Mediterranean Journal of Clinical Psychology

ISSN 2282-1619

MJCP

Volume 9, n 3, 2021

Section

Psychometric properties of Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) in Colombia and Peru

Eliana Ortiz [1], Alba Lucía Meneses Báez [1], Adriana Trujillo [2], César Núñez [3], Magda Yaneth Acevedo Rodríguez [4], Ángela Pilar Albarracín Rodríguez [4], Sergio Dominguez-Lara [5, 6]*

Abstract

This study attempts to adapt and validate the Spanish version of The Clinical Outcomes in Routine Evaluation–Outcome Measure CORE-OM questionnaire for Colombian and Peruvian populations. A linguistic and cultural adaptation of the CORE-OM Spanish version was created via cognitive interviews of potential users (Colombians n=28; Peruvians n=8). Then the questionnaire was applied along with the SOS and OQ-45 tests to 270 participants, divided in clinical (n=60), and non-clinical (n=210) samples, from three Colombian cities and one city in Peru. Some students made a retest application (n=26) 15 to 30 days after the first one. The CFA conducted showed a four-factor structure (χ^2/df=2.09; RMSEA=0.064; RMR=.069) that included well-being, problems/symptoms, general functioning, and risk. The Spanish version of CORE-OM showed excellent acceptability (96%; omission rate 0.14%), good internal consistency (α >.70 in most of the scales and samples), acceptable test-retest reliability (Spearman rho>0.70; p<.001) and convergent validity (all domains r > .4; p < 0.01, except subjective well-being) in both populations. Significant differences were found in the scores between the clinical and non-clinical sample; although, the clinical sample presented higher scores. These results validate the 1-point cut-off score of between clinical and non-clinical populations.

[1] Facultad de Psicología. Universidad Católica de Colombia, Bogotá, Colombia

[2] Facultad de Psicología. Universidad El Bosque, Bogotá, Colombia

[3] Programa de Psicología. Universidad de Medellín, Medellín, Colombia

[4] Facultad de Psicología. Universidad Pontificia Bolivariana, Seccional Bucaramanga, Colombia

[5] Instituto de Investigación de Psicología, Universidad de San Martín de Porres, Lima, Peru

[6] Facultad de Ciencias de la Salud, Universidad Privada San Juan Bautista., Lima, Perú

E-mail corresponding author: eiortiz@ucatolica.edu.co

Keywords:
CORE-OM; Psychometric properties; Outcome measure; Latin-America.

Received: 30 April 2021
Accepted: 21 October 2021
Published: 30 December 2021

Citation: Ortiz, E., Meneses Báez, A. L., Trujillo, A., Núñez, C., Acevedo Rodríguez. M. Y., Albarracín Rodríguez, A. P., Dominguez-Lara, S. A. (2021). Psychometric properties of Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) in Colombia and Peru. *Mediterranean Journal of Clinical Psychology, 9*(3). https://doi.org/10.13129/2282-1619/mjcp-3056

1. Introduction

Recently, there has been an increase in the number of studies attempting to demonstrate the impact of psychological interventions. As a result, complementary paradigms have emerged, such as the evidence-based practice and the practice-based evidence approaches. Both require outcome measures that allow the assessment of systematic interventions (Barkham et al., 2001; Margison et al., 2000). Outcome measures focus on assessing patients' progress and the quality of the services received during treatment (Cantero-Braojos et al., 2019; Zeldovich & Alexandrowicz, 2019).

The Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM, System Trust, 2021) is an instrument widely used in research worldwide, it measures psychological distress and has been used specifically for the evaluation of behavioral change and the impact of psychological interventions in several contexts (clinical, educational, psychological counseling and more). CORE-OM is an outcome measure sensitive to broad levels of psychological distress, covering both positive attributes and pathological symptoms (Barkham et al., 1998). It is a free-use questionnaire, easily applied, and it is not linked to any particular theoretical model nor used as a diagnostic tool. It can be used both in clinical practice and research, as well as in psychoeducational settings. In the United Kingdom it has been implemented as an instrument of common use within the national health system, allowing the assessment and comparison of the efficacy of different mental health services (Barkham et al., 2001; Evans et al., 2003; Mellor-Clark et al., 2001; Stiles et al., 2015).

The CORE-OM has been adapted in 47 countries (CORE System Trust, 2021) and translated into 54 languages and dialects. Twenty-one (21) translations are largely comparable. These versions report similar psychometric results and analyses according to reliability, validity and objectivity (Zeldovich & Alexandrowicz, 2019) like the Kenyan (Falkenström et al., 2018), Italian (Palmieri et al., 2009; Ronconi, et al., 2018), Japanese (Uji et al., 2012), Finnish (Honkalampi et al., 2017; Juntunen et al., 2015), Dutch (Meerding et al., 2012), Norwegian (Skre et al., 2013), Swedish (Elfström et al., 2012), Portuguese (Sales et al., 2012), Icelandic (Kristjánsdóttir et al., 2013), Lithuanian (Viliū nienė et al., 2013), Croatian (Jokic-Begic et al., 2014), Czech (Juhová et al., 2018) and Spanish versions (Feixas el at., 2012; Trujillo et al., 2016). Additionally, to date, some adaptations had been made in Latin America, such as the Brazilian, Ecuadorian and Mexican versions (Paz et al., 2020; Santana et al., 2015¸ Sosa-Torralba et al., 2019).

A recent scoping paper analysed 207 studies from Latin America that used instruments to measure change (i.e., outcome measures) based on a priori inclusion criteria. The results of this study showed that more than three quarters of the studies were conducted in only three of the 20 countries that make up Latin America (Brazil, Chile and Mexico). The most commonly used measures were the Outcome Questionnaire, the Beck Depression Inventory-II, the Hamilton Rating Scale, the Beck Anxiety Inventory and the Yale-Brown Obsessive Compulsive Disorder Scale (Paz et al., 2021).

Adaptations of the CORE-OM for Latin American countries are still in progress. The Brazilian version (Santana et al., 2015) reports linguistic adaptations but does not provide additional information on validity and reliability tests beyond Cronbach's alpha (α). The study on the Ecuadorian population (Paz et al., 2020) focused on convergent validity and reliability analyses, as did the Spanish version of the first psychometric analysis (Trujillo et al., 2016). The Mexican study (Sosa-Torralba et al., 2019) assessed the dimensional structure of the test through exploratory and confirmatory factor analyses in a sample of university students; and the results showed three factor structures (psychological distress, mental health, and risky behaviour), with eight of the test items having differential item functioning (DIF).

Different versions of this instrument have proven to be useful in several areas, including primary and secondary care services (Barkham et al., 2001, 2005; Brand et al., 2021; Kontunen et al., 2016; Iqbal et al., 2021), students' mental health assessments (Kiziela et al., 2019), the efficacy of psychological interventions in university counseling centers (Biasi et al., 2017; Connell et al., 2007; Medlicott et al., 2021; Murray et al., 2016, 2018); the evaluation of therapy services for adults with disabilities from a social and individual perspective (Halacre & Jalil, 2017), the prediction of health-related quality of life and well-being capacity in the context of substance dependence (Peak et al., 2018), the assessment of distress during pandemics (Mækelæ et al., 2021; Zibetti el al., 2021), the screening of patients that are seeking help in the context of suicidal crisis (Saini et al., 2021) and the evaluation of effectiveness of treatments from a variety of approaches (Cattivelli et al., 2021; Jenkins et al., 2021; Stroud & Griffiths, 2021).

