuación en la Tabla 1.

Posteriormente, se calcularon los indicadores de ajuste del modelo. Entre estos se incluyeron: a) índices de ajuste global: prueba de chi-cuadrado (χ^2), donde se espera hallar un resultado no estadísticamente significativo, con una seguridad del 95% (Jordán-Muiños, 2021); b) índices de ajuste de parsimonia: chi-cuadrado normalizado,

Tabla 1
Dimensiones e ítems del instrumento de estrés LGBTI de Outland (2016).

Factores	Ítems
Ocultamiento de la identidad	1, 2, 3 y 4
Microagresiones / discriminación diaria	5, 6, 7 y 8
Anticipación de rechazo	9, 10, 11 y 12
Eventos de discriminación	13, 14, 15 y 16
Estigma internalizado	17, 18 y 19
Eventos de victimización	20, 21 y 22
Conexión con la comunidad	23, 24 y 25

en donde se toma como valor máximo aceptable el de 3 (Fernández-Pulido, 2008); c) índices de ajuste incremental: *GFI* y CFI, en donde se toma como punto de corte que estos igualen o superen un valor de .95 (Jordan-Muiños, 2021); d) índices de error del modelo: *RMSEA* y *SRMR*, con un punto de corte máximo de .08 (Ferrando & Anguiano-Carrasco, 2010).

Seguidamente, se calcularon los estadísticos descriptivos para las variables numéricas y se incluyeron las dimensiones de estrés minoritario. En esta oportunidad, se incluyeron medidas de tendencia central: media y mediana; medidas de dispersión: desviación típica; y medidas de forma: asimetría y curtosis. Asimismo, se estimaron los coeficientes de confiabilidad omega de McDonald y alfa de Cronbach: donde puntuaciones cercanas al 1 indican una adecuada precisión en las mediciones hechas con el cuestionario (Ventura-León & Caycho-Rodríguez, 2017).

Finalmente, se construyó una matriz de correlación producto-momento de Pearson entre las variables cuantitativas del estudio, en donde se incluyó tanto el puntaje total, como las dimensiones de estrés minoritario. Adicionalmente, para las asociaciones entre estrés minoritario, depresión, bienestar recordado e ideación suicida, se calcularon intervalos de confianza con una seguridad de 95%. Se esperaba que estas correlaciones fueran significativas al 5% y que los intervalos no incluyeran el valor de asociación nula (r = .0) en su recorrido.

Resultados

Los resultados obtenidos en el análisis factorial confirmatorio se presentan a continuación en la Tabla 2. Como se puede apreciar, todos los ítems presentan cargas factoriales positivas y en su mayoría son superiores a .6; por lo que estas son significativas con una seguridad del 99%. Esto permite indicar que los ítems cargan adecuadamente con su respectivo factor.

En relación con las pruebas de ajuste de este modelo, nuevamente se aprecian resultados que en su mayoría señalan un adecuado ajuste entre los datos y el modelo factorial propuesto. En la prueba de chi-cuadrado se obtuvo un resultado contrario a lo esperado: un valor significativo ($\chi^2_{(254)} = 551.2$; $p < .001$); sin embargo, esta es una situación habitual al tratarse de una prueba sensible al tamaño muestral (Jordan-Muiños, 2021).

En cambio, el chi-cuadrado normalizado permitió evidenciar la adecuación del modelo factorial ya que este toma en cuenta el número de observaciones en la muestra. De este modo, se encontró un valor inferior al criterio de máximo 3 ($\chi2/gl$ = 2.17). Otros indicadores del adecuado ajuste del modelo superaron el monto mínimo deseado de .9 (*GFI* = .97; *CFI* = .98; asimismo, uno de los indicadores del error en el modelo fue inferior al criterio máximo esperado de .08 (*RMSEA* [IC95%] = .07 [.06-.08)], aunque el segundo no lo fue (*SRMR* = .09).

En medidas generales, estos resultados permiten señalar que la estructura factorial del instrumento de estrés minoritario LGBT, en su versión en español de 25 ítems, se ajusta de manera adecuada al modelo propuesto por su autora

Tabla 2
Análisis factorial confirmatorio para el instrumento de estrés LGBTI (Outland, 2016) en español.

Constructo	Carga factorial estandarizada	*Z*	*p*
1. Ocultamiento de identidad			
1. Evito decirle a las personas acerca de ciertos elementos en mi vida porque podrían implicar que soy LGBT.	.92	33.83	<.001
2. Evito hablar acerca de mi vida romántica porque no me gusta que otros sepan que soy LGBT.	.89	34.14	<.001
3. No llevo a mi pareja a eventos sociales porque no quiero que otros sepan que soy LGBT.	.70	20.04	<.001
4. Limito lo que comparto en redes sociales o quién puede verlo, porque no quiero que otros sepan que soy LGBT.	.64	21.44	<.001
2. Microagresiones/discriminación diaria			
5. Se espera de mí que eduque a personas no-LGBT acerca de temas LGBT.	.55	17.48	<.001
6. Las personas han re-etiquetado mi identidad o se refieren a mí por nombres/pronombres que son diferentes a cómo yo me identifico.	.67	21.13	<.001
7. Cuando me encuentro en una organización o actividad que se clasifica por género, me siento fuera de lugar porque soy LGBT.	.78	23.39	<.001
8. Me han acusado de ser muy defensive o políticamente correcte cuando hablo de temas LGBT con alguien que no es LGBT.	.67	21.37	<.001
3. Anticipación de rechazo			
9. Cuando conozco a alguien nuevo, me preocupa que secretamente no les agrade porque soy LGBT.	.75	32.12	<.001
10. Me preparo mentalmente para ser tratado irrespetuosamente porque soy LGBT.	.86	39.44	<.001
11. Me encuentro a la espera de que otros no me acepten porque soy LGBT.	.86	40.74	<.001
12. Me preocupa qué sucederá si las personas descubren que soy LGBT.	.81	34.62	<.001
4. Eventos de discriminación			
13. He sido excluide de alguna organización (p. ej. grupo religioso, equipo deportivo, etc.) porque soy LGBT.	.77	28.23	<.001
14. He sido presionade a recibir servicios innecesarios o se me han negado servicios, por algún profesional de la salud porque soy LGBT.	.79	28.19	<.001
15. He recibido mala atención en algún negocio porque soy LGBT.	.81	32.0	<.001
16. He sido tratade injustamente por supervisores o profesores porque soy LGBT.	.78	28.85	<.001
5. Estigma internalizado			
17. Si se me ofreciera la oportunidad de ser alguien que no es LGBT, aceptaría la oportunidad.	.86	23.69	<.001
18. Desearía no ser LGBT.	.84	23.57	<.001
19. Envidio a las personas que no son LGBT.	.85	22.0	<.001
6. Eventos de victimización			
20. He sido acosade verbalmente o insultade, porque soy LGBT.	.92	40.64	<.001
21. Otros han amenazado con hacerme daño porque soy LGBT.	.84	36.09	<.001
22. He sufrido de bullying por ser LGBT.	.81	36.59	<.001
7. Conexión con la comunidad			
23. Siento que podría encontrar información sobre temas LGBT[a.]	.68	12.82	<.001
24. Siento que podría encontrar servicios profesionales para temas LGBT si los necesito[a.]	.82	13.22	<.001
25. Siento que podría encontrar espacios públicos que sean seguros para las personas LGBT[a.]	.71	12.99	<.001

Nota. *Z* = valor Z obtenido; *p* = significancia.
Se empleó el método de mínimos cuadrados ponderados diagonalmente como forma de estimación de los factores.
a Se invirtieron las puntuaciones en el ítem.

original (Outland, 2016). Por lo tanto, se podría considerar que existe validez de constructo para interpretar las puntuaciones del instrumento en sus siete dimensiones.

Seguidamente, se procedió a realizar un análisis descriptivo de las variables de estudio (ver Tabla 3). En el caso del estrés minoritario LGBT, se obtuvo un promedio de 2.11 puntos (*S* = .57; *CV* = 27%), con valores que se encuentran entre 1.12 y 4.08. Con respecto a la forma de la distribución, se observa una distribución coleada hacia la derecha, por lo que existe una mayor agrupación de datos en torno a los valores inferiores de la distribución (*As* = .98); al mismo tiempo, esta se presenta como leptocúrtica (*K* = .93).

En el caso de la depresión mayor, se presentaron valores medios de 20.74 (*S* = 12.35), que podrían catalogarse como depresión leve (Herrera-Lopéz et al., 2022) con valores que se encuentran entre 0 y 49. La asimetría positiva (*As* = .49) y la curtosis (*K* = -.85) permiten señalar que existe una agrupación platicúrtica de los datos con una mayor concentración hacia los niveles inferiores de la distribución.

En tercer lugar, los niveles obtenidos en el bienestar recordado se ubican en promedio en 7.11 (*S* = 1.62) con valores que se encuentran entre 1.45 y 10. La forma de la distribución se caracteriza por tener una agrupación homogénea de datos a la izquierda de la distribución (*As* = -1.11, *K* = 1.31).

Por último, con respecto a la ideación suicida, se obtuvieron valores promedio de 1.29 (*S* = 1.41) con un rango que se encuentra entre 0 y 6 puntos. La forma de la distribución se caracteriza por ser asimétrica positiva (*As* = .97) y tener una forma leptocúrtica (*K*= .23).

Finalmente, se procedió a analizar la asociación entre las variables cuantitativas del estudio. Esto se llevó a cabo con el fin de comprobar que el constructo de interés y el estrés minoritario LGBT se asociaban de manera similar a como se ha hallado en la literatura. A continuación se presentan las asociaciones obtenidas entre las puntuaciones totales de estrés minoritario LGBT, así como sus dimensiones, con el estrés, el bienestar recordado y la ideación suicida (ver Tabla 4).

Tabla 3
Estadísticos descriptivos y coeficientes de confiabilidad para las variables de estudio.

Variables	***Md***	***M***	***S***	***As***	***K***	ω	α
Estrés minoritario	2	2.11	.57	.98	.93	.88	.87
Ocultamiento de identidad	1.75	1.92	.86	.86	-.05	.81	.80
Microagresiones	2.25	2.52	.93	.67	-.01	.70	.70
Anticipación de rechazo	2	2.24	1.04	.89	.18	.85	.85
Eventos de discriminación	1.50	1.75	.82	1.32	1.43	.81	.80
Estigma internalizado	1.33	1.74	.91	1.25	.92	.81	.8
Eventos de victimización	2	2.21	.94	.75	.19	.85	.85
Conexión con la comunidad[a]	2.33	2.37	.89	.46	-.10	.73	.72
Depresión	17	20.74	12.35	.49	-.85	.94	.93
Bienestar recordado	7.46	7.12	1.62	-1.11	.16	.92	.87
Ideación suicida	1	1.29	1.41	.97	.16	.73	.70

Nota. *Md* = mediana; *M* = media aritmética; *S* = desviación estándar; *As* = asimetría de Fisher; *K* = curtosis centrada en 0; ω = coeficiente de confiabilidad omega de McDonald; α = coeficiente de confiabilidad alfa de Cronbach. Todos los descriptivos fueron calculados con *n* = 223.

Tabla 4
Matriz de correlaciones producto-momento de Pearson entre las variables de estudio.

Variable	1	2	3	4	5	6	7	8	9	10
1. Estrés minoritario										
2. Ocultamiento de identidad	.56**									
3. Microagresiones	.69**	.19**								
4. Anticipación de rechazo	.85**	.54**	.58**							
5. Eventos de discriminación	.72**	.07	.49**	.48**						
6. Estigma internalizado	.49**	.36**	.13*	.39**	.17**					
7. Eventos de victimización	.69**	.11	.53**	.49**	.73**	.06				
8. Conexión con la comunidad[a]	·21**	.03	-.12	-.03	.13*	.08	.05			
9. Depresión	.53**	.38**	.40**	.54**	.26**	.29**	.25**	.03		
10. Bienestar recordado	-.44**	-.39**	-.17*	-.48**	-.17*	-.44**	-.08	-.10	-.61**	
11. Ideación suicida	.50**	.25**	.50**	.47**	.35**	.17*	.33**	-.04	.58**	-.35**

Nota. * $p < .05$; ** $p < .01$.
Todos los cálculos se realizaron con $n = 223$.
[a]Puntajes invertidos.

Como se puede apreciar, existe una asociación estadísticamente significativa entre el estrés minoritario LGBT y la depresión ($r_{(223)} = .53$; $p < .001$; I*C* 95% = [.42; .61]), así como con la ideación suicida ($r_{(223)} = .50$; $p < .001$; *IC* 95% = [.39; .59]). En ambos casos, estas relaciones se consideran moderadas y positivas, por lo que, niveles altos de estrés minoritario se asocian con niveles altos en depresión e ideación suicida. Por su parte, el bienestar recordado también presentó una asociación estadísticamente significativa con el estrés minoritario ($r_{(223)} = -.44$; $p < .001$; *IC* 95% = [-.32; -.54]). Sin embargo, esta asociación resultó ser moderada y negativa; es decir, niveles elevados en estrés minoritario se relacionan con bajos niveles de bienestar recordado.

Discusión

Como objetivo de investigación se propuso realizar un análisis psicométrico de la versión en español de la *medida de estrés minoritario LGBT*, originalmente elaborada en inglés por Outland (2016), en una muestra venezolana. Para analizar la validez de las puntuaciones obtenidas con este instrumento, se hipotetizó que se asociaron significativamente con depresión e ideación suicida, así como con bienestar recordado. A su vez, se consideró que sería positiva la correlación con las dos primeras y negativa con esta última. En cambio, la confiabilidad sería estudiada a partir de coeficientes de consistencia interna.

En relación con las dimensiones subyacentes al instrumento, aunque el índice chi-cuadrado (χ^2) de ajuste general resultó ser significativo, situación que también se presentó en el estudio original de Outland (2016), a partir de la mayoría de indicadores considerados, se puede concluir que las puntuaciones obtenidas con esta versión en español del instrumento se ajustan de la manera esperada. Adicionalmente, se debe considerar que la prueba de χ^2 es una de las que se ve más afectadas por el tamaño de la muestra (Kline, 2016). Por estos motivos, tanto a partir de los indicadores estadísticos como del modelo planteado originalmente, se puede concluir que las puntuaciones obtenidas en el instrumento se ajustan de manera adecuada a la estructura factorial (Martin & Savage-McGlynn, 2013), conformada por siete dimensiones.

Al analizar la confiabilidad del instrumento, se aprecia una adecuada consistencia interna evidenciada a través de los coeficientes de confiabilidad omega de McDonald ($\omega = .88$) y alfa de Cronbach ($\alpha = .87$). Por lo tanto, puede considerarse que hay evidencia de confiabilidad para las puntuaciones registradas mediante esta prueba.

Con respecto al resto de las variables de estudio, se cumplieron las relaciones hipotetizadas. En concreto, se evidenciaron relaciones positivas, moderadas y significativas entre la depresión y la ideación suicida con el estrés minoritario, en donde mayores niveles de depresión y una mayor ideación suicida se asocian con mayores niveles de estrés minoritario. Esto coincide con investigaciones anteriores (Baams et al., 2015), donde se ha considerado que el estrés minoritario suele conllevar autopercepciones de ser una carga y una falta de sentido de pertenencia, que pueden generar depresión e ideación suicida en aquellas personas que lo experimentan.

Por otra parte, también se cumplió la hipótesis con respecto a la relación con el bienestar recordado dado que la asociación con el estrés minoritario resultó ser negativa, moderada y significativa. De esta manera, un menor bienestar recordado se asocia con mayores niveles de estrés minoritario. Nuevamente, estos resultados coinciden con lo hallado en la literatura (Meyer & Frost, 2013), donde se ha planteado que el estrés minoritario implica estigmas, prejuicios y discriminaciones en contra de estas minorías, por lo que estos factores son determinantes en la elección de hacer pública su orientación sexual y de poder disfrutar de su sexualidad (González-Rivera & Pabellón-Lebrón, 2018). Esto repercute negativamente en el bienestar de estos individuos.

Habiéndose encontrado este patrón diferencial de asociaciones, análogas a las obtenidas en la literatura, se puede considerar que las puntuaciones obtenidas mediante el instrumento de estrés minoritario LGBT presentan evidencia de validez de constructo (Martin & Savage-McFlynn, 2013) para ser empleadas en la población venezolana con el fin de medir el estrés minoritario.

A modo de síntesis, la versión en español del instrumento de *estrés minoritario LGBT* de Outland (2016) ha presentado evidencias de validez y de confiabilidad para ser empleado en la población venezolana. Igualmente, el análisis factorial confirmatorio permite concluir que las puntuaciones del instrumento pueden ser interpretadas a partir de las siete dimensiones propuestas: ocultamiento de la identidad, microagresiones, anticipación de rechazo, eventos de discriminación, estigma internalizado, eventos de victimización y conexión con la comunidad.

Por lo tanto, se recomienda la utilización de este instrumento para futuras investigaciones en el área de la psicología de la salud que busquen estudiar el poder predictivo y explicativo del estrés minoritario en personas LGBTI, así como posibles factores desencadenantes para así determinar la red nomológica del constructo y poder elaborar intervenciones más efectivas basadas en la evidencia.

Agradecimientos

A la organización no gubernamental de Venezuela, País Plural, por haber financiado la investigación y haber realizado la administración de los cuestionarios.

Referencias

Asociación Americana de Psicología (2017). *Ethical principles of psychologists and code of conduct*. https://www.apa.org/ethics/code/ethics-code-2017.pdf

Ato, M., López-García, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología/ Annals of Psychology, 29*(3), 1038-1059. https://doi.org/10.6018/analesps.29.3.178511

Baams, L., Grossman, A. H., & Russell, S. T. (2015). Minority stress and mechanisms of risk for depression and suicidal ideation among lesbian, gay, and bisexual youth. *Developmental Psychology, 51*(5), 688-696. https://doi.org/10.1037/a0038994

Bech, P., Rasmussen, N. A., Raabæk-Olsen, L., Noerholm, V., & Abildgaard, W. (2001). The sensitivity and specificity of the Major Depression Inventory, using the present state examination as the index of diagnostic validity. *Journal of Affective Disorders, 66*(2), 159-164. https://doi.org/10.1016/S0165-0327(00)00309-8

Bech, P., Timmerby, N., Martiny, K., Lunde, M., & Soendergaard, S. (2015). Psychometric evaluation of the Major Depression Inventory (MDI) as depression severity scale using the LEAD (Longitudinal Expert Assessment of All Data) as index of validity. *BioMed Central Psychiatry, 5*(190). https://doi.org/10.1186/s12888-015-0529-3

Borsa, J. C. Damásio, B. F., & Bandeira, D. R. (2012). Adaptação e validação de instrumentos psicológicos entre culturas: Algumas considerações. *Paidéia (Ribeirão Preto), 22*(53), 423-432. http://doi.org/10.1590/S0103-863X2012000300014

Chodzen, G., Hidalgo, M. A., Chen, D., & Garofalo, R. (2019). Minority stress factors associated with depression and anxiety among transgender and gender-nonconforming youth. *Journal of Adolescent Health, 64*(4), 467-471. https://doi.org/10.1016/j.jadohealth.2018.07.006

Conlin, S. E., Douglass, R. P., & Ouch, S. (2019). Discrimination, Subjective Wellbeing, and the Role of Gender: A mediation model of LGB Minority Stress. *Journal of Homosexuality, 66*(2), 238-259. http://doi.org/10.1080/00918369.2017.1398023

Dentato, M. P. (2012). The minority stress perspective. *Psychology & AIDS Exchange, 37*, 12-15. https://ecommons.luc.edu/socialwork_facpubs/2

Erkut, S. (2010). Developing multiple language versions of instruments for intercultural research. *Child Development Perspectives, 4*(1), 19-24. https://doi.org/10.1111/j.1750-8606.2009.00111.x

Feinstein, B. A., & Dyar, C. (2017). Bisexuality, minority stress, and health. *Current Sexual Health Reports, 9*, 42-49. https://doi.org/10.1007/s11930-017-0096-3

Fernández-Pulido, R. (2008). Modelos de medida y análisis factorial confirmatorio. En M. Verdugo, M. Crespo, M. Badía, B. Arias (Eds.), *Metodología en la investigación sobre discapacidad: Introducción al uso de las ecuaciones estructurales* (pp. 29-41). Publicaciones del INICO.

Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo, 31*(1), 18-33. https://www.papelesdelpsicologo.es/home

Fisher, C. M., Woodford, M. R., Gartner, R. E., Sterzing, P. R., & Victor, B. G. (2019). Advancing research on LGBTQ microaggressions: A psychometric scoping review of measures. *Journal of Homosexuality, 66*(10), 1345-1379. https://doi.org/10.1080/00918369.2018.1539581

Goldbach, J. T., Tanner-Smith, E. E., Bagwell, M., & Dunlap, S. (2014). Minority stress and substance use in sexual minority adolescents: A meta-analysis. *Prevention Science, 15*, 350-363. https://doi.org/10.1007/s11121-013-0393-7

González-Rivera, J. A., & Pabellón-Lebrón, S. (2018). Desarrollo y validación de un instrumento para medir discriminación percibida en la comunidad LGBT. *Revista Evaluar, 18*(2), 59-74. https://doi.org/10.35670/1667-4545.v18.n2.20809

Herrera-Lopéz, H. M., Hormoza-Moreno, A. M., & Martínez-Paz, L. J. (2022). Propiedades psicométricas del Inventario de Depresión Mayor (MDI) en población universitaria colombiana. *Informes Psicológicos, 22*(1), 217-234. http://doi.org/10.18566/infpsic.v22n1a13

Hervás, G., & Vázquez, C. (2013). Construction and validation of a measure of integrative well-being in seven languages: The Pemberton Happiness Index. *Health and Quality of Life Outcomes, 11*(66). https://doi.org/10.1186/1477-7525-11-66

Johns, M. M., Lowry, R., Haderxhanaj, L. T., Rasberry, C. N., Robin, L., Scales, L., Stone, D., & Suarez, N. A. (2020). Trends in violence victimization and suicide risk by sexual identity among high school students — Youth Risk Behavior Survey, United States, 2015–2019. *Morbidity and Mortality Weekly Report Supplements, 69*(1), 19-27. https://doi.org/10.15585/mmwr.su6901a3

Jordan-Muiños, F. M. (2021). Valor de corte de los índices de ajuste en el análisis factorial confirmatorio. *Psocial, 7*(1), 66-71. https://publicaciones.sociales.uba.ar/index.php/psicologiasocial/index

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guillford Press.

