SAPIENZA
UNIVERSITÀ EDITRICE

E-ISSN 2724-2943
ISSN 2723-973X

Psychology Hub (2024)
XLI, 3, 43-54

Article info

Submitted: 09 March 2024
Accepted: 07 September 2024
DOI: 10.13133/2724-2943/18460

Psychometric properties of the 28-item General Health Scale (GHQ-28). Analysis from the Classical Test and Item Response Theories in Ecuadorian university students

Rodrigo Moreta-Herrera[1, 2*], Alberto Rodríguez-Lorenzana[3], Guido Mascialino[4], Fabián Castro-Ochoa[5], Viviana Narváez-Pillco[5], Tomás Caycho-Rodríguez[6], and Marlon Mayorga-Lascano[2]

[1] *Dean's Office of Research and Liaison, Universidad de Las Américas, Ecuador*

[2] *School of Psychology, Pontificia Universidad Católica del Ecuador*

[3] *Faculty of Psychology, Universidad Pública de Navarra, Spain*

[4] *School of Psychology and Education, Universidad de Las Américas, Ecuador*

[5] *Universidad Católica de Cuenca, Ecuador*

[6] *Universidad Científica del Sur, Perú*

Abstract

Background: *Monitoring mental health in the university population is essential to be able to meet the demands and needs of this segment. For this, it is essential to have properly calibrated instruments to adequately describe reality and generate adequate decision making.*

Aim: *Confirm bi-factor model and measurement invariance by gender of the 28-item General Health Questionnaire (GHQ- 28) in university students in Ecuador according to Classical Test Theory (CTT). In addition, analyze the discrimination parameters, item difficulty and global reliability according to the Item Response Theory (IRT).*

Participants and procedure: *A descriptive and instrumental design of the GHQ-28 from a bi-factor model was used with 476 participants, of whom 61.5% were women and 38.5% men, aged 17 to 47 years (M = 21.2; SD = 3.7), from 34 higher education centers in Ecuador.*

Results: *The factorial validity of the GHQ-28 was confirmed based on a bi-factor model. In addition, it showed measurement invariance across gender of the participants according to the CTT. The items of the scale present adequate discrimination and difficulty, and the global reliability of the measure is correct based on IRT.*

Conclusions: *The GHQ-28 is a valid test that can be applied to a university population in Ecuador. The scale is essentially unidimensional in its assessment of psychological distress and the contribution of the IRT broadens and complements the information on the known items of the scale previously collected by the CTT.*

Keywords: intra-gender, ostracism, queen bee, trait mindfulness, worker bee

*Corresponding author.
Rodrigo Moreta-Herrera
Dean's Office of Research and Liaison,
Universidad de Las Américas,
Quito, Ecuador,
E-mail: carlos.moreta.herrera@udla.edu.ec
(R. Moreta-Herrera)

Introduction

Mental health is a fundamental component of people's quality of life (Suryavanshi et al., 2020). It involves not only the absence of illness, but the expression of biopsychosocial well-being through stress management, functionality, introspection, and working in the community. It involves the search for both psychological/behavioural and social balance within a specific environment that allows you to perceive levels of well-being and satisfaction. (World Health Organization [WHO], 2001) Despite its significance, optimal mental health can be compromised under various circumstances (Walton et al., 2021). An example of this is the aftermath of the Covid-19 pandemic (Caycho-Rodríguez et al., 2021; Larzabal-Fernandez et., 2023; Zumba-Tello & Moreta-Herrera, 2022). This situation led to the proliferation of disorders such as anxiety and depression, as well as an increase in psychological distress (Mena-Freire et al., 2023; Sánchez-Vélez & Moreta-Herrera, 2022).

The university population is a high-risk group for mental health problems (Lorca et al., 2021; Ruíz-Olarte et al., 2023; Vallejo et al., 2007). This vulnerability could be due to the transition from adolescence to adulthood and academic pressure, among other reasons (Moreta-Herrera et al., 2023; Wang et al., 2020). However, assessing mental health conditions in university students is challenging, especially in Ecuador, due to the limited availability of validated and adapted instruments. In this regard, for the study of psychological symptoms associated with probable mental order problems, there are a number of specific measures such as the Generalized Anxiety Scale (GAD-7, Spitzer et al., 2006), the Symptom Severity Scale-Revised (EGS-R; Echeburúa et al., 2016) for post-traumatic stress disorder or the Patient Health Scale for depression screening (PHQ-9; Kroenke & Spiltzer, 2002), as well as global ones such as the 90 Symptom Scale Revised (SCL-90R; Derogatis, 1994). However, many of these scales have certain limitations, such as functionality, since they analyze very specific areas, and extension, which may limit the operational capacity of the global evaluation due to time. It is also important to point out that several of these measures have not yet undergone validation processes in the regional context that includes Ecuador, which affects the exercise of psychological assessment and diagnosis. For this reason, psychometric research is needed to resolve this situation.

Within this framework of limitations, one of the available instruments that may be broad enough to measure different aspects of mental health and at the same time operational in terms of the number of items is the General Health Questionnaire (GHQ; Goldberg, 1978), which was designed to assess perceptions of general health in adults. It does this by configuring its 28 items into four factors that analyze somatization, insomnia/anxiety, social dysfunction, and depression in a general way. It is a powerful tool for rapid population screening in situations of limited health care or primary prevention. This has made it one of the most widely used tools worldwide, especially in the adult population (Vallejo et al., 2014). However, despite being a widely used tool, its operability still requires further psychometric validation to ensure its correct functioning, especially in the Ecuadorian and Latin American context.

Validity and new findings on the internal structure of GHQ-28

As mentioned above, the GHQ-28 is a widely used instrument that has undergone several processes of item revision and reduction, originally from 60 items (GHQ; Goldberg, 1972) to reduced versions until reaching the 28-item version that is the basis of this study (Goldberg, 1978). From this, different translations and adaptations have been made in different contexts and cultural and national groups. As a reference, there are works on adults in Greece (Kokkinis et al., 2017), India (Kumaranayake & Srimathi, 2016), Spain (Lobo et al., 1986) and Norway (Hjelle et al., 2019). Although validation studies have also been conducted in university populations, such studies are less common (Lorca et al., 2021; Vallejo et al., 2007). The GHQ-28 has a multidimensional character with four factors: somatization, anxiety/insomnia, social dysfunction, and severe depression, and presents high internal consistency among scores. Although there is relative consensus on the factorial structure of the GHQ-28, it has not been explored in more complex fit models that preserve the internal components and integrate the items' role in the construct's global variance. Building on this idea, Moreta-Herrera et al. (2021) identified a new agglutinating type factor that provided a unidimensional shape to the measure through a bi-factor model. The authors found that the items contributed not only to the variance of their specific factors (SF), but also to the reliable variance of the measure through a general factor (GF) that adjusts to the university population of Ecuador. Although this study represents a significant advance and innovation in the factorial interpretation of the GHQ-28, providing a broader internal structure, the measure still lacks sufficient evidence for generalization as there are no confirmatory studies conducted inside or outside Ecuador.