Colombia and Peru have very few standardized outcome measurement instruments to assess the efficacy of psychological services, which makes evidence-based research on treatment efficacy much more difficult. This is a psychometric study that aims to adapt the Spanish version of the CORE-OM questionnaire to Colombian and Peruvian populations, and analyze its psychometric properties, reliability and validity indicators in clinical and non-clinical samples, providing evidence for its utility as an outcome measure in these contexts.

2. Materials and Methods

2.1 Participants

The total sample consisted of 270 people who agreed to participate voluntarily. Sixty (60) were clients in outpatient clinical psychology services (44 Colombians; 16 Peruvians); 210 were students and community members (178 Colombians; 32 Peruvians). One hundred-sixty (160) were women (59.2%), aged between 18-70 years (M=29.5; DS=12.79), 110 (40.8%) were men, aged between 18-64 years (M=28.51; DS=11.6). The clinical sample consisted of users of psychological care services (Table 1). The data was gathered between September 2018 and March 2019.

Table 1. Sociodemographic data

Sample	Total (missing data for sex)	Females (%)	Males (%)	Mean age (SD)	Age range (years)
Non-clinical	210 (0)	123 (58.6)	87 (41.4)	27.9 (12.3)	18-70
Students	98 (0)	62 (63.3)	36 (36.7)	20.4 (2.6)	18-35
Community	112 (0)	61 (54.5)	51 (45.5)	34.5 (12.5)	18-70
Clinical (outpatients)	60 (0)	37 (61.7)	23 (38.3)	32.8 (13.7)	18-70

Note. SD=Standard deviation

Each sample consisted of different inclusion criteria. Below we present the specific set of criteria for each one:

Community. Adults without psychological or psychiatric care processes during data collection. Most of them became aware of the study through family, friends, relatives, undergraduate psychology students and coworkers.

University students. Adults pursuing undergraduate studies at a private university. Their participation was voluntary. They were invited to take the CORE-OM questionnaire twice. The second time occurred two or three weeks after the first.

Users of psychological care services. Adults who were in the initial phase of psychological treatment in private offices or psychological university service centers. These people were receiving psychological attention and may have already had a clinical diagnosis. Data were collected during their first or second session.

2.2 Instruments

Clinical Outcomes in Routine Evaluation-Outcome Measure Spanish Version (CORE-OM) (Feixas et al., 2012; Trujillo et al., 2016). This is a 34 item, self-reporting questionnaire scored on a 5-point Likert-type scale (0=Never, 4=always or usually). This questionnaire evaluates four dimensions:

Subjective Well-being (4 items) assesses personal feelings of discomfort; Problems/Symptoms (12 items) assesses anxiety, depression, trauma and physical symptoms; General functioning (12 items) assesses intimate relationships, social relationships and daily functioning level; and the Risk scale (6 items) assesses suicide attempts and self-harm or aggression to third parties (Evans et al., 2000; Lyne et al., 2006).

CORE-OM scores are interpreted as follows: high scores represent high levels of problems or symptoms. For the subjective well-being subscale, high scores also represent greater levels of discomfort. The instrument direct scores range from 0 to 136. To estimate the final score, the total score is divided over the number of items answered. For the scores of each dimension, the same logic must be followed, estimating the total score of the dimension and dividing it by the number of items answered (Evans et al., 2000).

The original CORE-OM has reliability indicators with British clinical and non-clinical samples. Cronbach's Alpha for all dimensions is between .75-.90; the test-retest for risk scale is .61, and the remaining dimensions between .81 and .91. There is also evidence of an exploratory principal-component analysis that supports three factors. These authors obtained evidence for convergent and discriminant validity, sensitivity and discrimination across clinical and non-clinical samples (Evans et al., 2002).

Sociodemographic Data Questionnaire. This questionnaire obtains data on the participants' date of birth, sex, marital status, educational level, type of work and present psychotherapeutic or pharmacological treatments.

Schwartz Outcome Scale-10 (SOS) (Blais et al., 1999). The SOS is a brief self-report scale that assesses the effectiveness of diverse mental health interventions among different levels of care. It has 10 items that are scored on a 7-point Likert-type scale (0= never to 6= all the time or almost all the time). It has a Spanish version (Rivas-Vasquez et al., 2001). Both versions have Cronbach's Alphas above 0.90 (Blais et al., 1999; Rivas-Vasquez et al., 2001).

Outcome Questionnaire (OQ-45) (Lambert et al., 2004). This is a self-report questionnaire that assesses treatment outcomes in mental health settings. It has three subscales: Symptom distress, Interpersonal relationships, and the social role performance. The English and Spanish versions have acceptable internal consistency (Cronbach Alpha; $\alpha > 0.74$) (Errázuriz et al., 2017; Lambert, et al., 2004; Lambert, 2015). This questionnaire also has a Colombian adaptation (Londoño et al., 2017) with a similar alpha (between 0.52-0.92 for each factor). The present study used the Colombian version in order to obtain evidence of concurrent validity.

2.3 Procedure

The research team included members from three Colombian cities (Medellín, Bogotá and Bucaramanga) and one Peruvian city (Lima) where the instruments were applied; and a founding member of the Core System Group in England who met with the team several times during the research process. The study was conducted in two phases:

Phase I. Linguistic and cultural adaptation of the CORE-OM Spanish version. In order to get a linguistic and cultural equivalent of the CORE-OM Spanish version (Feixas et al., 2012; Trujillo et al., 2016) for Colombian and Peruvian populations, the research team assessed the instrument via cognitive interviews of 36 potential users in Colombia (28) and Peru (8), to determine if the target population had a good understanding of the items on the questionnaire (instructions, words, items and response options). This sample included users of psychological care services (11), persons in condition of vulnerability (8); university students (8), and general population (9).

After analyzing the cognitive interviews, the researchers found that more than 80% of the potential users interviewed understood the instructions and items on the instrument. As a result of this process, we used two versions of item 17 "*I have felt overwhelmed by my problems*" (*me he sentido agobiado por mis problemas*) and "*I have felt worried about my problems*" (*me he sentido preocupado por mis problemas*).

Later on, in order to assess psychometric indicators, the final version of the CORE-OM questionnaire for Colombia and Peru was adjusted to 35 items in two versions (A and B). The two versions differ in the presentation order of items 17 and 35, since it was necessary to evaluate which, one was the most appropriate for the Colombian and Peruvian populations.

Phase II. Data collection and analysis CORE-OM 35 versions A and B. The questionnaire was applied along with the SOS and OQ-45 tests in Bogotá, Bucaramanga, Medellín and Lima, individually, in the context of psychological care services or community social programs; and in groups for students and workers.

The respondents received instructions asking them to please read and answer each question, thinking about how they have felt during the past seven days.

2.4 Statistical data analysis

Once the data had been collected and compiled, the authors analyzed them with the SPSS v.23.0 program, following the same analysis procedure used in the Spanish study to facilitate comparison with the Spanish data, which is described below.

First, the response omission rate for each questionnaire and the total response omission rate were estimated.

Second, the demographic characteristics of the participants were obtained through descriptive statistical analyses for sex, age and type of sample (Table 1).

