Psychology, Society & Education

Validation of the Purpose in Life Scale-Short Form in an Ecuadorian sample: factor analysis, reliability and graded response model

Rodrigo Moreta-Herrera* [1] [2], Alberto Rodríguez-Lorenzana[1], Guido Mascialino[1], Tomás Caycho-Rodríguez[3], Claudio López-Calle[4], Miguel Gallegos[5] [6], Mauricio Cervigni[7] [8], Pablo Martino[7] [9] & Lindsey W. Vilca[10]

[1] *Universidad de Las Américas, Quito (Ecuador)*

[2] *Pontificia Universidad Católica del Ecuador, Quito (Ecuador)*

[3] *Universidad Científica del Sur, Lima (Peru)*

[4] *Universidad de Cuenca, Cuenca (Ecuador)*

[5] *Universidad Católica del Maule, Talca (Chile)*

[6] *Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte (Brazil)*

[7] *Universidad Nacional de Rosario, Rosario (Argentine)*

[8] *CONICET (National Council for Scientific and Technical Research, Argentine)*

[9] *Universidad Norbert Wiener, Lima (Peru)*

[10] *Universidad Nacional de San Luis, San Luis (Argentine)*

KEYWORDS Difficulty Discrimination Reliability Purpose in life-Short Form Factor analysis Meaning in life	**ABSTRACT** Introduction: The purpose in life is analyzed within psychological well-being, it deals with the meaning of life that people have, which has implications in areas such as mental health. Currently, precise instrumentation is required for its assessment of the Ecuadorian population, which at the moment is limited. Objective: To corroborate the unidimensional structure of the Purpose in Life Scale-Short version (PIL-SF), the equivalence of measurement according to gender, the reliability and the parameters of discrimination and difficulty in a sample of Ecuadorian adults. Methods: Instrumental study of the PIL-SF under the precepts of Classical Test Theory and Item Response Theory. Participants: 743 Ecuadorian adults (64.2% women and 45.8% men) aged 18 to 73 years ($M = 24.7$, $SD = 7.9$). Results: The unifactorial structure of the PIL-SF with adequate adjustment was confirmed. In addition, the PIL-SF is invariant at the strong level based on gender and no differences are presented at the level of latent means. The internal consistency reliability is acceptable per group and the discrimination and item difficulty parameters are adequate. Conclusion: The PIL-SF is a valid, reliable, and accurate tool to measure the Purpose in Life construct in Ecuadorian adults.

Validación de la Versión Corta de la Escala de Propósito en la Vida en muestra ecuatoriana: análisis factorial, confiabilidad y modelo de respuesta gradual

PALABRAS CLAVE Dificultad Discriminación Fiabilidad Propósito en la vida-Forma breve Análisis factorial Significado de vida	**RESUMEN** Introducción: El propósito de vida se analiza dentro del bienestar psicológico, se trata del sentido de vida que tienen las personas, lo que tiene implicaciones en áreas como la salud mental. Actualmente se requiere instrumentación precisa para su valoración de la población ecuatoriana, la cual por el momento es limitada. Objetivo: corroborar la estructura unidimensional de la Escala de Propósito en la Vida-Versión Corta (PIL-SF), la equivalencia de medida según género, la confiabilidad y los parámetros de discriminación y dificultad en una muestra de adultos ecuatorianos. Métodos: Estudio instrumental del PIL-SF bajo los preceptos de la Teoría Clásica de Test y la Teoría de Respuesta al Ítem. Participantes: 743 adultos ecuatorianos (64.2% mujeres y 45.8% hombres) de 18 a 73 años ($M = 24.7$, $DT = 7.9$). Resultados: Se confirmó la estructura unifactorial del PIL-SF con ajuste adecuado. Además, el PIL-SF es invariante a nivel fuerte en función del género y no se presentan diferencias a nivel de medias latentes. La confiabilidad de la consistencia interna es aceptable por grupo y los parámetros de discriminación y dificultad de los ítems son adecuados. Conclusión: El PIL-SF es una herramienta válida, confiable y precisa para medir el constructo Propósito en la Vida en adultos ecuatorianos.

The construct of purpose in life (PIL) or meaning in life (MIL), which has been widely explored in psychology since the 1960s, can be traced back to Crumbaught and Maholick's (1969) operationalization of Victor Frankl's logotherapy. Frankl (1953; 1984) posited that the primary driving force in human life is the search for meaning, more so than other motivations, however basic they may seem. An existentially frustrated person, one without meaning in life, would be susceptible to many psychological illnesses. The MIL is also fundamental for understanding psychological well-being, as it is part of the construct that allows us to assess the positive aspects of life (Casas & González-Carrasco, 2021; Ryff, 1989; Ryff et al., 2006; Torres-Salazar et al., 2020). The MIL has a relevant importance in psychology, which is why it gives the individual reasons and motivations to exist and feel an important and relevant part of life. In addition, it is an element that allows us to understand other phenomena with which it is associated, such as mental health and psychopathology, well-being, cognitive functioning, burnout among health professionals, employability, inflammation, marital interacion, and physical activity, among many others (Galek et al., 2015; Kim et al., 2019; Sharpley et al., 2021), so its monitoring and exploration is relevant. Psychology and related fields have traditionally been interested in developing measures that allow for empirical estimation of the MIL. This can be a complex aspect due to the difficulty of operationalizing constructs that start from a strong subjectivity that seeks to identify the purpose or reason for living on the part of individuals. In this sense, the Purpose in Life Test (PIL-Test) (Crumbaugh & Maholick, 1964) was developed to know the level of perception of purpose an individual has about the life he/she has.

The PIL-Test (Crumbaugh & Maholick, 1964), is the first questionnaire that allowed to obtain quantitative and qualitative results regarding the perception of the constructs "purpose in life". The test is divided into two parts: Part A is composed of 20 items and provides quantitative information (this part is the most relevant for the study as it allows a more objective measurement of the MIL) with items like: "Presence of clear goals in life", while Parts B and C have a qualitative component. In Part A, the elements of the PIL-Test are used through a 7-point Likert-type response format (where 1 is *strongly disagree*, 4 *neutral* and 7 *totally agree*). The total score on the test is obtained by adding the scores on each of the items. Lower scores indicate a perception of less meaning/purpose in life. There is ample evidence of the validity and reliability of the PIL-Test in different population segments in different countries, including Spanish-speaking countries (Brunelli et al., 2012; Caycho-Rodríguez et al., 2022a; Garcia-Alandete et al., 2017; Simkin et al., 2018), supporting the existence of a unifactorial dimensional structure.