Even so, Moreta-Herrera et al.'s (2021) findings open up new possibilities about the evaluative abilities of the measure, such as the existence of a GF that allows a global assessment of health perception, without representing the presence of dominant dimensions or factors (Rodriguez et al., 2016). This, in turn, allows for calculations and interpretations of total scores on the test. However, this finding also raises concerns about the usefulness of the SFs in comparison to the GF and their contribution to the reliable variance of the measure. It is unclear whether the SFs complementarily contribute to the GF or if they are solely dependent on the GF, making the measure essentially unidimensional.

Current limitations in the instrumental examination of the GHQ-28

The verification of this property is essential for multigroup research (*cross-gender* or *cross-cultural* studies) because in order to find differences ($p < .05$) between groups it is unavoidable that the measure retains its factorial configuration characteristics so that the probable existing differences come from changes in the particular characteristics of the groups (differences in variance) and not from differences in the factorial structure of the measure (Brown, 2015; Caycho-Rodríguez et al., 2022; Merlyn Sacoto et al., 2022).

The GHQ-28 has been researched extensively, with several studies examining its validity, including internal structure and

convergent validity with depression (Galindo et al., 2017), as well as its reliability. However, there are other properties based on Classical Test Theory (CTT) that have yet to be addressed. One such property is measurement equivalence (ME), which refers to the degree of invariance of the factor structure across different groups within a sample, such as gender, marital status, nationality, and culture (Asparouhov & Muthén, 2014; Meade et al., 2008). It is essential to verify this property for multigroup research, such as cross-gender or cross-cultural studies, because any differences found between groups must be attributable to differences in the characteristics of the groups rather than to differences in the factorial structure of the measure (Brown, 2015; Caycho-Rodríguez et al., 2022; Merlyn Sacoto et al., 2022).

Research shows that psychological symptoms and psychopathology differ between genders, with women reporting higher levels of somatization, anxiety/insomnia, social dysfunction, depression, and general distress, according to studies using the GHQ-28 (Canal-Rivero et al., 2022; Tsukamoto et al., 2021; Walton et al., 2021; Zumba-Tello & Moreta-Herrera, 2022). However, other studies show that men report higher levels of social dysfunction and depression than women (Moreta Herrera et al., 2021). These findings are inconclusive and may be affected by the sensitivity of the measure's structure to change, given that ME has not been verified accross genders for the GHQ-28. As a result, it is crucial to determine its invariance for cross-gender research in order to properly understand the phenomenon.

Another limitation of current research on the GHQ-28 is that, like other measures, evidence of its validity and reliability is primarily based on CTT. According to CTT, the validity of a measure is determined based on a latent common factor in a reference population (De Champlain, 2010). In this mathematical approach, the weight assigned to each item is similar because it focuses more on the analysis of measurement error, which is why the factor or factors that make up the items are explained globally without considering the variation specific to the test items. This limitation does not allow us to look at the level of individual contribution that each item has to capture the construct being evaluated. In this aspect, the CTT does not analyze the particularities of the items or the individual's ability or 'latent trait' to know if they respond correctly (Hambleton & Swaminathan, 2013). That is why, as a counterpart, the Item Response Theory (IRT) appears, which corresponds to a probabilistic model that explains the responses of respondents to the items based on the properties of the item and the latent trait (theta). The IRT analyzes specific parameters of each of the items such as discrimination (a), difficulty (b) and guessing (c) of the items. (Harvey & Hammer, 1999) to know the degree of contribution and precision that each of them has to identify the probability of interpretation of the latent variable that is analyzed. Unlike CTT, in IRT the analysis is much more detailed.

With regard to the IRT and the GHQ-28, psychometric studies have not yet been conducted, and their use could contribute to a better understanding of the performance of the items in detecting mental health problems and how they affect the precision and scoring of the latent variable. . The use of IRT can help complement the work carried out from a CTT framework, since they deal with theoretical models that are not exclusive but complementary.

Objectives of the study

Based on the above, the objectives of the study include: a) From the CTT, verify the bi-factor model of the GHQ-28 in university students and in turn confirm the equivalence of measurement based on gender; and b) From the IRT, analyze the performance of the scale items based on the discrimination and difficulty parameters to know their precision and contribution, as well as establish the reliability of the scale in a general way.

Method

Participants

The study was carried out with 476 participants. 61.5% of the sample correspond to women and the remaining 38.5% to men. Their ages range from 17 to 47 years (M = 21.2; SD = 3.7). Ethnically, 91.2% self-identify as half blood, while 3.4% as indigenous, 1.1% Afro-ecuadorian, 3.2% white and 1.6% without ethnic definition. 78% of the sample came from urban areas and 22.1% from rural areas. In addition, the participants were studying in 34 higher education centers (universities and technological institutes) in 10 provinces of Ecuador.

Participants were selected by means of non-probabilistic sampling based on inclusion criteria. The criteria were: a) being over 18 years of age; b) being a student at a higher education center; c) willingness to participate voluntarily; and d) signing a letter of consent for participation.

Instruments

General Health Questionnaire (GHQ-28; Goldberg, 1978) in the version adapted for university students in Ecuador (Moreta-Herrera et al., 2021). It consists of 28 items answered on a four-option Likert scale (from better than usual to much worse than usual) and divided into four dimensions: a) somatization; b) anxiety/insomnia; c) social dysfunction; and d) depression. Regarding the psychometric properties of the scale, the Ecuadorian version showed factorial validity through a bi-factor fit model, and also high internal consistency with $\omega_{(Somatization)}$ = .89 [.87 - .90], $\omega_{(Anxiety/insomnia)}$ = .957 [.951 - .963], $\omega_{(Social\ dysfunction)}$ = .90 [.88 - .91], $\omega_{(Depression)}$ = .958 [.952 - .964] for the SF and $\omega_{(Global)}$ = .96 [.95 - .97] for the GF.

Procedure

Data collection was done virtually due to the Covid-19 pandemic, during which students participated in classes remotely. For this purpose, an evaluation form was created in Google Forms through the link https://forms.gle/CAgtTbc8FaYWMh4C6,

which included informed consent, a sociodemographic form, and the study instruments. The informed consent stated the theme and objectives of the research, role of the participants, and information regarding anonymity and confidentiality. Data collection was carried out during the first semester of 2022; the approximate evaluation time was approximately 15 minutes. Once the evaluation was concluded, the data were cleaned and systematized; this was followed by statistical analyses, consideration of the hypotheses, and formulation of the respective conclusions. The research project from which this work derives adhered to the criteria and ethical norms for the development of research with human beings according to the Helsinki Convention. Furthermore, the project from which it is derived complies with the institutional regulations regarding the ethical management of research of the Pontificia Universidad Católica del Ecuador and was approved by the Institutional Review Commission.

Data analysis

Data analysis was carried out in three blocks. The first was the preliminary study of the items. The individual behavior of the items was investigated through measures of central tendency (arithmetic mean [*M*]), dispersion (standard deviation [*SD*]) and distribution (skewness [*g1*] and kurtosis [*g2*]). The univariate normality assumption was also verified, which is fulfilled when *g1* and *g2* are within the range ±1.5 (Ferrando & Anguiano-Carrasco, 2010). Likewise, multivariate normality is studied through Mardia's test (1970), which is fulfilled when the results in g_1 and g_2 do not present statistical significance ($p > .05$).