Third, we conducted normality tests for the CORE-OM domain scores and the referential instruments scores (OQ-45 y SOS-10) using Kolmogorov-Smirnov normality test. Most of the scores did not exhibit normality, therefore, Spearman rho was computed to estimate the existing correlations between variables: test-retest reliability of the CORE-OM questionnaire; correlations between the scores of the OQ-45 and SOS-10 questionnaires and those of the CORE-OM as evidence for convergent validity having as criterion an $r \geq .45$ (Howitt & Cramer, 2017); and also we estimated correlations between domains scores of the CORE-ON questionnaire for clinical and non-clinical samples and correlations between age and domains of the CORE-OM for clinical and non-clinical samples.

Fourth, the authors conducted a linguistic evaluation of the Spanish versions of the 34 items of the CORE-OM (Trujillo et al., 2016), so it was necessary to include two versions of item 17 (17 and 35 in this study). These were evaluated using two independent confirmatory factor analyses (CFA), including two data samples (clinical and non-clinical) with AMOS v.24, one for item 17 and another for item 35, to determine which item was better. One for item 17 and one for item 35, to determine which item was the best. The CFA allows for a theoretical evaluation through the conceptual structure of the instrument to test whether the hypothesized model fits the data collected (Collier, 2020). This analysis requires goodness-of-fit tests based on Chi-square ($\chi 2$), and Table 2 lists these estimation criteria according to Hair et al. (2014).

Fifth, we estimated the internal consistency coefficients (Cronbach's alpha and Omega with their respective Confidence Intervals) for the complete CORE-OM questionnaire and for each of its domains (Subjective well-being, Problems/Symptoms, Life/social functioning, and Risk) for the subsamples. There were 9 missing data; this data was replaced by the mean of the item that was left blank; and we estimate the Confidence Interval (CI) for the alphas using the two-way mixed model intraclass correlation coefficient in SPSS (Baumgartner & Chung, 2001; Bravo & Potvin, 1991). The comparison of Cronbach's alpha coefficients between samples were estimated using the cocron interface based on the Chi-Squared test (Diedenhofen & Musch, 2016).

Sixth, estimations of the standardized measure of Cohen's d were computed to assess the effect size between sex for the clinical and nonclinical samples using as criteria .2 = low effect; .3 = moderate effect; and .8 or more = high effect (Howitt & Cramer, 2017).

Seventh, cut-off score estimates for men and women were calculated for all domains of the CORE-OM to identify clinically significant change using a c-criterion that uses a cutoff point which is a weighted midpoint between two-distribution means, based on the contrast between dysfunctional and general population samples (Jacobson & Truax, 1991). This criterion considers data from both populations and allows us to identify whether there is change after treatment.

Table 2. Model fit indexes criteria

Measure	Threshold
CFI	More than .95 is great; more than .90 is traditional; more than .80 is acceptable
GFI	More than .95
TLI	More than .95
NFI	More than .95
PNFI	More than .95
AGFI	More than .80
RMR	Less than .08
RMSEA	Less than .05 is good; between .05 and .10 is moderate; more than .1 is bad
PCLOSE	Should be more than .05

3. Results

3.1 Acceptability

All the questionnaires have a maximum of two items missing to allow prorating for a usable overall score. Fifty-eight (58) (96.7%) participants of the clinical and 203 (97%) of the non-clinical samples returned completed data. The overall omission rate was 0.14%.

3.2 Confirmatory Factor Analysis (CFA)

The results of the first CFA (conducted with item 35) show that the factor loading of item 35 was not significant for the model (β=-.078; p>.005). The second CFA (conducted with item 17), shows that the item was significant for the model (β=.681; p<.005) and had a strong contribution to the Subjective wellbeing factor, so we chose the model with item 17 for the rest of the analysis.

The CORE-OM data were subjected to CFA with the structure specified by Evans et al. (2002), and Trujillo et al. (2016) using SPSS AMOS v.21. The hypothesis tested is that CORE-OM is a measure with a four-factor structure that included well-being (4 items), problems/symptoms (12 items), general functioning (12 items), and risk (6 items). Table 3 presents the results for

goodness-of-fit model indices for the four-factor model of the CORE-OM. The fit indices for the one-factor model do not meet criteria for a well-fitted model (Hair et al.,2014). The model showed a slightly lower value than the recommended cutoff criteria for all, except for RMSEA, RMR, and the relation χ^2/df.

Table 3. Model fit information for a 4-factor CFA of the CORE-OM

Statistics	χ^2 (p)	χ^2/df	RMSEA	CFI	TLI	NFI	PNFI	GFI	AGFI	RMR
Criteria	>.05	< 3	< .08	≥.95	≥.95	≥.95	≥.95	≥.95	≥.80	< .08
Model	.000	2.09	.064	.839	.825	.734	.675	.797	.766	.069

Note. RMSEA=Root Mean Square Error of Approximation; CFI=Comparative Fit Index; TLI=Tucker–Lewis index; NFI=Normed Fit Index; PNFI=Parsimony Normed Fit Index; GFI=Goodness-of-fit index; AGFI=Adjusted Goodness-of-Fit Index; RMR=Root Mean Square Residual; χ^2 = 1078.124; df = 516

3.3 Internal consistency

All domains showed appropriate internal reliability in both samples. Although the Risk domain has the lower alpha across samples, and the subjective well-being domain has the lower alpha in the students' sample (Table 4).

Table 4. Coefficient α (95% CI) denoting internal consistency for non-clinical and clinical samples CORE-OM

Domains	Students (n=98)		Community (n=111)		Pooled non-clinical sample (n=210)		Pooled clinical sample (n=60)	
	α	CI 95%	α	CI 95%	α	CI 95%	α	CI 95%
Subjective well-being	0.59	(0.45, 0.71)	0.77	(0.69, 0.83)[a]	0.71	(0.64, 0.77)	0.75	(0.63, 0.84)
Problems/ Symptoms	0.70	(0.61, 0.78)	0.89	(0.86, 0.92)[b]	0.83	(0.80, 0.87)	0.87	(0.81, 0.91)
Functioning	0.77	(0.69, 0.83)	0.84	(0.79, 0.88)	0.81	(0.77, 0.85)	0.84	(0.77, 0.89)
Risk	0.69	(0.58, 0.78)	0.51	(0.35, 0.64)	0.60	(0.51, 0.68)	0.74	(0.62, 0.83)
Non-risk	0.86	(0.82, 0.89)	0.94	(0.92, 0.95)[b]	0.92	(0.90, 0.93)	0.93	(0.90, 0.95)
All-items	0.80	(0.82, 0.90)	0.94	(0.92, 0.96)[c]	0.92	(0.90, 0.93)	0.94	(0.91, 0.96)

Note. CI= Confidence interval [a] p = 0.031 (significantly higher α in the community sample in comparison with the students' sample). [b] p<0.000 significantly higher α in the community sample in comparison with students' sample. [c] p =0.0001 significantly higher α in the community sample in comparison with students' sample.

The students' sample in this study shows slightly lower alphas when compared to the Spain data, and only the confidence interval of the subjective well-being domain of both studies included the alpha value of the other study. In our study, the domains subjective well-being and risk on the community sample have an alpha slightly lower in respect to the Spain data, but the confidence interval of each study includes the alpha values of the other. The domain Functioning has the same alpha and confidence interval of the Spain study; the other domains (Problem/Symptoms, Non-risk and All-items) have slightly higher alphas covered for the confidence interval of the referential study.