PIL-SF: implications and psychometric needs

Despite the good psychometric properties reported for the PIL-Test, it has sometimes been criticized for the fact that some of its items may be indicative of constructs other than meaning in life, such as depression (Schulenberg & Melton, 2010; Steger 2006). Considering these and other criticisms, such as

reservations about the unidimensionality of the scale (Brunelli et al., 2012; Ochoa et al., 2018; Ortiz et al., 2012), Morgan and Farsides (2009) developed a study to identify possible factors within the scale. In this way, the authors generated the Exciting life factor of the expanded scale and the Purposeful life factor composed of the same scale. Furthermore, with this evidence on the behavior of the items, Schulenberg et al. (2011) developed the shortened version of the PIL scale (PIL-SF) consisting of four items. One of the main reasons for the development of this shortened version, in addition to the obvious issue of application time, was to counteract the criticism that had previously been reflected on the question of the unidimensionality of the scale or the unspecificity of some of its items (Steger, 2006).

Thus, the PIL-SF was defined by items 3, 4, 8, and 20 of the original PIL-Test (Schulenberg et al., 2011). For the authors, the selection of these items allows for a more direct examination of the MIL construct and identifies the presence of clear life goals, the perception of whether life has meaning, the presence of life goals/purposes, and the achievement of these goals/purposes (Schulenberg et al., 2010). The study establishes a one-dimensional structure of four items with an adequate internal consistency of $\alpha = .86$ and it shows divergence with the construct of psychological distress in a moderate way ($r = -.49$). The PIL-SF has also been translated and validated to a limited extent, in which it has been found to have adequate psychometric properties (mainly factor validity and reliability) (Pacak-Vedel et al., 2021) like in splanish undergraduates (Rubio-Belmonte et al., 2023) and the Latin American context (Caycho-Rodríguez et al., 2022b; Weber et al., 2022), in which the unidimensional factorial structure is confirmed, in addition to an adequate internal consistency reliability. It should be noted that in this process of psychometric exploration of the PIL-SF there are no validity studies in relation to other variables to determine the convergence of the scale with other relevant variables for the context such as mental health or symptoms of disturbance conditions such as depression, for example.

Continuing the instrumental research of the PIL-SF is necessary, on the one hand, to determine the feasibility of the scale to accurately assess the MIL construct in different populations and nations, such as Ecuador. And, on the other hand, in exploring other less analyzed traits that are necessary to estimate its efficacy. Currently, much of the instrumental evaluation of the PIL-SF is based on Classical test theory (CTT) considerations (factorial validity, reliability, others), but the verification of measurement equivalence (ME) is almost nonexistent. Only the work of Caycho-Rodríguez et al. (2022b), who tested ME as a function of national origin in Latin America, stands out. However, for other variables of interest, such as gender (given its relevance in social and psychological research), there is no current evidence of its application in the case of the PIL-SF, although some results suggest the existence of differences (Xi et al., 2022), in which women score higher than men. Currently, the gender presents scientific, social, and cultural importance, with debates about the psychic response in various attributes that sometimes present prejudices and inaccuracies (Upchurch, 2020) and feed narra-

tives loaded with ideology that in this aspect science and psychology should help to clear with evidence But in the case of the sciences, preliminary findings of gender differences without precise and reliable instruments are questionable, since they do not guarantee that the calibration represents equality of measurement between the groups. In this sense, is common in multi-group research that, when conducting comparative analyses, no prior verification of the ME of a measure is carried out in order to estimate inter-group discrepancies because, if they exist, it is unlikely to determine whether they are due to divergences in the interpretation of the items in the structure of the scale or to the characteristics of the groups themselves (Asparouhov & Muthén, 2014; Larzabal-Fernandez et al., 2023; Moreta-Herrera et al., 2021; Van de Schoot et al., 2015). The inability to account for the exact origin of the difference between groups leads to bias in the estimation of the final conclusions. Therefore, it is necessary to examine analyses of this type that help to clarify information about the presence or absence of this property.

PIL-SF and the Item Response Theory (IRT)

It is usual the analyses of psychometric properties of the tests are focused on the CTT approaches, which explore the global properties of a scale based on a normative sample (Chávez Vega & Rodríguez Méndez, 2022; Muñiz Fernández, 2010), leaving aside analyses of other types of psychometric properties, such as those based on IRT, which ask about the particular performance of the items of a scale, such as discrimination (a), difficulty (b), and pseudo-azar (c), rather than about people's ability to respond correctly (Cai et al., 2016; Moreta-Herrera et al., 2023). IRT-based research with the PIL-SF is no exception, and interest in this aspect is low. The only exception is the work of Caycho-Rodríguez et al. (2022b), which concludes that the PIL-SF items have adequate discrimination and difficulty, especially the item "Presence of goals/purpose in life" (item 20), which is the best discriminator and shows greater precision in the latent variable. The use of IRT allows a better understanding of the mechanism by which a measure manages to assess a specific construct (Hambleton et al., 2010; Moreta-Herrera et al., 2022). Thus, the use of CTT and IRT provides more precise information about the evaluative capacity of a measure.