The second segment was the psychometric analysis of the GHQ-28 based on CTT, comprising Confirmatory Factor Analysis (CFA) testing four fit models (unidimensional, four correlated factors, hierarchical and bi-factor) and the multigroup CFA (CFA-MG) from the bi-factor model verified using Diagonally Weighted Least Squares (DWLS) estimation, given that multivariate normality is not present and the response scale of the items are ordinal in nature (Li, 2016). The factorial validity was verified according to the fit obtained in the indicators comprising absolute indices such as Chi-square ($\chi 2$), normed Chi-square ($\chi 2/df$) and Standardized Root Mean Residual (*SRMR*), as well as relative indicators such as Comparative Fit Index (*CFI*) and Tucker-Lewis Index (*TLI*), and non-centrality based indicator such as Root Mean Square Error of Approximation (*RMSEA*). A factor model is deemed adequate when the $\chi 2$ is not significant ($p > .05$) or the $\chi 2/df < 4$, the *CFI* and *TLI* > .9, the *SRMR* and *RMSEA* < .08, and for item saturations (λ) > to .40 (Brown, 2015; Byrne, 2008; Dominguez-Lara, 2018; Mueller & Hancock, 2018; Wolf et al., 2013). Since we worked on a bi-factor model, we used specific indices such as the hierarchical Omega (ω*H*) and the factor-specific Omegas (ω*s*), the Explained Variance Common (*EVC*), the Percentage of Uncontaminated Correlation (*PUC*) and the Hancock and Mueller coefficient (*H*). To validate this structure, the specific indices must show that the ω*H* and ECV are > .70, the *PUC* > .60, the *H* > .90 and the ω*s* > .30 (in this case to demonstrate that the reliable variance is attributed by the factors themselves or failing that by the GF) (Dominguez-Lara & Rodriguez, 2017; Moreta-Herrera et al., 2022; Reise et al., 2013; Smits et al., 2015; Rodriguez et al., 2016).

In this same segment, the measurement invariance by gender is performed with the AFC-MG with DWLS estimation. After the independent CFAs, the differences in the $\Delta\chi 2$ were analyzed and no significance ($p > .05$) was expected to establish similarity between groups. We then proceeded to place constraints on the model (metric, hard and hard) and measured the change (Δ) in the fit indicators $\chi 2$, *CFI* and *RMSEA* between each constraint time. The change found is not expected to be high (Asparouhov & Muthén, 2014; Brown, 2015). If at least strong invariance was verified, latent means were analyzed for construct differences between groups (Δ*K*). For this purpose, the intercepts of the men's group are set to 0, while the women's group is left unchecked; if no significance is found ($p < .05$), the groups are considered to be similar at the latent level.

The last block was devoted to IRT-based analyses of item discrimination (*a*) and difficulty (*b*). For this, the Graded Response Model (GRM) wase used, which is an extension of the *2-Parameter Logistic Model* (2-PLM) for ordered polytomous items (Hambleton et al., 2010; Samejima, 1997). For *a* we checkd the slope at which item responses change as a function of individuals' ability and allow us to discriminate individuals between those who do and do not possess ability to respond correctly to an item. A value of [a] > 1 is expected to consider moderate and acceptable discrimination of the items, while it is considered high when it is [a] > 1.5 or mul high when it is [a] > 2 (Baker & Kim, 2017). In the case of *b,* the manner in which an item behaves along the ability scale (four options) is analyzed. The difficulty is determined at the point of average probability (50%) of the individuals' ability to obtain a correct answer. As it is a four-choice scale, 3 estimates of difficulty are considered (1 for each threshold, b1 - b3). For a more detailed view of these parameters, the information curves of both the items as the Item Information Curve (IIC) and the Test Information Curve (TIC) were reviewed.

Statistical management was performed by means of the R programming language in version 4.2.2 (R Core Team, 2022) with the packages foreign, lavaan, ltm, MBESS, MVN for the AFC, AFC-MG and GRM; as well as the application Bifactor Index Calculator (Dueber, 2017) for the calculation of bi-factor estimates.

Results

Preliminary item analysis

Table 1 shows the behaviour of the GHQ-28 items, in general, as well as by gender. In all these cases, the average scores are homogeneous. Regarding the distribution measures, these remain equally uniform (except items 1 and 25 in the group of women) and are within the range ±1.5, so for these cases, the assumption of univariate normal distribution is met. This is not the case in the multivariate normality, since the values of the Mardia test show statistical significance ($p < 0.05$) in the data reported in general, as well as by the groups based on gender.

Tab. 1. Preliminary analysis of GHQ-28 items

Items	*General*				*Men*				*Women*			
	M	*SD*	*g1*	*g2*	*M*	*SD*	*g1*	*g2*	*M*	*SD*	*g1*	*g2*
Item 1	2.00	0.57	0.53	1.91	1.95	0.60	0.47	1.44	2.03	0.55	0.62	2.35
Item 2	2.13	0.91	0.45	-0.57	2.02	0.88	0.56	-0.40	2.20	0.92	0.38	-0.64
Item 3	2.20	0.92	0.38	-0.68	2.06	0.88	0.51	-0.44	2.28	0.94	0.29	-0.77
Item 4	1.96	0.86	0.58	-0.37	1.88	0.82	0.65	-0.14	2.01	0.88	0.52	-0.49
Item 5	2.18	0.91	0.39	-0.64	2.00	0.83	0.46	-0.42	2.30	0.95	0.30	-0.79
Item 6	1.95	0.95	0.65	-0.63	1.81	0.88	0.83	-0.17	2.03	0.99	0.53	-0.85
Item 7	1.68	0.84	1.01	0.11	1.63	0.82	1.07	0.23	1.70	0.85	0.98	0.06
Item 8	2.21	1.02	0.36	-1.00	2.17	1.03	0.47	-0.92	2.24	1.02	0.30	-1.04
Item 9	2.04	1.02	0.60	-0.80	2.10	1.04	0.54	-0.88	1.99	1.01	0.64	-0.74
Item 10	2.22	0.99	0.29	-0.99	2.07	0.95	0.44	-0.81	2.31	1.00	0.19	-1.05
Item 11	2.17	0.99	0.39	-0.91	2.02	0.93	0.60	-0.52	2.27	1.02	0.26	-1.06
Item 12	1.85	0.95	0.83	-0.40	1.77	0.87	0.93	0.02	1.90	1.00	0.75	-0.62
Item 13	2.23	1.01	0.37	-0.95	2.03	0.94	0.63	-0.48	2.36	1.03	0.21	-1.10
Item 14	2.07	1.01	0.54	-0.85	1.89	0.94	0.84	-0.22	2.18	1.04	0.37	-1.07
Item 15	2.14	0.80	0.69	0.30	2.04	0.82	0.75	0.36	2.20	0.79	0.69	0.33
Item 16	2.34	0.81	0.44	-0.24	2.25	0.78	0.54	0.13	2.40	0.83	0.37	-0.40
Item 17	2.07	0.71	0.71	0.95	1.99	0.71	0.74	1.08	2.12	0.71	0.71	0.94
Item 18	2.13	0.77	0.46	0.05	2.04	0.75	0.55	0.30	2.19	0.77	0.41	-0.04
Item 19	2.13	0.85	0.61	-0.05	2.07	0.81	0.76	0.47	2.17	0.87	0.52	-0.28
Item 20	2.01	0.84	0.63	-0.05	1.92	0.81	0.78	0.37	2.07	0.85	0.55	-0.21
Item 21	2.16	0.80	0.53	0.04	1.98	0.76	0.65	0.51	2.28	0.80	0.45	-0.12
Item 22	1.85	1.01	0.90	-0.42	1.80	0.96	0.97	-0.11	1.87	1.04	0.85	-0.58
Item 23	1.71	0.95	1.17	0.29	1.74	0.91	1.08	0.28	1.70	0.97	1.22	0.31
Item 24	1.67	0.94	1.22	0.35	1.70	0.95	1.17	0.26	1.65	0.93	1.26	0.43
Item 25	1.55	0.85	1.53	1.56	1.62	0.90	1.36	0.92	1.51	0.81	1.66	2.12
Item 26	1.81	0.93	0.94	-0.11	1.72	0.90	1.21	0.69	1.86	0.95	0.79	-0.45
Item 27	1.68	0.94	1.25	0.49	1.67	0.94	1.26	0.53	1.68	0.94	1.25	0.50
Item 28	1.61	0.90	1.40	1.00	1.63	0.92	1.36	0.85	1.60	0.89	1.44	1.14
Mardia			9469.5***	66.2***			7715.9***	34.7***			7743.2***	41.9***

