Trends in Psychology (2022) 30:111–128
https://doi.org/10.1007/s43076-021-00107-w

ORIGINAL ARTICLE

Spanish Version of the Revised Mental Health Inventory-5 (R-MHI-5): New Psychometric Evidence from the Classical Test Theory (CTT) and the Item Response Theory Perspective (IRT)

Lindsey W. Vilca[1] · Blanca V. Chávez[1] · Yoselin Shara Fernández[1] · Tomás Caycho-Rodríguez[2]

Accepted: 21 September 2021 / Published online: 11 October 2021

Abstract

The study's objective was to evaluate the psychometric properties of the revised version of the MHI-5, for which a sample of 1002 university students of both sexes (41.4% men and 58.6% women) between the ages of 17 to 35 years (M = 21.4; SD = 3.4) was collected. Along with the R-MHI-5, other instruments were applied to measure anxiety and depression. Regarding the results, it was evidenced that the model with two related factors presents better fit indices (CFI = .99; TLI = .99; RMSEA = .071) compared to a one-dimensional model (CFI = .74; TLI = .48; RMSEA = .422). Also, it was found that the factorial structure of the MHI-5 did not show evidence of being strictly invariant for the group of men and women. However, it did show evidence of partial invariance for the group of adolescents and adults. Furthermore, the latent relationships model showed that psychological well-being is negatively related to anxiety (− .25) and depression (− .37), and psychological distress is positively related to anxiety (.85) and depression (.87). From the IRT perspective, all items present adequate discrimination indices, with item 4 being the most accurate item to assess psychological well-being. Regarding psychological distress, items 3 and 5 are the most accurate to assess this dimension. It is concluded that the scale that the R-MHI-5 is an instrument with robust psychometric evidence from the perspective of CTT and IRT.

Keywords Mental health · IRT model · CFA model · Factorial invariance · Mental Health Inventory-5

✉ Lindsey W. Vilca
lwquiro@gmail.com

[1] Departamento de Psicología, Universidad Peruana Unión, Carretera Central Km. 19, Ñaña, Lima, Peru

[2] Universidad Privada del Norte, Lima, Peru

Introduction

The COVID-19 pandemic has an important impact on people's mental health worldwide, generating implications for public health, the economy, and social dynamics in general (Zolotov et al., 2020). In this context, university students are considered a vulnerable population to suffer from mental health problems (Husky et al., 2020) since they are generally in a transition stage in their academic, professional, and personal life (Acharya et al., 2018). Numerous studies examining mental health in college students during the COVID-19 pandemic found that it damages their mental health, as there is increased stress, psychological distress, symptoms of anxiety and depression, and difficulties concentrating at work academic (Cao et al., 2020; Kecojevic et al., 2020; Jungmin Lee et al., 2021; Rogowska, et al., 2020a, b; Savitsky et al., 2020). Likewise, increases in anxiety and depression symptoms in college students are associated with a higher perceived risk of the disease (Feng et al., 2020). To this must be added that the closure of universities, the change to virtual learning platforms, difficulties in technological access, economic problems, social isolation, fear of infection, and the death of family members due to COVID-19 have exacerbated the risk of mental health problems in this group significantly (Kecojevic et al., 2020; Sahu, 2020).

Faced with this, adequate measurement of mental health is necessary to prevent and implement intervention programs in university students (Visser & Law-van Wyk, 2021). According to the dual-factor model of mental health (Greenspoon & Saklofske, 2001), it is essential to use mental health instruments that measure both positive and negative emotions. An instrument developed under this dual perspective is the Mental Health Inventory (MHI, Veit & Ware, 1983), made up of 38 items that assess psychological well-being and distress in the general population. Subsequently, a short, five-item version of the MHI was developed, called MHI-5 (Berwick et al., 1991). The MHI-5 is as effective as the extended version (Rivera-Riquelme et al., 2019) and other instruments such as the General Health Questionnaire (Marques et al., 2011) and is even a better measure than the Hopkins Symptom Checklist (Strand et al., 2003). Also, due to its brevity, ease of response, as well as its evidence of validity and reliability, the MHI-5 has been used in different groups and cultures such as Portuguese teenagers (Marques et al., 2011), the general population of Brazil (Damásio et al., 2014), Australia (Milner et al., 2020) and Finland (Elovanio et al., 2020), American women with and without risk of atrial fibrillation (Whang et al., 2012), patients with chronic heart failure (Mo et al., 2020), university students (Almeneessier et al., 2015), people with spinal cord injury (Verwer et al., 2016), children and Spanish adolescents (Rivera-Riquelme et al., 2019), among others.

The MHI-5 has been used as a general measure of mental health problems and used in surveys of general health and quality of life in the non-psychiatric population (Rivera-Riquelme et al., 2019) since it assesses both psychological well-being like psychological distress. Furthermore, the MHI-5 has shown high sensitivity for detecting depressive, anxiety, or panic disorders in the general

population and primary care patients (Means-Christensen et al., 2005; Rumpf et al., 2001; Thorsen et al., 2013).

Few studies aimed to evaluate psychometric properties. A study with Portuguese adolescents reported that the MHI-5 presents a single factor that explained 59.88% of the total variance, adequate reliability (α = 0.82), item-test correlations that vary from 0.78 to 0.81, and evidence of convergent validity with other measures of hopeful thinking, life satisfaction, and self-esteem (Marques et al., 2011). Another research carried out in the general population of Brazil supported the presence of a single factor, adequate reliability (Cronbach alpha = 0.86; composite reliability = 0.82), evidence of convergent and discriminant validity with subjective happiness, satisfaction with life, and general health (Damásio et al., 2014). A more recent study with Finland's general population also indicated that the MHI-5 has good psychometric properties, with good reliability (α = 0.89) and a unidimensional factorial structure. Furthermore, all the items showed adequate discrimination indices, and increasing difficulty as the symptoms became more severe (Elovanio et al., 2020).