Li, CH. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*, 936-949. https://doi.org/10.3758/s13428-015-0619-7

Martin, C. R., & Savage-McGlynn, E. (2013). A 'good practice' guide for the reporting of design and analysis for psychometric evaluation. *Journal of Reproductive and Infant Psychology*, *31*(5), 449-455. http://doi.org/10.1080/02646838.2013.835036

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin, 129*(5), 674-697. https://doi.org/10.1037/0033-2909.129.5.674

Meyer, I. H., & Frost, D. M. (2013). Minority stress and the health of sexual minorities. In C. J. Patterson & A. R. D'Augelli (Eds.), *Handbook of psychology and sexual orientation* (pp. 252-266). Oxford University Press.

Molina, V. (2022). Abortion, same-sex marriage, and gender identity during the Pink Tide: Venezuela compared to Latin American trends. *The International Journal of Human Rights, 26*(7), 1293-1312. https://doi.org/10.1080/13642987.2022.2088521

Mongelli, F., Perrone, D., Balducci, J., Sacchetti, A., Ferrari, S., Mattei, G., & Galeazzi, G. M. (2019). Minority stress and mental health among LGBT populations: An update on the evidence. *Minerva Psichiatrica, 60*(1), 27-50. https://doi.org/10.23736/S0391-1772.18.01995-7

Ogunbajo, A., Iwuagwu, S., Williams, R., Biello, K. B., Kahler, C. W., Sandfort, T. G. M., & Mimiaga, M. J. (2020). Validation of depressive symptoms, social support, and minority stress scales among gay, bisexual, and other men who have with men (GBMSM) in Nigeria, Africa: A mixed methods approach. *BMC Public Health, 20*, 1023. https://doi.org/10.1186/s12889-020-09127-0

Outland, P. L. (2016). Developing the LGBT Minority Stress Measure [Master's thesis, Colorado State University]. Mountain Scholar Digital Collections of Colorado https://mountainscholar.org/items/54cd061c-1505-4122-acf9-2494d6465a68

Peña, G. (2017). *Estadística inferencial: Una introducción para las ciencias del comportamiento* (2nd ed.). AB ediciones.

Pineda-Roa, C. A. (2019). Factores de riesgo de ideación suicida en una muestra de adolescentes y jóvenes colombianos autoidentificados como homosexuales. *Revista Colombiana de Psiquiatría, 48*(1), 2-9. https://doi.org/10.1016/j.rcp.2017.06.001

Poetar, CR., & Criț, C.I. (2024). Minority stress and anxiety and depressive symptoms in sexual minority adults: Irrationality and self-compassion as serial mediators. *Sexuality Research and Social Policy*. https://doi.org/10.1007/s13178-023-00929-6

Proyecto Jamovi (2023). Jamovi (versión 2.3). [Software de estadística]. Recuperado de https://www.jamovi.org

Tan, K. K. H., Treharne, G. J., Ellis, S. J., Schmidt, J. M., &

Veale, J. F. (2019). Gender minority stress: A critical review. *Journal of homosexuality, 67*(10), 1471-1489. https://doi.org/10.1080/00918369.2019.1591789

Vázquez-López, N., Fernández-Rodríguez, M., García-Vega, E., & Guerra-Mora, P. (2020). Perceived social support in transgender people: A comparative study with cisgender people. Interacciones, *Revista de Avances en Psicología, 6*(2), e126. https://doi.org/10.24016/2020.v6n2.126

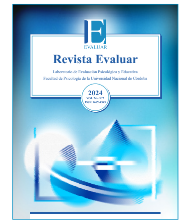
Vázquez, C., & Hervás, G. (2013). Addressing current challenges in cross-cultural measurement of well-being: The Pemberton Happiness Index. In H. Knoop & A. Delle Fave (Eds.), *Well-being and cultures* (pp. 31-49). Springer. https://doi.org/10.1007/978-94-007-4611-4_3

Ventura-León, J. L., & Caycho-Rodríguez, T. (2017). El coeficiente Omega: Un método alternativo para la estimación de la confiabilidad. Revista Latinoamericana de Ciencias Sociales, *Niñez y Juventud, 15*(1), 625-627. https://revistaumanizales.cinde.org.co

EVALUAR

2023, Vol. 24, No. 2
ISSN 1667-4545
Recuperado de https://revistas.unc.edu.ar/index.php/revaluar

Laboratorio de Evaluación Psicológica y Educativa
Facultad de Psicología - Universidad Nacional de Córdoba



Argentine Adaptation of the Counterproductive Work Behavior Checklist - Short Version

Adaptación Argentina del Counterproductive Work Behavior Checklist - Short Version

María Laura Lupano-Perugini * [1] , Melissa Patricia Fabara-Torres [2]

1 - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad de Palermo, Buenos Aires, Argentina.
2 - Universidad de Palermo, Buenos Aires, Argentina.

Recibido: 20/02/2023 **Revisado:** 24/04/2023 **Aceptado:** 18/05/2023

Abstract

Two studies were conducted within the framework of undesirable behavior in the workplace. The objective of the first study was to validate the Counterproductive Work Behavior Checklist – Short Version (CWB-C-10) for research purposes in Argentina. It was carried out with a sample of 874 workers (54.7% women, 44.6% men) from Argentina with a mean age of 37.5 years old (SD = 12.2). The analyses carried out confirmed the one-dimensional structure of the test. Likewise, the internal consistency through alpha and omega coefficients was adequate. The objective of the second study was to identify psychological and organizational variables (Dark Triad personality traits, engagement, and job satisfaction) that allow predicting the development of counterproductive work behavior. It was carried out with a sample of 103 active workers (60.9% women, 39.1% men) from Argentina with a mean age of 33 years old (SD = 10.7). As a result, it was observed that the Machiavellianism trait and the job satisfaction level were the variables with the greatest predictive power.

Keywords: *counterproductive work behavior, Dark Triad, engagement, job satisfaction, adaptation*

Resumen

Los autores realizaron dos estudios. El objetivo del primero fue validar el Counterproductive Work Behavior Checklist - Short Version (CWB-C-10) para uso en investigación en Argentina. Contó con una muestra de 874 trabajadores/as (54.7% mujeres, 44.6% varones) activos/as de Argentina con una media de edad de 37.5 años (DE = 12.2). Los análisis efectuados permitieron confirmar la estructura unidimensional de la prueba. Asimismo, la consistencia interna mediante coeficiente alfa y omega resultó adecuada. El objetivo del segundo estudio fue identificar variables psicológicas y organizacionales (rasgos de personalidad de la Tríada Oscura, engagement, y satisfacción laboral) que permiten predecir el desarrollo de comportamiento laboral contraproducente. Contó con una muestra de 103 trabajadores/as (60.9% mujeres, 39.1% varones) activos/as de Argentina con una media de edad de 33 años (DE = 10.7). Como resultado, se observó que el rasgo maquiavelismo y el nivel de satisfacción laboral resultaron las variables de mayor poder predictor.

Palabras clave: *comportamiento laboral contraproducente, Tríada Oscura, engagement, satisfacción laboral, adaptación*

***Correspondence to:** Mario Bravo 1259 (1175), Autonomous City of Buenos Aires, Argentina. E-mail: mllupano@hotmail.com

How to cite: Lupano-Perugini, L., M., & Fabara-Torres, P., M., (2024): Argentine Adaptation of the Counterproductive Work Behavior Checklist - Short Version. *Revista Evaluar, 24*(2), 39-52. Recuperado de https://revistas.unc.edu.ar/index.php/revaluar

Authors' Notes: The current study is conducted with the support of PICT-2020-SERIEA-02184 grant: "Analysis of positive and negative personality traits in the workplace. A proposal for an alternative assessment to classical models".

Participaron en la edición de este artículo: Agustina Mangieri, Pablo Carpintero, Eugenia Barrionuevo, Florencia Ruiz, Ricardo Hernandez, Jorge Bruera.

Introduction

In organizations, behaviors driven by various factors take place. When behavior undermines efficiency instead of promoting it, it is known as counterproductive work behavior (CWB), the outcome of which negatively impacts both the individuals within an organization and the organization itself (Bennett & Robinson, 2000; Spector & Fox, 2002; Spector, 2006; Spector et al., 2010).

The relevance of studying these behaviors lies in several reasons. On the one hand, they can lead to economic harm. Additionally, they can damage the company's reputation, given the spread of detrimental rumors (Vélez-Vega, 2022), potentially involving the company in legal conflicts (Morf et al., 2017). Moreover, these behaviors impact the mental health and well-being of employees who are victims of them (Bowling & Michel, 2011; Ones & Dilchert, 2013; Mercado et al., 2018; Spector, 2006), resulting in losses of productivity and resources. The reason why the study of CWB becomes complex and necessary is that these behaviors tend to occur in a concealed or secretive manner (Spector, 2001). Therefore, it is crucial to investigate them anonymously in order to identify them in workplace environments.

Originally, these behaviors were examined in isolation, encompassing phenomena such as turnover intentions, absenteeism, aggression, theft, and so forth. In the mid-1990s, Robinson and Bennett (1995) initiated a comprehensive exploration, categorizing them collectively under the umbrella term 'counterproductive work behaviors' (CWBs). Subsequently, additional attitudes were incorporated into this conceptual framework, including bullying, retaliatory behaviors, and destructive leadership, among others. Broadly speaking, CWB denotes intentional actions and behaviors by employees that yield adverse consequences for both the organizational well-being and its stakeholders (Ones & Dilchert, 2013; Spector & Fox, 2002; Spector, 2006). Concerning generational factors, empirical findings indicate a higher prevalence of these behaviors among younger employees (Ng & Feldman, 2008; Pletzer, 2021).

Spector (2006) and Spector et al. (2010) developed an instrument to measure these behaviors: The Counterproductive Work Behavior Checklist (CWB-C). There are three versions of this test. The first version, with 45 items, assesses two dimensions of CWB (toward the organization and individuals). The second version, with 32 items, consists of five subscales: abuse, production deviance, sabotage, theft, and withdrawal or absenteeism. The third version, with 10 items, was developed to obtain an overall score, although, according to the authors, half of its items are focused on the organization and the other half on the individual. In its three versions, this instrument has been translated into several languages, including Spanish, English, German, Italian, and Polish (for more information, see https://paulspector.com/assessments/pauls-no-cost-assessments/counterproductive-work-behavior-checklist-cwb-c/counterproductive-work-behavior-checklist-cwb-c-translations/). However, the instrument has been validated and adapted only in a few countries, with publications in Italy (45-item version; Barbaranelli et al., 2013) and Pakistan (32-item version; Rauf & Farooq, 2014).

In this study, we aimed to validate the 10-item version for research purposes. Additionally, we were interested in analyzing the factors that may influence the occurrence of these behaviors in work environments. Therefore, we examined the relationship between CWB and the following variables: the Dark Triad personality traits, work engagement, and job satisfaction.

The Dark Triad of personality (Paulhus & Williams, 2002), composed of the traits of Machiavellianism, narcissism, psychopathy, and later sadism (Paulhus, 2014), may provide a valid justification for the presence of undesirable behaviors in the workplace. According to a meta-analysis by O'Boyle et al. (2012) and subsequent studies (e.g., DeShong et al., 2015; Filipkowski & Derbis, 2020; Junça-Silva & Silva, 2023; Miller, 2017; Uysal et al., 2023), the Dark Triad positively correlates with CWB. For example, Rehman and Shahnawaz (2018) conducted a study on managers and determined that the Machiavellianism trait was significantly associated with CWB because individuals with this trait were less likely to adhere to common workplace norms. Similarly, Blickle and Schütte (2017) found that high levels of psychopathy, along with low interpersonal influence, led to an increase in counterproductive behaviors directed toward the organization. Given the dissimilar results regarding the traits most associated with these behaviors (Miller, 2017), it is important to provide evidence of how this relationship unfolds in samples from diverse cultural contexts.

Research suggests that work engagement, considered a psychological presence in the role that includes attention, absorption, and energy directed toward work tasks (Rothbard & Patil, 2012), is negatively associated with CWB and negatively predicts these behaviors (e.g., Bilal et al., 2020; Filipkowski & Derbis, 2020; Malik & Zahra, 2022). Additionally, some studies have highlighted the mediating/moderating role of engagement in the generation of CWB. For instance, Lebron et al. (2018) found that engagement plays a mediating role in the leader-member exchange (LMX) and the level of CWB. Similarly, Chen et al. (2020) used a mediation-moderation model and demonstrated that engagement reduces levels of CWB in individuals with high levels of emotional stability and responsibility.

Finally, job satisfaction refers to individuals' attitudes toward their work and encompasses different facets (e.g., satisfaction with the supervisor, coworkers, remuneration, promotion opportunities, and the job in general) (Medrano et al., 2018; Spector, 1997, 2022). Therefore, it tends to be a determining factor in the actions people take in their jobs. Research has shown an inverse relationship between CWB and this variable (e.g., Mercado et al., 2018; Sackett, 2002). Álvarez-Escalante et al. (2021) reported that employees experiencing low job satisfaction and immersion in stressful work situations are more likely to engage in counterproductive behaviors. Likewise, Selvarajan et al. (2019) argue that one of the reasons why employees engage in CWB is dissatisfaction with organizational responses to certain work situations that do not fulfill employees' expectations (e.g., work-family balance). Similarly, De Clercq et al. (2019) postulate that the less attention and the greater pressure employees receive, the more likely they are to express their dissatisfaction through direct or indirect actions that harm the organization. Therefore, it is particularly important to analyze the predictive role of this variable in the development of CWB.

In light of the above considerations, this research aimed to: 1) validate the Counterproductive Work Behavior Checklist – Short Version (CWB-C-10) for research purposes in Argentina; 2) analyze individual differences in psychological and organizational variables (Dark Triad personality traits, work engagement, and job satisfaction) by dividing the sample according to the level of CWB; and 3) identify psychological and organizational variables (Dark Triad personality traits, work engagement, and job satisfaction) that predict the development of CWB.

Methodology

Participants

Study 1 Sample. This study included an intentional sample of 874 workers from Argentina. The average age was 37.5 years (*SD* = 12.2, *Min.* = 18, *Max.* = 75). Regarding sex, 54.7% (*n* = 478) of the participants were women, 44.6% (*n* = 390) were men, 0.5% (*n* = 4) were non-binary, and 0.2% (*n* = 2) preferred not to respond. The place of residence was as follows: 46.6% (*n* = 407) lived in the Autonomous City of Buenos Aires, 35% (*n* = 306) in Greater Buenos Aires, and the remaining 18.4% (*n* = 161) lived in other provinces. Regarding educational level, 13.9% (*n* = 121) had completed primary and secondary education, 72.6% (*n* = 635) had completed tertiary education, and 13.5% (*n* = 118) had completed postgraduate studies. Self-perceived socio-economic status was as follows: 16.1% (*n* = 141) perceived themselves as lower-middle class, 65.8% (*n* = 575) as middle class, and 12.1% (*n* = 106) as upper-middle class.

Study 2 Sample. The intentional sample consisted of 103 workers from Argentina. Their average age was 33 years (*SD* = 10.7, *Min*. = 19, *Max*. = 68). Regarding sex, 60.9% (*n* = 63) of the participants were women and 39.1% (n = 40) were men. Regarding place of residence, 83.5% (*n* = 86) lived in the Autonomous City of Buenos Aires, 13.6% (*n* = 14) in Greater Buenos Aires, and 2.9% (*n* = 3) in other provinces. Concerning educational level, 9.8% (*n* = 10) had completed primary and secondary education, 76.5% (*n* = 79) had completed tertiary education, and 13.7% (*n* = 14) had completed postgraduate studies. Regarding self-perceived socio-economic status, 25.5% (*n* = 26) perceived themselves as lower-middle class, 58.8% (*n* = 61) as middle class, and 15.7% (*n* = 16) as upper-middle class. Concerning organizational variables, 85.1% (*n* = 88) of the employees were working in private companies and 14.9% (*n* = 15) in public companies. The size of the companies was as follows: 53.2% (*n* = 55) of the participants were working for large companies, 36.2% (*n* = 38) for medium-sized companies, and 10.6% for small companies (*n* = 10). Most participants did not have subordinates (80.9%; *n* = 83), while 19.1% (*n* = 20) held leadership positions.

Instruments

Counterproductive Work Behavior Checklist – Short Version (CWB-C-10; Spector et al., 2010). This is a short version of the CWB-C (Spector et al., 2006), designed to assess Counterproductive Work Behaviors (CWB) in work environments. Although half of the items evaluate behaviors directed toward the organization and the other half toward individuals, the authors propose the use of a global score. Therefore, in this study, the fit to a one-dimensional structure was verified through confirmatory factor analysis, which yielded adequate indices. Additionally, the internal consistency data obtained from alpha and omega coefficients exceeded .70, which also resulted in optimal values (see Results section). The test consists of 10 items answered on a five-point Likert scale ranging from *never* to *always*.

Scale of Work Engagement (EACT; Lupano-Perugini et al., 2017). This test was designed to assess work engagement, based on the theoretical proposal of Rothbard and Patil (2012). It comprises two cognitive dimensions (attention and absorption) and one physical dimension (energy). The validation process in Argentina resulted in an 11-item version (e.g., *When I am working, I often lose track of time*), with responses on a Likert scale ranging from 1 (*completely disagree*) to 5 (*totally agree*). The scale has demonstrated good internal consistency and adequate evidence of convergent and discriminant validity. Additionally, a three-factor structure was confirmed through exploratory and confirmatory factor analyses. Only the total score was used in this study. Cronbach's alpha and omega coefficients for the total scale were calculated from the Study 2 sample: α = .91 and ω = .92, respectively.

Dark Triad Scale (DTS; Jones & Paulhus, 2014 / Argentine adaptation by Salessi & Omar, 2018). This measurement instrument consists of 24 items assessing traits of the Dark Triad of Personality (Paulhus & Williams, 2002). It comprises three dimensions: Machiavellianism (e.g., *Most people can be manipulated*); narcissism (e.g., *I demand that people treat me with the respect I deserve*); and psychopathy (e.g., *I could say anything to get what I want*). Each item is rated on a 5-point Likert scale (1 = *totally disagree* to 5 = *totally agree*). Validation studies conducted in Argentina through exploratory and confirmatory factor analyses confirmed the three-factor structure, which remained invariant across genders. Cronbach's alpha and omega coefficients estimated for Study 2 were as follows: Machiavellianism (α = .82, ω = .83), narcissism (α = .75, ω = .76), and psychopathy (α = .72, ω = .74).

In addition, two surveys were conducted to assess some of the variables considered in the second study:

Job Satisfaction Survey (Lupano-Perugini, 2017). A survey designed for a previous study (Lupano-Perugini, 2017) was employed. It consists of six items with a Likert response scale ranging from 1 (*completely dissatisfied*) to 7 (*completely satisfied*), evaluating the individuals' self-perceived satisfaction with their job in general and particular aspects, such as salary, supervisors, colleagues, workplace, and career. An example item is *How satisfied am I with the salary I receive?* The choice of areas to be assessed (e.g., salary, supervisors, colleagues) was based on aspects analyzed in previous instruments (e.g., Balzer et al., 1997). A higher score indicates a higher level of satisfaction. Cronbach's alpha and omega coefficients calculated from Study 2 were α = .89 and ω = .90, respectively.

Organizational and Individual Performance Survey (Lupano-Perugini, 2017). A survey designed for a previous study (Lupano-Perugini, 2017) was employed. This survey was designed according to the performance indicators proposed by Cameron et al. (2004) in their research on positive variables and performance (i.e., efficiency, innovation, growth, quality, employee and customer retention, satisfaction, and adaptation). The first part of the survey consists of 10 items with a Likert response scale ranging from 1 (*Little*) to 6 (*Much*), aimed at evaluating organizational performance according to employees' perceptions. The second section, intended for the assessment of individual performance, consists of six items with the same response scale (1 = *Little* to 6 = *Much*). An example item is *To what extent do you believe high-quality results were obtained?* A higher score indicates a higher level of perceived performance. In this study, only the second part aimed at evaluating individual performance was used. Cronbach's alpha and omega coefficients for this second part were α = .91 and ω = .92, respectively.

Procedure

The design was non-experimental and cross-sectional, employing a non-probabilistic sampling method. Data were collected by students conducting research practice at a private university in Buenos Aires, Argentina. The participants were volunteers who did not receive any compensation for their collaboration. Surveys were administered online using SurveyMonkey. The survey homepage requested participant consent, ensured data anonymity, and clarified its exclusive use for research purposes. Participants were required to be over 18 years old and employed in an organization with at least 10 employees.

The data collection was supervised by a researcher. The research complied with international ethical guidelines (APA and NC3R) and those of the National Council for Scientific and Technical Research (CONICET) for ethical behavior in the Social Sciences and Humanities (Resolution No. 2857, 2006) and was approved by the corresponding ethics committees.