In the community sample the Subjective well-being, Problem symptoms, Non-risk, and All-items domains had the highest confidence interval limits for alpha (p<0.05). The Risk domain showed the lower alpha confidence interval for the samples (Table 4).

Also, the confidence interval between clinical and non-clinical samples shows that alpha values have a confidence interval higher in general for the clinical sample, however their differences were not significant (p>0.05).

3.4 Test–retest stability

CORE-OM Test–retest reliability was estimated with 26 students with ages between 18 and 28 years (M=20.9; SD=2.4). Correlations were strong (Spearman rho>0.70; p<0.001) within domains in the non-clinical data. We used Spearman rho since the scores' distribution did not fit a normal curve. The highest one was All-items with 0.88, followed by all Non-risk items with 0.84, Problems/symptoms with 0.81, Risk with 0.78, and Functioning and Subjective well-being with 0.72.

3.5 Convergent validity

The correlations estimated (Table 5) between domain scores and the domains of the OQ-45 and SOS-10 questionnaires, evidence an acceptable convergent validity with exception of the domain subjective well-being and OQ45-IR (interpersonal relationship) that show the lowest correlation magnitude although significant (p>0.01).

Table 5. *CORE-OM correlations with referential measures in clinical sample*

Sample	Domains					
(n=52)	W	P	F	R	ALL-R	ALL
OQ45-IR	0.33*	0.55**	0.44**	0.42**	0.49**	0.51**
OQ45-SD	0.58**	0.74**	0.49**	0.40**	0.68**	0.68**
OQ45-SR	0.55**	0.70**	0.52**	0.51**	0.67**	0.67**
OQ-45	0.56**	0.76**	0.52**	0.46**	0.69**	0.70**
SOS-10	-.35*	-0.50**	-0.46**	-0.40**	-0.50**	-0.51**

Note. **p<0.01 two tails; *p<0.05 two tail.W: Subjective well-being; P: Problem symptoms; F: Functioning; R: Risk; ALL-R, Non-risk

3.6 Differences between clinical and non-clinical samples

Results show that there are significant differences between clinical and non-clinical samples in all domains with the highest scores for the clinical sample. The effect sizes values of the differences of this study domains are between 0.3 and 0.7. The lowest effect size value corresponds to the Non-risk domain, and the highest value is for the subjective well-being domain (Table 6).

Table 6. Mean and standard deviations for clinical and non-clinical samples

Domains	Present study						Trujillo et al. (2016) *					
	Non-clinical n = 210		Clinical n = 60		95% CI		Non Clinical (n=452)		Clinical (n=192)		95% CI	
	M	SD	M	SD	Difference	d[a]	M	SD	M	SD	Difference	d[a,b]
Subjective well-being	1.18	0.78	1.80	0.88	0.39, 0.85	0.7 (0.68, 0.87)	1.18	0.76	2.41	0.95	1.08, 1.36	1.5 (1.31, 1.68)
Problems/ symptoms	1.26	0.68	1.63	0.77	0.17, 0.57	0.5 (0.45, 0.61)	0.99	0.62	1.98	0.87	0.86, 1.10	1.4 (1.22, 1.59)
Functioning	0.99	0.57	1.30	0.69	0.13, 0.49	0.5 (0.45, 0.59)	0.74	0.52	1.56	0.75	0.71, 0.92	1.3 (1.19, 1.55)
Risk	0.23	0.36	0.39	0.54	0.04, 0.28	0.5(0.45, 0.59)	0.11	0.27	0.48	0.66	0.29, 0.44	0.8 (0.69, 1.04)
Non-risk	1.13	0.59	1.51	0.69	0.20, 0.56	0.3 (0.25, 0,40)	0.91	0.55	1.86	0.78	0.84, 1.05	1.5 (1.32, 1.70)
All-items	0.97	0.53	1.31	0.64	0.18, 0.50	0.6 (0.55, 0.68)	0.77	0.48	1.62	0.71	0.75, 0.94	1.5 (1.33, 1.71)

Note. [a]Cohen effect size parameter. [b] Cohen's d has been calculated with the data provided by the Spanish study.
*Reproduced with permission from Trujillo et al. (2016).

Regarding age, the non-clinical sample had negative and significant correlations (p<0.05) in all domains except for Risk. However, those correlations were low with the highest (ϱ=-0.23, p=0.001) being for Functioning and Non-risk, and the lowest for Subjective Wellbeing with (ϱ=-0.14, p=0.041). In the clinical sample the correlations were not significant.

There were significant differences between males and females for all domains (Table 7) in the non-clinical and clinical samples. The effect sizes were small for all of the samples, being smaller for the clinical (between -0.39 and 0.36) and slightly higher for the non-clinical (between -0.02 and 0.2). The Risk domain had the highest effect size in both samples.

Table 7. Sex differences in score for non-clinical and clinical samples

Domains	Non-clinical						Clinical					
	Male n=87		Female n=123		95% CI		Male (n=23)		Female (n=37)		95% CI	
	M	SD	M	SD	Difference	d[a]	M	SD	M	SD	Difference	d[a]
Subjective well-being	1.07	0.76	1.25	0.79	-0.36, 0.39	-0.2 (-0.34, -0,13)	1.72	1.02	1.85	0.78	-0.34, 0,60	-0.15 (-0.27, -0.03)
Problems/ symptoms	1.28	0.73	1.24	0.66	-0.23, 0.15	-0.06 (-0.03, 0.15)	1.44	0.88	1.74	0.68	-0.11, 0.71	-0.39 (-0.50, -0.29)
Functioning	0.99	0.59	1,00	0.56	-0.14, 0.18	-0.02 (-0.09, 0.06)	1.28	0.87	1.32	0.56	-0.33, 0,41	-0.06 (-0.15, -0.04)
Risk	0.28	0.38	0.20	0.35	-0.18, 0.02	0.2 (0.18, 0.27)	0.51	0.69	0.32	0.39	-0.48, 0.84	0.36 (0.29, 0.43)
Non-risk	1.13	0.62	1.14	0.59	-0,15, 0.18	-0.02 (-0.10, 0.06)	1.41	0.85	1.58	0.57	-0.20, 0.53	-0.25 (-0.34, -0.15)
All-items	0.98	0.54	0.97	0.52	-0.15, 0,14	0.02 (-0.05, 0.09)	1.26	0.81	1.35	0.51	-0.24, 0.44	-0.14 (-0.23, -0.05)

M: Mean; SD: Standard deviation; CI: Confidence interval. [a]Cohen effect size parameter.

3.7 Correlations between domain scores

The correlations between all domain scores are significant and strong, as was expected (Table 8). However, the correlation between the risk domain score and the other domain scores were the lowest in both samples; although these correlations were slightly lower in the non-clinical sample.