In Spain, the MHI-5 has been validated in children and adolescents between 10 and 15 years old (Rivera-Riquelme et al., 2019). Unlike previous studies, in the validation in Spanish, a two-factor structure was obtained that explains 69.2% of the total variance (factor 1 = psychological distress; factor 2 = psychological well-being), adequate reliability for the total scale (α = 0.71) and the psychological distress (α = 0.71) and psychological well-being (α = 0.70) subscales. In addition, the scale showed a significant relationship with symptoms of anxiety and depression. Furthermore, this Spanish version of the MHI-5 (R-MHI-5) presents a simplified response format of four alternatives (never, sometimes, several times, and always) and not six as originally proposed. The reduction of response alternatives is justified due to the lower cognitive demand necessary to complete the inventory and the evidence that there are no differences in the psychometric properties between measures that use 4, 5, or 6 alternatives (Jihyun Lee & Paek, 2014). In Peru, the MHI-5 was recently studied in a small sample of 75 students from a private secondary educational institution (Merino-Soto et al., 2019), whose results indicated that a two-factor model presents a better fit and has adequate evidence of reliability due to internal consistency (alpha = 0.70). However, the smallest number of participants makes it difficult to generalize the results. Also, the version with six answer options was used and not the one with four options.

As seen in the literature review, only the study by Rivera-Riquelme et al. (2019) evaluates the scale's psychometric properties in university students. Therefore, there is little evidence of the internal structure and other psychometric properties of the scale in university students that allow an adequate evaluation of mental health in the context of the COVID-19 pandemic.

Most of the previous psychometric studies of the MHI-5 have used TCT models. Only one previous study by Elovanio et al. (2020) evaluated the psychometric properties of the MHI-5 based on SEM and TRI models. However, none research was performed with the revised version of 4 response alternatives (R-MHI-5). Using both procedures will allow more robust results since the participants' characteristics do not influence the psychometric findings derived from TRI models, while the evidence produced by TCT models is (Lin et al., 2020). The TRI approach will also

allow estimating the difficulties of the items, the reliability of the people and items, and the standard errors, providing more stable results (Magno, 2009). Thus, using both approaches (TCT and TRI) will allow us to corroborate the previous findings and provide a better perspective of the psychometric properties of R-MHI-5. On the other hand, evaluating the MI will demonstrate that the R-MHI-5 measures the same construct in the same way for different groups (Vandenberg & Lance, 2000). In the present study, the MI evidence would indicate that men and women would have the same conceptualization of the latent variable and would have the same expected score on the R-MHI-5. It would also indicate that the relationships between the observable variables (items of the R-MHI-5) and the latent variable are independent of belonging to one group or another (Lubke et al., 2003). Finally, having evidence of MI is a prerequisite for comparing the measured variable between different groups (Caycho, 2017).

In this sense, this study aimed to evaluate the psychometric properties of the revised version of the MHI-5 (Rivera-Riquelme et al., 2019) in Peruvian university students, using the Classical Test Theory (TCT) and the Item Response Theory (TRI). Specifically, the validity evidence based on the construct, validity evidence based on the relationship with other variables, reliability, the discrimination and difficulty parameters, and the measurement invariance (MI) according to sex and age were evaluated.

Method

Participants

Non-probabilistic sampling was used to collect the data, using the following inclusion criteria: (a) informed consent of the participants, (b) age not older than 40 years, (c) ability to read and write in Spanish, and (d) being enrolled in a university program. A sample of 1002 university students of both sexes (41.4% men and 58.6% women) between 17 to 35 years (M = 21.4; SD = 3.4) was collected. Undergraduate students are from Peru, and the majority came from the highlands (47.5%), 26.2% came from the coast, and an equal percentage came from the jungle. Soper's (2020) online calculator was used, taking into account the following criteria: five observed variables, two latent variables, anticipated effect size of 0.30 (minimum lambda value for factorial models), desired probability of 0.05, and a power level statistic of 0.95. The minimum size required was 288 cases. Therefore, the present study collected a sample of participants that far exceeds the minimum required.

Instruments

Mental Health Inventory-5 (R-MHI-5)

Developed by Berwick et al. (1991) and adapted into Spanish by Rivera-Riquelme et al. (2019). The R-MHI-5 is made up of five items that assess the presence of

psychological well-being (items 2 and 4) and psychological distress (inverse items 1, 3, and 5). Furthermore, the Spanish version has four response categories ranging from "never" (0) to "always" (3), where a higher score indicates a better state of mental health.

Generalized Anxiety Disorder Scale (GAD-7)

Developed by Spitzer et al. (2006) and adapted to Spanish by García-Campayo et al. (2010), this scale is made up of seven items that have four response categories ranging from "never" (0) to "almost every day" (3), where a higher score indicates a greater presence of the disorder. In the present study, the one-dimensional model presented adequate indices of reliability ($\alpha = 0.93$; $\omega = 0.89$) and validity based on internal structure ($\chi^2 = 174.73$; df = 14; $p < 0.001$; CFI = 0.99; TLI = 0.99; RMSEA = 0.071; SRMR = 0.033).

Patient Health Questionnaire (PHQ-9)

Developed by Spitzer et al. (1999) and adapted into Spanish by Zhong et al. (2014), this questionnaire consists of nine items that have four categories ranging from "not at all" (0) to "almost every day" (3). In the present study, the one-dimensional model presented adequate indices of reliability ($\alpha = 0.92$; $\omega = 0.90$) and validity based on internal structure ($\chi^2 = 132.39$; df = 27; $p < 0.001$; CFI = 0.99; TLI = 0.98; RMSEA = 0.062; SRMR = 0.032).

Procedure

The study obtained the approval of the ethics committee of the Center for Research and Innovation in Health of the Universidad Peruana Unión (N° 00,131–2020), and the standards of the Helsinki Declaration were met (World Medical Association, 2013). The data was collected through a virtual form, using the Google Forms digital platform. In the first part of the virtual form, the study's objectives were explained, the time required to complete the form, and the informed consent was presented. The confidentiality of the information was ensured, and the possibility that the participants could withdraw at any time. Only participants who gave their informed consent could complete the following sections of the form.

Data Analysis

A confirmatory factor analysis (CFA) was carried out in this study using the weighted least squares with mean and variance adjusted (WLSMV) estimator since the items are at the ordinal level (Brown, 2015). The chi-square test (χ^2), the RMSEA index, and the SRMR index were used to evaluate the fit of the model, in which case values less than 0.05 indicate a good fit and between 0.05 and 0.08 is considered acceptable (Kline, 2015). Also, the CFI and TLI index were used; for these cases, values greater than 0.95 indicate a good fit and higher than 0.90, an acceptable fit (Schumacker & Lomax,

2015). The Cronbach's alpha coefficient (Cronbach, 1951) and the omega coefficient (McDonald, 1999) were used to evaluate the internal consistency of the scale, where a value of >0.80 is adequate (Raykov & Hancock, 2005). The internal consistency of the scale with the composite reliability index was also used. Values greater than 0.70 are generally considered acceptable (Viladrich et al., 2017).