Note: *** *p< .001; SD: standard deviation; g1: asymmetry; g2: kurtosis*

Tab. 2. Confirmatory Factor Analysis of GHQ-28

Models	*χ2*	*df*	*χ2/df*	*CFI*	*TLI*	*SRMR*	*RMSEA*	RMSEA 90% C.I.
Unidimensional	3256.2***	350	9.30	.904	.897	.112	.132	.128; .136
Hierarchical	1360.1***	349	3.90	.967	.964	.068	.078	.074; .082
Four factors	1376.7***	344	4.00	.966	.963	.057	.079	.075; .084
Bi-factor	803.3***	332	2.41	.996	.996	.051	.056	.051; .061

Note: *** *p < .001; χ2: Chi-square; df = degree free; χ2/df: normed Chi-square; CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; SRMR: Standardized Root Mean Residual; RMSEA: Root Mean Squared Error of Approximation*

Given the lack of multivariate normality in the item responses and the ordinal nature of the items, future factor analyses should be based on the use of robust estimates such as the WLSMV for better precision in reporting results.

Confirmatory Factor Analysis

Table 2 shows the CFA of the GHQ-28 considering four fit models (unidimensional, four correlated factors, hierarchical and bi-factor). As the results show, the model that presents the best fit of those tested is the one that corresponds to the bi-factor model of the GHQ-28 composed of a general factor and four specific factors (somatization, anxiety/insomnia, social dysfunction and depression).

Moreover, on the specific bi-factor indicators, the measure is essentially unidimensional with ω*H* = .903; *PUC* = .778; *CVE* = .72 and *H* = .969; and even the reliable variance of the SFs ω*s1* = .193; ω*s2* = .179; ω*s3* = .206; ω*s4* = .362 are more attributed to the GF than to the SFs themselves, so they are not independent.

Multigroup Confirmatory Factor Analysis by gender

Once this process was completed of factorial validity, the CFA-GM by gender was performed. Table 3 shows, in the first part, the independent CFAs (baselines by gender), which are adequate. In addition, when analyzing the change (Δχ2) between groups, no significant differences were found (Δχ2 = 135.66; p > .05), so they were considered similar. The second part comprised the nesting of the model considering restrictions in the saturations, intercepts and residuals. As can be observed, as the restrictions increase changes are presented in the fit indicators (χ2, CFI and RMSEA). These changes do not exceed the tolerance points at any of the restriction levels, so factorial invariance exists and at a strict level.

Tab. 3. Gender equivalence analysis of the GHQ-28

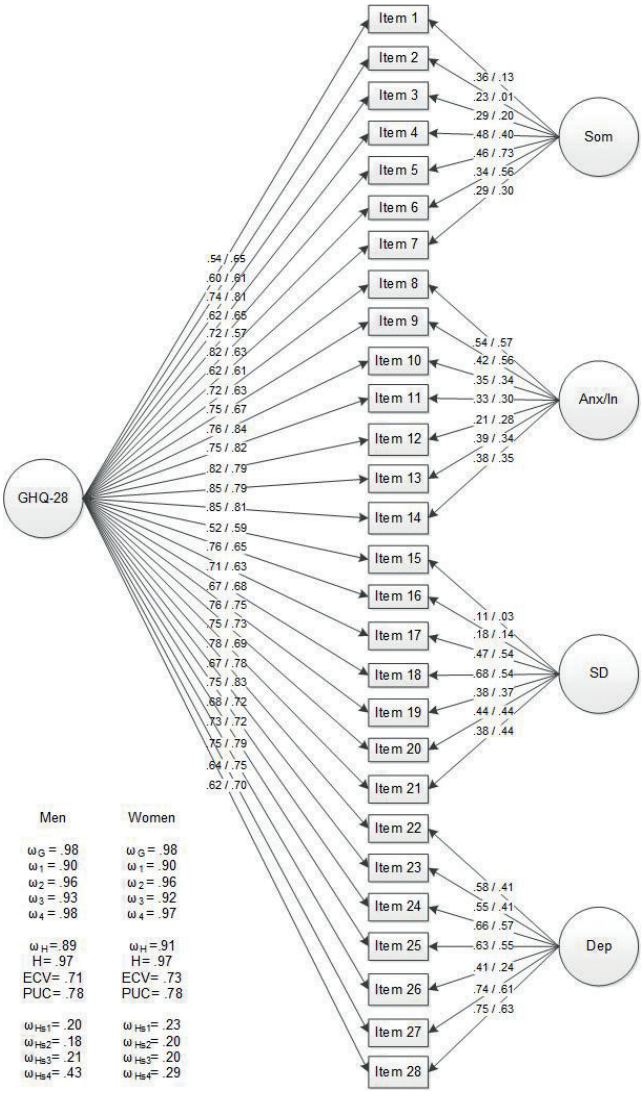
Restrictions	d.f.	*χ2*	*CFI*	*RMSEA*	*Δd.f.*	*Δχ2*	*ΔCFI*	*ΔRMSEA*
Baseline men	322	425.5	.998	.042		-	-	-
Baseline women	322	561.1	.997	.050		-	-	-
Unrestricted	644	986.6	.998	.047		-	-	-
Metrics	695	1338.4	.995	.063	51	351.8***	.002	.015
Fort	746	1113.6	.997	.046	51	224.8	.002	.017
Strict	751	1246.3	.997	.053	5	132.6***	.001	.007

Note: *** *p< .001; χ2: Chi-square; CFI: Comparative Fit Index; RMSEA: Root Mean Squared Error of Approximation; Δ: Delta (differentials)*

Figure 1 shows the internal structure of the GHQ-28 with item saturations by groups (left men and right women). The reported values of the items for both the GF and the SF respond better for the unidimensionality of the measure than for the multidimensionality. Complementarily, the specific bi-factor indicators show that, even when segmented by gender, the essentiality of unidimensionality of the measure is maintained, since the reliable variances for the specific SFs are predominantly attributed to the GF, so they are not independent.