Table 8. Spearman rho correlations for clinical and non-clinical samples

Domains	W	P	F	R	ALL-R
Non-clinical (n = 210)					
W	1				
P	0.78**	1			
F	0.70**	0.71**	1		
R	0.55**	0.45**	0.63**	1	
ALL-R	0.88**	0.93**	0.88**	0.58**	1
ALL	0.87**	0.92**	0.90**	0.64**	0.99**
Clinical (n = 60)					
W	1				
P	0.72**	1			
F	0.70**	0.72**	1		
R	0.43**	0.49**	0.37**	1	
ALL-R	0.84**	0.92**	0.91**	0.48**	1
ALL	0.84**	0.92**	0.89**	0.54**	0.99**

Note: **p<0.01 two tails; *p<0.05 two tail.W: Subjective well-being; P: Problem symptoms; F: Functioning; R: Risk; ALL-R, Non-risk

3.8 Clinically significant change

The estimated values to establish a clinically significant change (Table 9) were calculated according to the c-criterion, which is a weighted midpoint between the two-distributions means (Jacobson & Truax, 1991). In this case, the data from the clinical and nonclinical sample allow us to establish cut-off scores for women and men and to identify whether the change after treatment was significant.

Table 9. Male and female cutoff scores between clinical and non-clinical populations

Domains	Present Study		Trujillo et al.*		Evans et al.**	
	Male	Female	Male	Female	Male	Female
Subjective well-being	1.34	1.55	1.46	1.82	1.37	1.77
Problems/symptoms	1.35	1.49	1.33	1.43	1.44	1.62
Functioning	1.11	1.16	1.06	1.07	1.29	1.30
Risk	0.36	0.26	0.24	0.21	0.43	0.30
Non-risk	1.25	1.36	1.24	1.33	1.36	1.50
All-items	1.09	1.16	1.06	1.13	1.19	1.29

Note: *Reproduced with permission from Trujillo et al. (2016) **Reproduced with permission from Evans et al. (2002)

4. Discussion

This paper presents the psychometric properties of a Spanish version of the CORE-OM questionnaire for Colombian and Peruvian populations in clinical and non-clinical samples. The linguistic and cultural adaptation showed that the instructions and items were understood by the populations, supporting its use.

Our results support the acceptability, internal consistency, test-retest stability and convergent validity of the CORE-OM in the Colombian and Peruvian population. Although the samples were not large, they were diverse and this might support the use of the questionnaire in different populations. Nonetheless it is necessary to conduct studies on these populations using invariance measurements or DIF analysis. As for acceptability, the results were close to those obtained in the original English questionnaire (Evans et al., 2002) and superior to those obtained in the Spanish adaptation (Trujillo et al., 2016) for clinical and non-clinical samples.

Regarding internal consistency, although almost all domains showed appropriate internal reliability when the non-clinical sample was pooled (Risk didn't), their alpha values were lower than in the studies from Spain and Ecuador (Paz et al., 2020; Trujillo et al., 2016), and their confidence intervals did not include the alpha values of the present study. The Risk domain had the lowest alpha in the clinical and non-clinical samples. These results are consistent with the

analyses conducted in other adaptation and validation studies, such as the ones from Sweden and Spain (Elfström et al., 2012; Trujillo et al., 2016). The students' sample in this study shows slightly lower alphas in the Well-being domain compared to the data from Spain. In the clinical sample, the alpha values of our study were slightly lower than those of the Spain study (Trujillo et al., 2016), but their confidence intervals include the alpha values of all domains, except for the Well-being.

In general, we can conclude that test-retest reliability indicate that the CORE-OM data is stable and satisfactory in all domains, although the sample was small and only included non-clinical population. The correlations are slightly lower than the Spain and the original UK data. It should be noted that in our study the correlation for the Risk domain was better than in the Swedish, British and Spanish versions (Elfström, et al., 2012; Evans et al., 2002; Trujillo et al., 2016).

Likewise, our results suggest that the current data is consistent with other measures (SOS, OQ-45), therefore, supporting the convergent validity of our scale, except for the subjective well-being domain and OQ45-IR (interpersonal relationship) that shows the lowest significant (p>0.01) correlation magnitude.

According to our findings, there are significant differences between clinical and non-clinical samples in all domains as in UK, Finland, and Spain (Evans et al., 2002; Honkalampi et al., 2017; Trujillo et al., 2016). For the effect sizes, the lowest value corresponds to the Non-risk domain items, and the highest value to the subjective well-being domain, unlike the Spain and UK studies, where the value was smaller in the risk dimension. Nonetheless, the CIs values for all domains are not included in the CIs values for UK and Spain (Evans et al., 2002; Trujillo et al., 2016).

Regarding age and sex differences, the highest effect size in both samples was the Risk domain. Compared to the UK and Spain data (Evans et al., 2002; Trujillo et al., 2016) the effect sizes were smaller.

Our findings show that the correlations between all domains are significant and strong. The correlations between risk and other domain scores were the lowest in both samples, but they were slightly lower in the non-clinical sample, which is consistent with the Sweden, UK, Finland and Spain data (Elfström et al., 2012; Evans et al., 2002; Honkalampi et al., 2017; Trujillo et al., 2016).

The CFA tested a model with a four-factor structure that barely met the criteria, but is considered acceptable. These results coincide with those found in Japan (Uji et al., 2012), Norway (Skre et al., 2013) and Croatia (Jokic-Begic et al., 2014), where RMSEA, GFI, AGFI

and CFI values were similar. Likewise, the Mexican study (Sosa-Torralba et al., 2019) performed a CFA based on three factors obtained in an EFA. The authors report that CFA results partial resemble the structure of the original version of the CORE-OM, but they do not present the RMSEA value. So far, the evidence from CFA studies does not have a robust result on the factor structure of the CORE-OM that supports the model proposed by Evans et al. (2002) because most of the studies that conducted CFA have reported the existence of four different dimensions for this test.

In fact, Zeldovich and Alexandrowicz (2019) compared different studies on the psychometric properties of CORE-OM, and regarding its factorial structure, none of the studies that applied CFA or PCA found a model that fitted properly, highlighting the need for further studies addressing the structural aspects of the questionnaire. However, in none of the versions included in this study was the questionnaire modified; neither the number of items nor their form of qualification. All versions, including the version used in this study, assume this deficiency in the questionnaire.

Regarding the linguistic and cultural adaptation, it is important to state that the cognitive interviews showed that the Spanish translation of the CORE-OM used was understandable and comprehensible to both Peruvian and Colombian populations with the exception of one item. Therefore, we created two different versions of the same item, and applied both to the two samples. This decision allowed us to evaluate the behavior of the test with each item and to choose the one with the most acceptable performance. It is possible that this CORE-OM version could be used in other Latin American populations such as Bolivian or Chilean. Nonetheless before using it, it is necessary to conduct a linguistic and cultural assessment to determine if there are any differences on this regard. Paz et al. (2020) used the same procedure with Ecuadorians and their conclusion on this aspect is similar.

Although there are other ways of assessing psychological distress and change in psychotherapy, there are no other valid scales for Colombia and Peru that show adequate psychometric properties and that take into account the four dimensions contemplated by the CORE-OM.

Therefore, we consider that the Spanish version of CORE-OM is a useful, simple and brief questionnaire that can be used to assess the progress of therapy, as it provides information to both the therapist and patient about the process. This test is an indicator of change used worldwide in psychotherapy and helps to accomplish one of the most important and complex objectives in clinical work and psychotherapy research. It can be useful for both practices.

The CORE OM besides being an instrument that allows assessing the change occurred by psychotherapy, is an instrument that allows comparing the results of different cases and services,

as is the evaluation of the quality of health care, in combination with specific population or diagnostic instruments that focus on specific aspects associated with certain health conditions, such as the Patient Health Questionnaire (PHQ-9) the Beck Anxiety or Depression Inventory, and the Hamilton Rating Scale, among others (Sales et al., 2018).