A sequence of hierarchical models of variance was proposed, which were increasingly restrictive to evaluate the scale's invariance according to sex and age. First, the configural invariance (reference model) was evaluated, followed by the metric invariance (equality of factor loads), scalar invariance (equality of factor loads and intercepts), and finally, the strict invariance (equality of factor loads, intercepts, and residuals). First, a formal statistical test was used in the study to compare the sequence of models, for which the chi-square difference ($\Delta\chi^2$) was used where non-significant values ($p>0.05$) suggest invariance between the groups. Second, a modeling strategy was employed, using differences in the CFI (ΔCFI) where values less than <0.010 evidence model invariance between groups between the groups (Chen, 2007). Additionally, the RMSEA (ΔRMSEA) was used, where differences less than <0.015 show the model invariance between the groups (Chen, 2007).

For the Item Response Theory (IRT), a graduated response model (GRM, Samejima, 1997) was used, specifically an extension of the 2-parameter logistic model (2-PLM) for ordered polytomous items (Hambleton et al., 2010). For each item, two parameters were estimated: discrimination (a) and difficulty (b). The discrimination parameter determines the slope at which the responses to the items change as a function of the latent trait level, whereas the difficulty parameters of the item determine how much of the latent trait the item requires to be answered. Since scales have four response categories, there are three difficulty estimates, one per threshold. The estimates for these three thresholds indicate the latent variable's level at which an individual has a 50% chance of obtaining a score equal to or greater than a particular response category. The item information curves (IIC) and test information curves (TIC) were also calculated.

Regarding the validity of the MHI-5 relative to other variables, a structural equation model was proposed. In this model, the degree of psychological well-being and psychological distress is related to anxiety and depression. The WLSMV estimator was used to estimate the model, and the same adjustment indicators performed in the confirmatory factor analysis were taken into account.

All statistical analyzes were performed using the "lavaan" package (Rosseel, 2012) for the CFA, the "semTools" package (Jorgensen et al., 2018) for factorial invariance, and the "ltm" package for the GRM (Rizopoulos, 2006). In all cases, the RStudio environment (RStudio Team, 2018) was used for R (R Core Team, 2019).

Results

Descriptive Analysis

Table 1 shows that item 4 (During the last month, how often have you felt happy?) presents the highest average score in the total sample ($M=1.61$) and the different

groups of men (M = 1.63), women (M = 1.59), adolescents (M = 1.59), and adults (M = 1.62). It is also appreciated that item 5 (During the last month, how often have you felt so sad that nothing could cheer you up?) presents the lowest average score in the total sample (M = 0.72) and the different groups of males (M = 0.68), females (M = 0.75), adolescents (M = 0.75), and adults (M = 0.70). Furthermore, it is appreciated that the items present adequate asymmetry and kurtosis indexes (± 1.5) in the total sample and all the specific groups.

Validity Based on the Internal Structure

Table 2 shows that in the total sample, a one-dimensional model does not show adequate fit indices ($\chi^2 = 894.34$; df = 5; CFI = 0.74; TLI = 0.48; RMSEA = 0.422 [IC90% 0.398–0.445]). Similarly, a unidimensional model with reversing negative items does not present adequate fit indices ($\chi^2 = 894.34$; df = 5; CFI = 0.74; TLI = 0.48; RMSEA = 0.422 [IC90% 0.398–0.445]). In contrast, a model with two related factors presents better fit indices ($\chi^2 = 24.03$; df = 4; < 0.001; CFI = 0.99; TLI = 0.99; RMSEA = 0.071 [IC90% 0.045–0.099]), where the relationship between both factors is acceptable (− 0.30).

Also, it can be seen that the model of two related factors shows adequate adjustment indices in the specific groups: men (CFI = 0.99; TLI = 0.99; RMSEA = 0.051), women (CFI = 0.99; TLI = 0.98; RMSEA = 0.086), adolescents (CFI = 0.99; TLI = 0.98; RMSEA = 0.082), and adults (CFI = 0.99; TLI = 0.99; RMSEA = 0.055). Furthermore, it can be seen that in the total sample and in the specific groups, the factorial weight of the latent variable with each of its items is high and significant (see Table 3).

Factorial Invariance According to Sex and Age

Table 2 shows that the factorial structure of the MHI-5 did not show evidence of being strictly invariant for the group of men and women in the sequence of invariance models proposed: metric invariance ($\Delta\chi^2 = 9.56$, $p = 0.022$; ΔCFI = − 0.016), scalar ($\Delta\chi^2 = 18.58$, $p < 0.001$; ΔCFI = − 0.036), and strict ($\Delta\chi^2 = 11.16$, $p = 0.048$; ΔCFI = − 0.008). Regarding the group of adolescents and adults, the MHI-5 in the sequence of metric invariance models ($\Delta\chi^2 = 5.34$, $p = 0.148$; ΔCFI = − 0.005) and scalar ($\Delta\chi^2 = 3.92$, $p = 0.269$; ΔCFI = − 0.002) showed evidence of factorial invariance. However, it did not show evidence of strict invariance ($\Delta\chi^2 = 15.86$, $p = 0.007$; ΔCFI = − 0.024).