The present study

As previously mentioned, the PIL-SF emerges as a brief measure that calculates the MIL with greater precision, unlike the extended version, which also other different constructs such as depression and these elements as suggested (Schulenberg et al., 2011) should be analyzed independently. Hence, it is necessary to analyze this construct with other measures aimed at psychological distress, depression, and others. On the other hand, although the data on the psychometric properties of the PIL-SF in the Ecuadorian population are presented through a cross-cultural study, the evidence is still insufficient (at the national and regional level), so it would be considered relevant

to also establish specific studies in this type of population on the validity and applicability of the measure. In the context of psychological research, the Ecuadorian population has little tradition of study than other groups in the region and the information is limited. An instrumental study of the PIL-SF through a specific study would significantly contribute to learning more about this population group, would allow monitoring of the measure in relation to the reference population, and, in turn, would contribute to the current discussion on the employability of measure. Therefore, the objectives of this study are: a) to verify the unifactorial structure of the PIL-SF in a sample of adults from Ecuador; b) check the equivalence of measurement based on gender; c) confirm the reliability of the internal consistency; and d) estimate the discrimination and difficulty parameters of the items using the IRT. For this, it is proposed as a hypothesis that the one-dimensional structure of the PIL-SF is adequate (H_1); the equivalence of measurement according to gender is strict (H_2); the reliability of the test is adequate (H3); and the discrimination and difficulty parameters of the items are suitable (H_3).

Method

Design

The present research is methodologically based on the work of Ato et al. (2013), in which the psychometric properties of the PIL-SF were determined using CTT (i. e., factorial validity and internal consistency reliability), and IRT (i. e., item discrimination and difficulty).

Participants

A total of 743 Ecuadorian people (64.2% women) participated in the study. Their age ranged from 18 to 73 years ($M = 24.7$; $SD = 7.9$). Of the participants, 68.4% had higher education (Bachelor's degree or higher), 31.2% had secondary education, and the remaining 0.4% had primary education or no education at all. The selection of the participants was done through non-probabilistic intentional sampling, based on the intentional adhesion of the participants through the link published in the different diffusion networks and who gave their consent to participate.

Instruments

Purpose in Life Scale-Short Form (PIL-SF; Crumbaugh & Maholick, 1969; Schulenberg et al., 2011). This scale is designed to identify the level of purpose and meaning people have in life. The short scale consists of four items derived from the original 20-item scale. Items 3, 4, 8, and 20 of the Spanish form were used in this four-item version (Simkin et al., 2018), which are answered on a 7-point Likert scale. The scale ranges from 4 to 28 points, with the assumption that the higher the score, the greater the presence of purpose in life. In the original version the reliability is adequate with α = .86.

Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001; Spitzer et al., 1999) in the Spanish version of Urtason et al. (2017). Designed to assess the condition associated with symptoms of depression in people with primary care needs. It consists of nine items that are answered on a 4-option Likert scale between 0 = *never day* to 3 = *every day*. The psychometric properties of the scale show that its internal consistency is adequate with values of ω = .89 [.88 - .9].

Procedure

The study was conducted virtually because of the conditions of social isolation in Ecuador due to COVID-19, which forced the suspension of face-to-face activities, including this research. Potential study participants were contacted through the dissemination of the project on social networks (Facebook, Instagram, Twitter, and others) and by e-mail. They voluntarily clicked the link to the psychological evaluation in order to be able to answer it. Before starting the study, the collaborators read the project information and the conditions of participation, gave their consent, and proceeded to fill out the tests. Once this process was completed, the information was collected and systematized for statistical treatment. It should be noted that 756 data were collected, but 13 were discarded because they were incomplete in their completion, so the final sample was 743 cases. The hypothesis was tested, and the conclusions were analyzed. Finally, the research reports were written. The guidelines and protocols used in the present study followed the ethical and legal norms regulated by the APA and the Helsinki Convention for research with human subjects and approved by the ethics committee of the National University of Rosario (Argentina). In addition, it had the support of the Rosario Neuroscience Research Center (CINR) and the Cognition and Emotion Laboratory (LABce) of the School of Psychology of the National University of Rosario.

Data analysis

The results were analyzed in three blocks. The first includes a descriptive analysis of the items, including measures of central tendency (arithmetic mean [*M*]), dispersion (standard deviation [*SD*]), and distribution (skewness [g_1] y kurtosis [g_2]). The distribution measures are used to test for compliance with the univariate normality assumption, where the values of g_1 and g_2 are within the range ±1.5 (Ferrando & Anguiano-Carrasco, 2010). In addition, multivariate normality is examined through Mardia's test on g_1 and g_2, and it is expected that there is no statistical significance (*p* > .05) for compliance (Mardia, 1970). Finally, the matrix of polychoric correlations of the PIL-SF items is presented to find the performance of the items within the measure configuration (Jin & Cao, 2018).

The second block is an analysis of the psychometric properties of the PIL-SF based on TCT. These include the AFC, the AFC-MG, and the internal consistency reliability. For the AFC (see Figure 1), the single-factor model of the measure is tested through weighted mean and variance adjusted maximum likelihood (WLSMV) estimation, as the items do not exhibit multivariate normality (Li, 2016). The validity of the factor model is verified according to the degree of fit using fit indicators: a) absolute (Chi-square [χ^2] and Standardized Root Mean Square Residual [SRMR]); b) relative (Comparative Fit Index [CFI] and Tucker-Lewis Index [TLI]); and c) not based on centrality (Root Mean Squared Error of Approximation [RMSEA]). The established cut-off points are that the χ^2 is not significant (*p* > .05), the CFI and TLI are greater than .9, and the SRMR and RMSEA are less than .08. Additionally, in the case of item saturation (λ), they are expected to be greater than .50 (Brown, 2015; Byrne, 2008; Moreta-Herrera et al., 2020; Mueller & Hancock, 2018; Wolf et al., 2013). Attached to this analysis, the ME by gender is studied with the AFC-MG with WLSMV estimation (see Table 2) to know the presence or not of invariance of the factorial model by segmenting the sample into groups from a referential variable (gender). For this, independent CFAs are performed for each group, from which it is expected that the differences in the chi-squares ($\Delta\chi^2$) are not significant (*p* > .05). We then proceed to increase the restrictions on the model (metric, hard, and hard) and measure the difference (*Δ*) of the χ^2, CFI, and RMSEA values between each increase in restriction on the model. It is expected that the differential found will not be significantly high (Asparouhov & Muthén, 2014; Brown, 2015). If strong invariance is verified, the latent means are analysed for construct differences between groups (Δ_K). For this purpose, the intercepts of the male group are set to 0, while the female group is set to free; if no significance is found (*p* <. 05), it would be assumed that the groups are similar in terms of the manifestation of MIL. Finally, this section completes the AFC and the AFC-MG by examining the internal consistency reliability of the scale through the Cronbach's test (α) together with the 95% confidence intervals (IC 95%) and the convergence validity of the scale through the PHQ-9 by means of the Pearson correlation coefficient (*r*) both for the total group of the sample and for groups classified by gender.