Data Analysis

First, permission was obtained from the original test author through personal communication (Spector, 2022). This permission granted validation of the test for research purposes and noncommercial use. Subsequently, the test was translated from English to Spanish using direct translation. The translated version was subjected to a pilot study to ensure comprehension of items and instructions and to an expert review to analyze the appropriateness of item content for the evaluated construct. Psychometric properties were estimated from a sample of 874 participants through confirmatory factor analysis, and a polychoric matrix was used because of the polytomous nature of the items (Bandalos & Finney, 2018). Various fit indices were assessed to study model fit, including the Comparative Fit Index (CFI), Normed Fit Index (NFI), Incremental Fit Index (IFI), and Root Mean Square Error of Approximation (RMSEA). Expected values for CFI, NFI, and IFI indices should exceed .90 (Rial-Boubeta et al., 2006), while the RMSEA value should fall between .05 and .08 (Hu & Bentler, 1995). The regression weights for each item were also considered. The reliability of the scale was assessed using Cronbach's alpha and McDonald's Omega. Values above .70 are considered acceptable, and values surpassing .80 are deemed high (Kline, 2000). Finally, differences in psychological and organizational variables were analyzed using another sample of 103 participants and considering groups based on the level of counterproductive work behavior (CWB). Moreover, efforts were made to determine which variables included in the model could predict CBW development. Pearson's *r* test and hierarchical multiple regression were used for the analyses. Statistical analysis was conducted using EQS 6.2 and Jamovi 2.2.5 software within the *R* environment.

Results

Study 1. Validation of the Counterproductive Work Behavior Checklist – Short Version (CWB-C-10)

After permission was obtained from the author to validate the scale for research use (Spector, 2022), the translation process was initiated. The method employed was direct translation. Two researchers holding Ph.D. degrees in Psychology and having a good command of the English language participated in the process. They independently translated the original version and then compared their results. According to the translators' criteria, there were no notable differences between the two Spanish versions. Finally, adjustments were made to ensure comprehension, conceptual equivalence, and accuracy in the translation from English to Spanish. The translated version was tested with a pilot study involving 15 employees (8 women, 7 men), who suggested minor changes in the wording of some items.

The translated version was subjected to expert judgment, in which three judges assessed the content adequacy of the items. Two of the judges held Ph.D. degrees in Psychology and the other one was in the final stages of her doctorate. All three had experience in the field of psychological assessment and psychometrics. They were asked to indicate whether each item on the scale allowed the CWB construct to be evaluated. The overall agreement level was high, exceeding 90%.

Subsequently, the construct validity and reliability of the final Spanish version of the test were estimated. For these analyses, an initial sample of 874 participants was used. To obtain evidence of construct validity, a confirmatory factor analysis was conducted using the polychoric data matrix, given the Likert-type format of the responses. The robust maximum likelihood estimation method was used, which is appropriate for this type of data. Model fit was examined with various indices that showed very good fit (Schumacker & Lomax, 2016): $\chi^2 = 236.021$; $df = 45$; NFI = .901; CFI = .930; IFI = .930; RMSEA = .051 (90% CI = .035-.067).

All regression weights of the items (see Figure 1) were above .40 and statistically significant ($p < .001$) (Byrne, 2006). While the factor loadings were adequate, calculating the value of the average variance extracted (AVE) resulted in a value below .50 (AVE = .29), indicating low convergent validity according to Hair et al. (2010). Nevertheless, it should be clarified that this index is sensitive to the number of items per factor. As the number of items increases, the convergent validity measured deteriorates, while reliability increases. Therefore, more flexible criteria should be considered when many indicators per factor are involved. Moral de la Rubia (2019) suggests that for factors with more than nine items, AVE values greater than .25 can be considered acceptable as long as the factor loadings tend to exceed .50 and the omega coefficient is greater than .75 or .80, indicating acceptable convergent validity.

The internal consistency of the scale was examined through the calculation of Cronbach's alpha and McDonald's omega. The obtained values indicated that the scale is reliable: $\alpha = .75$, $\omega = .78$.

Considering the total score, the mean CWB in the sample used for validation was 14.3 (*SD* =

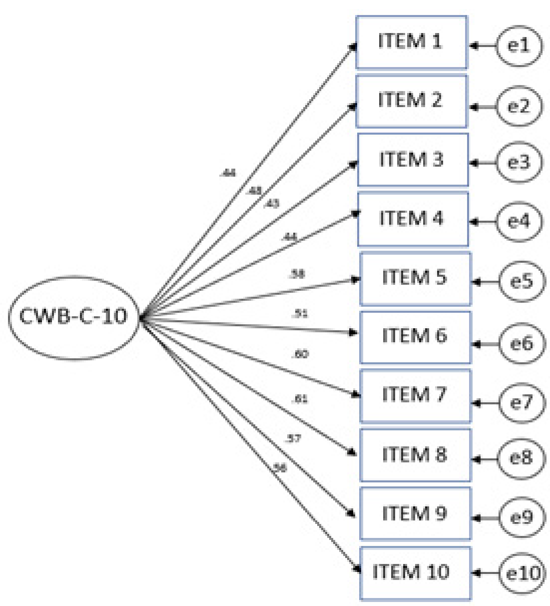


Figure 1
Confirmatory Factor Analysis of CWB-C-10.

3.7). No gender differences were observed [$t_{(866)}$ = -.815, p = .415]. A significant negative association was found between age and CWB (r = -.07, p = .046).

Study 2. Differences in Psychological Variables and Organizational Performance According to the Level of CWB. Predictors of CWB

For the second study, a sample of 103 employees was employed. Considering the total score, the mean CWB for this sample was 14.23 (SD = 3.66). No gender differences were observed [$t_{(101)}$ = .177, p = .860], and there was no relationship with age (r = -.002, p = .981). The level of CWB was also analyzed considering organizational variables. No difference was found in terms of organization size (small, medium, and large) [$F_{(11)}$ = .140, p = .871], nor was there a relationship with the employee's tenure in the organization (r = .301, p = .070). Finally, a difference in the level of CWB according to the type of organization was found, being higher in public capital organizations [$t_{(45)}$ = -2.240, p = .030; M_{Pub} = 16.4, M_{Priv} = 13.3].

Data from this sample were used to determine whether there were differences in certain psychological variables and organizational performance between employees with a low level of CWB and those with a moderate level. Subjects with a high level of CWB were not considered in the analysis because, in general, the analyzed sample, like that of the validation, did not show a mean representing a high presence of these behaviors, coinciding with international reports.

Two groups were formed according to standardized values in the total CWB score variable (cut-off score at z = 0). The group of employees with a low presence of CWB consisted of 64 subjects, while the group with a moderate level was composed of 39 subjects. Student's t-tests were calculated for these groups to determine whether there were differences in the variables included in the second study that, according to the literature, were related to the development or absence of CWB. Differences were found in negative personality traits, job satisfaction, job performance, and work engagement. As shown in Table 1, some significant differences were found, indicating that employees with a moderate level of CWB have, on the one hand, lower levels of job satisfaction and engagement, and on the other hand, higher levels of negative traits such as Machiavellianism and psychopathy in comparison with the low CWB group.

Subsequently, the psychological and organizational variables that predict the development of CWB were determined. First, correlations were calculated between CWB, negative personality traits, the level of work engagement, job satis-

Table 1
Differences according to groups configured based on the level of CWB.

	$t(gl)$	p	CWB	
			Employees with low CWB level	Employees with medium to moderate CWB level
Job satisfaction	3.29(101)	.002	4.80(1.56)	3.80(1.38)
Job performance	-1.20(101)	.232	4.30(1.15)	4.60(1.13)
Machiavellianism	-2.34(101)	.021	2.51(.59)	2.79(.57)
Psychopathy	-2.41(101)	.018	1.73(.58)	2.02(.56)
Narcissism	-1.49(101)	.147	2.56(.50)	2.70(.46)
Engagement	2.57(101)	.012	3.88(.70)	3.51(.69)

Table 2
Multiple hierarchical regression: CWB prediction.

	R^2	β *Standardized*	*p*
Block 1	.08		
engagement		-.25	.013
Block 2	.16		
engagement		-.23	.014
machiavellianism		.26	.017
psychopathy		.07	.498
Block 3	.26	.138	.006
engagement		-.09	.380
machiavellianism		.29	.006
psychopathy		-.06	.519
job satisfaction		-.36	< .001

faction, and job performance. Negative and significant correlations were found between CWB and job satisfaction (r = -.38, p < .001) and with engagement (r = .25, p = -.013). Positive and significant correlations were observed between CWB and Machiavellianism (r = .31, p = .002) and psychopathy (r = .24, p = .013). No significant correlations were found between CWB and job performance or the dark personality trait of narcissism (p > .05).

Next, a multiple hierarchical regression was calculated to identify the variables that increased the prediction of CWB. The selection criterion for introducing variables into the blocks considered the previously obtained correlations and excluded from the model those that were not significant. The overall work engagement score was entered in the first block; scores for the Dark Triad traits of Machiavellianism and psychopathy were entered in the second block; and the job satisfaction score was entered in the third block. In all cases, the tested models were statistically significant (p < .01).

Table 2 shows that the adjusted $R2$ of Block 1 was .08, [$F_{(1,101)}$ = 6.45, p = .013], with engagement being a statistically significant predictor. In Block 2, the adjusted $R2$ increased to .16, [$F_{(3,99)}$ = 6.08, p < .001], and this change was statistically significant (p < .001), representing an 8% increase in explained variance. Engagement remained a statistically significant predictor, and among the Dark Triad traits, Machiavellianism also emerged as a statistically significant predictor. Lastly, in Block 3, the adjusted R2 increased to .26, [$F_{(4,98)}$ = 8.63, p < .001], and this change was statistically significant (p < .001), indicating a 10% increase in explained variance. In this final block, engagement lost its significant predictive power. Instead, Machiavellianism and the level of job satisfaction emerged as significant predictors, with the latter being the most influential.

Discussion

One of the primary objectives of this study was to validate, for research purposes in Argentina, the Counterproductive Work Behavior Checklist – Short Version (CWB-C-10). The results of the analyses underscore that this version exhibits satisfactory psychometric properties of validity and reliability. Confirmatory factor analysis confirmed a good fit for the one-dimensional structure. In agreement with Stanek et al. (2017), examinations of various measures of counterproductive work behavior (CWB) proposed by different authors (e.g., Bennett & Robinson, 2000; Spector, 2006; Spector et al., 2010) reveal that, despite differences in item focus on CWB toward the organization and its members, these items are often highly correlated, suggesting a single-dimensional construct.

This instrument has the advantage of being a rapid and effective measure to evaluate counterproductive behaviors in workplace settings. It should be noted that, although a local test was designed in Argentina to measure such behaviors (Omar et al., 2012), having an internationally used instrument allows cross-cultural comparisons and offers a broader scope in studying the phenomenon.

As the test is validated for research use, it aids in understanding how this phenomenon occurs in large samples and, based on that information, it also helps to shape effective practices in organizations for a healthier environment. Contributions from Positive Organizational Scholarship (POS) and the Healthy and Resilient Organizations Model (HERO; Cameron & Spreitzer, 2012; Salanova et al., 2012) provide tools to counteract the development of negative behaviors and promote healthy resources and practices.

In light of the above considerations, conducting research such as the one presented here provides insight into potential predictors of CWB. As numerous studies have already explored, negative personality traits are associated with a higher likelihood of engaging in these detrimental actions (e.g., DeShong et al., 2015; Filipkowski & Derbis, 2020; Junça-Silva & Silva, 2023; Miller, 2017; Uysal et al., 2023), but, according to the obtained results, they do not independently explain the phenomenon. One of the major predictors of CWB is apparently linked to low levels of job satisfaction. Therefore, it is necessary to examine not only individual variables but also contextual factors that may influence the development of these behaviors. A toxic environment, where employees are under dysfunctional leadership, with competitive colleagues and a high-pressure climate, can be an influential factor. A recent study by Brassey et al. (2022), encompassing 15 countries, including Argentina, demonstrated that being immersed in a toxic environment is the primary negative predictor of optimal organizational outcomes.

Hence, in practical terms, there is a need to design instruments for assessing these aspects both in the selection processes and in evaluations conducted once employees are already part of an organization and the emergence of such behaviors becomes evident. Efforts should focus on developing techniques that effectively capture these behaviors. As Spector (2001) suggests, employees tend to engage in these behaviors covertly, and they may not be willing to disclose them in personnel selection or workplace climate assessments.

Finally, there is an urgent need to advance in the technological aspects of assessment processes. The progression of Information and Communication Technologies (ICT) increasingly enables the use of technological tools for assessing psychological phenomena in applied contexts such as organizational settings (Woods et al., 2020). For example, methodologies known as gamification, through the implementation of video games and simulated situations, allow the creation of motivating, novel, and unexpected work scenarios in a virtual reality environment. These situations place the assessed candidate or employee in a position where they must make decisions, enabling experts to evaluate the fit of the candidate with the position and organization (Fetzer et al., 2017). These technologies may facilitate the covert assessment of the propensity to engage in CWB and correct the bias of self-descriptive format instruments.

Limitations and Future Directions

As previously mentioned, the use of self-report inventories may pose a challenge in evaluating behaviors, such as those investigated in this study. Additionally, the sample size in Study 2 should be noted as a potential limitation that could impact the generalizability of the obtained results.

Regarding future lines of study, apart from the implementation of technological tools for assessment, the relevance of studying how these behaviors may manifest in new workplace environments is noteworthy. After the COVID-19 pandemic, many jobs have transitioned to virtual work, either partially or exclusively. For instance, Chong et al. (2020) observed an increase in levels of exhaustion, anxiety, and occurrences of counterproductive work behavior (CWB) in workers who engaged in telecommuting during the initial stages of the pandemic. Therefore, it is intriguing to examine whether telecommuting contributes to an escalation of such behaviors and whether they manifest differently in a virtual environment.

References

Álvarez-Escalante, J. C., Ruvalcaba-Coyaso, F. J., Romero de Loera, B. M., & Palacios-Montoya, J. C. (2021). Comportamiento antisocial en las organizaciones: Una revisión y análisis conceptual. *Revista Digital Internacional de Psicología y Ciencia Social, 7*(1), 23-41. https://doi.org/10.22402/j.rdipycs.unam.7.1.2021.330.23-41

Balzer, W. K., Kihm, J. A., Smith, P. C., Irwin, J. L., Bachiochi, P. D., Robie, C.,…Parra, L. F. (1997). User's manual for the Job Descriptive Index (JDI; 1997 Revision) and the Job in General (JIG) Scales. Bowling Green, OH: Bowling Green State University.

Bandalos, D. L., & Finney, S. J. (2018). Factor analysis: Exploratory and confirmatory. In G. R. Hancock, M. L. Stapleton & R.O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 98-122). Routledge. https://doi.org/10.4324/9781315755649

Barbaranelli, C., Fida, R., & Gualandri, M. (2013). Assessing counterproductive work behavior: A study on the dimensionality of CWB-Checklist. *Testing, Psychometrics, Methodology in Applied Psychology, 20*(3), 235-248. https://doi.org/10.4473/TPM20.3.3

Brassey, J., Coe, E., Dewhurst, M., Enomoto, K., Giarola, R., Herbig, B., & Jeffery, B. (May 27, 2022). Addressing employee burnout: Are you solving the right problem? McKinsey Health Institute. https://www.mckinsey.com

Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology, 85*(3), 349-360. https://doi.org/10.1037/0021-9010.85.3.349

Bilal, A. R., Fatima, T., & Imran, M. K. (2020). Why ostracized full-time faculty should not be labeled as "low performer"? A qualitative insight from higher education in Pakistan. *Journal of Applied Research in Higher Education, 12*(5), 805-827. https://doi.org/10.1108/JARHE-12-2018-0267

Blickle, G., & Schütte, N. (2017). Trait psychopathy, task performance, and counterproductive work behavior directed toward the organization. *Personality and Individual Differences, 109*, 225-231. https://doi.org/10.1016/j.paid.2017.01.006

Bowling, N. A., & Michel, J. S. (2011). Why do you treat me badly? The role of attributions regarding the cause of abuse in subordinates' responses to abusive supervision. *Work & Stress, 25*(4), 309-320. https://doi.org/10.1080/02678373.2011.634281

Byrne, B. M. (2006). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, application and programming* (2nd Ed.). Lawrence Erlbaum.

Cameron, K. S., & Spreitzer, G. M. (2012). What is positive about positive organizational scholarship? In K. S. Cameron & G. M. Spreitzer (Eds.), *The Oxford handbook of positive organizational scholarship* (pp. 1-14). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734610.013.0001

Cameron, K. S., Bright, D., & Caza, A. (2004). Exploring the relationships between organizational virtuousness and performance. *American Behavioral Scientist, 47*(6), 766-790. https://doi.org/10.1177/0002764203260209

Chen, H., Richard, O. C, Boncoeur, O. D., & Ford, D. L. (2020). Work engagement, emotional exhaustion, and counterproductive work behavior. *Jour-*

nal of Business Research, 114, 30-41. https://doi.org/10.1016/j.jbusres.2020.03.025

Chong, S., Huang, Y., & Chang, C. -H. (D.). (2020). Supporting interdependent telework employees: A moderated-mediation model linking daily COVID-19 task setbacks to next-day work withdrawal. *Journal of Applied Psychology, 105*(12), 1408- 1422. https://doi.org/10.1037/apl0000843

De Clercq, D., Haq, I. U., & Azeem, M. U. (2019). Why happy employees help: How meaningfulness, collectivism, and support transform job satisfaction into helping behaviours. *Personnel Review, 48*(4), 1001-1021. https://doi.org/10.1108/PR-02-2018-0052

DeShong, H. L., Grant, D. M., & Mullins-Sweatt, S. N. (2015). Comparing models of counterproductive workplace behaviors: The five-factor model and the dark triad. *Personality and Individual Differences, 74*, 55-60. https://doi.org/10.1016/j.paid.2014.10.001

Fetzer, M., McNamara, J., & Geimer, J. (2017). Gamification, serious games and personnel selection. In H. Goldstein, E. Pulakos, J. Passmore & C. Semedo (Eds.), *The Wiley Blackwell handbook of the psychology of recruitment, selection and employee retention* (pp. 293-309). Wiley. https://doi.org/10.1002/9781118972472.ch14

Filipkowski, J., & Derbis, R. (2020). The Dark Triad, goal achievement and work engagement as predictors of counterproductive work behaviors among IT employees. *Current Issues in Personality Psychology*, *8*(3), 197-210. https://doi.org/10.5114/cipp.2020.100095

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Prentice Hall.

Hu, L., & Bentler, P. (1995). Evaluating model fit. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 76-99). Sage.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment, 21*(1), 28-41. https://doi.org/10.1177/1073191113514105

Junça-Silva, A., & Silva, D. (2023). The buffering effect of micro-daily events on the relationship between the Dark Triad traits and counterproductive work behavior. *Management Research Review, 46*(5), 667-681. https://doi.org/10.1108/MRR-12-2021-0864

Kline, P. (2000). *Handbook of psychological testing*. Routledge.

Lebron, M., Tabak, F., Shkoler, O., & Rabenu, E. (2018). Counterproductive work behaviors toward organization and leader-member exchange: The mediating roles of emotional exhaustion and work engagement. *Organization Management Journal, 15*(4). https://scholarship.shu.edu/omj

Lupano-Perugini, M. L. (2017). Virtudes organizacionales, performance y satisfacción laboral. Diferencias según variables individuales y organizacionales en empleados argentinos. *Psicodebate, 17*(1), 35-50. http://doi.org/10.18682/pd.v17i1.637

Lupano-Perugini, M. L., de la Iglesia, G., Castro-Solano, A., & Fernández-Liporace, M. (2017). Validación de una escala sobre work engagement. Perfiles asociados a alta performance y satisfacción laboral. *Ciencias Psicológicas, 11*(2), 127-137. https://doi.org/10.22235/cp.v11i2.1482

Malik, N. I., & Zahra, A. (2022). Impact of job engagement on counterproductive work behaviors: Moderating role of gender among organisational employees. *Advances in Psychology and Behavioral Studies, 1*(1), 1-8. http://www.metisjournals.com/index.php/apbs

Medrano, L. A., Ortiz, A., Flores-Kanter, P. E., Dominguez-Lara, S., & Gibelli, S. (2018). Medición de recursos personales socio-cognitivos en organizaciones: Análisis psicométricos en trabajadores argentinos. *Revista Evaluar, 18*(2), 1-16. https://doi.org/10.35670/1667-4545.v18.n2.20799

Mercado, B., Dilchert, S., Giordano, C., & Ones, D. (2018). Counterproductive work behaviors. In D. S. Ones, N. Anderson & C. Viswesvaran (Eds.), *The SAGE handbook of industrial, work and organizational psychology* (pp. 109-210). Sage. https://doi.org/10.4135/9781473914940.n7

Miller, M. (2017). *The influence of dark personality on counterproductive work behaviors*. (Tesis de Maestría en Ciencias en Psicología Aplicada). University of Wisconsin-Stout, Wisconsin. https://doi.org/10.13140/RG.2.2.34488.96004

Moral de la Rubia, J. (2019). Revisión de los criterios para validez convergente estimada a través de la varianza media extraída. *Psychologia. Avances de la Disciplina, 13*(2), 25-41. https://doi.org/10.21500/19002386.4119

Morf, M., Feierabend, A., & Staffelbach, B. (2017). Variedad de tareas y comportamiento laboral contraproducente. *Revista de Psicología Gerencial, 32*(8), 581-592. https://doi.org/10.1108/jmp-02-2017-0048

Ng, T. W. H., & Feldman, D. C. (2008). The relationship of age to ten dimensions of job performance. *Journal of Applied Psychology, 93*(2), 392-423. https://doi.org/10.1037/0021-9010.93.2.392

O'Boyle, E. H., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the Dark Triad and work behavior: A social exchange perspective. *Journal of Applied Psychology, 97*(3), 557-579. https://doi.org/10.1037/a0025679

Omar, A., Vaamonde, J., & Delgado, H. (2012). Comportamientos contraproducentes en el trabajo: Diseño y validación de una escala. *Diversitas, 8*(2), 249-265. https://doi.org/10.15332/s1794-9998.2012.0002.04

Ones, D. S., & Dilchert, S. (2013). Counterproductive work behaviors: Concepts, measurement, and nomological network. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 643-659). American Psychological Association. https://doi.org/10.1037/14047-035

Paulhus, D. L. (2014). Toward a taxonomy of dark personalities. *Current Directions in Psychological Science, 23*(6), 421-426. https://doi.org/10.1177/0963721414547737

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of Research in Personality, 36*(6), 556-563. https://doi.org/10.1016/S0092-6566(02)00505-6

Pletzer, J. L. (2021). Why older employees engage in less counterproductive work behavior and in more organizational citizenship behavior: Examining the role of the HEXACO personality traits. *Personality and Individual Differences, 173*(1), 110550. https://doi.org/10.1016/j.paid.2020.110550

Rauf, K., & Farooq, A. (2014). Adaptation and validation of Counterproductive Work Behavior Checklist (45 and 32). *International Journal of Novel Research in Humanity and Social Sciences, 1*(1), 39-49. https://www.noveltyjournals.com/journal-details/IJNRHSS

Rehman, U., & Shahnawaz, G. (2018). Machiavellianism, job autonomy and counterproductive work behaviour among Indian managers. *Journal of Work and Organizational Psychology, 34*(2), 83-88. https://doi.org/10.5093/jwop2018a10

Rial-Boubeta, A., Varela-Mallou, J., Abalo-Piñeiro, J., & Lévy-Mangin, J. P. (2006). El análisis factorial confirmatorio. In J. P. Lévy-Mangin & J. Varela (Eds.), *Modelización con estructuras de covarianzas en ciencias sociales* (pp. 119-143). Netbiblo.

Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal, 38*(2), 555-572. https://doi.org/10.2307/256693

Rothbard, N., & Patil, S. (2012). Being there: Work engagement and positive organizational scholarship. In K. S. Cameron & G. M. Spreitzer (Eds.), *The Oxford handbook of positive organizational scholarship* (pp. 57-69). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734610.013.0005

Sackett, P. (2002). The structure of counterproductive work behaviors: Dimensionality and relationships with facets of job performance. *International Journal of Selection and Assessment, 10*(1-2), 5-11. https://doi.org/10.1111/1468-2389.00189

Salanova, M., Llorens, S., Cifre, E., & Martínez, I. M. (2012). We need a Hero! Toward a validation of the Healthy and Resilient Organization (HERO) model.

Group and Organization Management, 37(6), 785-822. https://doi.org/10.1177/1059601112470405

Salessi, S., & Omar, A. (2018). Tríada oscura de personalidad, satisfacción laboral y cinismo organizacional: Un modelo estructural. *Universitas Psychologica, 17*(3), 1-12. https://doi.org/10.11144/Javeriana.upsy17-3.tops

Schumacker, R. E., & Lomax, R. G. (2016). *A beginner's guide to structural equation modeling* (4th Ed.). Routledge.

Selvarajan, T. T., Singh, B., Cloninger, P., & Misra, K. (2019). Work–Family conflict and counterproductive work behaviors: Moderating role of regulatory focus and mediating role of affect. *Organisation Management Journal, 16*(1), 42-54. https://doi.org/10.1080/15416518.2019.1574547

Spector, P. E. (1997). *Job satisfaction: Application, assessment, causes and consequences.* Sage.

Spector, P. (2001). Research methods in industrial and organizational psychology: Data collection and data analysis with special consideration to international issues. En P. Spector (Ed.), *Handbook of industrial, work and organizational psychology - Volume 1: Personnel psychology* (pp. 10-26). Sage. https://doi.org/10.4135/9781848608320.n2

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods, 9*(2), 221-232. https://doi.org/10.1177/1094428105284955

Spector, P. E. (2022). *Job satisfaction: From assessment to intervention*. Routledge. https://doi.org/10.4324/9781003250616

Spector, P. E., Bauer, J. A., & Fox, S. (2010). Measurement artifacts in the assessment of counterproductive work behavior and organizational citizenship behavior: Do we know what we think we know? *Journal of Applied Psychology, 95*(4), 781-790. https://dx.doi.org/10.1037/a0019477

Spector, P. E., & Fox, S. (2002). An emotion-centered model of voluntary work behavior: Some parallels between counterproductive work behavior and organizational citizenship behavior. *Human Resource Management Review, 12*(2), 269-292. https://doi.org/10.1016/S1053-4822(02)00049-9

Spector, P. E., Fox, S., Penney, L. M., Bruursema, K., Goh, A., & Kessler, S. (2006). The dimensionality of counterproductivity: Are all counterproductive behaviors created equal? *Journal of Vocational Behavior, 68*(3), 446-460. https://doi.org/10.1016/j.jvb.2005.10.005

Stanek, K. C., Ones, D. S., & McGue, M. (2017). Counterproductive behaviors: Relations across life domains, etiology, and implications for applied practice. *International Journal of Selection and Assessment, 25*(2), 111-124. https://doi.org/10.1111/ijsa.12164

Uysal, H. T., Ak, M., & Koşar, A. (2023). Counterproductive work behaviors in organizations: A research on the effect of the Dark Triad. *AGATHOS, 14*(1), 247-259. https://www.agathos-international-review.com/index.html

Vélez-Vega, A. (2022). Psychometric properties of the Office Gossip Scale in Puerto Rico. *Revista Evaluar, 22*(1), 57-72. https://doi.org/10.35670/1667-4545.v22.n1.37443

Woods, S., Ahmed, S., Nikolaou, I., Costa, A., & Anderson, N. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology, 29*(1), 64-77. https://doi.org/10.1080/1359432X.2019.1681401

2024, Vol. 24, No. 2
ISSN 1667-4545
Recuperado de https://revistas.unc.edu.ar/index.php/revaluar

Laboratorio de Evaluación Psicológica y Educativa
Facultad de Psicología - Universidad Nacional de Córdoba

Assessment of Critical Thinking Skills in Primary Education: Validation of Challenges of Thinking Test

Evaluación de destrezas de pensamiento crítico en Educación Primaria: Validación de la prueba Desafíos del Pensamiento

Maria-Antonia Manassero-Mas * [1], Ángel Vázquez-Alonso [2]

1 - University of the Balearic Islands, Palma, Spain.
2 - University of the Balearic Islands, Palma, Spain.

Recibido: 22/03/2024 **Revisado:** 05/06/2024 **Aceptado:** 08/06/2024

Abstract

The global demand for 21st century competencies raises critical thinking (CT) as a priority educational objective, which in turn projects the need for CT evaluation. The lack of CT assessment instruments for youngsters justifies the aim of this study: to develop a CT test for Primary Education and unveil its psychometric properties. The methodology follows the test development prescriptions through the elaboration of the six-skill CT test, the test application to primary sixth-graders and the confirmatory factor analysis on the answers. Starting from a 48-item test form, the empirical analysis confirms a six-factor structure, interpret the six empirical factors in face of the postulated CT skills, confirm a one-dimension structure for the whole test and for each of the six factors (except Comparison), and describe the goodness-of-fit psychometric parameters that support the reliability and validity of the 31-item test form. Finally, the properties, utility, limitations, and prospective improvements, developments and applications of the test for education and CT research are discussed.

Keywords: *student evaluation, critical thinking test, validity, reliability, primary education*

Resumen

La demanda global de competencias del siglo XXI plantea el pensamiento crítico (PC) como un objetivo educativo prioritario, lo que a su vez proyecta la necesidad de la evaluación del PC. La falta de instrumentos de evaluación del PC para jóvenes justifica el objetivo de este estudio: desarrollar un test de PC para Educación Primaria y presentar sus propiedades psicométricas. La metodología sigue las prescripciones de desarrollo de pruebas a través de la elaboración de una prueba de PC con seis destrezas, la aplicación de la prueba a estudiantes de sexto grado de primaria y el análisis factorial confirmatorio de las respuestas. Partiendo de un formulario de prueba de 48 ítems, el análisis empírico confirma una estructura de seis factores, interpreta los seis factores empíricos frente a las destrezas de PC postuladas, confirma una estructura unidimensional para toda la prueba y para cada uno de los seis factores (excepto Comparación), y describe los parámetros psicométricos de bondad de ajuste que respaldan la confiabilidad y validez de una forma de la prueba con 31 ítems. Finalmente, se discuten las propiedades, utilidad, limitaciones y posibles mejoras, desarrollos y aplicaciones de la prueba para la educación y la investigación en PC

Palabras clave: *evaluación de estudiantes, prueba de pensamiento crítico, validez, confiabilidad, educación primaria*

***Correspondence to:** Ángel Vázquez-Alonso. E-mail: angel.vazquez@uib.es. Address: Universidad de las Islas Baleares, Edificio Guillem Cifre de Colonya. Carretera de Valldemossa, km. 7.5, 07122 - Palma, Spain.
Tel. +34 971 17 30 75. FAX: 34 971 17 31 90.

How to cite: Manassero-Mas, A., M., & Vázquez-Alonso, Á.,(2024): Assessment of Critical Thinking Skills in Primary Education: Validation of Challenges of Thinking Test. *Revista Evaluar, 24*(2), 53-75. Retrieved from https://revistas.unc.edu.ar/index.php/revaluar
Acknowledgments: EDU2015-64642-R (AEI/FEDER, UE) funded by the Spanish State Research Agency and the European Regional Development Fund.
Declaration of conflict of interest: none.
Participaron en la edición de este artículo: Stefano Macri, Eugenia Barrionuevo, Abigail Pérez, Florencia Ruiz, Jorge Bruera.

Introduction

Critical thinking is currently an overarching concept in psychology, philosophy, education and job. Since centuries philosophers adopted CT as their working tool to quality thinking that brings along high standards (precision, solidity, coherence, etc.) and to avoid error, fallacy and bias (ego-centrism and socio-centrism) (Ennis, 2018; Bailin et al., 1999; Facione, 1990). Psychologists developed CT as a set of higher order cognitive skills (e.g., inference, analysis, problem-solving, interpretation, creativity, decision-making, evaluation, etc.) along with a set of attitudinal dispositions (e.g., truth-seeking, self-confidence, curiosity, open-mindedness, etc.) that drive the adequate application of skills (Fisher, 2021; Halpern, 2003; Manassero-Mas et al., 2022).

Worldwide educational institutions and experts are claiming CT as a core part of the 21st century skills that citizens need to face the current challenges (globalization, accelerated development, ecological emergency, etc.) and their consequent personal, labor, and social impacts (Almerich et al., 2020; European Union, 2014; Fullan & Scott, 2014; International Society for Technology Education, n.d.; National Research Council, 2012; Organisation for Economic Co-operation and Development [OECD], 2018; UNESCO, 2016; Vincent-Lancrin et al., 2019). From the employers' view, most surveys reiterate CT at the top of skills required for future jobs (Whiting, 2021) and a key factor for people's success in the information age (Tremblay, 2013).

The mastery of CT is a key educational factor for significant and deep learning that characterizes educational excellence (Hattie, 2012; Valenzuela, 2008). In fact, CT aligns with Piaget's pioneering studies (Piaget & Inhelder, 1997) and the cognitive acceleration programs (Shayer & Adey, 2002), which have empirically demonstrated their significant impact on learning (effect size = 1.28), according to Hattie's (2009, 2012) meta-analysis of visible learning, which additionally assigns a large impact to some CT skills (meta-cognitive strategies, creativity, problem-solving, etc.).

The beneficial consequences of mastering CT for learning, personal development and success in job and social contexts justify the research attention to CT. However, teaching, assessing and making CT visible is still difficult for most elementary schools due to the lack of CT resources and tests for primary education, as justified below. In order to fill in these gaps, this study aims to develop a quantitative, valid and reliable CT test for primary education from an educational, diagnostic and evaluation perspective.

The conceptual framework on critical thinking

The literature research on CT developed along three basic lines since years: conceptualization, teaching, and evaluation, though each line has achieved a misbalanced development (Saiz, 2017).

The conceptualization of CT displays a lack of consensus, as the researchers display a big diversity of concepts and terms among that impede achieving a shared definition for CT (e.g., Ennis, 2018; Facione, 1990; Halpern, 2003; Paul & Nosich, 1993). The Ennis' (2018) conceptualization of CT, as reflective and reasonable thinking focused on deciding what to believe or do, and its expanded development in dispositions and skills involved in such decisions are widely cited. In the pursuit of consensus, a panel of experts from the American Philosophical Association (APA) proposed a definition of CT as the deliberate and self-regulating judgment for a specific purpose, employing evidence-based interpretation, analysis, evaluation, and inference, concepts, methods, criteria, and contexts to establish such judgments (Facione, 1990).

As an alternative way, some researchers choose to define CT by specifying its constitutive skills, yet this perspective has not achieved consensus either (Fisher, 2009). For instance, the aforementioned APA panel definition proposed the following skills: interpretation, analysis, evaluation, inference, interpretation, judgment, and self-regulation. On the other hand, the national plan for the assessment of CT proposed an 88-skill list, yet they were grouped in dimensions (Paul & Nosich, 1993).

Further, the practical function of CT assessment tests requires clearly specifying the skills they evaluate, so that the analysis of tests' specific skills might shed more light on CT conceptualization than CT definitions. However, the comparison of the different CT assessment tests shows the sets of skills are different across tests, again lacking relevant coincidences with each other. Thus, CT tests also lead to conceptual discrepancy and complexity about the CT construct along its constitutive skills. The different skills labels, the unequal number of test skills (ranging between 88 and 2), and the categories of skills add to the CT complexity (Manassero-Mas & Vázquez-Alonso, 2019).

Some synthetical taxonomies have been proposed to alleviate the lack of consensus on conceptualizing CT. Dwyer et al. (2014) developed an integrated framework of educational objectives, cognitive processes (reflective judgment and self-regulation, and meta-cognition) and CT skills (analysis, evaluation, and inference), with memory and comprehension as necessary processes for CT application. Two recent theoretical frameworks coincidentally organize CT into four similar dimensions. Manassero-Mas and Vázquez-Alonso (2019) proposed CT as the foundational construct, which develops along four dimensions (creativity, reasoning and argumentation, complex processes, and evaluation and judgment), each containing categories of thinking skills (e.g., deductive, inductive, abductive, statistical thinking, problem-solving, decision-making) and other associated concepts (assumptions, standards, attitudinal dispositions, norms). Similarly, Fisher (2021) organized CT skills across four basic groups (interpretation, analysis, evaluation, and self-regulation), whose contents widely overlap with the previous taxonomy.

All in all, the CT conceptualization shows important differences across authors, both theoretical and practical. The synthetical and integrative taxonomies provide a balanced view of the CT, echo the CT skills involved in most CT tests and avoid a dysfunctional alphabet soup in the field. Herein, the Manassero-Masa and Vázquez-Alonso' (2019) taxonomy will be used as an overall reference to frame the researched CT test.

Teaching and evaluating critical thinking

In spite of the conceptual disagreements, all CT experts endorse the assumption that thinking can be taught and learnt, which lead to the development of a variety of teaching and assessment programs (Follmann et al., 2018; Saiz, 2017; Swartz et al., 2013). The recommendations (12 and 13) of APA experts' statement (Facione, 1990) advocated frequent, explicit, diagnostic and summative evaluations of CT, through valid, reliable and equitable tools, currently obvious features of tests (Muñiz & Fonseca-Pedrero, 2019). Further, Ennis (2018) provided many reasons to assess CT: diagnostic of students' skills, feedback on program progress, motivation to learn CT, information about teaching, investigate CT, counsel about study choice and stimulation to report results.

The evaluation of CT requires the construction of appropriate tests to assess valid and reliable measurements, plus some of those tests have been developed. Most tests focus on assessing a few CT skills (e.g., Facione et al., 1998; Halpern, 2010; Rivas & Saiz, 2012; Watson & Glaser, 2002), yet some are broader and ambitious (Madison, 2004). The analysis of the CT skills included in the tests lead to synthesizing the mentioned taxonomies (Manassero-Mas & Vázquez-Alonso, 2019; Ennis, 2009; Fisher, 2021).

However, the programs that have proven their effects through empirical evaluation are the exception rather than the rule (Saiz, 2017). Lipman's (1982) philosophy for children has been repeatedly evaluated (Colom et al., 2014), while others have only been occasionally appraised (e.g., thinking-based learning, Swartz et al., 2013), and still others lack evaluations (e.g., the reasoning program, Walton & Macagno, 2015). So, this paper tries to construct a valid and reliable test that can be deemed functional for feedback on the educational programs of primary education.

The vast majority of CT assessment instruments address adults and university students, and there are hardly any tests for young students, though some items of the Cornell tests may be adaptable to young people (Ennis & Millman, 2005a, 2005b), and other proposals require further development (Lopes et al., 2018). The review of Aktoprak and Hursen (2022) diagnoses the scarcity and weaknesses of CT research in primary education, proposes increasing it by intensifying the evaluation of CT skills in educational projects, and points out to use reliable tests to complement the predominant qualitative methods in primary (e.g., Gelerstein et al., 2016). Thus, the new test developed here adheres to this proposal on quantitative and functional CT assessment for primary education.

In sum, the scant attention paid to the youngest students' teaching and assessment of CT lead to the inadequate of the previously mentioned tests for children. The growing importance of CT in education, as a key constituent of 21st century skills, points out a return to this situation and justifies the development of a new test to evaluate young people's CT. This aim involves the test content being adapted to the developmental ability of the primary students: focus on some specific, appropriate, and functional skills, adequate the item cognitive demand and make the test independent of the curricular knowledge. Further, the test must meet a balanced development of each of the dimensions of the CT taxonomy that has been adopted as a reference (Manassero-Mas & Vázquez-Alonso, 2019), to make thinking visible at early educational stages. Consequently, the objective of this study is to develop an assessment instrument to quantitatively diagnose the CT skills of young primary school students, investigate their relationship with learning (represented by school grades as empirically validated external criteria), and apply confirmatory analytical methods to support the psychometric validity and reliability of the instrument.

This study continues a work that has already developed some previous stages: the construction of a wide bank of Spanish items on CT skills, which allow the development of some pilot applications involving many items and small samples of sixth-graders, in order to classify the tested items according to their difficulty, their fit within CT skills and their mutual correlations (Manassero-Mas & Vázquez-Alonso, 2020a, 2020b). On this basis, a previous study developed and evaluated a test form that achieved hopeful results with sixth graders, yet the five-skill final test still left room for its validation improvement (Manassero-Mas & Vázquez-Alonso, 2024). Thus, applying the usual recommendations for test development to the former five-skill final test, a new 48-item and six-skill test was raised, as the starting point of this validation study (Ferrando et al., 2022; Muñiz & Fonseca-Pedrero, 2019). Likewise, the results of this 48-item test were applied as a diagnostic evaluation of the Spanish primary school students' thinking skills (Manassero-Mas & Vázquez-Alonso, 2023).

Methodology

A new 48-item and six-skill test (mentioned above) is the starting form of this new validation study through its application to assess the CT skills of elementary students. The methodology and results are presented here.

Participants

The complete 48-item test was applied to a varied convenience sample comprising 655 students (320 male and 335 female) sixth graders of Primary Education, aged between 10 and 13 years (average 11.1 years). The students attended fourteen public and public-funded schools, which were located in a variety of places in three regions. The schools were willing towards teaching and learning thinking and students were tested by their own teachers within their entire natural groups, as a classroom activity on thinking assessment. The database was cleaned taking into account responses that were potentially biased, highly empty or lacking attention (as reported by teachers).

Instrument

The Challenges of Thinking Test (CoTT_PE6) is designed to measure six CT skills (Manassero-Mas & Vázquez-Alonso, 2019): prediction and logical reasoning (from the reasoning dimension), comparison (creativity dimension), classification (evaluation dimension), and decision-making and problem-solving (complex processes dimension). Classification assesses the ability to group or separate different elements according to the appraisal of various common or differential features. Prediction and Comparison assess the ability to verify a conclusion from inductive reasoning or from the creative contrast between several statements, respectively. Decision-making and Problem-solving measure the ability to identify the best (worst) decisions/solutions in a particular situation. Logical reasoning assesses simple (simple syllogism) and complex (several pieces of information and conclusions involved simultaneously) deductive abilities.

The skills were agreed by the researchers and the teachers of the participating schools considering their fit to the usual cognitive demands in the sixth grade of primary school (PE6). The researchers selected the test items from the five-skill test previously validated and the piloted items of the item bank through the following criteria: simplicity of wording, ease reading comprehension, balance between item cognitive demand and target students' cognitive development, and motivating and interesting challenge for students (e.g., a simple story on futurist planet exploration, many items with figurative contents and logical reasoning on pencils, books and colors). Then, the researchers assigned each item to the skill that presented the greatest congruence with the item content (Table 1).

Initially, many CT items were selected, translated and adapted from some original standardized CT tests (Ennis & Millman, 2005a, 2005b; Halpern, 2010) and from scholar CT publications (https://www.criticalthinking.com) affordable for

Table 1
Specifications of the two tests applied (CoTT_PE6) in this study to evaluate thinking skills in sixth-grade primary education PE6.

Thinking skills	Item Source	Type	Number of items	
			Initial (48)	Final (31)
Prediction (PREDIC)	Ennis & Millman, 2005a	Verbal	9	6
Comparison (COMPA)		Verbal	7	3
Classification (CLASSIF)	Author elaboration*	Figurative	6	5
Problem-solving (PROBL)	Halpern (2010)	Verbal	6	4
	Author elaboration*	Figurative	4	1
Decision-making (DECISION)	Author elaboration*	Mixed	9	5
Logical reasoning (LOG-REAS)	Author elaboration*	Figurative	-	2
	Ennis & Millman, 2005b	Verbal	7	5

Note. * Translated and adapted from open materials (https://ww.criticalthinking.com).

primary students. The publications provided the figurative items and both made explicit the cognitive demands and the specific item skills. Further, the researchers' professional judgment consensually scrutinized and selected a bunch of items again, by reviewing the best fit between item content and skills and between item's cognitive demand and primary students' cognitive stage. The subsequent item set was piloted and the analysis of the results set up the 48-item CoTT, which is analyzed here (Table 1). All in all, the scholarly nature of the managed item sources and the selection and pilot processes warrant an accurate item-skill correspondence and a sound contribution to assure the content validity of CoTT *ab initio*.