Scale Reliability

It can be seen in the lower part of Table 3, that in the total sample, the psychological well-being dimension (ω = 0.75) and psychological distress (ω = 0.79) present adequate reliability indices. Similarly, it occurs in specific groups: men (ω = 0.82 and ω = 0.79), women (ω = 0.71 and ω = 0.78), adolescents (ω = 0.70 and ω = 0.78), and adults (ω = 0.78 and ω = 0.81). An adequate level of composite reliability is also

Table 1 Descriptive analysis of the items in the entire simple and in specific groups

Items	Total sample ($n=1002$)				Men ($n=415$)				Women ($n=587$)				Adolescence ($n=499$)				Early adulthood ($n=503$)			
	M	SD	g1	g2	M	SD	g1	g2	M	SD	g1	g2	M	SD	g1	g2	M	SD	g1	g2
M1	.93	.68	.43	.32	.87	.69	.61	.64	.97	.66	.31	.17	.97	.69	.47	.43	.89	.66	.37	.16
M2	1.46	.78	.04	.32	1.57	.83	−.17	−.51	1.39	.74	.17	−.22	1.43	.73	.10	−.25	1.50	.83	−.03	−.54
M3	1.04	.72	.33	−.08	.93	.72	.38	−.17	1.11	.71	.31	.02	1.09	.69	.27	−.08	.99	.74	.41	−.09
M4	1.61	.79	−.08	−.43	1.63	.84	−.26	−.46	1.59	.76	.59	−.41	1.59	.78	.03	−.45	1.62	.81	−.19	−.41
M5	.72	.77	.83	.05	.68	.74	.84	.16	.75	.81	.82	−.04	.75	.79	.83	.11	.70	.77	.84	.00

M, mean; *SD*, standard deviation; *g1*, skewness; *g2*, kurtosis

Table 2 Fit indices of the related two-dimensional model and invariance models according to sex and age

Models	χ^2	df	*p*	SRMR	TLI	CFI	RMSEA	$\Delta\chi^2$	Δdf	*p*	ΔCFI	ΔRMSEA
Total sample												
Unidimensional model	894.34	5	< .001	.191	.48	.74	.422	–	–	–	–	–
Unidimensional model corrected[a]	894.34	5	< .001	.191	.48	.74	.422					
Two-related factor model	24.03	4	< .001	.023	.99	.99	.071	–	–	–	–	–
By sex												
Men	8.24	4	.083	.021	.99	.998	.051	–	–	–	–	–
Women	21.22	4	< .001	.030	.97	.990	.086	–	–	–	–	–
Configural	16.16	8	.040	.017	.96	.986	.045	–	–	–	–	–
Metric	29.12	11	.002	.026	.94	.970	.057	9.56	3	.022	− .016	.012
Scalar	53.72	14	< .001	.033	.91	.934	.075	18.58	3	< .001	− .036	.018
Strict	63.54	19	< .001	.040	.92	.926	.068	11.16	5	.048	− .008	− .007
By age												
Adolescence (17 to 20 years old)	17.33	4	.002	.028	.98	.992	.082	–	–	–	–	–
Early adulthood (21 to 35 years old)	10.18	4	.037	.021	.99	.997	.055	–	–	–	–	–
Configural	14.19	8	.077	.016	.97	.990	.039	–	–	–	–	–
Metric	20.47	11	.039	.021	.97	.985	.042	5.34	3	.148	− .005	.003
Scalar	24.82	14	.036	.023	.97	.983	.039	3.92	3	.269	− .002	− .003
Strict	44.70	19	.001	.035	.95	.959	.052	15.86	5	.007	− .024	.013

[a]Unidimensional model with reversing negative items; χ^2, chi-square; *df*, degrees of freedom; *SRMR*, standardized root mean square residual; *TLI*, Tucker-Lewis index; *CFI*, comparative fit index; *RMSEA*, root mean square error of approximation; $\Delta\chi^2$, differences in chi-square; *Δdf*, differences in degrees of freedom; *ΔRMSEA*, change in root mean square error of approximation; *ΔCFI*, change in comparative fix index

Table 3 Standardized factorial loads for the items and validity of the scale based on sex, age and for the total sample

Items	Total sample ($n=1002$)		Men ($n=415$)		Women ($n=587$)		Adolescence ($n=499$)		Early adulthood ($n=503$)	
	Factor 1 λ (error)	Factor 2 λ (error)	Factor 1 λ (error)	Factor 2 λ (error)	Factor 1 λ (error)	Factor 2 λ (error)	Factor 1 λ (error)	Factor 2 λ (error)	Factor 1 λ (error)	Factor 2 λ (error)
M2	.75 (.43)		.75 (.43)		.71 (.48)		.68 (.53)		.84 (.29)	
M4	.92 (.15)		.99 (.01)		.90 (.18)		.91 (.16)		.89 (.20)	
M1		.72 (.48)		.80 (.35)		.65 (.56)		.64 (.59)		.80 (.35)
M3		.85 (.27)		.81 (.33)		.87 (.23)		.85 (.27)		.84 (.28)
M5		.84 (.29)		.83 (.29)		.85 (.27)		.87 (.24)		.82 (.31)
Relativity										
α	.82	.85	.85	.85	.78	.83	.77	.83	.86	.87
ω	.75	.79	.82	.79	.71	.78	.70	.78	.78	.81
CR	.83	.85	.87	.86	.80	.84	.79	.84	.86	.87

λ, factor loadings; *Factor 1*, psychological well-being; *Factor 2*, psychological distress; *α*, Cronbach's alpha coefficient; *ω*, omega coefficient; *CR*, composite reliability

appreciated in the total sample and the specific groups in the dimensions of psychological well-being (CR ≥ 0.79) and psychological distress (CR ≥ 0.89).

Item Response Theory Model: Gradual Response Model (GRM)

Two gradual response models (GRM) were fitted, specifically a 2PLM model for each scale's dimension. Table 4 shows that all the discrimination parameters of the psychological well-being and psychological distress dimensions are above the value of 1, generally considered good discrimination (Hambleton et al., 2010). Regarding the difficulty parameters, in both dimensions, all the threshold estimators increased monotonically, as expected.

Figure 1 shows the information curves for the items and dimensions (IIC and TIC, respectively). Regarding the psychological well-being dimension, the IIC shows that item 4 is the most accurate to evaluate the latent trait. The TIC also shows that the factor is more reliable (accurate) in the scale range between − 2 and 1.5. Regarding the psychological distress dimension, the IIC shows that items 3 and 5 are the most accurate for evaluating the latent trait. Furthermore, the TIC shows that the factor is more reliable (accurate) in the scale range between − 1.5 and 3.

Validity Based on the Relationship to Other Constructs

Considering the literature review, we proposed a model to evaluate the relationship between the two dimensions of the construct (psychological well-being and psychological distress) and the level of anxiety and depression. It can be seen in Fig. 2 that the structural model presents adequate adjustment indices (RMSEA = 0.064; CFI = 0.97; TLI = 0.97), and the measurement models are adequately represented by their items.