The third and last block consists of the analysis of the psychometric properties of the PIL-SF by means of IRT, which includes the (a) discrimination and (b) difficulty parameters. The Graded Response Model (GRM), which is an extension of the 2-parameter logistic model (2-PLM) for ordered polytomous items (Hambleton et al., 2010; Samejima, 1997), is used to test a and b. Parameter a examines the slope with which item responses change as a function of the individual's skill level (item scores are expected to be greater than 1 for adequate discrimination), while parameter b analyses how the item behaves along the ability scale, determined at the point of average probability (50%) of the individual's ability to obtain a correct response. Since the scale has seven items, there are six estimates of difficulty (1 for each threshold). For a more detailed view, the information curves of both the scale (Test Information Curve [TIC]) and the items (Item Information Curve [IIC]) are examined.

Data management was performed with the R programming language, version 4.2.1 (R Core Team, 2022), using the packages: *foreign, Lavaan, SemPlot, lmr, MBESS*, and *MNV.*

Results

Descriptive analysis

Table 1 shows the performance of the PIL-SF items. It can be seen that the mean scores vary between PIL-SF$_{(item\ 8)}$ = 5.3; SD = 1.31 and PIL-SF$_{(item\ 3)}$ = 5.93; SD = 1.25 with a difference of ΔPIL-SF$_{(item\ 8\ -\ item\ 3)}$ = 0.63 points between them, which is considered homogeneous. Regarding the behavior of g_1 and g_2, they are in all cases within the range ±1.5, so that in this aspect the assumption of univariate normality is present; however, at the multivariate level this assumption is not fulfilled, since there is statistical significance ($p < .05$) in the Mardia tests for g_1 and g_2. This establishes that the use of robust estimators is recommended for CFA. Finally, the matrix of polychoric correlations of the items shows that the association between items is marked, which indicates contribution between items to the construct and non-redundancy (multicollinearity).

Confirmatory Factor Analysis

Figure 1 shows the CFA with WLSMV estimation of the PIL-SF, which confirms that the measure is unifactorial according to the proposed theoretical model. The goodness of fit indicators (χ^2, CFI, TLI, SRMR, and RMSEA) indicate that the model has an adequate fit for its assessment in Ecuadorian adults. Likewise, in the analysis of the factor loadings, all of them are above $\lambda > .50$, which establishes that the items are consistent within the proposed model and contribute significantly to the construct and allow an adequate explanation of the variance. Finally, regarding the internal consistency reliability of the PIL-SF, the results show the presence of adequate consistency both in the α value and in its CIs.

Invariance by gender

Table 2 shows the ME analysis of the PIL-SF considering gender. The AFCs for men and women show adequate fit indi-

Table 1

PIL-SF descriptive analysis

Ítems	Descriptives				Correlation matrix			
	M	SD	g_1	g_2	1	2	3	4
1. Presence of clear goals in life	5.93	1.25	-1.3	1.45	1			
2. Life has meaning	5.8	1.4	-1.22	1.01	.68**	1		
3. Fulfillment of vital goals	5.3	1.31	-0.67	0.26	.59**	.60**	1	
4. Presence of goals/purpose in life	5.74	1.36	-1.18	1.07	.70**	.71**	.61**	1
Mardia			819.6***	35.9***				

Note. M = arithmetic mean; *SD* = Standard Deviation; g_1 = asymmetry; g_2 = kurtosis.
** $p < .05$, *** $p < .001$.

Figure 1

Confirmatory Factor Analysis of PIL-SF with WLSMV estimation
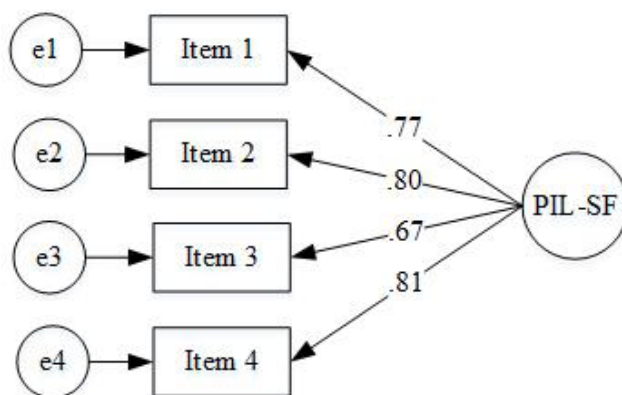
$\chi^2 = .057$; $df = 2$; $p < .001$; CFI = .999; TLI = .998; SRMR = .003; RMSEA = .001 [.000 - .005]
$\alpha = .85$, IC$_{95\%}$ [.83 - .87]

Note. χ^2/df = Normalized chi-square; CFI = Comparative adjustment index; TLI = Tucker-Lewis Index; RMSEA = Approximate root mean square error; SRMR = Standardized root mean square; α = Cronbach's alpha; IC$_{95\%}$ = 95% confidence intervals.

cators for the established groups. Already in the AFC-MG of the PIL-SF it is observed through the anova test that the values in the χ^2 for both men and women do not show significant differences ($\Delta\chi^2$ = -0.12; p > .05). In addition, as the factor model increases in the restrictions (metric, strong, and strict), the differential (Δ) in the CFI and RMSEA are not significant, so the equivalence of the PIL-SF is present at the strong level.

In addition, Table 3 shows the loadings of the CFAs by groups. As in the global model, the λ are greater than .50, so it is considered that they contribute significantly to the factorial model classified by groups. The differential between loadings is not significant. With these aspects, the equivalence of the PIL-SF measure is corroborated.

On the other hand, regarding the internal consistency segmented by groups, it is observed to be adequate for the sample of Ecuadorian adults; moreover, the differences between the α are statistically significant, since the CIs include zero.

Finally, in the convergence analysis, the PIL-SF is analyzed with the PHQ-9, which assesses the perception of health in general. It is observed that the PIL-SF converges moderately with the PHQ-9 with the total sample, while, by gender, the convergence is slightly higher in the group of women, the correlation being moderate, while in the group of men, it is slight. In all these cases, the relationship is negative.