Finally, the internal consistency (ω) shows that the reliability for the GF as well as for the SFs are high for both groups, so it is considered adequate for its evaluation in Ecuadorian university students.

Fig. 1. Internal structure of the GHQ-28 from a bi-factor fit model segmented by gender



Note: ω= omega coefficient; ωH: hierarchical omega; ωs: specific omega; H: Hancock and Mueller coefficient; CVE: common variance explained; PUC: percentage of uncontaminated correlation.

Latent mean difference analysis

Faced with the presence of measurement invariance by gender of strict type and containing strong invariance, it was possible to identify differences between groups in their latent means. Women present greater presence of symptoms than men in the SF of anxiety/insomnia $\Delta K = .213$; $p < .001$; $t = -2.26$ and small effect $\Delta K^* = .21$; and Depression with $\Delta K = .216$; $p < .001$; $t = .06$ and no effect $\Delta K^* = .01$. Whereas, in the GF there was more general psychological distress in men than in women $\Delta K = -.259$; $p < .001$; $t = -2.18$ and small effect $\Delta K^* = .20$. In the SFs of somatization and social dysfunction there were no differences ($p > .05$).

Graded response model

Once the assumption of unidimensionality and independence of the GHQ-28 has been verified by means of the AFC, the GRM is used. As shown in Table 4, in the case of [a], it is mentioned that this parameter is appropriate for all the items of the scale, because its values exceed [a] >1 and generally indicate adequate discrimination. Items 10, 11, 12, 13, 14, 22, 23, 26 and 27 present a very high discrimination as they are above [a] > 2. This is evidence that these Items and their response categories are powerful enough to allow better differentiation between those who present and those who do not present a significant symptom burden. On the other hand, concerning [b], the estimators of the thresholds from b1 to b3 for all items increase jointly and monotonically, making this parameter equally suitable (the level of difficulty increases as the next appears response level). At the extreme-left threshold, it is observed that the difficulty levels of the GHQ-28 are less than 2 SD away from the average (θ), indicating that the level of ability to respond to the items correctly is low; while on the far-right the reported values are close to 1 SD of θ and even items 1 and 7 exceed 3 SD. From this it follows that to achieve extreme scores the level of difficulty required is average and high (pointing out that to predict the latent trait the symptomatology must be widely manifest). In general, the average difficulty thresholds (b(avg)) between the items indicate that the GHQ-28 presents average difficulty, with values slightly above θ, which indicates that to increase the

probability of adequate detection of the latent trait medium or higher skill is required from the participant.

Figure 2a shows the informaton curves for the 28 items of the GHQ-28. In the IIC it is observed that the items that have the greatest relevance for evaluating the latent variable correspond to items 10, 14, 13 and 23. The shape of the curve indicates that these items have both greater discriminatory capacity, as well as similar skill levels to θ, making them particularly effective in counting more accurately to identify the latent trait, especially against medium or high skill levels.

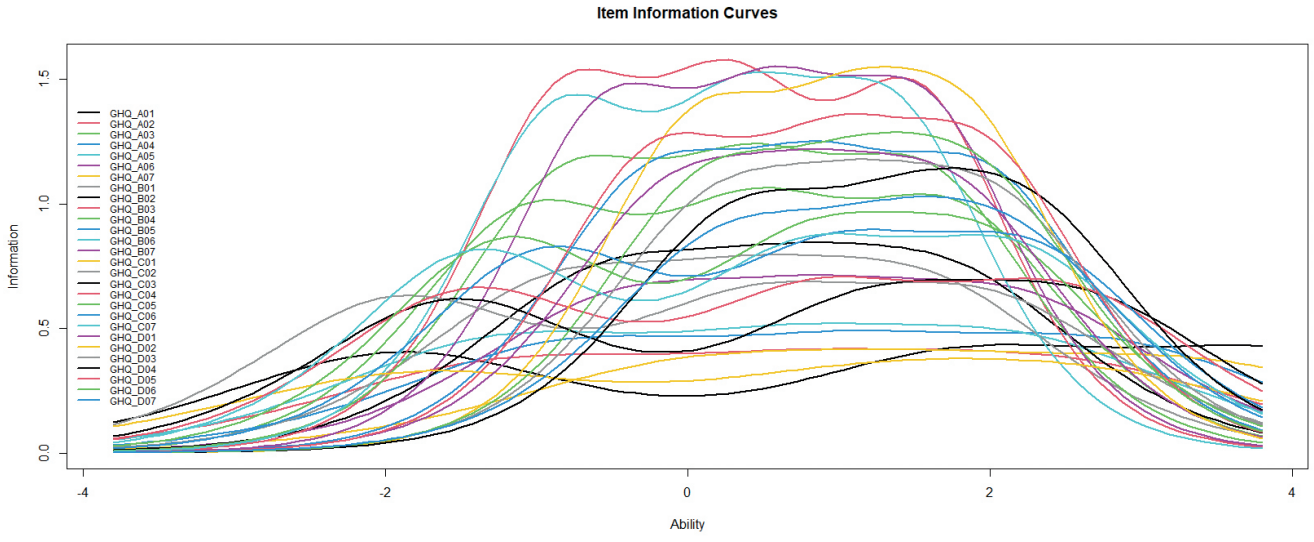
Finally, in the global TIC (figure 2b), the GHQ-28 curve has an adequate amount of information, which indicates acceptable levels of precision, especially in the ranges -1 to 2, which is when the most information is presented. Based on this, it is estimated that the scale is reliable and much more so at skill levels close to the average.