Discussion

University students are a population vulnerable to mental health problems resulting from the COVID-19 pandemic (Son et al., 2020). In this sense, a quantitative measure of mental health is needed that is useful for the development of prevention programs. Therefore, this study's objective was to evaluate the psychometric

Table 4 Discrimination and difficulty parameters for the items of each dimension

Dimensions	Item	a	b_1	b_2	b_3
Psychological well-being	M2	2.21	− 1.67	.05	1.74
	M4	4.02	− 1.58	− .16	1.27
Psychological distress	M1	1.90	− .91	1.33	2.95
	M3	3.22	− .87	.83	2.30
	M5	2.66	− .13	1.19	2.42

a, discrimination parameters; *b*, difficulty parameters

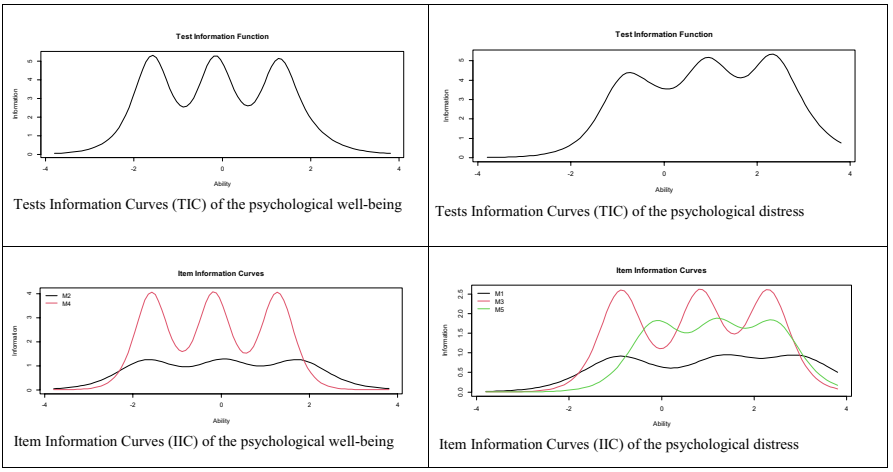

Fig. 1 Item and test information curves for the scale

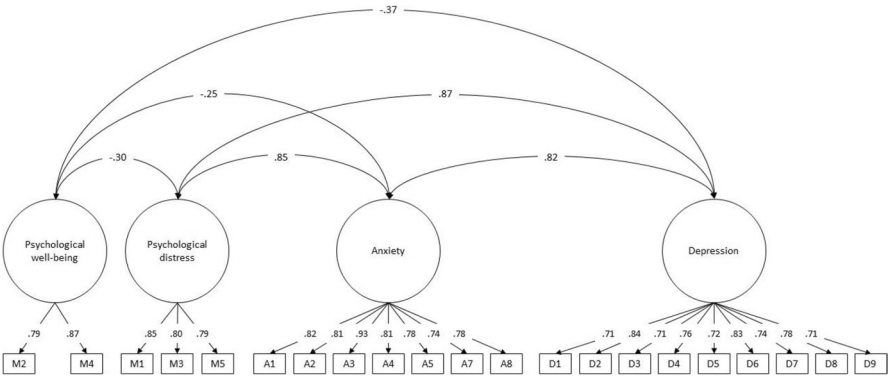

Fig. 2 Relationship model with other constructs

properties of R-MHI-5 in a relatively large population of university students, based on traditional methods, such as TCT, and modern, such as TRI analysis.

The results confirmed a 2-factor model, which assesses well-being and psychological distress. The finding coincides with the study carried out in Spain (Rivera-Riquelme et al., 2019) and the version of 6 response options carried out in Peru (Merino-Soto et al., 2019); however, it is different from the one-dimensional structure found in Portugal (Marques et al., 2011), Brazil (Damásio et al., 2014), and Finland (Elovanio et al., 2020). As mentioned, the scores of items 1, 3, and 5 must be inverted to calculate the total score. These results suggest that the combination of direct and inverse items in the same measure produces that Spanish-speaking people have a different understanding of some mental health indicators compared to people who have another mother tongue (Suárez-Alvarez et al., 2018).

Furthermore, this result is consistent with the theoretical model under which the scale was developed (Veit & Ware, 1983). The dual-factor model of mental health explains that well-being and the absence of psychopathological symptoms are not opposites within a single dimension but rather constitute two different factors of mental health that are negatively related (Antaramian et al., 2010). Under this theoretical model, the subjective well-being dimension is the positive indicator of mental health, and the psychological distress dimension is the negative indicator of mental health. Therefore, evaluating these two factors is essential to have a comprehensive understanding of mental health (Wang et al., 2011).

Although all the items have relatively high factor loadings, item 4 (During the last month, how often have you felt happy?) is the one with the highest value and, therefore, is the indicator that best represents psychological well-being.

Furthermore, from the IRT perspective, item 4 is the most accurate indicator to assess psychological well-being. This result is not surprising since happiness is considered one of the most influential factors in psychological well-being due to positive feelings and the absence of negative feelings (Lyubomksky et al., 2005). On the other hand, items 3 and 5 are the most accurate for evaluating psychological distress. This result is essential since discouragement and deep sadness indicators are fundamental to measure the presence or absence of psychological distress (Wang et al., 2011). Especially in the context of the pandemic, since they are one of the most prevalent problems reported in university students (Generali et al., 2021; Martínez Arriaga et al., 2021). In a complementary way, the results based on IRT indicate that university students require a higher presence of the latent trait (greater well-being or psychological distress) to respond to the higher response categories of the R-MHI-5.

Regarding the test information curves (TIC), in the dimension of psychological well-being, most of the information is in the range of − 2 to 1.5, which indicates that the scale is useful and reliable, especially to identify people with low levels of tranquility, peace, and happiness. In the psychological distress dimension, most of the information is in the range of − 1.5 to 3, which indicates that the scale is handy for identifying people with low and high levels of discouragement, sadness, and anxiety.

Finally, the reliability was also adequate and similar to that reported by Rivera-Riquelme et al. (2019) and Merino-Soto et al. (2019).