Finally, because of the presence of ME at the strong level of the PIL-SF, we analyzed the latent means of the PIL-SF

segmented by groups. Once the men's group was placed in the intercepts at 0 and the women's group at free, it was found that there were no differences at the latent level with values of Δ_K = .036; p = .438. In addition, the differences between groups at the level of direct variables were t =-0.47; p= .639 and without effect size Δ_K^* =-0.03. In this sense, the groups are invariant with respect to the PIL.

Graded response model

Table 4 shows the values corresponding to the (a) discrimination of the items. In all cases, the PIL-SF items have an adequate discrimination since their values are greater than 1. On the other hand, regarding the (b) difficulty, as observed in the estimators of the thresholds from b1 to b6, they increased monotonically, so it is considered that the difficulty of the items is adequate.

Figure 2 shows the information curves for the four items of the PIL-SF (IIC) and for the scale in general (TIC). In the IIC, the item with the greatest relevance and precision for assessing the latent variable is item 20 (Presence of goals/purpose and Presence of goals/purpose in life), which has the greatest discriminatory capacity, followed by items 4 and 3, while the item with the least capacity is item 8 (Fulfillment of life goals). Finally, in the overall ICT scale, the scale is reliable and accurate, especially in the -2 and 1 ranges.

Table 2

Analysis of equivalence of measurement according to gender of the PIL

Restrictions	(*df*) χ^2	CFI	RMSEA	Δ(*df*) χ^2	ΔCFI	ΔRMSEA
Base males	(2) 0.44	.999	.000	-	-	-
Base Females	(2) 0.96	.998	.000	-	-	-
Unrestricted	(4) 0.24	.999	.000	-	-	-
Metrics	(7) 2.47	.998	.020	(3) 4.63	.002	.020
Strong	(10) 5.1	.988	.036	(3) 6.77	.009	.026
Strict	(11) 5.62	.999	.000	(1) 0.12	.011	.036

Note. $\chi2$ = Chi-square; *df* = Degrees of freedom; CFI = Comparative Adjustment Ratio; RMSEA = Mean square error of approximation; Δ = Delta.

Table 3

Factor loadings analysis of the PIL-SF and reliability according to gender

| Ítems	λ_{Males}	$\lambda_{Females}$	$\Delta\lambda$
1. Presence of goals/purpose in life	.83	.87	-.04
2. Fulfillment of vital goals	.65	.76	-.11
3. Life has meaning	.82	.85	-.03
4. Presence of clear goals in life	.82	.81	.01
Internal consistency	α_{Males} IC$_{95\%}$	$\alpha_{Females}$ IC$_{95\%}$	Δ_α IC$_{95\%}$
PIL-SF	.82 [.78 - .86]	.87 [.85 - .89]	-.05 [-.12 - .07]
Convergence validity			
PHQ-9	-.382***	-.460***	-.429***

Note. α = Cronbach's alpha; IC$_{95\%}$ = Confidence intervals at 95%; λ = Saturations; $\Delta\lambda$ = Delta Lambda; PHQ-9 = Patient Health Questionnaire-9.

Table 4

PIL-SF step response analysis

Ítems	a	Difficulty					
		b1	b2	b3	b4	b5	b6
Ítem 3.	2.641	-3.319	-2.418	-1.998	-1.307	-0.677	0.214
Ítem 4.	2.864	-2.730	-2.194	-1.688	-1.102	-0.531	0.193
Ítem 8.	1.970	-3.273	-2.541	-1.837	-0.901	0.042	1.112
Ítem 20.	3.095	-2.674	-2.257	-1.722	-1.110	-0.494	0.366

Note. a = Difficulty; b = Discrimination.

Figure 2

Information curves of the items and of the PIL-SF in general
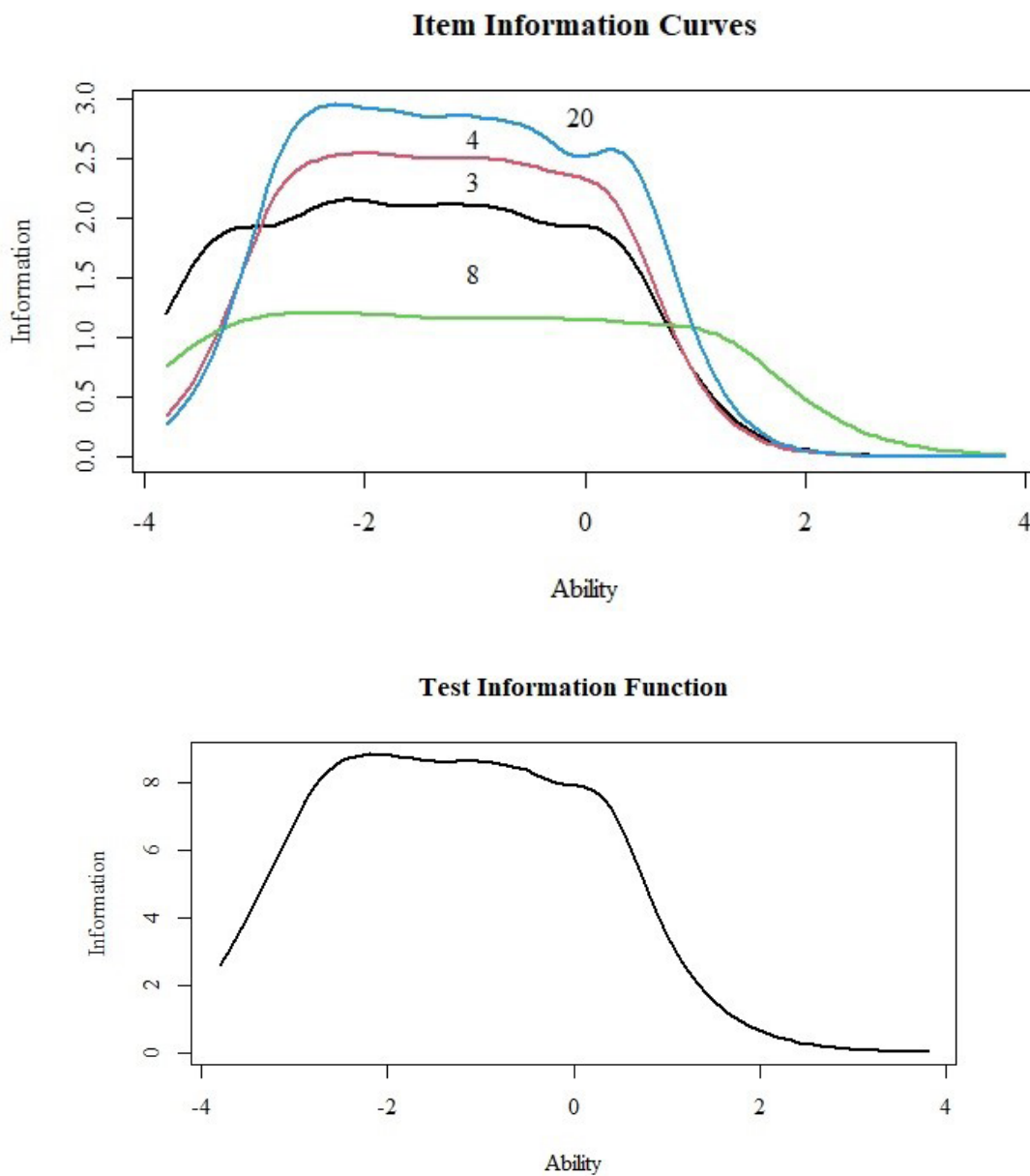
Note. IIC = Item Information Curves; TIC = Test Information Curves.