Tab. 4. GHQ-28 Graded Response Model

	a	*b1*	*b2*	*b3*	*b(prom)*
Item 1	1.266	-1.902	1.844	3.909	1.283
Item 2	1.195	-1.192	0.774	2.439	0.673
Item 3	1.961	-1.045	0.442	1.761	0.386
Item 4	1.309	-0.802	1.065	2.781	0.681
Item 5	1.338	-1.228	0.655	2.173	0.533
Item 6	1.557	-0.448	0.868	2.184	1.167
Item 7	1.185	0.019	1.581	3.408	1.669
Item 8	1.639	-0.871	0.425	1.564	0.372
Item 9	1.682	-0.544	0.666	1.728	0.617
Item 10	2.392	-0.818	0.282	1.515	0.326
Item 11	2.096	-0.792	0.413	1.578	0.792
Item 12	2.088	-0.238	0.851	1.970	0.861
Item 13	2.330	-0.858	0.367	1.359	0.289
Item 14	2.340	-0.539	0.555	1.562	0.526
Item 15	1.128	-1.769	1.218	2.589	0.679
Item 16	1.572	-1.908	0.457	1.949	0.166
Item 17	1.564	-1.543	1.235	2.590	0.761
Item 18	1.610	-1.434	0.869	2.482	0.639
Item 19	1.841	-1.204	0.825	1.961	0.527
Item 20	1.792	-0.936	0.963	2.269	0.765
Item 21	1.792	-1.397	0.787	2.175	0.521
Item 22	2.023	-0.111	0.835	1.767	0.830
Item 23	2.284	0.053	1.074	1.845	0.955
Item 24	1.982	0.184	1.122	2.035	1.113
Item 25	1.947	0.330	1.505	2.259	1.364
Item 26	2.176	-0.205	0.970	1.958	0.907
Item 27	2.069	0.128	1.157	1.934	1.073
Item 28	1.841	0.247	1.376	2.136	1.253

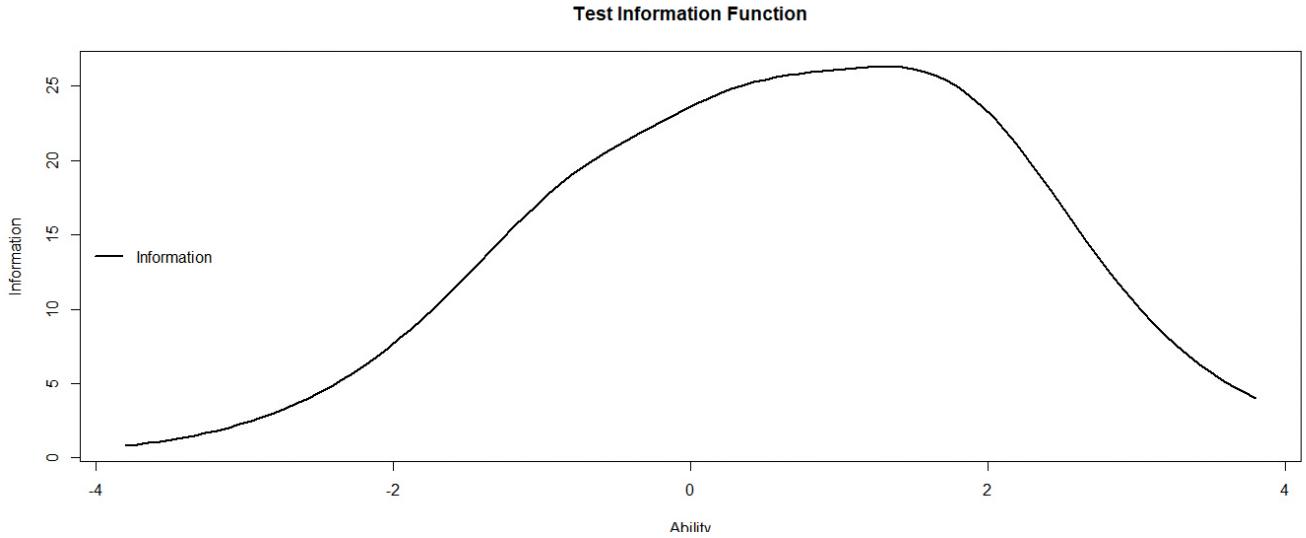
Note: *a: discrimination; b: difficulty*

Fig. 2a. Information curves of the items and of the GHQ-28 in general



Note: *IIC: Item Information Curves;*

Fig. 2b. Information curves of the items and of the GHQ-28 in general



Note: TIC: Test Information Curves

Discussion

The objectives of the study were to verify the bi-factor structure, the measurement invariance by gender, and the internal consistency reliability of the GHQ-28 according to the postulates of the CTT as well as the discrimination and difficulty parameters of the items according to IRT in a sample of university students from Ecuador.

Regarding its internal structure, the AFC shows that the GHQ-28 fits a bi-factor factorial model considering four SFs and, moreover, one GF in the sample under analysis. The fit values ($\chi 2$, $\chi 2/df$, CFI, TLI, SRMR and RMSEA) are within the tolerance parameters to estimate adequate (Brown, 2015; Byrne, 2008; Dominguez-Lara, 2018; Mueller & Hancock, 2018; Wolf et al., 2013). Moreover, the specific bi-factor indices (ωH, ωs, *CVE*, *PUC* and *H*), point out that the proposed model itself is essentially unidimensional, given that the reliable variance of the SFs is attributed more to the GF than to the SFs themselves ($\omega s < .30$), so these are not independent (Dominguez-Lara & Rodriguez, 2017; Moreta-Herrera et al., 2022; Reise et al., 2013; Smits et al., 2015; Rodriguez et al., 2016). The results regarding the bi-factor structure agree with the referential work of Moreta-Herrera et al. (2021) in Ecuadorian university students. Regarding the evidence supporting the unidimensionality of the measure, at the moment there are no other contributions with which it can be contrasted.

Regarding the measurement invariance by gender, the GHQ-28 is invariant; the values of the change (Δ) in the fit indices ($\chi 2$, *CFI*, *RMESEA*), the restrictions in the saturations, intercepts and residuals increase (Asparouhov & Muthén, 2014; Brown, 2015) and do not vary widely, so the measurement invariance is strict. The interpretation of the items in the factor structure by the participants was not significant regardless of group belonging. There are no similar reports with which to contrast the results shown. Thus, these results are a significant advance in the study of the measurement invariance of the scale. Likewise, in a complementary manner, by identifying strong invariance we investigated the differences in the latent means of the SF and the GF of the GHQ-28. Women show more presence of anxiety/insomnia and depression than men; while men show more presence of global psychological distress than women. In all these cases, the differences are significant ($p < .5$) and the effect sizes are small ($\Delta K^* \geq .20$). The presence of gender differences in aspects of mental health has already been reported in preliminary studies (Canal-Rivero et al., 2022; Tsukamoto et al., 2021; Walton et al., 2021; Zumba-Tello & Moreta-Herrera, 2022) although the contrast with these results can be complex due to the differences in the methods used, since using CFA to compare groups analyzes latent relationships considering the measurement errors of the items, while the use of bivariate tests such as Pearson's coefficient does not. In the future, specific studies of differences between groups are required.

Finally, on the discrimination and difficulty parameters, all items of the GHQ-28 show adequate discrimination (*a*) since they meet the criterion of adequacy. And in the case of *b*, the thresholds from 1 to 3 of all items also increase monotonically noting that the difficulty is adequate. In the IICs, the items with the highest relevance and accuracy are 10, 14, 13 and 23 which show the highest relevance and accuracy for assessing the latent variable. The TIC globally shows the scale to be reliable and accurate. The fact that these findings are the first in this field of study makes it impossible to contrast them with other studies. We consider the current work to represents an important advance and an innovation by broadening the assessment criteria of a measure such as the GHQ-28 using the postulates of IRT. The items of the GHQ-28 have different information power, although some show a higher level than others, with certain items being more relevant for the detection of mental health problems. In general, all items are located at the average of the highest positive levels of θ, so the scale is more useful for evaluating average and high levels of mental health problems (which is why its use is mainly recommended in primary care with some symptom burden present), but is not powerful enough to screen for low or very low levels of mental health problems (prodromal or subclinical conditions).