In addition, clinicians in all settings are increasingly considering the use of outcome measures as a strategy to address existing changes in patients' primary concerns; however, the results obtained by such change measures must be analyzed in a relational context that includes both contextual and cultural aspects (Paz et al., 2020). Furthermore, Börjesson and Boström (2020) note that when clients perform a routine outcome measure (ROM) using the CORE-OM, they experienced it as a part of their therapy. Meanwhile, other authors recommend the CORE-OM for assessing people's quality of life, as well as for evaluating the effectiveness and cost-effectiveness of mental health interventions (Wickramasekera & Tubeuf, 2020).

However, caution is needed as more tests are required to obtain evidence on the factor structure of the CORE-OM, its reliability and DIF analysis. The psychometric properties of this scale should be studied further, especially in larger samples from different regions and groups, to determine the level of stability and generality of the CORE-OM in Colombia and Peru.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any potential conflict of interest.

References

1. Barkham, M., Evans, C., Margison, F., McGrath, G., Mellor-Clark, J. & Milne, D. (1998). The rationale for developing and implementing core batteries in service settings and psychotherapy outcome research. *Journal of Mental Health*, *7*, 35 – 47. https://doi.org/10.1080/09638239818328
2. Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark & J., Evans, C. (2001) Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, *69*, 184–196. http://dx.doi.org/10.1037/0022-006X.69.2.184
3. Barkham, M., Gilbert, N., Connell, J., Marshall, C. & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presentig problems in psychological therapy services based in primary and secondary care settings. *The British Journal of Psychiatry, 186*(3), 239 – 246. https://doi.org/10.1192/bjp.186.3.239
4. Baumgartner, T. A., & Chung, H. (2001) Confidence limits for intraclass reliability coefficients. *Measurement in Physical Education and Exercise Science*, *5*(3), 179-188. https://doi.org/10.1207/S15327841MPEE0503_4
5. Biasi, V., Patrizi, N., Mosca, M. & De Vincenzo, C. (2017). The effectiveness of university counselling for improving academic outcomes and well-being. *British Journal of Guidance & Counselling, 45*(3), 248–257. https://doi.org/10.1080/03069885.2016.1263826
6. Börjesson, S., & Boström, P. K. (2020). "I want to know what it is used for": Clients' perspectives on completing a routine outcome measure (ROM) while undergoing psychotherapy. *Psychotherapy Research, 30*(3), 337-347. https://doi.org/10.1080/10503307.2019.1630780
7. Blais, M. A., Lenderking, W. R., Baer, L., de Lorell, A., Peets, K., Leahy, L. & Burns, C. (1999). Development and initial validation of a brief mental health outcome measure. *Journal of Personality Assessment, 73*(3), 359–373. https://doi.org/10.1207/S15327752JPA7303_5
8. Brand, C., Ward, F., MacDonagh, N., Cunningham, S., & Timulak, L. (2021). A national evaluation of the Irish public health counselling in primary care service–examination of initial effectiveness data. *BMC Psychiatry*, *21*, 227. https://doi.org/10.1186/s12888-021-03226-x
9. Bravo, G., & Potvin, L. (1991) Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: Toward the integration of two traditions. *Journal of Clinical Epidemiology, 44*(4-5), 381-390. https://doi.org/10.1016/0895-4356(91)90076-L
10. Cantero-Braojos, M. Á., Cabrera-León, A., López-González, M. A., & Saúl, L. A. (2019). Intervención grupal desde un enfoque sensoriomotriz para reducir la intensidad del dolor crónico. *Atención Primaria*, *51*(3), 162-171. https://doi.org/10.1016/j.aprim.2017.07.006
11. Cattivelli, R., Guerrini Usubini, A., Manzoni, G. M., Vailati Riboni, F., Pietrabissa, G., Musetti, A., Franceschini, C., Varallo, G., Spatola, C. A. M., Giusti, E., Castelnuovo, G. & Molinari, E. (2021). ACTonFood. Acceptance and Commitment Therapy-Based Group Treatment Compared to Cognitive Behavioral Therapy-Based Group Treatment for Weight Loss Maintenance: An Individually Randomized Group Treatment Trial. *International Journal of Environmental Research and Public Health*, *18*(18), 9558. https://doi.org/10.3390/ijerph18189558

12. Collier, J. E. (2020). *Applied structural equation modeling using AMOS. Basic to advanced techniques*. Routledge Taylor & Francis Group. https://doi.org/10.4324/9781003018414
13. Connell, J., Barkham, M., & Mellor-Clark, J. (2007). CORE-OM mental health norms of students attending university counselling services benchmarked against an age-matched primary care sample. *British Journal of Guidance & Counselling, 35*(1), 41-57. https://doi.org/10.1080/03069880601106781
14. CORE System Trust (2021, June 02). *About CORE and CST*. https://www.coresystemtrust.org.uk/home/about-cst/
15. Diedenhofen, B. & Musch, J. (2016). cocron: A web interface and R package for the statistical comparison of Cronbach's alpha coefficients. *International Journal of Internet Science, 11*, 51- 60. https://www.ijis.net/ijis11_1/ijis11_1_diedenhofen_and_musch.pdf
16. Elfström, M. L, Evans, C., Lundgren, J., Johansson, B., Hakeberg, M., & Carlsson, S. G. (2012). Validation of the Swedish version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM). *Clinical Psychology & Psychotherapy, 20*(5), 447- 455. https://doi.org/10.1002/cpp.1788
17. Errázuriz, P., Opazo, S., Behn, A., Silva, O., & Gloger, S. (2017). Spanish Adaptation and Validation of the Outcome Questionnaire OQ-30.2. *Frontiers in Psychology,* 8, 673. https://doi.org/10.3389/fpsyg.2017.00673
18. Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J. & McGrath, G. (2000). CORE: Clinical outcome measure in routine evaluation. *Journal of Mental Health*, *9*(3), 247-255. https://doi.org/10.1080/jmh.9.3.247.255
19. Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardized brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry, 180*(1), 51-60. https://doi.org/10.1192/bjp.180.1.51
20. Evans, C., Connell, J., Barkham, M., Marshall, C. &, Mellor-Clark, J. (2003). Practice-based evidence: benchmarking NHS primary care counselling services at national and local levels. *Clinical Psychology & Psychotherapy, 10,* 374 -388. https://doi.org/10.1002/cpp.384
21. Falkenström, F., Kumar, M., Zahid, A., Kuria, M. & Othieno, C. (2018). Factor analysis of the Clinical Outcomes in Routine Evaluation–Outcome Measures (CORE-OM) in a Kenyan sample. *BMC Psychology,* 6,48. https://doi.org/10.1186/s40359-018-0260-1
22. Feixas, G., Evans, C., Trujillo, A., Saúl, L. A., Botella, L., Corbella, S., González, E., Bados, A., García-Grau, E., & López-González, M. A. (2012). La versión española del CORE-OM: Clinical Outcomes in Routine Evaluation - Outcome Measure. *Revista de Psicoterapia*, *23*(89), 109-135. https://doi.org/10.33898/rdp.v23i89.641
23. Hair, J. F., Black, W. C., & Barry, B. J. (2014). *Multivariate Data Analysis* (7th ed.) Pearson. https://www.pearson.com/uk/educators/higher-education-educators/program/Hair-Multivariate-Data-Analysis-Pearson-New-International-Edition-7th-Edition/PGM1052102.html
24. Halacre, M., & Jalil, R. (2017). Holistic therapy with disabled adults from a social and individual perspective: A service evaluation feasibility study. *Counselling and Psychotherapy Research*, *17*(4), 320-329. https://doi.org/10.1002/capr.12137