The items pose authentic and motivating thinking challenges for students through a variety of scenarios and situations, where information is communicated by several means of representation (verbal, numerical, and figurative), and their cognitive demands fit the represented skill and the students' evolutive stage. Further, the item contents are independent of the school curricula (for example, they do not involve numerical calculations), so achieving the correct answer just requires applying reasoning to the existing information and does not need any previous knowledge. Therefore, CoTT is considered a culture-free test, as its challenges are not mediated by academic knowledge, as is often the case. For example, the Science CT test requires primary science curriculum knowledge to answer correctly (Mapeala & Siew, 2015).

The response formats are mostly closed and the four items asking for a short open answer were dealt with a simple rubric to code them as correct/wrong. This format allows for a standardized, fast, valid, and reliable evaluation of thinking skills, for the development of diagnostic baselines to objectively compare different research, programs, and teaching methodologies, and for practical use by teachers. Correct answers scored one point and incorrect answers zero, the score of each skill is the sum of the correct answers achieved in their assigned items, and the overall score is the sum of the total correct answers, which is considered an estimate of the students' global CT performance.

Data collection and analysis

The CoTT was applied to the students by their teachers within their class group, as an ordinary regular activity of school evaluation to stimulate the students' efforts and motivation on thinking. To ensure the application consistency the authors supervised the class applications, which followed the usual standardized guidelines for tests, without any time limit (usually a class period); the guidelines were written at the first screen of the test digital device and were read aloud by teachers and students.

The procedures involved a two-stage action. The first stage performed the construction of the 48-item CoTT on the basis of the CT item bank, its application to a large sample of sixth-grade primary school students and the analysis of their responses through exploratory (EFA) and confirmatory factor analysis (CFA) of the restricted 46-item CoTT_PE6_46 (two items of a triplet were eliminated). The second stage involved analyzing the former results to eliminate some items with inadequate psychometric traits to leave a final shorter 31-item form CoTT_PE6_31, which is again scrutinized through EFA and CFA to set up its psychometric properties.

Data were processed with the programs SPSS (25), Amos 23.0.0 and Factor (12.01.02). SPSS and Amos are well-known statistical tools and Factor provides computation of tetrachoric correlations (appropriate for test dichotomous scores) and develops EFAs and CFAs that extract factors with a robust method of unweighted least squares (RULS), parallel analysis, bootstrap sampling, Promin rotations and several indices of reliability (Ferrando & Lorenzo-Seva, 2017, 2018; Lorenzo-Seva & Ferrando, 2019).

Results

The statistical descriptors of the items of the two tests in the two stages of the study, obtained from the students' answers, are summarized in Table 2.

The second column of Table 2 presents the correct answer average for the 48 items that constitute the starting point of this study. Most of the items (39) achieve an intermediate mean of correct answers (.30 – .70), a few items (3) are very easy ($M > .70$), and others (6) are difficult ($M < .30$). The item distribution by quartiles is as follows: 12 (25.0%) items are in the lower quartile (1), 9 (18.7%) items are in the lower-middle quartile (2), 13 (27.1%) items are in the upper middle quartile (3) and 14 (29.2%) items are in the upper quartile (4). The overall average of correct answers is close to 50% ($M = .485$), which confirms the moderate difficulty of the test, as befits this kind of test.

Lastly, three items of the initial 48-item CoTT_PE6 instrument displayed quite high correlations among them to consider they form a triplet. Thus, two of them (PROBL11 and PROBL12) were eliminated, and the subsequent analyses refer to the remaining 46 items (Ferrando et al., 2022).

Factor analysis

The analysis of the 46 items with the RULS method and tetrachoric correlations have got an unfavorable value for the Kaiser-Meyer-Olkin (KMO) parameter (0.206). However, a solution of six empirical factors (as theoretically required) produced quite acceptable goodness-of-fit parameters through minimum rank parallel CFA (Table 3). However, the rotated six-factor solution did not allow a theoretically consistent interpretation of the factors, as it displayed many items with low factor loads or with overlaps on several factors.

To increase the model coherence, the 46 items were scrutinized to remove the items that attain most of the following psychometric parame-

Table 2
Proportion of item average correct answers (difficulty index) of the CoTT_PE6 instrument in a sample of 6th-grade students (n = 655).

Initial items (48)		**Average Correct Answers (0-1)**	**Standard deviation**	**Quartile**	**Final items (31)**
V1	PREDIC1	.623	.485	4	PREDIC1
V2	PREDIC2	.431	.496	2	*
V3	PREDIC3	.499	.500	3	PREDIC3
V4	PREDIC4	.400	.490	2	*
V5	PREDIC5	.791	.407	4	PREDIC5
V6	PREDIC6	.736	.441	4	PREDIC6
V7	PREDIC7	.708	.455	4	PREDIC7
V8	PREDIC8	.380	.486	2	PREDIC8
V9	PREDIC9	.638	.481	4	*
V10	COMPA1	.441	.497	2	COMPA1
V11	COMPA2	.565	.496	3	*
V12	COMPA3	.431	.496	2	*
V13	COMPA4	.521	.500	3	*
V14	COMPA5	.499	.500	3	COMPA5
V15	COMPA6	.501	.500	3	COMPA6
V16	COMPA7	.356	.479	1	*
V17	CLASSIF1	.635	.482	4	PROBLº
V18	CLASSIF2	.553	.498	3	CLASSIF2
V19	CLASSIF3	.559	.497	3	CLASSIF3
V20	CLASSIF4	.653	.476	4	CLASSIF4
V21	CLASSIF5	.663	.473	4	CLASSIF5
V22	CLASSIF6	.640	.480	4	CLASSIF6
V23	PROBL1	.649	.478	4	*
V24	PROBL2	.562	.497	3	PROBL2
V25	PROBL3	.485	.500	3	*
V26	PROBL4	.325	.469	1	PROBL4
V27	PROBL5	.627	.484	4	PROBL5
V28	PROBL6	.621	.485	4	PROBL6
V29	DECISION1	.351	.478	1	*
V30	DECISION2	.282	.451	1	DECISION2
V31	DECISION3	.298	.458	1	*
V32	DECISION4	.328	.470	1	DECISION4
V33	DECISION5	.426	.495	2	DECISION5
V34	DECISION6	.185	.388	1	DECISION6
V35	DECISION7	.145	.352	1	*
V36	DECISION8	.646	.479	4	*
V37	DECISION9	.464	.499	2	DECISION9
V38	PROBL9	.211	.408	1	LOG-REASº
V39	PROBL10	.426	.495	2	PROBL10
-	PROBL11	.536	.499	3	-
-	PROBL12	.377	.485	2	-
V40	LOG-REAS1	.540	.499	3	LOG-REAS1
V41	LOG-REAS2	.554	.497	3	LOG-REAS2
V42	LOG-REAS3	.305	.461	1	*
V43	LOG-REAS4	.588	.493	4	LOG-REAS4
V44	LOG-REAS5	.235	.424	1	*
V45	LOG-REAS6	.574	.495	3	LOG-REAS6
V46	LOG-REAS7	.327	.469	1	LOG-REAS7

Note. - Items eliminated for being part of a triplet (observed correlations 0.853; 0.957; 0.896).
* Items eliminated in the validation process of the final instrument.
º New skill of the items that have changed its final skill assignment through the validation process.

ters: negative values in the standardized tetrachoric correlation matrix; sampling adequacy measure values less than .50; negative, null or crossed factor loads in the rotated matrix; low standardized regression loads between the item and the empirical factors; MIREAL (Mean of Item REsidual Absolute Loadings) parameter greater than .30; approximately meeting the optimal standard that 75% items of the item pool achieve an intermediate range (.40 – .60) of the relative difficulty index, and the remaining are evenly distributed in both tails (Ferrando & Lorenzo-Seva, 2017).

The joint qualitative application of the former criteria leads to the identification of 12 items that showed deficiencies in several criteria, thus deciding their elimination to improve the test (PREDIC2, PREDIC4, PREDIC9, COMPA3, COMPA4, PROBL1, DECISION1, DECISION3, DECISION8, LOG-REAS3, LOG-REAS5, COMPA7).

The resulting set of 34 items was again analysed with the RULS method and tetrachoric correlations. The results showed a better but still low value of the KMO parameter (0.545), although other parameters improved, and some were excellent. The average difficulty index was now .503, and the reliability factor was high (α = .83). The six empirical factors explained 46% of the variance, and the excellent goodness-of-fit indices showed that the data adequately fit an empirical six-factor model (RMSEA= .036 CFI = .973, GFI = .956, RMSR = .05), which also shows closeness to unidimensionality assessment (MIREAL = .193) (Ferrando & Lorenzo-Seva, 2018). Despite these good parameters, three items presented zero or slightly negative factor loadings on all the factors of the rotated matrix and nonsignificant standardized regression coefficients. Further, two of them also displayed multiple negative correlations, and still another item showed a high difficulty index. Therefore, these three items were also removed (COMPA2, PROBL3, DECISION7).

The set of the remaining 31 items was again reanalyzed to get the new values of CFA parameters that may validate the test final form (second column, Table 3), although the KMO parameter (.578) is still moderate. The average difficulty index (.514) and the reliability coefficient are also good (α = .838). The scree-plot displays a main eigenvalue that explains 19% of the total explained variance and a main elbow respect the following eigenvalues; then, the soft decrease displays a smaller elbow between the eigenvalues 6 and 7, where the first six make contributions to the variance well over 5%, and the whole six first factors explain 49%, while the following decreasingly contribute less than 4%. Thus, a CFA was applied to the six-factor model, whose rotated loading matrix showed no cross-factor loads greater than .30 between the factors, only two factor loadings were less than .20 (COMPA1 and LOG-REAS7) and an interpretable structure of factors. The CFA goodness-of-fit indices show an excellent fit for the six-factor empirical model (MIREAL=.186, RMSEA = .032, CFI = .982, GFI = .965, RMSR = .049). Further, the CFA analysis of the six-factor empirical model is close to one-dimension and the reliability parameters (ORION) of factors are also good (.735 – . 999).

To test whether the covariance structure of 31 items can also be satisfactorily explained by a single general factor, a general factor model was analyzed through CFA (right column, Table 3). The comparison of six-factor and general factor models through the likelihood ratio test (chi^2 = 1391.096, *df* = 140, *p* < .000) is significant, which means both models are different. Further, the six-factor model significantly increments its scores of RMSEA, NNFI, CFI, GFI and residuals in relation to those of the general factor model, and attains the thresholds prescribed for these parameters, while the general factor model does

Table 3
Robust statistical parameters of the confirmatory goodness of fit of the contrasted factor models for the initial test (46 items) and the final test (31 items).

	Contrasted models		
	46 items		31 items
Extracted factors	6	6	1
Kaiser-Meyer-Olkin	.206	.578	.578
Bartlett (Sig.)	-	.000	.000
Explained variance	.382	.489	.192
Goodness of fit			
RMSEA*	.041	.032	.071
Chi-square	1633.367	489.959	1880.985
Chi-square (p)	.000	.000 (*df* = 294)	.000 (*df* = 434)
NNFI**	.934	.971	.857
CFI***	.951	.982	.867
GFI****	.932	.965	.866
Residuals			
RMSR[o]	.058	.049	.097
WRMSR[oo]	.036	.030	.053
Reliability			
EAP-GLB[ooo]	.981	.963	.963
Omega	.850	.822	.822
Cronbach Alpha	.853	.838	.838
ORION[a]-Factor1	.913 (CLASSIF)	.922 (CLASSIF)	
ORION[a]-Factor2	.809 (PROBL)	.999 (PREDIC)	
ORION[a]-Factor3	.859 (DECIS)	.815 (COMPA)	
ORION[a]-Factor4	.738 (COMPA)	.845 (REAS)	
ORION[a]-Factor5	.857 (REAS)	.741 (DECIS)	
ORION[a]-Factor6	.860 (PREDIC)	.735 (PROBL)	

Note.* Root Mean Square Error of Approximation.
** Normed Fit Index.
*** Comparative Fit Index.
**** Goodness of Fit Index .
[o] Root Mean Square of Residuals (acceptable close to .048).
[oo] Weighted Root Mean Square of Residuals (acceptable fit < 1.0).
[ooo] Expected a Posteriori (EAP) Greatest Lower Bound (GLB) for reliability.
[a] Overall Reliability of fully-Informative prior Oblique N-Expected a Posteriori scores.

not (Calderón-Garrido et al., 2019). These results point out the six-factor model represents the data better than the general factor model.

As the six-factor empirical model identified corresponds to a structure whose constituent elements allow a reasonable interpretation of the model according to the theoretical proposal presented at the beginning of this study, the same names of the factors were retained, namely Classification, Prediction, Comparison, Reasoning, Decision, and Problem. The results of the previous analysis eliminated 17 initial items, due to the detection of some empirical dysfunctions, thus decreasing sharply the number of items that form the empirical factors that are retained for the final form of the test; however, only two items switched their theoretically factor assigned initially, as a consequence of the CFA validation of the final empirical factors. For instance, the CLASSIF1 item was initially and theoretically assigned to the scale Classification and was empirically allocated into the Problem final factor (CLASSIF1_PROBL); the PROBL9 item was initially and theoretically assigned to the scale Problem, yet it was empirically allocated into the final empirical factor Reasoning (PROBL9_REAS). The mixed denominations of these elements, which include both the initial theoretical dimension and their final empirical factor, try to reflect their switched situation (Table 4).

Figure 1 represents the structural equation model corresponding to the loading matrix of Table 4 and the excellent parameters of the CFA (Table 3). The diagram shows the standardized regression coefficients among the latent variable and with the observable variables of the instrument, as well as the proportions of empirically explained variance for each variable. The model depicts strong relationships of five latent scales (Problem solving achieves the highest standardized coefficient) and also suggests some weakness of the Comparison latent scale, probably due to the drastic reduction of its length to just three items.

The model incorporates four residuals correlations that were added because of their high modification indices and the gain in the model fit, as the model without the residual correlations attained worst fit parameters (e.g., higher Chi-square = 677.915, NNFI and CFI < .90) than those reported for the final model (Table 3). Further, the residual correlations may also have some basis due to theoretical similarities of the items.

Analysis of the closeness to the single-dimension of the factors

Each of the six groups of items that make up the six empirical factors obtained from the previous CFAs for the 31-item test (Table 4) were submitted separately to a confirmatory RULS analysis to verify their one-dimensional nature. The overall results obtained in the six factors show adequate goodness-of-fit indices, explained variance and reliability, but also suggest some improvements (Table 5).

Parallel analyses with the RULS method and tetrachoric correlations based on the minimum rank factor analyses confirmed a single-dimension model for the six factors, because the MIREAL parameter presents acceptable values (< .30), with a moderate exception in Decision. These results allow us to consider these six factors as one-dimensional, and, consequently, justify that their scores validly and reliably measure each of the skills operationalized by the items that make up the empirical factors.

The proportion of variance explained by each of the six unique factors is high (.57 –.39), and both reliability values, omega (.873 –.638) and Cronbach's alpha (.870 –.624) are good. Although four factors display good KMO values (.823 –.658), the Prediction and Comparison factors show low KMO values (.535 –.501). Almost all the loadings of the constituent items of empirical factors reach scores greater than .30, with the sole exception of COMPA1 item, which may be the source of the problems of this factor, together with the small number of items that form it (3). The goodness-of-fit parameters of the CFA show that the data obtained for the six factors adequately fit the one-dimension structure as their scores for the six factors are excellent: RMSEA (.03 – .09), CFI (.954 – . 996), GFI (.969 – .997), and RMSR (.019 – . 098).

Empirical analysis of the test validity

The process of CoTT_EP6 construction stemmed from scholar and credible item sourc-

Table 4
Factor loading matrix of the reduced CoTT_PE6_31 test (31 items; Promin rotation).

Variables			**Empirical**	**Factors**		
	Classification	Prediction	Comparison	Reasoning	Decision	Problem
PREDIC1		.415				
PREDIC3		.369				
PREDIC5		1.014				
PREDIC6		.515				
PREDIC7		.339				
PREDIC8		.274				
COMPA1			.126			
COMPA5			.810			
COMPA6			.830			
CLASSIF2	.974					
CLASSIF3	.737					
CLASSIF4	.601					
CLASSIF5	.609					
CLASSIF6	.861					
CLAS1_PROBL						.268
PROBL2						.666
PROBL4						.633
PROBL5						.560
PROBL6						.440
PROBL10						.276
DECISION2					.436	
DECISION4					.440	
DECISION5					.327	
DECISION6					.622	
DECISION9					.757	
PROBL9_REAS				.233		
LOG-REAS1				.805		
LOG-REAS2				.677		
LOG-REAS4				.821		
LOG-REAS6				.570		
LOG-REAS7				.154		
Number of items	5	6	3	6	5	6
Reliability (ORION[a])	.922	.999	.815	.845	.741	.735
Explained variance	.192	.258	.319	.378	.434	.489

Note. Loadings below .30 were eliminated (except for six loadings that correspond to items theoretically assigned to a factor).
[a] Overall Reliability of fully-Informative prior Oblique N-Expected a Posteriori scores.

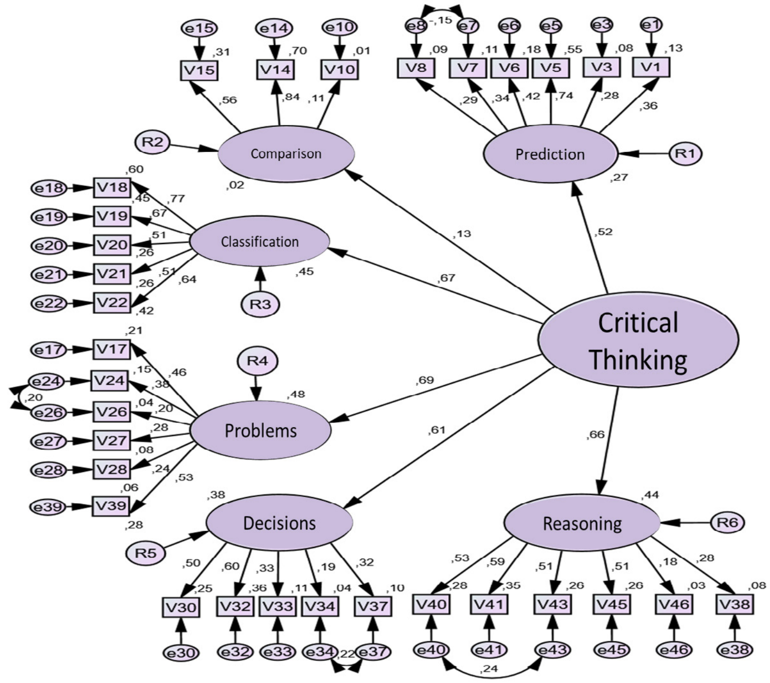


Figure 1
Diagram of structural equations corresponding to the CoTT_PE6_31 instrument.

Table 5
Statistical parameters of the robust goodness of fit of confirmatory factor analysis (CFA) for the closeness to single-dimensional model of each of the six empirical scales resulting from the factorization of the CoTT_PE6_31 instrument.

CFA Statistics	**Prediction**		**Comparison**		**Classification**		**Problem**		**Decision**		**Reasoning**	
	Item	Load	Item	Load	Item	Load	Item	Load	Item	Load	Item	Load
	PRE1	.457	COM1	.133	CLA2	.915	CLA1	.310	DEC2	.585	PRO9	.355
	PRE3	.378	COM5	1.000	CLA3	.766	PRO2	.627	DEC4	.638	REAS1	.741
	PRE5	1.000	COM6	.664	CLA4	.645	PRO4	.645	DEC5	.394	REAS2	.685
	PRE6	.534			CLA5	.658	PRO5	.556	DEC6	.440	REAS4	.744
	PRE7	.346			CLA6	.807	PRO6	.385	DEC9	.583	REAS6	.599
	PRE8	.333					PRO10	.411			REAS7	.194
Kaiser-Meyer-Olkin	.535		.501		.823		.700		.658		.753	
Bartlett (P-Sig.)	.000		.000		.000		.000		.000		.000	
Explained variance	.390		.567		.661		.408		.425		.404	
One-dimensionality												
MIREAL*	.297		.293		.229		.250		.330		.215	
Goodness of fit												
RMSEA**	.061		-		.051		.040		.091		.031	
RM Ji Squared	3.831		-		13.625		1.221		31.973		22.943	
Ji Square(P)	.000		-		.197		.071		.000		.064	
NNFI ***	.956		-		.992		.978		.908		.990	
CFI****	.974		-		.996		.989		.954		.993	

CFA Statistics	Prediction	Comparison	Classification	Problem	Decision	Reasoning
GFI*****	.970	.997	.996	.991	.972	.969
Residuals						
RMSR°	.055	.098	.055	.045	.019	.074
WRMSR°°	.038	.063	.039	.041	.012	.043
Reliability						
EAP-GLB°°°	.792	.696	.902	.686	.750	.812
Omega	.670	.679	.873	.638	.663	.743
Cronbach Alpha	.658	.555	.870	.624	.656	.730

Note.* Mean of Item Residual Absolute Loadings (unidimensional < .30).
** Root Mean Square Error of Approximation.
*** Normed Fit Index.
**** Comparative Fit Index.
***** Goodness of Fit Index.
° Root Mean Square of Residuals (acceptable model if close to .048).
°° Weighted Root Mean Square of Residuals (acceptable fit < 1.0).
°°° Expected a Posteriori (EAP) Greatest Lower Bound (GLB) for reliability.

es that provided quality and affordable CT items. The items were scrutinized, analyzed, selected and piloted by researchers, in order to test their cognitive demand to PE6 students and prepare the elaboration of the 48-item CoTT_PE6. The reliability of the sources and the selection processes warrant some ab initio content validity of CoTT through sound fitness between item contents, skills, cognitive demands and students' abilities. This section aims to further develop the empirical CoTT validity.