Discussion

The objectives of the study were to confirm the unidimensional structure of the PIL-SF in a sample of Ecuadorian adults, as well as the properties of ME by gender and reliability based on the TCT, and finally the parameters of difficulty and discrimination based on the IRT.

In the analysis of the factorial structure, the AFC confirms that the PIL-SF, structured in a unidimensional way, represents an adequate adaptation to interpret the MIL construct in Ecuadorian adults. The fit indices reported to establish factorial validity are within acceptable ranges (Brown, 2015; Byrne, 2008; Mueller & Hancock, 2018; Wolf et al., 2013). The unifactorial structure of the PIL-SF in the reference sample is consistent with both the original version (Schulenberg et al., 2011) and other validation work (Caycho-Rodríguez et al., 2022b; Pacak-Vedel et al., 2021; Weber et al., 2022). This supports the hypothesis of unidimensionality of the measure.

Regarding the gender-based ME of the PIL-SF, the results show that the measure is invariant at a strong level according to the acceptance criteria (Asparouhov & Muthén, 2014; Brown, 2015). The evidence shows that regardless of gender status, the interpretation of items within the factor structure is similar. Furthermore, in the AFC, the criteria for the fit of the measure are adequate regardless of the groups. With regard to the results of the MS, no preliminary studies have been reported with which the present results can be compared. In this sense, the conclusions reached are pioneering for instrumental research on PIL-SF and EM. Furthermore, with this evidence, the development of multi-group comparative studies is feasible for the Ecuadorian sample in order to confirm this property. Given the strong ME of the PIL-SF, the differences in the latent means between groups were analyzed. In this analysis, men and women do not differ significantly from each other ($p > .05$), so they are considered equal. In this sense, this result differs from that reported by Xi et al. (2022), who mentioned the presence of differences. It is likely that this equality is due to specific cultural factors of the participating sample, in which the social construction of the MIL does not depend on parenting styles defined by the gender roles to be played, however, this can be rushed to mention without further ado supporting evidence.

In the case of internal consistency reliability, the results achieved by groups are acceptable for each of them. In addition, the differences in α are not significant so they are considered equal. The confirmation of internal consistency in the present study agrees with previous reports analyzing this aspect (Caycho-Rodríguez et al., 2022b; Pacak-Vedel et al., 2021; Rubio-Belmonte et al., 2023; Weber et al., 2022). Likewise, regarding the validity of the relationship with other variables, the PIL-SF converges in a negative sense with the PHQ-9, indicating a relevant relationship with measures that assess the symptoms of depression. This is a representative aspect since no evidence of the relationship with other PIL-SF variables is collected, which implies a contribution from the study of the Ecuadorian sample, which is not collected with other populations. In addition, as he points out Schulenberg et al. (2011), the PIL-SF and depression are analyzed as independent variables, the same ones that show relationship validity and that could not be fully identified with the PIL-Test.

Finally, the analysis of the discrimination and difficulty parameters according to IRT, after verifying the assumption of unidimensionality and local independence, showed that the PIL-SF items represent an adequate parameterization of the analyzed criteria. The behavior of item 20 stands out, as it is the item with the best discrimination and precision for the assessment of the latent variable. In general, the scale is reliable and accurate in assessing the MIL construct. These results are consistent with the work of Caycho-Rodríguez et al. (2022b), although in this case the discrimination of item 20 is significantly more pronounced and different from the rest of the PIL-SF items.

The current study has limitations, especially in the co-sampling of the sample, since it was conducted exclusively with Ecuadorian adults; therefore, future validation studies are recommended for other populations such as adolescents, older adults, clinical populations, and others. Regarding the implications of the study, the current work provides evidence of the internal structure of the PIL-SF, for which there is still insufficient research in Ecuador and in the world. In addition, we have a precise measure to evaluate the PIL-SF adapted to Ecuadorian adults, with which we can carry out extensive studies in the population, not only at the prevalence level, but also comparative, given the verification of the ME by gender. This will also lay the groundwork for future studies in other populations and even in other national contexts within the region. In the case of ME verification, the contribution shown is unprecedented in other similar studies, making it a significant and pioneering contribution. The development of a short version of 4 items in relation to the 20-item one also has a practical implication when making quick and precise measurements, but also a methodological one, since it allows the PIL-SF to be more precise in the assessment of the MIL and with fewer measurement errors than the PIL-test, since this apparently measures with other attributes.