25. Honkalampi, K., Laitila, A., Juntunen, H., Lehmus, K., Piiparinen, A., Törmänen, L., Inkinen, M., & Evans, C. (2017). The Finnish Clinical Outcome in Routine Evaluation Outcome Measure: Psychometric exploration in clinical and non-clinical samples. *Nordic Journal Psychiatry*, *71*(8), 589 - 597. https://doi.org/10.1080/08039488.2017.1365378
26. Howitt, D., & Cramer, D. (2017). *Research methods in psychology* (5th ed.). Pearson. https://www.pearson.com/uk/educators/higher-education-educators/program/Howitt-Research-Methods-in-Psychology-5th-Edition/PGM1240430.html
27. Iqbal, Z., Airey, N. D., Brown, S. R., Wright, N. M., Miklova, D., Nielsen, V., Webb , K. & Sajjad, A. (2021). Waiting list eradication in secondary care psychology: Addressing a National Health Service blind spot. *Clinical Psychology & Psychotherapy*. https://doi.org/10.1002/cpp.2551
28. Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19. https://doi.org/10.1037/0022-006X.59.1.12
29. Jenkins, P. E., Smith, L., & Morgan, C. (2021). Can early change in eating disorder psychopathology predict outcome in guided self-help for binge eating?. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *26*, 2143-2152. https://doi.org/10.1007/s40519-020-01059-3
30. Jokic,Begic, N., Korajlija, A. L., Jurin, T., & Evans, C. (2014). Faktorska struktura, psihometrijske karakteristike i kritična vrijednost hrvatskoga prijevoda CORE-OM upitnika. *Psihologijske Teme, 23*(2), 265–288. https://hrcak.srce.hr/125093.
31. Juhová, D., Řiháček, T., Cígler, H., Dubovská, E., Saic, M., Černý, M., Dufek, J., & Evans, C. (2018). Česká adaptace dotazníku CORE-OM: vybrané psychometrické charakteristiky. *Ceskoslovenska Psychologie*, *62*(1), 59-74.http://cspsych.psu.cas.cz/result.php?id=1007
32. Juntunen, H., Piiparinen, A., Honkalampi, K., Inkinen, M. & Laitila, A. (2015). CORE OM mittarin suomalainen validointitutkimus yleisväestössä. *Psykologia*,*50*(4), 257–276. http://www.psykologia.fi/arkisto/vanhat-numerot/vanhat-numerot-ghost/351-2015-04/tieteelliset-artikkelit/418-core-om-mittarin-suomalainen-validointitutkimus-yleisvaestossa.
33. Kiziela, A., Viliūnienė, R., Friborg, O., & Navickas, A. (2019). Distress and resilience associated with workload of medical students. *Journal of Mental Health*, *28*(3), 319-323. https://doi.org/10.1080/09638237.2018.1521922
34. Kontunen, J., Timonen, M., Muotka, J., & Liukkonen, T. (2016). Is interpersonal counselling (IPC) sufficient treatment for depression in primary care patients? A pilot study comparing IPC and interpersonal psychotherapy (IPT). *Journal of Affective Disorders*, *189*(1), 89-93. https://doi.org/10.1016/j.jad.2015.09.032
35. Kristjánsdóttir, H., Sigurðsson, B. H., Salkovskis, P., Ólason, D., Sigurdsson, E., Evans, C., Gylfadóttir, E. D., & Sigurðsson, J. F. (2013). Evaluation of the psychometric properties of the Icelandic version of the Clinical Outcomes in Routine Evaluation–Outcome Measure, its transdiagnostic utility and cross-cultural validation. *Clinical Psychology & Psychotherapy, 22*(1), 64-74. https://doi.org/10.1002/cpp.1874
36. Lambert, M. J., Morton, J., Hatfield, D., Harmon, C., Hamilton, S. & Reid, R. C. (2004) *Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire) (3rd ed.).* American Professional Credentialing Services LLC.

37. Lambert, M. J. (2015). Outcome Research: methods for improving outcome in routine care. In: O. Gelo, A. Pritz, & B. Rieken (Eds.), *Psychotherapy Research: Foundations, process and outcome* (pp. 593-610). Springer. https://doi.org/10.1007/978-3-7091-1382-0_29
38. Londoño, N. H., Agudelo, D., Martínez, E., Anguila, D., Aguirre, D., & Mogollón, C. (2017). Cuestionario de efectividad de la psicoterapia Outcome Questionnaire. Validación en muestra clínica colombiana. *Psychologia*, *11*(1), 97-108. https://doi.org/10.21500/19002386.3105
39. Lyne, K. J., Barrett, P., Evans, C., & Barkham, M. (2006). Dimensions of variation on the CORE-OM. *The British Journal of Clinical Psychology*, *45*(2), 185-203. https://doi.org/10.1348/014466505x39106
40. Mækelæ, M. J., Reggev, N., Defelipe, R. P., Dutra, N., Tamayo, R. M., Klevjer, K., & Pfuhl, G. (2021). Identifying resilience factors of distress and paranoia during the COVID-19 outbreak in five countries. *Frontiers in Psychology*, *12*, 2160. https://doi.org/10.3389/fpsyg.2021.661149
41. Margison, F.R., Barkham, M., Evans, C., McGrath, G., Clark, J.M., & Audin, K. (2000). Measurement and psychotherapy evidence-based practice and practice-based evidence. *The British Journal of Clinical Psychology*, *177,* 123-130. https://doi.org/10.1192/bjp.177.2.123
42. Medlicott, E., Phillips, A., Crane, C., Hinze, V., Taylor, L., Tickell, A., Montero-Marin, J., & Kuyken, W. (2021). The Mental Health and Wellbeing of University Students: Acceptability, Effectiveness, and Mechanisms of a Mindfulness-Based Course. *International Journal of Environmental Research and Public Health*, *18*(11), 6023. https://doi.org/10.3390/ijerph18116023
43. Meerding, W. J., Van't Spijker, A., & Van Riessen, M. (2012). Monitoren van behandelresultaat met de CORE-OM. *Tijdschrift Psychotherapie*, *38*, 355–367. https://doi.org/10.1007/s12485-012-0041-x
44. Mellor-Clark, J., Connell, J., Barkham, M., & Cummins, P. (2001). Counselling outcomes in primary health care: a CORE system data profile. *European Journal of Psychotherapy & Counselling, 4,* 65–86. https://doi.org/10.1080/13642530110040118
45. Murray, A. L., McKenzie, K., Murray, K. R., & Richelieu, M. (2016). An analysis of the effectiveness of university counselling services. *British Journal of Guidance & Counselling*, *44*(1), 130-139. https://doi.org/10.1080/03069885.2015.1043621
46. Murray, A. L., McKenzie, K., Murray, K. R., & Richelieu, M. (2018). Examining response shifts in the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM). *British Journal of Guidance & Counselling*, *48*(2), 276-288. https://doi.org/10.1080/03069885.2018.1483007
47. Palmieri, G., Evans, C., Hansen, V., Brancaleoni, G., Ferrari, S., Porcelli, P., Reitano, F., & Rigatelli, M. (2009). Validation of the Italian version of the clinical outcomes in routine evaluation outcome measure (CORE-OM). *Clinical Psychology & Psychotherapy, 16*, 444-449. https://doi.org/10.1002/cpp.646
48. Paz, C., Adana-Díaz, L., & Evans, C. (2020). Clients with different problems are different and questionnaires are not blood tests: A template analysis of psychiatric and psychotherapy clients' experiences of the CORE-OM. *Counselling and Psychotherapy Research*, *20*(2), 274-283. https://doi.org/10.1002/capr.12290
49. Paz, C., Mascialino, G., & Evans, C. (2020). Exploration of the psychometric properties of the Clinical Outcomes in Routine Evaluation-Outcome Measure in Ecuador. *BMC Psychology, 8*, 94. https://doi.org/10.1186/s40359-020-00443-z