Parallel tests are often used by researchers to trial validity through external criteria. However, the scarcity of CT tests for primary students, which mainly motivates this study, makes this way impractical. Instead, each of the six theoretical skills that conform to CoTT are taken as external criteria for the remaining skills, so that the analysis of skill intercorrelations develops a validity test. Regardless of the debate about the importance of the general or specific context of CT education, the most widespread argument in favor of CT is its impact on learning (e.g., O'Hare & McGuinness, 2015). Thus, learning is operationalized here through students' subject school grades at the end of the school year, which are numerically assessed by teachers (1-10) and used here as an external criterion to test the CoTT_PE6 validity. Finally, the correlations and variances of the empirical factors are examined to add evidence on CoTT validity.

Correlations among critical thinking skills

The descriptive statistics and the correlations among the six CoTT theoretical skills are displayed and analyzed here assuming that all of them are significant and positive, considering that all of them measure an aspect of the CT construct (Table 6). All six skills display cases attaining their maximum and minimum scores, which display the range (9-43 and 4-31) for the total score of the two CoTT forms. The mean scores of skills show that Classification and Prediction tend to display the highest mean score, while Decision and Logical Reasoning have got the lowest scores. The asymmetry and kurtosis (not shown in Table 6) have got normal scores for the six skills and the total score.

Table 6

Descriptive statistics of the six theoretical skills (top) emerging from the empirical factorization of CoTT_PE6_46 and CoTT_PE6_31 (*n* = 655) and their Pearson correlation coefficients (bottom), where the upper triangle corresponds to the skills of CoTT_PE6_31 (columns) and the lower triangle to the CoTT_PE6_46 skills (rows).

	Descriptive Statistics of CoTT_PE6_46						
	Prediction9	Comparison7	Classification6	Problem8	Decision9	Log-Reason7	Total46
Range	9	7	6	8	9	7	34
Minimum	0	0	0	0	0	0	9
Maximum	9	7	6	8	9	7	43
Mean	5.206	3.313	3.702	3.907	3.125	3.124	22.377
Error std.	0.075	0.055	0.074	0.065	0.072	0.064	0.251
Std. deviation	1.924	1.417	1.889	1.672	1.854	1.631	6.416

	Descriptive Statistics of CoTT_PE6_31						
	Prediction6	Comparison3	Classification5	Problem6	Decision5	Log-Reason6	Total31
Range	6	3	5	6	5	6	27
Minimum	0	0	0	0	0	0	4
Maximum	6	3	5	6	5	6	31
Mean	3.737	1.441	3.067	3.197	1.686	2.794	15.922
Error std.	0.058	0.040	0.068	0.061	0.052	0.064	0.198
Std. deviation	1.474	1.024	1.728	1.573	1.334	1.641	5.056

CoTT_PE6_46 Skills	**CoTT_PE6_31 Skills**					
	Prediction6	Comparison3	Classification5	Problem6	Decision5	Log-Reason6
Prediction9		.063	.205	.265	.206	.142
Comparison7	.220**		.235	.190	.031	.384**
Classification6	.362**	.213**		.377**	.275*	.037
Problem8	.182**	.143**	.300**		.340*	.341*
Decision9	.282**	.162**	.373**	.263**		.060
Log-Reas7	.282**	.209**	.286**	.252**	.216**	

Note.* Correlation significant at the .05 level (bilateral).
** Correlation significant at the .01 level (bilateral).

All Pearson correlations between the theoretical skills are positive and significant ($p < .01$) for the CoTT_PE6_46. The CoTT_PE6_31 inter-skill correlations display lower scores than the former, though all them are still positive (Table 6). Thus, the inter-skills correlations are positive and mostly significant as expected to justify the internal validity of CoTT. The correlations among the empirical factors of both CoTT forms are overall higher than the correlations between the theoretical sub-scales (Table 6), which may justify much better than the correlations of table 6 both, the factors' high reliability and the relative weakness of the Comparison factor (correlations close to zero).

Correlations with school subject grades as external criterion

The empirical analysis of CoTT validity in regard to external criteria (school grades) applies correlational methods. To this aim, an incidental subsample of participants (n = 52), those whose final-course grades were available, is used. In spite of the size, the subsample is diverse, as it comes from three schools (two public and one public-funded), which are located at a small town and at the center and the periphery of a large city. Table 7 displays the descriptive statistics and Pearson correlations of the school grades and skills.

The descriptive statistics of grades show that the distribution of students' grades is quite homogeneous across subjects, and the asymmetry and kurtosis scores stay within acceptable ranges (not shown). It is worth highlighting that only Natural Science, Catalan Language and Math display 4 as minimum grade, which means that some sixth-graders have got negative final grades (under 5). Further, Physical Education showed the highest average grade and the minimum standard deviation, whilst Catalan Language displays the lowest average grade (Table 8).

Table 7

Pearson intercorrelations among school subject grades (top of the table) and the descriptive statistics of grades (bottom) for the incidental subsample (n = 52).

Subjects	Natural Sc.	Social Sc.	Catalan L	Spanish L.	Art Ed.	Physical Ed.	Math	Religion-Values	English L.
Natural Sciences	-	.858**	.819**	.840**	.708**	.544**	.733**	.527**	.742**
Social Sciences		-	.857**	.863**	.559**	.449**	.700**	.700**	.648**
Catalan Language			-	.899**	.590**	.466**	.794**	.696**	.681**
Spanish Language				-	.637**	.470**	.796**	.723**	.768**
Art Education					-	.630**	.541**	.343*	.693**
Physical Education						-	.394**	.344*	.499**
Mathematics							-	.665**	.694**
Religion-Values								-	.545**
English Language									-
	Descriptive statistics of grades (range of grade scores 1-10; grades under 5 are negative)								
Range	6	5	5	5	5	5	6	5	5
Minimum	4	5	4	5	5	5	4	5	5
Maximum	10	10	9	10	10	10	10	10	10
Mean	7.58	7.92	7.31	7.46	8.21	8.5	7.92	8.31	8
Std. deviation	1.719	1.453	1.515	1.578	1.637	1.111	1.747	1.489	1.521

Table 8
Pearson intercorrelations between school subject grades and CT theoretical skills of the two forms of CoTT for the incidental subsample (*n* = 52). The correlations in bold pinpoint the only skill that significantly enters the subject grade prediction model of the lineal regression analysis.

CoTT Skills	Correlations								
	Natural Sc.	Social Sc.	Catalan L	Spanish L.	Art Ed.	Physical Ed.	Math	Religion-Values	English L.
CoTT_PE6_46									
Prediction9	.225	**.319***	.236	**.392****	.022	.02	.265	.255	.243
Comparison7	.228	.182	.255	.316*	.145	.121	.323*	.16	**.412****
Classification6	.177	.055	.08	.241	**.306***	.059	.259	.023	.223
Problem8	.219	.128	**.300***	.332*	.219	.203	**.479****	.213	.256
Decision9	.164	.13	.225	.271	.23	.251	.264	.061	.117
Log-Reas7	.032	.06	.158	.121	-.039	-.071	.249	.203	.207
Total46	.294*	.247	.356**	.476**	.249	.161	.529**	.265	.408**
Explained variance(%)***	8.6	10.2	12.7	22.7	9.4	2.6	28.0	7.0	16.6
CoTT_PE6_31									
Prediction6	.113	.179	.107	.255	-.047	-.048	.156	.127	.140
Comparison3	-.042	-.150	-.034	-.055	.061	.125	-.051	-.236	.056
Classification5	.146	-.003	.010	.192	**.323***	.109	.216	-.075	.200
Problem6	**.329****	.257	**.415****	**.457****	.224	.103	**.520****	**.361****	**.321***
Decision5	.082	.162	.256	.247	.176	**.291***	.235	.155	.058
Log-Reas6	-.035	-.053	.051	.052	-.107	-.246	.185	.060	.135
Total31	.182	.122	.236	.355**	.177	.056	.397**	.134*	.287
Explained variance***	10.8	6.6	17.2	20.9	10.4	8.5	27.0	13.0	10.3

Note. * Correlation significant at the .05 level (bilateral).
** Correlation significant at the .01 level (bilateral).
*** Shared variance between subject grades and skills (computed through lineal regression analysis of grades and CT-skills as the square of the subject's bold correlation coefficient, due to the single-skill prediction model obtained).

The correlations between subjects are all positive and significant, and it is worth noting the highest correlations between Catalan, Spanish, Social and Natural Sciences, whilst Physical Education and Art Education display the lowest correlations with the others.

The correlations between CT skills and school grades are mainly positive as expected, although only a few are statistically significant. The total score of both CoTT forms correlates positively with all subjects, which means CoTT total scores predict grades and thus support the predictive validity of both forms and inform the amount of explained variance for each grade (top 28% for Math). From the point of view of the specific skills, Problem-solving displays the highest correlations with the subject grades, and significantly correlates with three/six subjects (depending on the form) for the two forms. At the opposite extreme, Decision and Logical Reasoning tend to display the lowest set of correlations with subjects. Problem solving of CoTT_PE6_31 largely displays the highest correlations with almost all subjects, while CoTT_PE6_46 displays a wider distributed pattern of skills than CoTT_PE6_46.

Further, the correlations show some specific correlation patterns of the subjects for both forms of CoTT. Overall, CoTT_PE6_46 tends to display higher correlations than CoTT_PE6_31, where a few correlations are negative yet non-significant. The highest and significant correlations ($p < .01$) correspond to Mathematics (maximum) and the language subjects (Catalan, Spanish and English), whereas the lowest (non-significant) correlations correspond to Physical Education (minimum). From the perspective of the subjects, the leading correlation of each subject makes sense of the subject curriculum; for instance, Problem Solving is the top correlated skill for Mathematics and Natural Sciences, where problem solving is a core learning activity. Again, Decision Making is the top correlated skill for Physical Education, where sports continuously practice decision making, Classification (mainly made of figurative items) is top for Art and Prediction (causes and consequences) for Social. Comparison and Problem Solving are top for language subjects. This association between subjects and its leading skills also put forward a qualitative predictive validity of CoTT.

Finally, in order to discriminate the predictive power of the different single skills (predictors) on each subject grade, a forward stepwise lineal regression analysis was performed. Of course, the subjects lacking significant correlations have not got significant predictors and the subjects that display only one significant skill correlation this skill is the only predictor. However, the subjects having two or more significant skill correlations have got a prediction equation with only one predictor too (the highest correlation). The predictors of each subject are bold in Table 8 and this qualitative association between subjects and one specific skill also adds to the predictive validity of CoTT.

All in all, the correlations between CT skills and school grades are mainly positive as expected, which suggest the predictive validity of the CoTT in front of an external criterion as school subject grades. The above correlational profiles suggest specific trendy associations between subjects and skills, which could be rationally justified. However, these trends should be better elucidated through large samples and empirical CFA.

Discriminant validity of the model

The discriminant validity of the factor model was verified through the application of the Fornell-Larcker criterion, which requires the average factor variances to be higher than the correlations of each factor with the other factors (Fornell & Larcker, 1981). The average factor variances are computed from the data of Figure 1 and the comparison with the inter-factor correlations show that the Fornell-Larcker criterion is satisfied by the model. Thus, the model's discriminant validity is confirmed.

In sum, the positive and significant inter-skill correlations support the concurrent validity (relatively higher correlations across skills, as they all belong to the CT construct). Then, the computed variances of the factors are higher than the inter-correlations and confirm the discriminant validity of CoTT_EP6. Further, the correlations between CT skills and a theoretically-related external criterion (school subject grades) confirm that both constructs are positively and significantly correlated, where the subject correlations with the total CT score are especially high, also underlining the transversal importance of CT for school learning. Thus, the above correlational analyses support the validity of the CoTT_EP6 test.

Discussion

This study provides evidence about the validity and reliability of the Challenges of Thinking (CoTT_PE6) instrument to assess CT, a culture-free test (independent of the school curriculum) that is adapted to the evolutive and learning stage of sixth graders (11 year old). The contributions of CoTT arise from the inner transversal impact of CT on learning and the increasing extension of CT teaching in schools, with the consequent need to evaluate the educational results, as well as the lack of CT assessment instruments, which are adapted to younger students and appropriate for use in the classroom (Aktoprak & Hursen, 2022; Ennis, 2009).

The study follows the general prescriptions of test development to establish the psychometric properties of CoTT_PE6 (Ferrando et al., 2022; Muñiz & Fonseca-Pedrero, 2019). The study validates a six-factor empirical model of the final 31-item CoTT, which confirms a parsimonious and coherent interpretation of the theoretical factor structure postulated for CoTT_PE6_31 (Prediction, Comparison, Classification, Problem solving, Decision making and Logical Reasoning). The CFA goodness-of-fit parameters for the six-factor model are excellent: Chi-square (489.959, p = .000), RMSEA (.032), NNFI (.971), CFI (.982), GFI (.965), and RMSR (.049). In addition, the reliability indices of the whole CoTT_PE6_31 (.838) and each of the six identified empirical factors reach good scores (ORION: .999, .815, .922, .735, .741, .845), following the order of factors of the previous paragraph. The one-dimension structure for each of the six factors of the model is also supported by their CFA individual goodness-of-fit parameters, so they can be independently used in measurements. Their individual reliability is also acceptable (alpha: .658, .555, .870, .624, .656 and .730), despite some lower values, possibly due to the structural effect of shortening the length of each factor.

Validation evidence has been widely displayed along the results section through several confirmatory milestones, such as the credibility of the scholar sources that provided the starting bank of items, the previous piloting of many items that lead to construct the first CoTT form, and the correlational validity tests that were performed through external criteria (school subject grades), internal criteria (correlations of different factors of CoTT) and the computation of factor variances and. All in all, the CoTT validity results through the predictive validity of CT skills on grades, as well as the higher average variance of factors than the correlations among factors, advocate the claims for the transversal relevance of CT in regard of learning (European Union, 2014; OECD, 2018; UNESCO, 2016).

The psychometric validation of the CoTT_PE6, together with its simplicity of application and scoring, endorses its direct and practical application in primary education: this useful and functional tool makes thinking and its progress visible in primary classrooms and educational research. Educators and researchers can easily diagnose and evaluate CT to test the effectiveness of CT intervention programs (Colom et al., 2014; Saiz, 2017). In addition, CoTT_PE6 allows monitoring the progress of CT skills in longitudinal studies of the educational system, so that it can assess the impact of skills on learning and vice versa, which are crucial aspects of the quality of education (Hattie, 2012; OECD, 2018; UNESCO, 2016).

The CoTT_PE6 instrument also has some limitations that arise from its design. The first limitation is the obvious restriction to the six skills it contains, as they are considered appropriate skills for primary education. Overall, the validation process displays some limitations and suggests some corresponding future actions, such as performing test-retest reliability or increasing the sample of the predictive validity through grades. Another limitation is the modest values of some KMO indices, especially for Prediction and Comparison skills, which reflect a tension between opposing psychometric and difficult-to-balance demands; on the one hand, each item must provide differential variance from other items to obtain excellent KMO values; on the other hand, this differential variance partially opposed to the most basic principle of each item, namely, contributing to measuring the common cognitive ability of each skill. Thus, a balanced combination of new items and new samples is required to improve KMO test scores. Moreover, the small number of items that compose the Comparison factor (3) possibly harms the overall goodness of fit, despite its GFI parameter is still excellent (.997), yet a future stronger set of items may address this weakness. Consequently, the refinement of items and applications to new non-convenience samples is expected to overcome these limitations (Ferrando et al., 2022; Muñiz & Fonseca-Pedrero, 2019).

Future application of the CoTT_PE6_31 is expected to provide additional evidence for increasing the instrument's validity and reliability. In particular, the analysis of effects and consequences across different groups of students (gender, age, etc.) and across time (test-retest stability) may provide higher response variability and contribute to improve the test validity and reliability, and to consolidate its educational functionality in aspects such as the standardization in different groups, the relationship with other cognitive measures of CT and school grades, as well as the predictive validity between them, and, in short, to increase the visibility of thinking in education (Lopes et al., 2018).

Conclusions

The 31-item final form of the newer Challenges of Thinking test evaluates six CT skills (Prediction, Comparison, Classification, Problem-Solving, Decision-Making and Logical Reasoning) in 6th graders. This test evaluates six empirical factors that constitute genuine cognitive skills of CT, overlap and fit the theoretical description of CT skills and are independent of the previous knowledge and unidimensional, thus allowing the independent evaluation of each skill, unlike other similar instruments that sometimes mix skills, dispositions and knowledge. Its validation results show that the test is valid, reliable, functional, useful, its response processes are standardized, the internal structure shows quite good CFA parameters and adequate evidence of factor discriminant validity as well as a positive relationship with school grades, as an external variable. Finally, the test application satisfies conditions of time and material economy to easily evaluate, research and make thinking visible in primary education classrooms.

References

Aktoprak, A., & Hursen, C. (2022). A bibliometric and content analysis of critical thinking in primary education. *Thinking Skills and Creativity, 44*, 101029. https://

doi.org/10.1016/J.TSC.2022.101029

Almerich, G., Suárez-Rodríguez, J., Díaz-García, I., & Orellana, N. (2020). Structure of the competences of the XXI century in students of the educational field. Personal influential factors. *Educación XX1, 23*(1), 45-74. https://doi.org/10.5944/educXX1.23853

Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies, 31*(3), 285-302. https://www.tandfonline.com/toc/tcus20/31/3?nav=tocList

Colom, R., García-Moriyón, F., Magro, C., & Morilla, E. (2014). The long-term impact of philosophy for children: A longitudinal study (Preliminary results). *Analytic Teaching and Philosophical Praxis, 35*(1), 50-56. https://journal.viterbo.edu/index.php/atpp

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity, 12*, 43-52. https://doi.org/10.1016/J.TSC.2013.12.004

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi, 37*, 165-184. https://doi.org/10.1007/s11245-016-9401-4

Ennis, R. H. (2009). An annotated list of critical thinking tests. https://criticalthinking.net/how-can-critical-thinking-skills-be-tested

Ennis, R. H., & Millman, J. (2005a). *Cornell Critical Thinking Test Level X*. The Critical Thinking Company. https://www.criticalthinking.com/cornell-critical-thinking-tests.html

Ennis, R. H., & Millman, J. (2005b). *Cornell Critical Thinking Test Level Z*. The Critical Thinking Company. https://www.criticalthinking.com/cornell-critical-thinking-tests.html

European Union. (2014). *Key competence development in school education in Europe. KeyCoNet's review of the literature: A summary*. Key Competence Network. European Schoolnet. http://keyconet.eun.org

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations*. American Philosophical Association. https://eric.ed.gov/?id=ED315423

Facione, P. A., Blohm, S. W., Howard, K. L., & Giancarlo, C. A. (1998). *California Critical Thinking Skills Test: Manual (Revised)*. California Academic Press.

Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema, 29*(2), 236-240. https://doi.org/10.7334/psicothema2016.304

Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement, 78*(5), 762-780. https://doi.org/10.1177/0013164417719308

Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A., & Muñiz, J. (2022). Decalogue for the factor analysis of Test Items. *Psicothema, 34*(1), 7-17. https://doi.org/10.7334/psicothema2021.456

Fisher, A. (2009). *Critical thinking. An introduction*. Cambridge University Press.

Fisher, A. (2021). What critical thinking is. In J. A. Blair (Ed.), *Studies in critical thinking* (2nd ed., pp. 7-26). University of Windsor.

Follmann, D., Mattos, K. R. C., & Güllich, R. I. da C. (2018). Teaching strategies of sciences and the promotion of critical thinking in Portugal. *Tecné, Episteme y Didaxis* (Extra). https://revistas.upn.edu.co/index.php/TED/issue/view/583

Fullan, M., & Scott, G. (2014). *Education PLUS*. Collaborative Impact SCT. https://michaelfullan.ca

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.2307/3151312

Calderón-Garrido, C., Navarro-González, D., Lorenzo-Seva, U., & Ferrando-Piera, P. J. (2019). Multidimensional or essentially unidimensional? A multi-faceted factor-analytic approach for assessing the dimensionality of tests and items. *Psicothema, 31*(4), 450-457. https://doi.org/10.7334/PSICOTHEMA2019.153

Gelerstein, D., del Río, R., Nussbaum, M., Chiuminatto, P., & López, X. (2016). Designing and implementing a test for measuring critical thinking in primary school. *Thinking Skills and Creativity, 20*, 40-49. https://doi.org/10.1016/J.TSC.2016.02.002

Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking (4th ed.)*. Lawrence Erlbaum.

Halpern, D. F. (2010). *Halpern Critical Thinking Assessment*. SCHUHFRIED. https://www.schuhfried.com

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

International Society for Technology Education. (n.d.). ISTE Standards: For Educators. International Society for Technology Education. https://iste.org/standards/educators

Lipman, M. (1982). Philosophy for children. *Thinking: The Journal of Philosophy for Children, 3*(3/4), 35-44. https://doi.org/10.5840/thinking1982339

Lopes, J., Silva, H., & Morais, E. (2018). Critical thinking test for elementary and secondary students. *Revista de Estudios e Investigación en Psicología y Educación, 5*(2), 82-91. https://doi.org/10.17979/reipe.2018.5.2.3339

Lorenzo-Seva, U., & Ferrando, P. J. (2019). Robust promin: A method for diagonally weighted factor rotation. LIBERABIT. *Revista Peruana de Psicología, 25*(1), 99-106. https://doi.org/10.24265/liberabit.2019.v25n1.08

Madison, J. (2004). James Madison Critical Thinking Course. The Critical Thinking Company. https://www.criticalthinking.com/james-madison-critical-thinking-course.html

Manassero-Mas, M. A., & Vázquez-Alonso, Á. (2019). Taxonomía de las destrezas de pensamiento: Una herramienta clave para la alfabetización científica. En M. D. Maciel & E. Albrecht (org.), *Ciência, Tecnologia & Sociedade: Ensino, Pesquisa e Formação,* (pp. 17-38). UNICSUL.