That is why it would be advisable to use other measures that are even more sensitive or perhaps with a positive and not a negative orientation of mental health assessment, such as the GHQ-28 for these cases.

These advances in the instrumental research of the GHQ-28 are explained by a broader mechanics of interpretation of the measure through a more complex and integrative (bi-factor) fit model. In practice, this gives the measure a global scale for the interpretation of the health perception construct in a direct way and allows it to obtain general scores for diagnostic categorizations and not only by dimensions. In addition, the study evidences measurement invariance based on gender, which allows it to delve into the differences by group when assessing mental health in the future. Finally, with the use of IRT-based criteria, the particular functioning of the items that compose it is described in more detail, including which of them contribute more to the interpretation of the latent variable, which may allow in an individual or group evaluation more sensitive indicators of detection of alterations in mental health and discomfort.

Limitations

One of the main limitations of the study is that it worked only with a university population, so the results shown cannot be generalized to other population groups. Thus, similar validation studies are recommended for future studies, but aimed at other groups such as adolescents in school, the general population or others. It is also important to mention that given the length of the study, other analyses based on the CTT were not considered, such as the validity of the relationship with other variables, discrimination or reliability analysis based on temporal stability.

Ethical Approval
The project from which this product is derived followed the ethical criteria for research established by the Helsinki Convention. The participants were informed and signed the respective consent to participate in the study. Furthermore, this work followed the policies and regulations regarding the development of institutional research and was approved by the institutional review commission of the Pontificia Universidad Católica del Ecuador Sede Ambato.

Data availability statement
The data associated with these results are available to those interested and can be requested from the corresponding author of the study. It is important to note that the use of the same will be used solely for research and academic purposes and under no pretext for commercial purposes

Funding/Financial Support
The author(s) reported there is no funding associated with the work featured in this article.

Author Contributions
Conceptualization: RMH, TCR; Review and analysis of instruments: RMH, TCR, MML; Methodological design: TCR, ARL, GM, RMH; Data production: FCO; VNP; Data analysis: RMH, TCR, ARL; Original manuscript writing: ARL, MML, FCO; VNP; Manuscript review and editing: RMH, TCR; ARL, GM; Funding: No available; Project Administration: RMH, MML.

Conflict of Interests
The authors report no conflict of interest

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

Baker, F., & Kim, S. (2017). *The basics of item response theory using R.* Springer.

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

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

Canal-Rivero, M., Montes-García, C., Garrido-Torres, N., Moreno-Mellado, A., Reguera-Pozuelo, P., Ruiz-Veguilla, M., & Crespo-Facorro, B. (2022). The impact of COVID-19 pandemic on the psychological well-being among health care workers: a 6-month cohort longitudinal survey study. *Revista de Psiquiatría y Salud Mental*, En prensa. https://doi.org/10.1016/j.rpsm.2022.08.001

Caycho-Rodríguez, T., Valencia, P., Vilca, L., Lee, S., Carbajal-León, C., Vivanco-Vidal, A., . . . Arias Gallegos, W. (2021). COVID-19 Bereavement in ten Latin American countries: Measurement invariance of the Pandemic Grief Scale and Its Relation to Suicidal Ideation. *OMEGA—Journal of Death and Dying*, 1-29. https://doi.org/10.1177/00302228211048566

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

De Champlain, A. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical education,, 44*(1), 109-117.

Derogatis, L. (1994). *SCL-90-R. Symptom Checklist-90-R. Administration, Scoring and.* National Computer System.

Dominguez-Lara, S. (2018). Propuesta de puntos de corte para cargas factoriales: una perspectiva de fiabilidad de constructo. *Enfermería Clínica, 28*(6), 401-402. https://doi.org/10.1016/j.enfcli.2018.06.002.

Dominguez-Lara, S., & Rodriguez, A. (2017). Índices estadísticos de modelos bifactor. *Interacciones, 3*(2), 59-65. https://doi.org/10.24016/2017.v3n2.51

Dueber, D. M. (2017). Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models. https://doi.org/10.13023/edp.tool.01

Echeburúa, E., Amor, P., Sarasua, B., Zubizarreta, I., Holgado-Tello, F., & Muñoz, J. (2016). Escala de Gravedad de Síntomas Revisada (EGS-R) del Trastorno de Estrés Postraumático según el DSM-5: propiedades psicométricas. *Terapia psicológica, 34*(2), 111-128. https://doi.org/10.4067/S0718-48082016000200004

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

Galindo, O., Meneses, A., Herrera, Á., Cabrera, P., Suchil, L., & L., R. (2017). Propiedades psicométricas del Cuestionario General de salud de Goldberg –GHQ-28– en cuidadores primarios informales de pacientes con cáncer. *Psicooncologia, 14*, 71-82. https://doi.org/10.5209/PSIC.55812

Goldberg, D. (1972). *The Detection of Psychiatric Illness by Questionnaire.* Oxford University Press.

Goldberg, D. (1978). *Manual of General Health Questionnaire.* Windsor: NFER Publishing.

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 advances. En H. o. models, M. Nering, & R. Ostini (Edits.).

Hambleton, R., & Swaminathan, H. (2013). *Item response theory: Principles and applications.* Springer Science & Business Media.

Harvey, R., & Hammer, A. (1999). Item response theory. *The Counseling Psychologist, 27*(3), 353-383.

Hjelle, E., Bragstad, L., Zucknick, M., Kirkevold, M., Thommessen, B., & Sveen, U. (2019). The General Health Questionnaire-28 (GHQ-28) as an outcome measurement in a randomized controlled trial in a Norwegian stroke population. *BMC psychology, 7*(1), 1-11. https://doi.org/10.1186/s40359-019-0293-0

Kokkinis, N., Galanaki, E., & Malikiosi-Loizos, M. (2017). Factor structure and internal consistency of the Greek version of the General Health Questionnaire–28 (GHQ-28). *Mental Health & Prevention, 7*, 21-27. https://doi.org/10.1016/j.mhp.2017.07.004

Kroenke, K., & Spitzer, R. (2002). The PHQ 9: A new depression diagnostic andseverity measure. *Psychiatric Annals, 32*(9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Kumaranayake, A., & Srimathi, N. (2016). Evaluation of internal consistency and factor structure of General Health Questionnaire (GHQ-28) on a South Indian sample. *International Journal of Social Science and Humanities Research, 4*(1), 281-291.

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*, on-line first. 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

Lobo, A., Pérez-Echeverría, M., & Artal, J. (1986). Validity of the scaled version of the General Health Questionnaire (GHQ-28) in a Spanish population. *Psychological medicine, 16*(1), 135-140. https//doi.org/10.1017/S0033291700002579.