50. Paz, C., Mascialino, G., Proaño, K., & Evans, C. (2021). Psychological intervention change measurement in Latin America: Where from? Where to?. *Psychotherapy Research*, *31*(1), 132-141. https://doi.org/10.1080/10503307.2020.1750728
51. Peak, J., Goranitis, I., Day, E., Copello, A., Freemantle, N., & Frew, E. (2018). Predicting health-related quality of life (EQ-5D-5 L) and capability wellbeing (ICECAP-A) in the context of opiate dependence using routine clinical outcome measures: CORE-OM, LDQ and TOP. *Health and quality of life outcomes*, *16*(106), 1-11. https://doi.org/10.1186/s12955-018-0926-7
52. Rivas-Vazquez, R.A., Rivas-Vazquez, A., Blais, M.A., Rey, G.J., Rivas-Vazquez, F., Jacobo, M., & Carrazana, E. J. (2001). Development of a Spanish Version of the Schwartz Outcome Scale-10: A Brief Mental Health Outcome Measure. *Journal of Personality Assessment*, *77*(3), 436-446. https://doi.org/10.1207/S15327752JPA7703_05
53. Ronconi, L., Giannerini, N., Testoni, I., Zulian, M., Guglielmin, MS. A. (2018). A structural model of well-being, spontaneity and self-efficacy: Italian validation between adolescents and young adults. *Trends in Psychiatry and Psychotherapy*, *40*(2), 136-143. http://dx.doi.org/10.1590/2237-6089-2017-0072
54. Saini, P., Chopra, J., Hanlon, C. A., & Boland, J. (2021). A Case Series Study of Help-Seeking among Younger and Older Men in Suicidal Crisis. *International Journal of Environmental Research and Public Health. 18*, 14, 7319. https://doi.org/10.3390/ijerph18147319
55. Sales, C. M. D., Moleiro, C. M. M., Evans, C., & Alves, P. C. G. (2012). Versão Portuguesa do CORE-OM: tradução, adaptação e estudo preliminar das suas propriedades psicométricas. *Revista de Psiquiatria Clínica*, *39*(2), 54–59. https://doi.org/10.1590/S0101-60832012000200003
56. Sales, C. M., Neves, I. T., Alves, P. G., & Ashworth, M. (2018). Capturing and missing the patient's story through outcome measures: A thematic comparison of patient-generated items in PSYCHLOPS with CORE-OM and PHQ-9. *Health Expectations*, *21*(3), 615-619. https://doi.org/10.1111/hex.12652
57. Santana, M. R. M., Silva, M. M. da, Moraes, D. S. de, Fukuda, C. C., Freitas, L. H., Ramos, M. E., Fleury, H. J., & Evans, C. (2015). Brazilian Portuguese version of the CORE-OM: cross-cultural adaptation of an instrument to assess the efficacy and effectiveness of psychotherapy. *Trends in Psychiatry and Psychotherapy*, *37*(4), 227-231. https://doi.org/10.1590/2237-6089-2015-0002
58. Skre, I., Friborg, O., Elgarøy, S., Evans, C., Myklebust, L. H., Lillevoll, K., Sørgaard, K., & Hansen, V. (2013). The factor structure and psychometric properties of the Clinical Outcome in Routine Evaluation–Outcome Measure (CORE-OM) in Norwegian clinical and non clinical samples. *BMC Psychiatry, 13*,99. https://doi.org/10.1186/1471-244X-13-99
59. Sosa-Torralba, J. E., Romero-Mendoza, M. P., Medina-Mora-Icaza, M. E., Méndez-Ríos, E., & Blum-Grynberg, B. (2020) Estructura dimensional de la Escala Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) en universitarios mexicanos. *Revista Evaluar, 20*(1), 49-66. https://revistas.unc.edu.ar/index.php/revaluar
60. Stiles, W. B., Barkham, M., & Wheeler, S. (2015). Duration of psychological therapy: Relation to recovery and improvement rates in UK routine practice. *The British Journal of Clinical Psychology*, *207*(2), 115-122. https://doi:10.1192/bjp.bp.114.145565

61. Stroud, J., & Griffiths, C. (2021). An evaluation of compassion-focused therapy within adult mental health inpatient settings. *Psychology and Psychotherapy: Theory, Research and Practice*, *94*(3), 587–602. https://doi.org/10.1111/papt.12334

62. Trujillo, A., Feixas, G., Bados, A., García-Grau, E., Salla, M., Medina, J. C., Montesano, A., Soriano, J., Medeiros-Ferreira, L., Cañete, J., Corbella, S., Grau, A., Lana, F., & Evans, C. (2016). Psychometric properties of the Spanish version of the Clinical Outcomes in Routine Evaluation - Outcome Measure. *Neuropsychiatric Disease and Treatment*, 12, 1457–1466. https://doi.org/10.2147/NDT.S103079

63. Uji, M., Sakamoto, A., Adachi, K., & Kitamura, T. (2012). Psychometric properties of the Japanese version of the Clinical Outcomes in Routine Evaluation–Outcome Measure. *Comprehensive Psychiatry*, *53*(5), 600-608. https://doi.org/10.1016/j.comppsych.2011.09.006

64. Viliū nienė, R., Evans, C., Hilbig, J., Pakalniškienė, V., Danilevičiūtė, V., Laurinaitis, E., & Navickas, A. (2013). Translating the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM) into Lithuanian. *Nordic Journal of Psychiatry*, *67*(5), 305-311. https://doi.org/10.3109/08039488.2012.745599

65. Wickramasekera, N., & Tubeuf, S. (2020). Measuring quality of life for people with common mental health problems. *Journal of Mental Health*. https://doi.org/10.1080/09638237.2020.1818190

66. Zeldovich, M., & Alexandrowicz, R. (2019). Comparing outcomes: The Clinical Outcome in Routine Evaluation from an international point of view. *International Journal of Methods in Psychiatric Research*, *28*(3), e1774. https://doi.org/10.1002/mpr.1774

67. Zibetti, M. R., Serralta, F. B., & Evans, C. (2021). Longitudinal Distress among Brazilian University Workers during Pandemics. *International Journal of Environmental Research and Public Health*, *18*(17), 9072. https://doi.org/10.3390/ijerph18179072

DOI: 10.13129/2282-1619/mjcp-3056