Manassero Mas, M. A., & Vázquez Alonso, Á. (2020a). Evaluación de destrezas de pensamiento crítico: Validación de instrumentos libres de cultura. *Tecné, Episteme y Didaxis, 47*, 15-32. https://doi.org/10.17227/ted.num47-9801

Manassero-Mas, M. A., & Vázquez-Alonso, Á. (2020b). Las destrezas de pensamiento y las calificaciones escolares en educación secundaria: Validación de un instrumento de evaluación libre de cultura. *Tecné, Episteme y Didaxis, 48*, 33-54. https://doi.org/10.17227/ted.num48-12375

Manassero-Mas, M. A., Moreno-Salvo, A., & Vázquez-Alonso, Á. (2022). Development of an instrument to assess young people's attitudes toward critical thinking. *Thinking Skills and Creativity, 45*, 101100. https://doi.org/10.1016/J.TSC.2022.101100

Manassero-Mas, M. A., & Vázquez-Alonso, Á. (2023). Evaluación de las destrezas del pensamiento crítico: Un diagnóstico de los estudiantes de primaria. *Revista Evaluar, 23*(2), 40-56. https://doi.org/10.35670/1667-4545.v23.n2.42069

Manassero-Mas, M. A., & Vázquez-Alonso, Á. (2024). Visibilizar las destrezas de pensamiento en educación primaria: Desarrollo psicométrico de un instrumento de evaluación. *Bordón. Revista de Pedagogía, 76*(1), 119-139. https://doi.org/10.13042/BORDON.2024.95702

Mapeala, R., & Siew, N. M. (2015). The development and validation of a test of science critical thinking for fifth graders. *SpringerPlus, 4*(1). https://doi.org/10.1186/s40064-015-1535-0

Muñiz, J., & Fonseca-Pedrero, E. (2019). Ten steps for test development. *Psicothema, 31*(1), 7-16. https://doi.org/10.7334/psicothema2018.291

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. The National Academies Press. https://doi.org/10.17226/13398

Organisation for Economic Co-operation and Development [OECD]. (2018). *Future of education and skills*. https://www.oecd.org/en/topics/future-of-education-and-skills.html

O'Hare, L., & McGuinness, C. (2015). The validity of critical thinking tests for predicting degree performance: A longitudinal study. *International Journal of Educational Research, 72*, 162-172. https://doi.org/10.1016/j.ijer.2015.06.004

Paul, R., & Nosich, G. M. (1993). A model for the national assessment of higher order thinking. In R. Paul (Ed.), *Critical thinking: What every student needs to survive in a rapidly changing world* (pp. 78-123). Foundation for Critical Thinking. https://www.criticalthinking.org/data/pages/40/fe9f23bd821fcc3ac920a3ce58352412513525db13333.pdf

Piaget, J., & Inhelder, B. (1997). Psicología del niño [*The psychology of the child*]. Morata.

Rivas, S. F., & Saiz, C. (2012). Validation and psychometric properties of the PENCRISAL Critical Thinking Test. *Revista Electrónica de Metodología Aplicada, 17*(1), 18-34. https://reunido.uniovi.es/index.php/Rema/issue/view/766

Saiz, C. (2017). Pensamiento crítico y cambio [*Critical thinking and change*]. Pirámide.

Shayer, M., & Adey, P. S. (Eds.). (2002). *Learning intelligence: Cognitive acceleration across the curriculum from 5 to 15 years*. Open University Press.

Swartz, R. J., Costa, A. L., Beyer, B. K., Reagan, R., & Kallick, B. (2013). *Thinking-based learning*. SM.

Tremblay, K. (2013). OECD Assessment of Higher Education Learning Outcomes (AHELO). En Blömeke, S., Zlatkin-Troitschanskaia, O., Kuhn, C., Fege, J. (eds.) Modeling and Measuring Competencies in Higher Education. Professional and Vet Learning, vol 1 (pp 113–126). Sense Publishers. https://doi.org/10.1007/978-94-6091-867-4_8

UNESCO. (2016). *Education 2030: Incheon Declaration and Framework for Action for the implementation of Sustainable Development Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all*. https://unesdoc.unesco.org/ark:/48223/pf0000245656

Valenzuela, J. (2008). Thinking and deep learning skills. *Revista Iberoamericana de Educación, 46*(7), 1-9. https://doi.org/10.35362/rie4671914

Vincent-Lancrin, S., González-Sancho, C., Bouckaert, M., de Luca, F., Fernández-Barrerra, M., Jacotin, G., Urgel, J., & Vidal, Q. (2019). *Fostering students' creativity and critical thinking*. OECD. https://doi.org/10.1787/62212c37-en

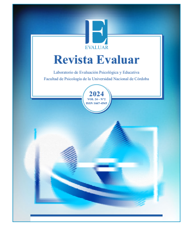
Walton, D., & Macagno, F. (2015). A classification system for argumentation schemes. *Argument and Computation, 6*(3), 219-245. https://doi.org/10.1080/19462166.2015.1123772

Watson, G., & Glaser, E. M. (2002). *Watson-Glaser Critical Thinking Appraisal-II Form E*. Pearson.

Whiting, K. (Oct 21, 2021). These are the top 10 job skills of tomorrow – and how long it takes to learn them. World Economic Forum. https://www.weforum.org/agenda/2020/10/top-10-work-skills-of-tomorrow-how-long-it-takes-to-learn-them

2024, Vol. 24, No. 2
ISSN 1667-4545
Recuperado de https://revistas.unc.edu.ar/index.php/revaluar

Laboratorio de Evaluación Psicológica y Educativa
Facultad de Psicología - Universidad Nacional de Córdoba



Creativity in the Workplace: Psychometric Properties of the CPPC-17 Scale

Creatividad en el Trabajo: Propiedades Psicométricas de la Escala CPPC-17

Ramón Rodríguez Montalbán [1], Adam Rosario Rodríguez [2], Sofía Rodríguez Santiago [3], William Figueroa Romero [2], Thaymis Báez-Luciano [3], Yarí Marini Santiago [4], Yamirca Carrión [4], Aileen Dávila Rodríguez [4], Obed Santiago [4], Cristina Quiñones [4], Mariana Figueroa [4], Carla Harrison [4]

1 - University of Puerto Rico, Cayey.
2 - Albizu University, San Juan Campus.
3 - University of Puerto Rico, Río Piedras Campus.
4 - Albizu University Mayagüez, University Center.

Recibido: 20/12/2023 **Revisado:** 01/03/2024 **Aceptado:** 12/07/2024

Abstract

This study aims to analyze the psychometric properties of the Creative Potential and Creative Practice (CPPC-17) scale. A heterogeneous sample of 1021 workers from different workplaces in Puerto Rico was used to meet this objective. This study used a quantitative method with an instrumental design. The findings reveal that the CPPC-17 scale has construct validity with a factorial structure of three dimensions with Cronbach's alpha values that fluctuated between .91 and .96. In conclusion, the CPPC-17 scale turned out to be a robust instrument to measure Organizational Creativity, along with these three factors: creative potential, creative practice, and organizational support, in the workplace in Puerto Rico.

Keywords: *creative potential, creative practice, organizational support*

Resumen

El propósito de esta investigación es analizar las propiedades psicométricas de la Escala de Potencial Creativo y Práctica Creativa (CPPC-17). Para cumplir con este objetivo, se utilizó una muestra heterogénea de 1021 personas trabajadoras de distintos sectores laborales de Puerto Rico. Este estudio empleó un método cuantitativo, con un diseño instrumental. Los hallazgos revelan que la escala CPPC-17 posee validez de constructo con una estructura factorial de tres dimensiones, con índices alfa de Cronbach que fluctuaron entre .91 y .96. En conclusión, la escala CPPC-17 resultó ser un instrumento de medición robusto para medir la creatividad organizacional junto a sus tres factores: potencial creativo, práctica creativa y apoyo organizacional, en el contexto laboral de Puerto Rico.

Palabras clave: *potencial creativo, práctica creativa, apoyo organizacional*

***Correspondence to:** Dr. Ramón Rodríguez Montalbán, Assistant Professor in the Department of Social Sciences at the University of Puerto Rico at Cayey. E-mail address: ramon.rodriguez14@upr.edu.

How to cite: Montalbán, R., R., Rodríguez, R., A., Rodríguez Santiago, S., Figueroa Romero, W., Báez-Luciano, T., Carrión Y., Davila Rodríguez, A., Santiago, O., Quiñones, C., Figueroa, M., Harrison, C.,(2024): Creativity in the Workplace: Psychometric Properties of the CPPC-17 Scale. *Revista Evaluar, 24*(2), 76-88. Retrieved from https://revistas.unc.edu.ar/index.php/revaluar

Participaron en la edición de este artículo: Gloria Nieve, Vanesa Toledo, Pablo Carpintero, Florencia Ruiz, Jorge Bruera, Benjamín Casanova.

Introduction

> "*The creative adult is the child who survived*", Ursula Leguin

Human beings are constantly searching for new ways of doing, innovating, and evolving in multiple facets of life. The work context is no stranger to the concept of creativity. DiLiello and Houghton (2008) state that the value of organizational creativity is related to novel ideas to increase organizational efficiency, solve complex problems, and improve effectiveness. Recent studies show that creativity in organizations is generated and fostered by both individual and organizational variables. As background, organizational creativity is related to intrinsic factors, engagement, leadership styles, group support, and emotional intelligence, among other variables (Amabile, 1998; da Costa et al., 2015; Cavaliere et al., 2015; Luu et al., 2019; Mubarak & Noor, 2018; Ramos et al., 2018). Consequently, organizational creativity is associated with employee well-being, innovation, success, and competitiveness of firms, among others (Anderson et al., 2014; Helzer & Kim, 2019; Gu et al., 2015).

Due to the scarcity of literature on work-related creativity in the Puerto Rican organizational context, it would be necessary to apply DiLiello and Houghton's (2008) model to the Puerto Rican work environment. At the same time, although the importance of creativity at the organizational level is well-known from various research studies, other research must focus on its significance at the individual level. As a consequence of the constant social changes and new organizational demands, it has been essential for organizations to employ new ways of fostering creativity to adapt to the new realities. Therefore, it is necessary to analyze the psychometric properties of the creativity scale in the working environment in Puerto Rico to foster creative ideas that help organizations adapt to the needs of a changing world.

Given the above, we ask ourselves the following questions: *How does creativity manifest itself in the Puerto Rican work context? How can work creativity be measured? What are the psychometric properties of DiLiello and Houghton's (2008) work creativity scale?*

To answer these questions, we explore the Theoretical Model of Creativity (DiLiello & Houghton, 2008) based on how individuals develop new ideas and practical skills and how they use their creative abilities in the company (Boada-Grau et al., 2014). This model also points out that individuals with strong and creative potential are more likely to practice creativity when they feel support from the organization (DiLiello & Houghton, 2008). We tested this model along with its three dimensions (creative potential, creative practice, and organizational support) to explore whether it can be used in the Puerto Rican work context. Finally, this study shows that organizational creativity in human behavior allows one to perform different work tasks dynamically and innovatively.

Creativity in the Workplace

According to Falco (2016), Georges de Mestral was an electrical engineer who enjoyed walking his dog in the countryside and noticed that Arctium seeds were constantly sticking to his clothes and his dog's fur. This fact contributed to de Mestral founding his own company and patenting Velcro in 1951 (Falco, 2016). Consequently, this author suggests that Velcro was an idea born in a context outside the norm, which is the typical scenario in which creativity dwells.

From a research perspective, Falco (2016) mentions that creativity is "the willingness to find new, spontaneous, surprising and effective ideas" (p. 59). For his part, Amabile (1988) defines creativity as the production of novel and valuable ideas by an individual or a group of people working together. In that sense, Córdoba et al. (2018) mention that creativity and innovation should coexist because they are two different constructs; they generate greater effectiveness when enhanced collectively. For this reason, these authors define creativity as "an ability that human beings have to generate ideas, problem solutions or offer different interpretations or solutions to different socioeconomic, social, and contextual realities" (p. 56).

In the organizational context, experts have given different definitions of creativity. Sousa et al. (2012) indicate that organizational creativity is a system for developing and channeling individual creativity through teams into monetary innovations of the company. Finally, DiLiello and Houghton (2008) mention that creativity in the organizational context has three interrelated dimensions: creative potential, creative practice, and perceived organizational support.

DiLiello and Houghton's Creativity Model (2008)

DiLiello and Houghton (2008) define *creative potential* as the individual's desire and ability to be creative. Boada-Grau et al. (2014) highlight that creative potential is what an individual performs to produce new ideas. DiLiello and Houghton define *creative practice* as the perceived opportunity to use creative skills and abilities at work. Creative practice is how individuals develop new ideas and practical skills and use their creative abilities in the organization (Boada-Grau et al., 2014). Finally, *organizational support* is the recognition the company gives its employees for being creative (Boada-Grau et al., 2014). Therefore, DiLiello and Houghton (2008) point out that individuals with strong creative potential are more likely to practice creativity when they feel organizational support. It is worth noting that this model of creativity helps employees generate new ideas for their work tasks through the three dimensions mentioned above (Boada-Grau et al., 2014).

DiLiello and Houghton's Creativity Model (2008) stands out as a superior framework for understanding workplace creativity compared to other prominent models such as the *Componential Model of Creativity* (Amabile, 1998, the *Interactionist Model of Creativity* (Woodman & Schoenfeldt, 1990), and the *Systems Model of Creativity* (Csikszentmihalyi, 1988). This assertion relies on several key aspects that underscore the model's comprehensiveness and practical applicability.

Firstly, DiLiello and Houghton's model uniquely integrates the concepts of creative potential, practiced creativity, and organizational support. This triadic approach provides a comprehensive understanding of creativity, encompassing not only the individual's inherent ability to generate novel ideas but also the practical application of these ideas in a work setting and the critical role of organizational support. In contrast, models like the Componential Model of Creativity (Amabile, 1996) primarily focus on individual-level factors such as domain-relevant skills, creativity-relevant processes, and task motivation without sufficiently addressing the organizational context that can either facilitate or hinder creative expression.

Secondly, DiLiello and Houghton's model emphasizes the dynamic interaction between an individual's creative potential and the organizational environment. This aspect aligns with the Interactionist Model of Creativity (Woodman & Schoenfeldt, 1990), which also takes into consideration the interplay between personal traits and environmental factors. However, DiLiello and Houghton (2008) delve into the concept and explicitly define practiced creativity as the perceived opportunity to use creative skills at work, thus providing a clearer operationalization of how organizational support can translate potential into actual creative output. This clear delineation of practiced creativity and the role of organizational support offers actionable insights for managers aiming to cultivate a creative workforce.

Moreover, the model's emphasis on organizational support as a facilitator of practiced creativity directly addresses the limitations of the Systems Model of Creativity (Csikszentmihalyi, 1988), highlighting the broader sociocultural context; however, it lacks specific guidance on how organizations can nurture individual creativity. By focusing on organizational practices such as recognizing and rewarding creativity, DiLiello and Houghton's Creativity Model (2008) provides concrete strategies for fostering an environment that supports and enhances creative efforts.

In summary, DiLiello and Houghton's Creativity Model (2008) offers a more integrated and actionable framework for understanding and fostering workplace creativity. Its comprehensive approach, which combines individual capabilities with organizational support mechanisms, provides a robust foundation for enhancing creative performance in organizational settings.

Creative Potential and Practiced Creativity Scale (DiLiello & Houghton, 2008)

The Creative Potential and Practiced Creativity (CPPC-17) scale is a comprehensive instrument designed to measure two crucial aspects of creativity in the workplace: creative potential and practiced creativity. The scale comprises 17 items distributed across three factors: Creative Potential, Practiced Creativity, and Perceived Organizational Support. *Creative Potential* (6 items) assesses an individual's self-perceived ability to generate novel ideas and solve problems creatively, while *Practiced Creativity* (5 items) evaluates the opportunities individuals have to apply their creative skills in their work environment. *Perceived Organizational Support* (6 items) measures the degree to which an organization fosters and recognizes employee creativity.

The CPPC-17 demonstrates robust psychometric properties, with reliability coefficients (Cronbach's alpha) ranging from .84 to .94 across the three factors (English version) and from .80 to .90 (Spanish version), indicating high internal consistency. Confirmatory factor analysis (CFA) supports the scale's three-factor structure in both the original (DiLiello & Houghton, 2008) and subsequent studies (Boada Grau et al., 2014). Construct validity is shown by significant correlations with related constructs such as workaholism, irritation, burnout, and personality traits (Boada Grau et al., 2014). The CPPC scale was constructed through a rigorous process that included item development based on DiLiello and Houghton's (2008) conceptual framework of creativity, followed by exploratory and confirmatory factor analyses to refine the instrument and ensure its validity and reliability (DiLiello & Houghton, 2008).

The translation of the CPPC-17 scale from English to Spanish was conducted by Boada Grau et al. (2014) in Spain using the back-translation method. Initially, bilingual experts translated the 17-item scale into Spanish. According to the authors, the process followed guidelines for test adaptation, ensuring the Spanish version retained the psychometric properties of the original scale, including reliability and construct validity. For the present study, we used the Spanish version translated by Boada Grau et al. (2014). We analyzed each instrument item, word by word, for linguistic equivalence, and found it equivalent to the Spanish used in Puerto Rico.

Antecedents and Consequences of Creativity at Work

Creativity at work is influenced by various individual and organizational factors. At the individual level, motivational factors such as intrinsic motivation (Amabile, 1998; da Costa et al., 2015), thriving at work (Zhang et al., 2023), and work engagement (Gonlepa et al., 2023; Mubarak & Noor, 2018) are crucial. Social factors, such as leadership styles like *authentic* (Luu et al., 2019; Mubarak & Noor, 2018), *transformational* and *charismatic* (Luu et al., 2019), as well as *leadership self-deprecating humor* and *identification* (Huang, 2023) foster follower creativity. *Leaders' creativity* also enhances *team creativity* (Li & Yue, 2019). *Psychological empowerment* (Mubarak & Noor, 2018) and *self-efficacy* (da Costa et al., 2015) significantly boost creativity, with *creative self-efficacy* building *confidence* and *innovative thinking* (DiLiello & Houghton, 2008). Other factors like *emotional intelligence*, *expressiveness*, and *positive affect* are also important (da Costa et al., 2015).

Creativity in the workplace has numerous positive outcomes. At the individual level, creativity can enhance *well-being* by providing a flexible response to stress (Helzer & Kim, 2019). It also helps individuals *stay focused on goals* and continuously improve their *skills* (Sajid et al., 2017). At the organizational level, leadership that promotes creativity can lead to beneficial emotional states among employees, fostering *problem-solving*, *questioning existing methods*, *idea generation*, and *positive discussions* (Shalley et al., 2015). When supervisors encourage creativity, *staff vision*, and *self-confidence* improve, enhancing *leadership performance* (García-Vidal et al., 2019).

Justification

According to Blomberg et al. (2017), competition, economic situation, and urgency to change have given way to organizational creativity. These authors mention that organizational creativity is emerging as a distinct space for academic research. They also recommend that organizations be adaptive, flexible, and innovative, according to the requirements that have brought organizational creativity to the center of managerial interests in recent years. Therefore, this study aims to analyze the psychometric properties of the validity and reliability of the CPPC-17 inferences to support and contribute to the development of research, the process of organizational diagnosis, and the practice of Industrial Organizational Psychology. This management will be relevant for organizations, employees, clients, and professionals of Industrial Organizational Psychology to enhance creativity in the work context.

Purpose of the Study

This research analyzes the psychometric properties of the Creative Potential and Creative Practice Scale (CPPC-17) in the Puerto Rico laboral context. The specific objectives of the study are:

1. To analyze the factor structure of the Creative Potential and Creative Practice Scale of Boada-Grau et al. (2014) employing confirmatory factor analysis with structural equations.

2. To analyze the discriminatory ability of the CPPC-17 items.
3. To analyze the reliability of the CPPC-17 and its factors using Cronbach's alpha internal consistency index and composite reliability.
4. To analyze the convergent and divergent validity of the CPPC-17 factors using the extracted mean analysis of variance.

Method

Design

This research used an instrumental study design (Montero & León, 2007) to analyze the psychometric properties of the CPPC-17 (Boada Grau et al., 2014) through confirmatory factor analysis. Consequently, we tested the instrument's factor structure and, in this way, we met the proposed objectives.

Participants

We obtained approval for this research from the Institutional Review Board (IRB) of Albizu University in San Juan, Puerto Rico. We collected sociodemographic information on gender, age, academic preparation, marital status, number of years working for the company, type of company, and type of industry.

The final sample of this study was composed of 1021 participants, 65% of whom indicated that they were female and 35% male, with ages ranging from 21 to 78 (M = 35.91, SD = 11.04). Regarding academic preparation, 38.8% of participants indicated having a high school or bachelor's degree.

Measure

Spanish version of the Creative Potential and Creative Practice Scale (CPPC-17). DiLiello and Houghton (2008) originally developed this scale and adapted it to Spanish by Boada-Grau et al. (2014). This instrument has 17 items, which measures three factors linked to creativity: six items of creative potential (α = .84; e.g., *"I feel comfortable trying new ideas"*), five items of creative practice (α = .84; e.g., *"At work, my creative abilities are used to the fullest"*), and six items of perceived organizational support (α = .94; e.g., *"In my company, creative work is appreciated"*). The scale has a Likert-type response format with anchors ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

General Procedures

For this study, we contacted participants electronically (e.g., via Facebook and email) and invited them to participate in the research by sharing the link to answer the survey. We collected the responses through the Survey Monkey® platform and downloaded them to a data matrix for further analysis.

Data Analysis Procedures

For this purpose, we performed different statistical analyses to assess the psychometric properties of the CPPC-17: 1) Multivariate normality analysis of the data; 2) Confirmatory factor analysis using structural equations with correction of the adjustment indexes using the corrections of Satorra and Bentler (2001); 3) Item discrimination analysis; 4) Correlation analysis (direct and factor scores of the scale); 5) Reliability analysis; and 6) Analysis of discriminant and convergent validity. The statistical programs used were AMOS Graphics version 26, IBM SPSS Statistics version 26, Stata version 14, and *R* version 3.3.3.