The latent relationship model reported that psychological well-being was negatively related to anxiety and depression, as suggested in the previous literature (Contreras et al., 2017; Lew et al., 2019; Yüksel & Bahadir-Yilmaz, 2019). In this sense, psychological well-being decreased as symptoms of anxiety and depression increased. On the other hand, psychological distress was positively related to symptoms of anxiety and depression. Previous studies reported similar results (Dyrbye et al., 2006; Sharp & Theiler, 2018). Therefore, anxiety, depression, and psychological distress are considered important predictors of university students' psychological well-being (Yüksel & Bahadir-Yilmaz, 2019). All these results suggest the importance of increasing psychological well-being and reducing anxiety, depression, and psychological distress in these students by the universities' psychological counseling centers. Finally, the observed relationships give evidence of validity based on the relationship with other variables.

On the other hand, no evidence of measurement invariance was reported according to sex, which would indicate that men and women understand psychological well-being and psychological distress differently. This result could represent a difficulty in using the R-MHI-5 in studies that aim to compare these mental health indicators between both genders. Non-invariance is not an expected result and is considered a statistical problem that must be solved as a previous step to carry out other studies (Vandenberg & Lance, 2000). However, there is evidence that studies with large samples and good psychometric instruments make it possible to identify non-invariance with greater probability (Meade & Bauer, 2007). On the other hand, the R-MHI-5 has the same factorial structure between adolescents and early adults and evidence partial scalar invariance. The previous allows us to suggest that although the R-MHI-5 shows differences in some individual items, it can be used in studies that compare well-being and psychological distress between groups of different ages.

The study has limitations. First, convenience sampling was used that is relatively biased in favor of women and those residing in the highlands of Peru, limiting the results' generalizability. It is important to note that it has been difficult to recruit participants due to social distancing measures and movement restrictions for people in general. Therefore, it is recommended to carry out other studies that use more representative samples of Peruvian university students. Second, self-report measures were used, which could generate insufficient or excessive responses to current symptoms due to social desirability.

Despite the limitations, the findings support using the Spanish version of the R-MHI-5 in clinical and research settings. Furthermore, there are important practical implications. First, studies conducted during the COVID-19 pandemic could include a brief mental health assessment. In these studies, mental health could be considered as an outcome measure or also an explanatory factor. Second, the scale would make it possible to identify psychological well-being and distress levels in the university context and examine their relationships with demographic variables. These results would be beneficial to health professionals and decision-makers in the university context to identify those students most likely to have mental health problems during the COVID-19 pandemic or others that could appear in the future and promote the development of psychoeducational interventions aimed at groups at potential risk.

Conclusion

In conclusion, the results show that the R-MHI-5 is an instrument with good psychometric evidence, based on classical and modern techniques. It was also shown that the R-MHI-5 is not invariant between men and women, but it can be useful to significantly compare the scores between groups of different ages without compromising the inventory's psychometric properties. Without checking for the presence of invariance, it cannot be assumed that the results of comparisons between different groups are valid (Chen, 2008). Therefore, the results are expected to motivate other researchers to assess measurement invariance before comparing well-being and psychological distress, measured by the R-MHI-5, between different age and sex

groups. Finally, the study aims to fill a gap in the measurement, identification, and investigation of well-being and psychological distress in the context of the COVID-19 pandemic.

Data Availability Upon request authors are prepared to send relevant documentation or data in order to verify the validity of the results.

Declarations

Ethics Approval The study obtained the approval of the ethics committee of the Center for Research and Innovation in Health of the Universidad Peruana Unión (N° 00131–2020), and the standards of the Helsinki Declaration were met (World Medical Association, 2013).

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent for Publication Does not apply.

Conflict of Interest The authors declare no competing interests.

References

Acharya, L., Jin, L., & Collins, W. (2018). College life is stressful today–Emerging stressors and depressive symptoms in college students. *Journal of American College Health, 66*(7), 655–664. https://doi.org/10.1080/07448481.2018.1451869

Almeneessier, A. S., Al Saadi, M. M., Nooh, R. M., & Al Ansary, L. A. (2015). Family violence among female medical students: Its prevalence and impact on their mental health status - A cross-sectional study. *Journal of Taibah University Medical Sciences, 10*(1), 33–39. https://doi.org/10.1016/j.jtumed.2015.01.002

Antaramian, S. P., Scott Huebner, E., Hills, K. J., & Valois, R. F. (2010). A dual-factor model of mental health: Toward a more comprehensive understanding of youth functioning. *American Journal of Orthopsychiatry, 80*(4), 462–472. https://doi.org/10.1111/j.1939-0025.2010.01049.x

Berwick, D. M., Murphy, J. M., Goldman, P. A., Ware, J. E., Barsky, A. J., & Weinstein, M. (1991). Performance of a five-item mental health screening test. *Medical Care, 29*(2), 169–176. https://doi.org/10.1097/00005650-199102000-00008

Brown, T. A. (2015). *Confirmatory factor analysis for applied research 43* (5). The Guilford Press

Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., & Zheng, J. (2020). The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Research, 287*, 112934. https://doi.org/10.1016/j.psychres.2020.112934

Caycho, T. (2017). Importancia del análisis de invarianza factorial en estudios comparativos en Ciencias de la Salud. *Revista Cubana de Educacion Medica Superior*, *31*(2), 5–7. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412017000200004

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Contreras, M., de León, A. M., Martínez, E., Peña, E. M., Marques, L., & Gallegos, J. (2017). Psychopathological symptoms and psychological well-being in Mexican undergraduates. *International Journal of Social Science Studies, 5*(6), 30. https://doi.org/10.11114/ijsss.v5i6.2287

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Damásio, B. F., Borsa, J. C., & Koller, S. H. (2014). Adaptation and psychometric properties of the Brazilian version of the five-item mental health index MHI-5. *Psicologia: Reflexao e Critica, 27*(2), 323–330. https://doi.org/10.1590/1678-7153.201427213

Dyrbye, L. N., Thomas, M. R., & Shanafelt, T. D. (2006). Systematic review of depression, anxiety, and other indicators of psychological distress among U.S. and Canadian medical students. *Academic Medicine, 81*(4), 354–373. https://doi.org/10.1097/00001888-200604000-00009

