Neuropsychiatric Disease and Treatment

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Psychometric properties of the Spanish version of the Clinical Outcomes in Routine Evaluation – Outcome Measure

This article was published in the following Dove Press journal:
Neuropsychiatric Disease and Treatment
21 June 2016
Number of times this article has been viewed

Adriana Trujillo,[1,2] Guillem Feixas,[1,2] Arturo Bados,[1] Eugeni García-Grau,[1] Marta Salla,[1] Joan Carles Medina,[1] Adrián Montesano,[1,2] José Soriano,[3] Leticia Medeiros-Ferreira,[4] Josep Cañete,[5] Sergi Corbella,[6] Antoni Grau,[7] Fernando Lana,[8] Chris Evans[9]

[1]Department of Personality, Assessment and Psychological Treatments, Faculty of Psychology, [2]Institute for Brain, Cognition and Behaviour, University of Barcelona, [3]Hospital of the Holy Cross and Saint Paul, [4]Nou Barris Mental Health Center, Barcelona, [5]Hospital of Mataró, Sanitary Consortium of Maresme, Mataró, [6]FPCEE, Blanquerna, Universitat Ramon Llull, [7]Institute of Eating Disorders, Barcelona, [8]MAR Health Park, CAEMIL, Santa Coloma de Gramenet, Spain; [9]East London NHS Foundation Trust, NPDDNet, London, UK

Correspondence: Guillem Feixas
Department of Personality, Assessment and Psychological Treatments, Faculty of Psychology, University of Barcelona, Passeig Vall d'Hebron, 171, 08035 Barcelona, Spain
Tel +34 93 312 5100
Email gfeixas@ub.edu

Objective: The objective of this paper is to assess the reliability and validity of the Spanish translation of the Clinical Outcomes in Routine Evaluation – Outcome Measure, a 34-item self-report questionnaire that measures the client's status in the domains of Subjective well-being, Problems/Symptoms, Life functioning, and Risk.

Method: Six hundred and forty-four adult participants were included in two samples: the clinical sample (n=192) from different mental health and primary care centers; and the nonclinical sample (n=452), which included a student and a community sample.

Results: The questionnaire showed good acceptability and internal consistency, appropriate test–retest reliability, and acceptable convergent validity. Strong differentiation between clinical and nonclinical samples was found. As expected, the Risk domain had different characteristics than other domains, but all findings were comparable with the UK referential data. Cutoff scores were calculated for clinical significant change assessment.

Conclusion: The Spanish version of the Clinical Outcomes in Routine Evaluation – Outcome Measure showed acceptable psychometric properties, providing support for using the questionnaire for monitoring the progress of Spanish-speaking psychotherapy clients.

Keywords: CORE-OM, outcome measure, reliability, validity, psychometric validation

Introduction

This paper reports the psychometric properties of the Spanish version of Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM). The CORE-OM was designed mainly for practice-based evidence (a complement to evidence-based practice).[1] We expected that the Spanish translation of the CORE-OM would be a useful, reliable, and valid instrument suitable to be widely used for research and practice in Spain and in some countries in which Spanish is spoken in similar form to that used in Spain. The translation should also prove a useful base, with the original English version, for countries where local Spanish usage is sufficiently different from that in Spain that a somewhat different translation will be needed.

From its origin, the measure was designed to be pan-theoretical (not associated with a school of therapy) and pan-diagnostic (not focused on a single presenting problem), and was driven by what practitioners and clients considered to be the most important generic aspects of psychological well-being, and change in therapies, to be measured. It is recommended to be used before and at the end of therapy.

The CORE-OM measure is copyleft; that is, it can be reproduced without payment of any license fee if it is not changed in any way.[2] Translations were done following the

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/NDT.S103079

CORE System Trust (CST) protocol, and with the supervision and guidance of Chris Evans (CE). Copyright violations are illegal, but CST and CE welcome collaboration on new translations to the protocol. All CORE instruments are available to download from,[3] which provides more information about the system, instruments, and translation protocol. Information focused on the CORE-OM in Spanish is at.[4]

There are many fields in which CORE-OM has demonstrated its utility having been used in areas as varied as benchmark studies,[5,6] assessment of outcome of psychological therapies in primary and secondary settings,[7–10] studies of treatment processes,[11–13] assessment of the psychological well-being of individuals in nonclinical occupational settings,[14] and examination of psychological health among university students who were receiving university counseling.[15,16] Acceptability and psychometric properties have been demonstrated with diverse samples, for example, older people and patients with eating disorders.[17,18] Though designed more for practice-based evidence, the CORE-OM has been used in randomized controlled trials.[19–21]

CORE-OM has been translated, following a clear and thorough protocol,[22] into over 20 languages, with that number growing. Evaluation of the psychometric properties of the translated measure has been completed showing comparable psychometric properties to that found for the English version in the UK for a growing number of languages including Italian,[23] Portuguese,[24] Swedish,[25] Lithuanian,[26] Icelandic,[27] and Croatian,[28] and many others are nearing completion including a Catalan version. All forms are widely used as a routine change measure in a range of health care settings in the UK and increasingly in other languages and countries.[13,29–32]

This study was designed to assess the psychometric properties of the Spanish translation of the CORE-OM, and hence its suitability to be used in routine assessment of mental health interventions in Spain and perhaps other Spanish-speaking countries.