Results

Item Descriptive Analysis

As part of the analyses, we obtained means and standard deviations for all CPPC-17 items to analyze the distributional properties of the scale. Means ranged from 3.55 to 4.32, with standard deviations from 1.26 to 1.73. We conducted an analysis of multivariate normality of the data using the Mardia, *Doornik-Hansen*, and *Henze-Zirkler M-statistical* tests (Doornik & Hansen, 2008), which also did not indicate this type of normality in the scale: M for skewness = 53.70121, χ^2 (969) = 8765.239, p < .001; M for kurtosis = 562.4389, χ^2 (1) = 21654.437, p < .001; Henze-Zirkler = 6.926755, χ^2 (1) = 5.96, p < .001; Doornik-Hansen χ^2(100) = 2073.329, p < .001. Since normality assumptions were not met, we used the corrections of Satorra and Bentler (2001) to calculate the fit of the structural equation models tested.

Factorial Analysis of the Scale

To analyze the factor structure of the CPPC-17, we tested two models using the confirmatory factor analysis with structural equations and the maximum likelihood estimation method. The first model tested was a base model (M1) in which the 17 CPPC-17 items represented a single latent factor. The results of the confirmatory factor analysis for M1 did not show an adequate fit to the data: χ^2 = 7181.882 (119) p < .001, *RMSEA* = .24, *CFI* = .59, *NFI* = .58, *IFI* = .59, *AIC* = 7249.882, corrected χ^2 = 542.4706 (116), p < .001, corrected *CFI* = .59, corrected *NFI* = .29, corrected *IFI* = .59, corrected *AIC* = 7184.412. This fact indicates that the CPPC-17 may consist of more than one factor. Then, we proceeded to test a model (M2) consisting of three factors (creative potential, creative practice, and perceived social support) as the scale design (M2) is the same as the one devised in the CPPC-17 originally. This second model showed a good fit to the data (Hu & Bentler, 1999) (see Table 1). The results of the three-factor M2 (see Figure 1) with an adequate fit to the data, χ^2 = 911.4705 (116) p < .001, *RMSEA* = .08, *CFI* = .95, *NFI* =.95, *IFI* = .95, *AIC* = 985.4705, corrected χ^2 = 542.4706 (116), < .001, corrected *CFI* = .96, corrected *NFI* =.96, corrected *IFI* = .96, corrected *AIC* = 914.4683. These indices comply with what Satorra and Blenter (2001) consider acceptable levels. Figure 1 shows the final version of the CPPC-17.

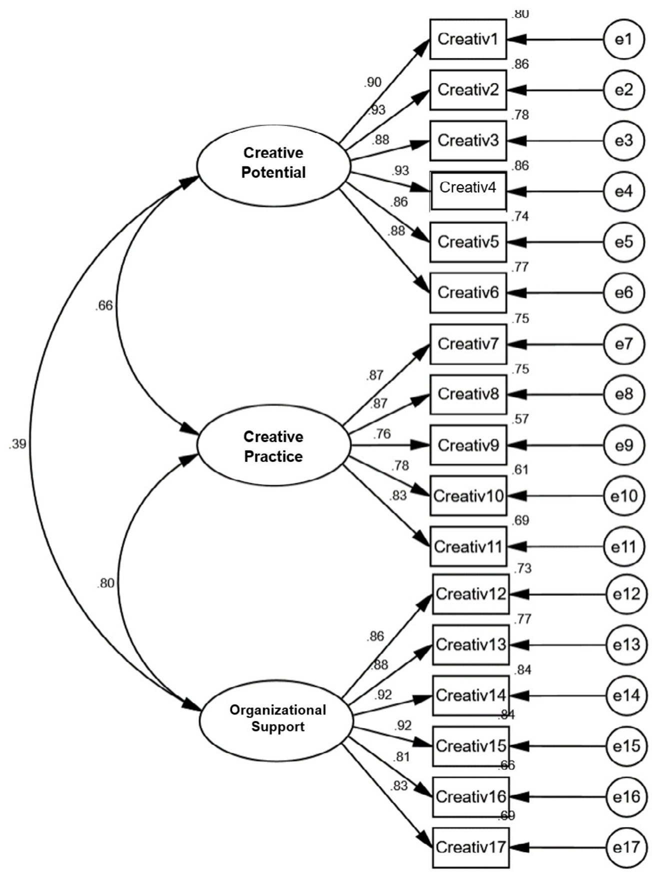


Figure 1.
A three-factor model of creativity.

Table 1
Discrimination indices and explained variance of the items in the final version of the CPPC-17.

Items	Discrimination index	R^2
1. I think I am good at generating innovative ideas.	.90	.80
2. I am confident in my ability to solve problems creatively.	.93	.86
3. I can further develop the ideas of others.	.88	.78
4. I am good at finding creative ways to solve problems.	.93	.86
5. I have the talent and skills to do my job well.	.86	.74
6. I feel comfortable trying new ideas.	.88	.77
7. At work, I have the opportunity to use my skills and creative abilities.	.87	.75
8. At work, people invite me to present ideas for improvement.	.87	.75
9. I have the opportunity to participate in teams.	.76	.57
10. I am free to decide how to carry out my tasks.	.78	.61
11. At work, my creative abilities are used to the fullest.	.83	.69
12. In my company, creative work is appreciated..	.86	.73
13. My company judges ideas fairly.	.88	.77
14. In my company, people are encouraged to solve problems creatively.	.92	.84
15. My company has good methods to encourage and develop creative ideas.	.92	.84
16. My company rewards innovative and creative ideas.	.81	.66
17. I think I am good at generating innovative ideas.	.83	.69

Note. R^2 = Variance explained; Items 1 to 6 = Creative Potential; Items 7 to 11 = Creative Practice; Items 12 to 17 = Organizational Support.

Following the recommendations proposed by Schumacker and Lomax (2010), the AIC_{corr} was used to compare the structural equation models since the χ values[2] were statistically significant. Given this, M2 presented a lower index (AIC_{corr} = 914.4683) than M1 (AIC_{corr} = 7,184.412). When comparing both models, M1 presented a greater difference in its AIC_{corr} (ΔAIC_{corr} = 6,269.9437).

Item Analysis

Discrimination of the 17 CPPC-17 items was analyzed using the total item correlation index. Item discrimination ranged from .64 to .80. Likewise, the factors' explained variance ranged between .37 and .54 (see Table 1). The discrimination indices in the CPPC-17 (M2) final version are above the recommended minimum value of .30 (Kline, 2005).

Reliability Analysis

As part of the objectives of this research, we analyzed reliability and composite reliability for the final three-factor version of the CPPC-17. Cronbach's alpha values for the CPPC-17 factors ranged from .91 to .96. The composite reliability of the factors (omega index) fluctuated between .91 and .96 (see Table 2). All indexes exceed the recommended minimum of .70 (Bagozzi & Yi, 2012).

Table 2
Means, standard deviations, alphas, composite reliability, mean variance extracted and correlations (N = 1,021).

	M	DE	α	FC	VME	1	2	3
1. Creative Potential	4.04	1.21	0.96	.96	.81	-	.66	.39
2. Creative Practice	4.52	1.39	0.91	.91	.68	.63	-	.80
3. Organizational Support	3.86	1.46	0.95	.95	.76	.38	.75	-

Note. *M* = mean; *SD* = standard deviation; α = Cronbach's alpha; *CF* = composite reliability; *SMV* = mean variance extracted. All correlations were significant at $p < .001$. Values above the diagonal represent correlations between latent factors, whereas values below the diagonal represent correlations of direct scores.

Convergent and Discriminant Validity Analysis

Convergent and discriminant validity were analyzed using the Average Variance Extracted (AVE), which measures the average variance explained by the construct in the items. High values in the AVE indicate lower error variance. Fornell and Larcker (1981) suggest that the variance shared between two factors is always less than the variance explained, thus fulfilling the discriminant validity criterion. The values obtained for the AVE of the CPPC-17 factors fluctuated between .68 and .80 (see Table 2). Fornell and Larcker (1981) suggested that values equal to or greater than .50 are acceptable. In addition, the relationship between the factors of the CPPC-17 was analyzed using Pearson's *r* correlation. The correlations between the factors fluctuated between .38 and .74 (see Table 2).

Discussion

The purpose of this research was to analyze the psychometric properties of CPPC-17. The three-factor structure proposed by DiLiello and Houghton (2008) was replicated. The discrimination ability of the items proved to be adequate. We performed Cronbach's alpha analysis and found the instrument to be reliable. Finally, we analyzed the convergent and discriminant validity of the instrument, and we determined that each of the instrument's dimensions measures a particular construct; therefore, there is no redundancy among the items.

Theoretical Implications

In light of our findings, the study helps us understand work creativity in the Puerto Rican organizational context from the theoretical model proposed by DiLiello and Houghton (2008). Similarly, the results show us that the Creative Potential and Creative Practice Scale (CPPC-17) is suitable for research in the Puerto Rican workplace because creativity is measured validly and consistently. At the same time, using this scale will allow us to study creativity about other variables, such as identifying antecedents and consequences in the Puerto Rican work environment. It is worth noting that, according to the results, no problems arise due to the findings, and it also guarantees its application in our work context. On the other hand, our research provides empirical evidence in the Puerto Rican work context that strengthens the theory of creativity proposed by DiLiello and Houghton (2008) with its three interrelated dimensions (creative potential, creative practice, and organizational support) that positively influence and mutually enhance each other.

Approaches established in the studies of Boada Grau et al. (2014) and DiLiello and Houghton (2008) support this empirical evidence.

Our psychometric results indicate that the subscales of the CPPC-17 demonstrated high levels of reliability for its dimensions, which are comparable to previous studies (Boada-Grau et al., 2014; DiLiello & Houghton, 2008). In comparison with the Spanish adaptation of Boada-Grau et al. (2014), our results are consistent in reliability and validity, suggesting that the CPPC-17 is a robust instrument for assessing creativity in different cultural contexts. These similarities reinforce the applicability of the CPPC-17 in diverse companies, indicating that it can successfully serve to identify and promote creative potential among employees. However, it is important to recognize that the type of sampling and the diversity of the sample in our study may limit the generalizability of these results. Therefore, there is a need for future research with large-scale and more diverse samples.

Practical Implications

The practical implications of this study enable us to make available a psychometric instrument whose inferences are valid and reliable to measure the three factors of work creativity (creative potential, creative practice, and perceived organizational support). In the same way, this study provides psychometric evidence of the CPPC-17 instrument to carry out research, as well as organizational diagnosis in Puerto Rican workplaces. This instrument allows us to evaluate work creativity at the individual level based on the model proposed by DiLiello and Houghton (2008). Similarly, this scale can be used in its entirety, with its three dimensions of work creativity and its particular dimensions, if necessary, under the criteria of research interest.

Limitations and Strengths

One of the limitations of this study is that the sampling is subject to availability, so the results do not apply to the whole population of Puerto Rico. However, the study sample is broad (1021 participants) and heterogeneous from different workplaces in the Puerto Rican context. Finally, as a strength, this research provides empirical information on work creativity in the same context.

Use of the Scale

The CPPC-17 Scale is divided into 17 items and three factors linked to creativity: creative potential, creative practice, and organizational support. The CPPC-17 could be used both in groups and individually. The scores of this instrument should be calculated for each of the factors mentioned earlier. The value obtained for each item should be summed to get the mean score per-factor. Then, the result should be divided by the number of items per-factor: creative potential (6 items), creative practice (5 items), and perceived organizational support (6 items). Each item is answered with a response anchor ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Calculating the mean scores of the scale factors provides an average score on the same response anchor as the CPPC-17. The average scores can fluctuate between 1 and 5.

Future Studies

Some suggestions for future research could address studying how factors in the work context promote creativity. That is, how the type of task, the structure of the job, and the work sector may affect workers' creativity. In turn, such new studies should analyze the influence of psychosocial factors such as mobbing or social support affect creativity at work likewise, how practices aimed at promoting the meaning of work (e.g., job crafting, the meaning of and in work) enhance creative practices. Finally, to analyze the organizational consequences of work creativity in different work environments (e.g., third-sector companies, governments, and private companies, among others).

General Conclusions

The literature review shows that creativity in the workplace catalyzes new ideas that help innovate and transform the work environment. An example of this is the discovery of Velcro by Georges de Mestral, who never thought that his invention would change the management of the textile industry.

In addition to providing a measurement instrument, this study demonstrated that creativity in the workplace manifests human behavior to be, rethink, and do work tasks differently, novelly, and dynamically. For this to be possible, people need to have the desire to create and put into practice their creative skills and abilities. In addition, they need to have the support from the company and other people who value, make, and enjoy the creative process, as Albert Einstein said, "*creativity is intelligence having fun*".

References

Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Research in Organizational Behavior, 10*(1), 123-167.

Amabile, T. M. (1998). *How to kill creativity. Harvard Business Review, 76*(5), 76-87.

Anderson, N., Potočnik, K., & Zhou, J. (2014). Innovation and creativity in organizations: A state of the science review, prospective commentary, and guiding framework. *Journal of Management, 40*(5), 1297-1333. https://doi.org/10.1177/0149206314527128

Bagozzi, R. P., & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science, 40*(1), 8-34. https://doi.org/10.1007/s11747-011-0278-x

Blomberg, A., Kallio, T., & Pohjanpää, K. (2017). Antecedents of organizational creativity: Drivers, barriers or both? *Journal of Innovation Management, 5*(1), 78-104. https://doi.org/10.24840/2183-0606_005.001_0007

Boada-Grau, J., Sánchez-García, J.C., Prizmic-Kuzmica, A.J., & Vigil-Colet, A. (2014). Spanish adaptation of the Creative Potential and Practiced Creativity scale (CPPC-17) in the workplace and inside the organization. *Psicothema, 26*(1), 55-62. https://doi.org/10.7334/psicothema2013.68

Cavaliere V., Lombardi S., & Giustiniano L. (2015). Knowledge sharing in knowledge-intensive manufacturing firms: An empirical study of its enablers. *Journal of Knowledge Management, 19*(6), 1124-1145. https://doi.org/10.1108/JKM-12-2014-0538

Córdoba-Zuñiga, E., Castillo-Toro, J. & Castillo-Quesada, N. (2018). Creatividad e innovación: Motores de desarrollo empresarial. *Lámpsakos (Revista Descontinuada), 1*(19), 55-65. http://dx.doi.org/10.21501/21454086.2663

Csikszentmihalyi, M. (1988). Society, culture, and person: A systems view of creativity. In R. J. Sternberg (Ed.), *The nature of creativity: Contemporary psychological perspectives* (pp. 325-339). Cambridge University Press.

da Costa, S., Páez, D., Sánchez, F., Garaigordobil, M., & Gondim, S. (2015). Personal factors of creativity: A second order meta-analysis. *Journal of Work and Organizational Psychology*, *31*(3),165-173. https://doi.org/10.1016/j.rpto.2015.06.002

DiLiello, T. C., & Houghton, J. D. (2008). Creative potential and practiced creativity: Identifying untapped creativity in organizations. *Creativity and*

Innovation Management, 17(1), 37-46. https://doi.org/10.1111/j.1467-8691.2007.00464.x

Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics, 70*(1), 927-939. https://doi.org/10.1111/j.1468-0084.2008.00537.x

Falco, M. (2016). Potenciando la creatividad en el ámbito organizacional. *Ciencias Administrativas, 8*, 53-60.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. https://doi.org/10.2307/3151312

García-Vidal, G., Sánchez-Rodríguez, A., Pérez-Campdesuñer, R., & Martínez-Vivar, R. (2019). The impact of self-confidence, creativity and vision on leadership performance: Perceptions at Ecuadorian SMEs owner/managers. *Serbian Journal of Management, 14*(2), 315-325. https://doi.org/10.5937/sjm14-17569

Gonlepa, M. K., Dilawar, S., & Amosun, T. S. (2023). Understanding employee creativity from the perspectives of grit, work engagement, person organization fit, and feedback. *Frontiers in Psychology, 13*, 1012315. https://doi.org/10.3389/fpsyg.2022.1012315

Gu, Q., Tang, T. L.-P., & Jiang, W. (2015). Does moral leadership enhance employee creativity? Employee identification with leader and leader-member exchange (LMX) in the Chinese context. *Journal of Business Ethics, 126*(3), 513-529. https://doi.org/10.1007/s10551-013-1967-9013-1967-9

Helzer, E. G., & Kim, S. H. (2019). Creativity for workplace well- being. *Academy of Management Perspectives, 33*(2), 134-147. https://doi.org/10.5465/amp.2016.0141

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huang, M.-J. (2023). Leader self-deprecating humor and employee creativity at workplace: A longitudinal study. *Review of Managerial Science, 17*(2), 467-492. https://doi.org/10.1007/s11846-022-00532-w

Kline, T. J. (2005). *Psychological testing: A practical approach to designand evaluation.* SAGE Publications.

Li, T., & Yue, C. (2019). Working with creative leaders: An examination of the relationship between leader and team creativity. *Social Behavior & Personality: An International Journal, 47*(6), 1-12. https://doi-org/10.2224/sbp.8084

Luu, T. T., Rowley, C., Dinh, C. K., Qian, D., & Le, H. Q. (2019). Team Creativity in public healthcare organizations: The roles of charismatic leadership, team job crafting and collective public service motivation. *Public performance & Management Review, 42*(6), 1448-1480. https://doi.org/10.1080/15309576.2019.1595067

Montero, I., & León, O. G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology, 7*(3), 847-862.

Mubarak, F., & Noor, A. (2018). Effect of authentic leadership on employee creativity in project-based organizations with the mediating roles of work engagement and psychological empowerment. *Cogent Business & Management, 5*(1), 1-12. https://doi.org/10.1080/23311975.2018.1429348

Ramos, M. A. W., Figueiredo, P. S., & Pereira- Guizzo, C. (2018). Antecedents of innovation industry: The impact of work environment factors on creative performance. *Innovation & Management, 15*(3), 269-285. https://doi.org/10.1108/INMR-05-2018-0032

Rahman Khattak, S., Saleem, Z., & Khan, H. (2017). Relationship between goal orientation and employee creativity: A mediating role of creative self-efficacy. *International Journal of Organizational Leadership, 6*(4), 434-443. https://doi.org/10.33844/ijol.2017.60338

Sajid, K., Zafar, S., & Hadayat, K. (2017). Relationship between goal orientation and employee creativity: A mediating role of creative self-efficacy. International Journal of Organizational Leadership, 6(4), 434-443. https://doi.org/10.33844/ijol.2017.60338

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507-514. https://doi.org/10.1007/BF02296192

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). Routledge/Taylor & Francis Group.

Shalley, C. E., Hitt, M. A., & Zhou, J. (Eds.). (2015). *The Oxford Handbook of Creativity, Innovation, and Entrepreneurship*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199927678.001.0001

Sousa, F., Pellissier, R., & Monteiro, P. (2012). Creativity, innovation and collaborative organizations. *The International Journal of Organizational Innovation, 5*(1), 26-64. http://hdl.handle.net/10400.1/2611

Woodman, R. W., & Schoenfeldt, L. F. (1990). An interactionist model of creative behavior. *The Journal of Creative Behavior, 24*(1), 10-20. https://doi.org/10.1002/j.2162-6057.1990.tb00525.x

Zhang, R., Kang, H., Jiang, Z., & Niu, X. (2023). How does workplace ostracism hurt employee creativity? Thriving at work as a mediator and organization-based self-esteem as a moderator. *Applied Psychology: An International Review, 72*(1), 211-230. https://doi.org/10.1111/apps.12374

Mis agradecimientos, Profe Edgardo Raúl Pérez

En el año 2000, tuve el privilegio de comenzar a trabajar con el Dr. Edgardo Raúl Pérez tras rendir un examen de técnica psicométrica. Hasta ese momento, no lo conocía ni como profesional ni como persona. Ese encuentro marcó un antes y un después en mi vida, transformando profundamente mi desarrollo profesional y personal.

Podría mencionar todas las contribuciones y el trabajo que hemos realizado juntos, los congresos, los cursos y demás experiencias compartidas. Sin embargo, en estas palabras quiero destacar lo que él significó para mí como profesor y como ser humano. Con él, aprendí no solo a trabajar, sino también a pensar, razonar y tomar decisiones en diversas áreas de mi formación. Su enfoque en el análisis estadístico y su incesante pasión por el aprendizaje me enseñaron la importancia de aprovechar cada oportunidad para profundizar en el conocimiento.

El legado y la formación que me brindó fueron pilares fundamentales en mi crecimiento como profesional, investigador y docente. Con el Dr. Pérez, comprendí que la relación entre maestro y discípulo es esencial, y que formar a otros es un acto invaluable. Aprendí de él que delegar no solo es necesario, sino también una manera de formar a quienes asumen nuevas responsabilidades, promoviendo el trabajo en equipo y la cooperación en la búsqueda del conocimiento.

Además, me enseñó que el conocimiento y la experiencia son herramientas esenciales, pero que lo más importante en la formación de otros es transmitir esos saberes con pasión y generosidad. Un profesor y un investigador tienen la obligación de aprender y compartir sus conocimientos sin reservas, siempre buscando impactar positivamente a sus discípulos, estudiantes y a la sociedad en general. Transmitir el conocimiento y contribuir al crecimiento de quienes nos rodean es una responsabilidad que ningún docente o investigador debería ignorar.

Agradezco profundamente todo lo que el Dr. Pérez me enseñó. Creo que nunca podré encontrar las palabras adecuadas para expresar mi gratitud. Considero que fue una de las personas más importantes en mi vida, alguien que realmente me marcó y me ayudó a encontrar mi camino en esta profesión y en otros aspectos de la vida. Estoy seguro de que su forma de pensar, su visión y su manera de vivir perdurarán en mí y en todos aquellos que trabajan a mi lado y que se están formando en esta disciplina.

Hasta siempre, y con mucho cariño, profe.

Marcos Cupani