Elovanio, M., Hakulinen, C., Pulkki-Råback, L., Aalto, A. M., Virtanen, M., Partonen, T., & Suvisaari, J. (2020). General health questionnaire (GHQ-12), Beck depression inventory (BDI-6), and mental health index (MHI-5): Psychometric and predictive properties in a Finnish population-based sample. *Psychiatry Research, 289*, 112973. https://doi.org/10.1016/j.psychres.2020.112973

Feng, Y., Dong, D., Zong, M., Yang, Z., & Qiao, Z. (2020). When altruists cannot help: The influence of altruism on mental health during COVID-19 pandemic. *Globalization and Health, 16*(61), 1–8. https://doi.org/10.21203/rs.3.rs-25715/v1

García-Campayo, J., Zamorano, E., Ruiz, M. A., Pardo, A., Pérez-Páramo, M., López-Gómez, V., Freire, O., & Rejas, J. (2010). Cultural adaptation into Spanish of the generalized anxiety disorder-7 (GAD-7) scale as a screening tool. *Health Qual Life Outcomes*, *8*(8). https://doi.org/10.1186/1477-7525-8-8

Generali, L., Iani, C., Macaluso, G. M., Montebugnoli, L., Siciliani, G., & Consolo, U. (2021). The perceived impact of the COVID-19 pandemic on dental undergraduate students in the Italian region of Emilia-Romagna. *European Journal of Dental Education, 25*(3), 621–633. https://doi.org/10.1111/eje.12640

Greenspoon, P. J., & Saklofske, D. H. (2001). Social indicators research. *Social Indicators Research, 54*, 81–108. https://doi.org/10.1023/A:1007219227883

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. In *Handbook of polytomous item response theory models* (pp. 21–42). Routledge/Taylor & Francis Group

Husky, M. M., Kovess-Masfety, V., & Swendsen, J. D. (2020). Stress and anxiety among university students in France during COVID-19 mandatory confinement. *Comprehensive Psychiatry, 102*, 152191. https://doi.org/10.1016/j.comppsych.2020.152191

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2018). *semTools: Useful tools for structural equation modeling. R package version 0.5–1*. https://cran.r-project.org/package=semTools

Kecojevic, A., Basch, C. H., Sullivan, M., & Davi, N. K. (2020). The impact of the COVID-19 epidemic on mental health of undergraduate students in New Jersey, cross-sectional study. *PLoS ONE, 15*(9), e0239696. https://doi.org/10.1371/journal.pone.0239696

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

Lee, J., & Paek, I. (2014). In search of the optimal number of response categories in a rating scale. *Journal of Psychoeducational Assessment, 32*(7), 663–673. https://doi.org/10.1177/0734282914522200

Lee, Jungmin, Jeong, H. J., & Kim, S. (2021). Stress, anxiety, and depression among undergraduate students during the COVID-19 pandemic and their use of mental health services. *Innovative Higher Education*, *0123456789*. https://doi.org/10.1007/s10755-021-09552-y

Lew, B., Huen, J., Yu, P., Yuan, L., Wang, D. F., Ping, F., Talib, M. A., Lester, D., & Jia, C. X. (2019). Associations between depression, anxiety, stress, hopelessness, subjective well-being, coping styles and suicide in Chinese university students. *PLoS ONE, 14*(7), 1–10. https://doi.org/10.1371/journal.pone.0217372

Lin, C. Y., Imani, V., Griffiths, M. D., & Pakpour, A. H. (2020). Validity of the Yale food addiction scale for children (YFAS-C): Classical test theory and item response theory of the Persian YFAS-C. *Eating and Weight Disorders*, 1–10. https://doi.org/10.1007/s40519-020-00956-x

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence, 31*(6), 543–566. https://doi.org/10.1016/S0160-2896(03)00051-5

Lyubomksky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology, 9*(2), 111–131. https://doi.org/10.1037/1089-2680.9.2.111

Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment*, *1*(1), 1–11. https://files.eric.ed.gov/fulltext/ED506058.pdf

Marques, S. C., Pais-Ribeiro, J. L., & Lopez, S. J. (2011). The role of positive psychology constructs in predicting mental health and academic achievement in children and adolescents: A two-year longitudinal study. *Journal of Happiness Studies, 12*(6), 1049–1062. https://doi.org/10.1007/s10902-010-9244-4

Martínez Arriaga, R. J., González Ramírez, L. P., de la Roca-Chiapas, J. M., & Hernández-González, M. (2021). Psychological distress of COVID-19 pandemic and associated psychosocial factors among Mexican students: An exploratory study. *Psychology in the Schools, 58*(9), 1844–1857. https://doi.org/10.1002/pits.22570

McDonald, R. P. (1999). *Test theory: A unified treatment*. London: Taylor & Francis.

Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling, 14*(4), 611–635. https://doi.org/10.1080/10705510701575461

Means-Christensen, A. J., Arnau, R. C., Tonidandel, A. M., Bramson, R., & Meagher, M. W. (2005). An efficient method of identifying major depression and panic disorder in primary care. *Journal of Behavioral Medicine, 28*(6), 565–572. https://doi.org/10.1007/s10865-005-9023-6

Merino-Soto, C., Cuba-Canales, Y., & Rojas-Aquiño, L. (2019). Inventario de Salud Mental – 5 (MHI-5) en adolescentes peruanos: estudio preliminar de validación. *Revista de Salud Pública, 21*(4), 1–5. https://doi.org/10.15446/rsap.v21n4.66357

Milner, A., Kavanagh, A., McAllister, A., & Aitken, Z. (2020). The impact of the disability support pension on mental health: Evidence from 14 years of an Australian cohort. *Australian and New Zealand Journal of Public Health, 44*(4), 307–312. https://doi.org/10.1111/1753-6405.13011

Mo, Y., Wang, H., Huang, G., & Chu, M. (2020). Effectiveness of nurse-led program on mental health status and quality of life in patients with chronic heart failure. *Medicine, 99*(33), e21746. https://doi.org/10.1097/MD.0000000000021746

R Core Team. (2019). *A language and environment for statistical computing* (R version 3.6.1). R Foundation for Statistical Computing. http://www.r-project.org/

Raykov, T., & Hancock, G. R. (2005). Examining change in maximal reliability for multiple-component measuring instruments. *British Journal of Mathematical and Statistical Psychology, 58*(1), 65–82. https://doi.org/10.1348/000711005X38753