Lorca, A., Lorca, M., Álvarez, J., García, J., & Armesilla, M. (2021). Psychometric Properties of the Patient Health Questionnaire-15 (PHQ-15) and the General Health Questionnaire-28 (GHQ-28). Validation in Spanish University Students During COVID-19 Outbreak. *International Journal of Psychology and Psychoanalysis, 049*, 6. https://doi.org/10.23937/2572-4037.1510049

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

Meade, A., Johnson, E., & Braddy, P. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *J. Applied Psych., 93*, 568–92.

Mena-Freire, M., Ruíz-Olarte, A., & Moreta-Herrera, R. (2023). Bienestar Subjetivo y Salud Mental en universitarios del Ecuador: análisis de relaciones latentes con SEM. *Actualidades en Psicología, 37*(135), 55-68. https://doi.org/10.15517/ap.v37i135.53576

Merlyn Sacoto, M., Díaz-Mosquera, E., & Moreta-Herrera, R. (2022). Psychometric properties of the ambivalent sexism Inventory (ASI) in ecuadorian adult population. *Actualidades en Psicología, 36*(132), 103-118. https://doi.org/10.15517/ap.v36i132.47618

Moreta Herrera, R., Zambrano Estrella, J., Sánchez Vélez, H., & Naranjo Vaca, S. (2021). Salud mental en universitarios del Ecuador: síntomas relevantes, diferencias por género y prevalencia de casos. *Pensamiento Psicológico, 19*(1), 1-26. https://doi.org/10.11144/Javerianacali.PPSI19.smue

Moreta-Herrera, R., Dominguez-Lara, S., Rodas, J., Sánchez-Guevara, S., Montes-De-Oca, C., Rojeab-Bravo, B., & Salinas, A. (2022). Examining psychometric properties and measurement invariance of the Emotion Regulation Questionnaire in an Ecuadorian Sample. *Psychological Thought, 15*(2), 57-74. https://doi.org/10.37708/psyct.v15i2.634

Moreta-Herrera, R., Dominguez-Lara, S., Vaca-Quintana, D., Zambrano-Estrella, J., Gavilanes-Gómez, D., Ruperti-Bonilla, E., & Bonilla, D. (2021). Psychometric properties of the General Health Questionnaire (GHQ-28) in Ecuadorian college students. *Psihologijske Teme, 30*(3), 573-590. https://doi.org/10.31820/pt.30.3.9

Moreta-Herrera, R., Núñez-Núñez, M., Lascano-Arias, G., Mascialino, G., & Rodríguez-Lorenza, A. (2024). Psychometric properties of the Escala de Gravedad de Síntomas Revisada del Trastorno de Estrés Postraumático (EGS-R) in adults in Ecuador after the Covid-19 pandemic. *European Journal of Trauma & Dissociation, 8*(2), 100400. https://doi.org/10.1016/j.ejtd.2024.100400

Moreta-Herrera, R., Zumba-Tello, D., de Frutos-Lucas, J., Llerena-Freire, S., Salinas-Palma, A., & Trucharte-Martínez, A. (2023). The role of negative affects as mediators in the relationship between stress and mental health in Ecuadorian adolescents. *Health Psychology Report, 11*(3), 241-251. https://doi.org/https://doi.org/10.5114/hpr/163484

Mueller, R. O., & Hancock, G. R. (2018). Structural equation modeling. In The reviewer's guide to quantitative methods in

the social sciences. En G. Hancock, L. Stapleton, & R. Mueller, *The Reviewer's Guide to Quantitative Methods in the Social Sciences* (págs. 445-456). Routledge.

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

Reise, S., Scheines, R., Widaman, K., & Haviland, M. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5-26. https://doi.org10.1177/00223891.2015.1089249.

Rodriguez, A., Reise, S., & Haviland, M. (2016). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137-150. https://doi.org/10.1037/met0000045.

Ruíz-Olarte, A., Mena-Freire, M., & Moreta-Herrera, R. (2023). Esperanza y salud mental en universitarios del Ecuador: Análisis de regresión estructural. *Psicoperspectivas. Individuo y Sociedad, 22*(3), 1-14. https://doi.org/10.5027/psicoperspectivas-Vol22-Issue3-fulltext-2906

Samejima, F. (1997). Graded response model. En H. o. theory, W. Van der Linden, & R. Hambleton (Edits.). Springer.

Sánchez-Vélez, H., & Moreta-Herrera, R. (2022). Miedo y Ansiedad al Covid-19, estrés y percepción de salud. Modelo de predicción en pacientes hospitalarios de atención primaria del Ecuador. *Anales de Psicología, 38*(3), 439–447. https://doi.org/10.6018/analesps.489761

Smits, I., Timmerman, M., Barelds, D., & Meijer, R. (2015). The Dutch symptom checklist-90-revised: is the use of the subscales justified? *European Journal of Psychological Assessment, 31*(4), 263-271. https://doi.org/10.1027/1015-5759/a000233.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097.

Suryavanshi, N., Kadam, A., Dhumal, G., Nimkar, S., Mave, V., Gupta, A., . . . Gupte, N. (2020). Mental health and quality of life among healthcare professionals during the COVID-19 pandemic in India. *Brain and behavior, 10*(11), e01837. https://doi.org/10.1002/brb3.1837

Tsukamoto, R., Kataoka, Y., Mino, K., Ishibashi, N., Shibata, M., Matsuo, H., & Fujiwara, H. (2021). Gender differences in anxiety among COVID-19 inpatients under isolation: a questionnaire survey during the first and second waves of the COVID-19 pandemic in Japan. *Frontiers in public health, 9*, 708965. https://doi.org/10.3389/fpubh.2021.708965

Vallejo, M. A., Rivera, J., Esteve-Vives, J., & Rodríguez-Muñoz, M. (2014). El cuestionario general de salud (GHQ-28) en pacientes con fibromialgia: propiedades psicométricas y adecuación. *Clínica y Salud, 25*(2), 105-110. https://dx.doi.org/10.1016/j.clysa.2014.06.005.

Vallejo, M., Jordán, C., Díaz, M., M.I., C., & J., O. (2007). Psychological Assessment via the Internet: A Reliability and Validity Study of Online (vs Paper-and-Pencil) Versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). *J Med Internet Res, 9*(1). https://doi.org/10.2196/jmir.9.1.e2

Walton, C., Rice, S., Gao, C., Butterworth, M., Clements, M., & Purcell, R. (2021). Gender differences in mental health symptoms and risk factors in Australian elite athletes. *BMJ open sport & exercise medicine, 7*(1), e000984. https://doi.org/10.1136/bmjsem-2020-000984

Wang, X., Hegde, S., Son, C., Keller, B., Smith, A., & Sasangohar, F. (2020). Investigating mental health of US college students during the COVID-19 pandemic: Cross-sectional survey study. *Journal of medical Internet research, 22*(9), e22817. https://doi.org/10.2196/22817

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.

World Health Organization [WHO]. (2001). *Fortaleciendo la prevención de salud mental.* Ginebra: World Organization Health.

Zumba-Tello, D., & Moreta-Herrera, R. (2022). Afectividad, Regulación Emocional, Estrés y Salud Mental en adolescentes del Ecuador en tiempos de pandemia. *Revista de Psicología de la Salud UHM, 10*(1), 117-129. https://doi.org/10.21134/pssa.v10i1.801