Conclusion

The present study contributes psychometric information derived from IRT, which not only helps to broaden the use of these analysis mechanisms but can also help to explain the reasons why the unidimensionality of the measure is criticized by several authors. In this sense, the use of analyses based on both TCT and IRT is relevant and allows for a broader approach to validation and confidence in the use of PIL-SF. In conclusion, we consider this study to be an important addition to the evidence base and discussion of the use of PIL-SF in the adult population of Ecuador and Latin America.

Funding

Conflict of interest

The authors declare that they have no conflict of interest in the publication of this article.

References

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 495-508. https://doi.org/10.1080/10705511.2014.919210

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

Brown, T. (2015). *Confirmatory factor analysis for applied research* (Vol. 2nd). Guilford Publications.

Brunelli, C., Bianchi, E., Murru, L., Monformoso, P., Bosisio, M., Gangeri, L., Miccinesi, M., Scrignaro, M., Ripamonti, C., & Borreani, C. (2012). Italian validation of the purpose in life (PIL) test and the seeking of noetic goals (SONG) test in a population of cancer patients. *Supportive Care in Cancer, 20*, 2775-2783. https://doi.org/10.1007/s00520-012-1399-6

Byrne, B. (2008). Testing for multigroup equivalence of a measuring instrument: a walk through the process. *Psicothema, 20*(4), 872-882.

Cai, L., Choi, K., Hansen, M., & Harrell, L. (2016). Item response theory. *Annual Review of Statistics and Its Application, 3*, 297-321. https://doi.org/10.1146/annurev-statistics-041715-033702

Casas, F., & González-Carrasco, M. (2021). Analysing Comparability of Four Multi-Item Well-being Psychometric Scales Among 35 Countries Using Children's Worlds 3rd Wave 10 and 12-year-olds Samples. *Child Indicators Research, 14*, 1829-1861. https://doi.org/10.1007/s12187-021-09825-0

Caycho-Rodríguez, T., Vilca, L., Cervigni M, G. M., Calandra M, R. A., López-Calle C, M.-H. R., … & Burgos-Videla, C. (2022a). Cross-national measurement invariance of the Purpose in Life Test in seven Latin American countries. *Frontiers in Psychology, 13*, Article 974133. https://doi.org/10.3389/fpsyg.2022.974133

Caycho-Rodríguez, T., Vilca, L., Cervigni, M., Gallegos, M., Martino, P., Calandra, M., … & Carbajal-León, C. (2022b). Cross-cultural measurement invariance of the purpose in life test-Short form (PIL-SF) in seven Latin American countries. *Current Psychology.* https://doi.org/10.1007/s12144-022-03465-5

Chávez Vega, R., & Rodríguez Méndez, A. (2022). Aplicación de la teoría clásica de test a la evaluación de preguntas de opción múltiple. *Educación Médica Superior, 36*(1), Article e2228.

Crumbaugh, J., & Maholick, L. (1964). An experimental study in existentialism: The psychometric approach to Frankl's concept of noogenic neurosis. *Journal of Clinical Psychology, 20*, 200-207.

Crumbaugh, J., & Maholick, L. (1969). *Manual of instructions for the Purpose-in-Life-Test.* Munster: Psychometric Affiliates.

Frankl, V. E. (1953). Logos and existence in psychotherapy. *American Journal of Psychoanalysis, 7*(1), 8-15. https://doi.org/10.1176/appi.psychotherapy.1953.7.1.8

Frankl, V. E. (1984). *Man's search for meaning.* Washington Square Press/Pocket Books.

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

Galek, K., Flannelly, K., Ellison, C., Silton, N., & Jankowski, K. (2015). Religion, meaning and purpose, and mental health. *Psychology of Religion and Spirituality, 7*(1), 1-12. https://doi.org/10.1037/a0037887

Garcia-Alandete, J., Marco, J., & Perez, S. (2017). Purpose-in-life test: Comparison of the main models in patients with mental disorders. *The Spanish Journal of Psychology, 20*(E31). https://doi.org/10.1017/sjp.2017.28

Hambleton, R. K., Van der Linden, W. J., & Wells, C. S. (2010). IRT models for the analysis of polytomously scored data: Brief and selected history of model building advance. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 21-42). Routledge/Taylor & Francis Group.

Jin, S., & Cao, C. (2018). Selecting polychoric instrumental variables in confirmatory factor analysis: An alternative specification test and effects of instrumental variables. *British Journal of Mathematical and Statistical Psychology, 71*(2), 387-413. https://doi.org10.1111/bmsp.12128

Kim, G., Shin, S., Scicolone, M., & Parmelee, P. (2019). Purpose in life protects against cognitive decline among older adults. *The American Journal of Geriatric Psychiatry, 27*(6), 593-601. https://doi.org/10.1016/j.jagp.2019.01.010

Kroenke, K., Spitzer, R., & Williams, J. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Larzabal-Fernandez, A., Pilco., K., Moreta-Herrera, R., & Rodas, J. (2023). Psychometric properties of the Kessler Psychological Distress Scale in a sample of adolescents from Ecuador. *Child Psychiatry & Human Development.* https://doi.org/10.1007/s10578-023-01501-4

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936-949. https://doi.org/10.3758/s1342

Mardia, K. (1970). Measures of Multivariate Skewness and Kurtosis with Applications Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika, 57*(3), 519-530. https://doi.org/10.2307/2334770

Moreta-Herrera, R., Caycho-Rodríguez, T., Salinas, A., Jiménez-Borja, M., Gavilanes-Gómez, D., & Jiménez-Mosquera, C. (2022). Factorial Validity, Reliability, Measurement Invariance and the Graded Response Model for the COVID-19 Anxiety Scale in a Sample of Ecuadorians. *OMEGA - Journal of Death and Dying.* https://doi.org/https://doi.org/10.1177/00302228221116515

Moreta-Herrera, R., Mayorga-Lascano, M., Larzabal-Fernandez, A., & Vaca-Quintana, D. (2020). Factor analysis, reliability and validity of a measurement model for the intent of alcohol consumption among Ecuadorian teenagers. *Health & Adictions/Salud y Drogas, 20*(1), 126-135. https://doi.org/10.21134/haaj.v23i1.741