Rivera-Riquelme, M., Piqueras, J. A., & Cuijpers, P. (2019). The revised mental health inventory-5 (MHI-5) as an ultra-brief screening measure of bidimensional mental health in children and adolescents. *Psychiatry Research, 274*, 247–253. https://doi.org/10.1016/j.psychres.2019.02.045

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 17*(5), 1–25. https://doi.org/10.18637/jss.v017.i05

Rogowska, A. M., Kuśnierz, C., & Bokszczanin, A. (2020a). Examining anxiety, life satisfaction, general health, stress and coping styles during COVID-19 pandemic in Polish sample of university students. *Psychology Research and Behavior Management, 13*, 797–811. https://doi.org/10.2147/prbm.s266511

Rogowska, A. M., Pavlova, I., Kuśnierz, C., Ochnik, D., Bodnar, I., & Petrytsa, P. (2020b). Does physical activity matter for the mental health of university students during the COVID-19 pandemic? *Journal of Clinical Medicine, 9*(11), 3494. https://doi.org/10.3390/jcm9113494

Rosseel, Y. (2012). lavaan : An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

RStudio Team. (2018). *RStudio: Integrated development environment for R*. RStudio, Inc. http://www.rstudio.com/

Rumpf, H. J., Meyer, C., Hapke, U., & John, U. (2001). Screening for mental health: Validity of the MHI-5 using DSM-IV Axis I psychiatric disorders as gold standard. *Psychiatry Research, 105*(3), 243–253. https://doi.org/10.1016/S0165-1781(01)00329-8

Sahu, P. (2020). Closure of universities due to coronavirus disease 2019 (COVID-19): Impact on education and mental health of students and academic staff. *Cureus, 12*(4), e7541. https://doi.org/10.7759/cureus.7541

Savitsky, B., Findling, Y., Ereli, A., & Hendel, T. (2020). Anxiety and coping strategies among nursing students during the COVID-19 pandemic. *Nurse Education in Practice, 46*, 102809. https://doi.org/10.1016/j.nepr.2020.102809

Schumacker, R. E., & Lomax, R. G. (2015). *A beginner's guide to structural equation modeling: Fourth edition* (4th ed.). Routledge

Sharp, J., & Theiler, S. (2018). A review of psychological distress among university students: Pervasiveness, implications and potential points of intervention. *International Journal for the Advancement of Counselling, 40*(3), 193–212. https://doi.org/10.1007/s10447-018-9321-7

Son, C., Hegde, S., Smith, A., Wang, X., & Sasangohar, F. (2020). Effects of COVID-19 on college students' mental health in the United States: Interview survey study. *Journal of Medical Internet Research, 22*(9), 1–14. https://doi.org/10.2196/21279

Soper, D. (2020). *Calculadora de tamaño de muestra para modelos de ecuaciones estructurales* (4.0). Retrieved on January 20, 2021 from https://www.danielsoper.com/statcalc/calculator.aspx?id=89

Spitzer, R. L., Kroenke, K., & Williams, J. B. W. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. Primary care evaluation of mental disorders Patient health questionnaire. *JAMA, 282*(18), 1737–1744. https://doi.org/10.1001/jama.282.18.1737

Spitzer, R. L., Kroenke, K., Williams, J. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Strand, B. H., Dalgard, O. S., Tambs, K., & Rognerud, M. (2003). Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nordic Journal of Psychiatry, 57*(2), 113–118. https://doi.org/10.1080/08039480310000932

Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema, 30*(2), 149–158. https://doi.org/10.7334/psicothema2018.33

Thorsen, S. V., Rugulies, R., Hjarsbech, P. U., & Bjorner, J. B. (2013). The predictive value of mental health for long-term sickness absence: The major depression inventory (MDI) and the mental health inventory (MHI-5) compared. *BMC Medical Research Methodology*, *13*(1). https://doi.org/10.1186/1471-2288-13-115

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–69. https://doi.org/10.1177/109442810031002

Veit, C. T., & Ware, J. E. (1983). The structure of psychological distress and well-being in general populations. *Journal of Consulting and Clinical Psychology, 51*(5), 730–742. https://doi.org/10.1037/0022-006X.51.5.730

Verwer, J. H., Van Leeuwen, C. M. C., Bolier, L., & Post, M. W. M. (2016). Feasibility of an online well-being intervention for people with spinal cord injury: A pilot study. *Spinal Cord, 54*(6), 473–477. https://doi.org/10.1038/sc.2015.165

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Anales De Psicología, 33*(3), 755–782. https://doi.org/10.6018/analesps.33.3.268401

Visser, M., & Law-van Wyk, E. (2021). University students' mental health and emotional wellbeing during the COVID-19 pandemic and ensuing lockdown. *South African Journal of Psychology, 51*(2), 229–243. https://doi.org/10.1177/00812463211012219

Wang, X., Zhang, D., & Wang, J. (2011). Dual-factor model of mental health: Surpass the traditional mental health model. *Psychology, 2*(08), 767–772. https://doi.org/10.4236/psych.2011.28117

Whang, W., Davidson, K. W., Conen, D., Tedrow, U. B., Everett, B. M., & Albert, C. M. (2012). Global psychological distress and risk of atrial fibrillation among women: The women's health study. *Journal of the American Heart Association, 1*(3), 1–7. https://doi.org/10.1161/jaha.112.001107

World Medical Association. (2013). World Medical Association Declaration of Helsinki ethical principles for medical research involving human subjects. *Clinical Review & Education Special, 27*(20), 2191–2194. https://doi.org/10.1001/jama.292.11.1359

Yüksel, A., & Bahadir-Yilmaz, E. (2019). Relationship between depression, anxiety, cognitive distortions, and psychological well-being among nursing students. *Perspectives in Psychiatric Care, 55*(4), 690–696. https://doi.org/10.1111/ppc.12404

Zhong, Q., Gelaye, B., Fann, J. R., Sanchez, S. E., & Williams, M. A. (2014). Cross-cultural validity of the Spanish version of PHQ-9 among pregnant Peruvian women: A Rasch item response theory analysis. *Journal of Affective Disorders, 158*, 148–153. https://doi.org/10.1016/j.jad.2014.02.012

Zolotov, Y., Reznik, A., Bender, S., & Isralowitz, R. (2020). COVID-19 fear, mental health, and substance use among Israeli university students. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-020-00351-8