Moreta-Herrera, R., Perdomo-Pérez, M., Reyes-Valenzuela, C., Torres-Salazar, C., & Ramírez-Iglesias, G. (2021). Invarianza factorial según nacionalidad y fiabilidad de la Escala de Afecto Positivo y Negativo (PANAS) en universitarios de Colombia y Ecuador. *Anuario de Psicología, 51*(2), 76-85. https://doi.org/10.1344/anpsic2021.51.9

Moreta-Herrera, R., Perdomo-Pérez, M., Reyes-Valenzuela, C., Gavilanes-Gómez, D., Rodas, J. A. & Rodríguez-Lorenzana, A. (2023). Analysis from the Classical Test Theory and Item Response Theory of the Satisfaction with Life Scale (SWLS) in an Ecuadorian and Colombian sample. *Journal of Human Behavior in the Social Environment.* https://doi.org/10.1080/10911359.2023.2187915

Morgan, J., & Farsides, T. (2009). Measuring meaning in life. *Journal of Happiness Studies, 10*, 197-214. https://doi.org/10.1007/s10902-007-9075-0

Mueller, R. O., & Hancock, G. R. (2018). Structural equation modeling. In G. Hancock, L. Stapleton & R. Mueller, *The Reviewer's Guide to Quantitative Methods in the Social Sciences* (pp. 445-456). Routledge.

Muñiz Fernández, J. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo, 31*(1), 57-66.

Ochoa, Y., Abreu, M., & Vea, H. (2018). Validation of the Purpose in Life Test (PIL) in adult patients with cancer. *Revista Habanera de Ciencias Médicas, 17*(5), 800-812.

Ortiz, E., Cano, Á., & Trujillo, C. (2012). Validación del Test de Propósito Vital (Pil Test-Purpose in Life Test) para Colombia. *Revista Argentina de Clínica Psicológica, 21*(1), 85-93.

Pacak-Vedel, A., Christoffersen, M., & Larsen, L. (2021). The Danish Purpose in Life test-Short Form (PIL-SF): Validation and age effects. *Scandinavian Journal of Psychology, 62*(2), 833-838. https://doi.org/10.1111/sjop.12776

R Core Team. (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Rubio-Belmonte, C., Mayordomo, T., & García-Alandete, J. (2023). Psychometric properties of the Purpose In Life-Short Form in Spanish population. *Journal of Clinical Psychology, 79*(4), 1099-1112. https://doi.org/10.1002/jclp.23461

Ryff, C. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57*, 1069-1081. https://doi.org/10.1037/0022-3514.57.6.1069

Ryff, C. D., Dienberg Love, G., Urry, H. L., Muller, D., Rosenkranz, M. A., Friedman, E. M., … & Singer, B. (2006). Psychological well-being and ill-being: do they have distinct or mirrored biological correlates? *Psychotherapy and Psychosomatics*, 75(2), 85-95.

Samejima, F. (1997). Graded response model. In W. J. Van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory.* Springer. https://doi.org/10.1007/978-1-4757-2691-6_5

Schulenberg, S. E., Schnetzer, L. W., & Buchanan, E. M. (2011). The purpose in life test-short form: Development and psychometric support. *Journal of Happiness Studies, 12*(5), 861-876. https://doi.org/10.1007/s10902-010-9231-9

Schulenberg, S., & Melton, A. (2010). A confirmatory factor-analytic evaluation of the purpose in life test: Preliminary psychometric support for a replicable two-factor model. *Journal of Happiness Studies, 11*(1), 95-111. https://doi.org/10.1007/s10902-008-9124-3

Sharpley, C., Murcell, N., Anderson, M., Bitsika, V., Fourie, P., & Agnew, L. (2021). An exploration of recent life stress, psychological resilience, purpose in life, and optimism as correlates of depression in social housing residents in rural Australia. *International Journal of Mental Health, 50*(3), 234-249. https://doi.org/10.1080/00207411.2021.1896079

Simkin, H., Matrangolo, G., & Azzollini, S. (2018). Argentine validation of the purpose in life test/Validación argentina del test de Propósito en la vida. Estudios de Psicología. *Estudios de Psicología, 39*(1), 104-126. https://doi.org/10.1080/02109395.2017.1407903

Spitzer, R. L., Kroenke, K., Williams, J. B., & Group Patient Health Questionnaire Primary Care Study (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA, 282*(18), 1737-1744. https://doi.org/10.1001/jama.282.18.1737

Steger, M. (2006). An illustration of issues in factor extraction and identification of dimensionality in psychological assessment data. *Journal of Personality Assessment, 86*(3), 263-272. https://doi.org/10.1207/s15327752jpa8603_03

Torres-Salazar, C., Moreta-Herrera, R., Ramos-Ramírez, M., & López-Castro, J. (2020). Sesgos cognitivo de optimismo y percepción de bienestar en una muestra de universitarios ecuatorianos. *Revista Colombiana de Psicología, 29*(1), 61-72. https://doi.org/10.15446/rcp.v29n1.75853

Upchurch, M. (2020). Gender bias in research. In N. A. Naples (Ed.), *Companion to Women's and Gender Studies* (pp. 139-154). Willey Ltd.

Urtasun, M., Daray, F. M., Teti, G. L., Coppolillo, F., Herlax, G., Saba, G., Rubinstein, A., Araya, R., & Irazola, V. (2019). Validation and calibration of the patient health questionnaire (PHQ-9) in Argentina. *BMC Psychiatry*, 19(1), 1-10. https://doi.org/10.1186/s12888-019-2262-9

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Measurement invariance. *Frontiers in Psychology, 6*, Article 1064. https://doi.org/10.3389/fpsyg.2015.01064

Weber, M., Furman, H., & Schulenberg, S. (2022). A Spanish Adaptation and Validation of the Purpose in Life Test–Short Form (PIL-SF). *{PSOCIAL} Journal of Research in Social Psychology, 8*(1), Article 5.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6), 913-934. https://doi.org/10.1177/0013164413495237

Xi, J., Lee, M., Carter, J., & Delgado, D. (2022). Gender differences in purpose in life: The mediation effect of altruism. *Journal of Humanistic Psychology, 62*(3), 352-376. https://doi.org/10.1177/0022167818777658