Method

The Spanish version

For the translation and adaptation of the CORE-OM to Spanish, we followed the steps established by international groups and the CST protocol including participation of a member of the group who designed the instrument (CE).[33] This process is congruent to the guidelines of the International Test Commission,[34] and it emphasizes the importance of translating the items according to their contextualized meaning in culture and the environment in which they will be applied, as well as making them understandable for the most varied range of possible potential users. It does not rely excessively on back-translation in order to avoid too literal translations. To seek for improvement of the resulting version, we requested the collaboration of 12 people from different parts of Spain, selected because of their high level of English proficiency. Ten of them responded to the request by providing a translation. Six of them were professionals in psychology, and four were lay people. With this material, a working session was organized with the participation of two of the professionals of psychology and two of the lay people who collaborated with the translation, along with a member of the CST (CE) who acted as a consultant or supervisor. In this session, each item was discussed taking into account the available translations. For each item, the best option was chosen by consensus. A first draft came out of this process which was reviewed by three experts in psychology with over 20 years of experience in clinical settings who made some modifications that were discussed by email with CE. This revised version was submitted to extensive scrutiny by a group of 64 people (between 16 and 76 years, all of them from different conditions and linguistic backgrounds, and fully proficient in Spanish, 12 of them were professionals of psychology, and 52 lay people) who were asked to read it carefully and to judge whether the items were understandable and clear. They were also encouraged to make all the comments they deemed appropriate with regard to the way items were written.

Afterward, the comments and observations made were discussed by the three experts mentioned, and issues that seemed to need discussion of the original English were shared with CE, until a final version was achieved. This version was delivered to an experienced, bilingual English–Spanish translator, with a degree in psychology and without access to the original version, to back-translate. Looking at this back-translation, neither the experts nor the member of the CST considered necessary to make any modification of the latest version, at which point it became the CST-approved translation into Spanish.

From that version,[33] the shorter versions (designed for routine use in therapy sessions, for screening and ongoing monitoring: CORE-SFA, CORE-SFB, CORE-10, and CORE-5, all in male and female versions[35]) were typeset and made available through webpages, initially[4] and now also.[22]

Participants and procedures

The study protocol was approved by the Bioethics Committee of the University of Barcelona (ref. IRB0003099) and by the ethical committees of the centers taking part in the study.

All the participants were informed of the implications of the study and signed an informed consent document before enrolling. The study included 644 adult participants in two samples (Table 1). The clinical sample (n=192) comprised patients from nine mental health centers and from some primary care centers in the Barcelona area. The CORE-OM was included in the routine pretreatment assessment of these centers, and it is this routine clinical data that are reported on in this paper. All patients who were referred for psychological treatment between March 2012 and May 2013 in the centers collaborating in the study were included in the study. Professionals were asked to exclude from these referrals inpatients and outpatients with severe psychological disorders. Another exclusion criterion was insufficient linguistic competence to communicate in Spanish.

The nonclinical sample (n=452) included a student and a community nonstudent sample, between 18 and 70 years of age (inclusion criteria) who were assessed in the same period from March 2012 to May 2013 and had sufficient linguistic competence to communicate in Spanish. The latter (n=127) consisted of volunteers and/or their relatives who were not receiving psychological treatment (exclusion criterion). The student sample (n=325) was drafted from the Faculty of Psychology of the University of Barcelona; 219 were undergraduate students from four different subject areas, and 106 were master-level students.

Forty-six participants of the community sample and 32 of the student sample agreed to take part in the test–retest survey completing the questionnaire twice; this second administration of the questionnaire took place between 15 and 30 days after the first one. For student, test and retest were made in their classrooms with a 2-week test–retest interval; for the community sample, all the participants who completed the first assessment were contacted by phone ~2 weeks later and were invited to participate in the retest survey. For those who accepted, the questionnaire was sent in an envelope, and they completed and returned it. Test–retest stability was not measured in the clinical sample as that would have involved significant interference with normal clinical management of these participants. This was in line with the UK original study where there was no test–retest stability examination in the clinical sample for the same reason.[36]

Instruments and measures

CORE-OM is a 34-item self-report questionnaire that assesses the client's status in the domains of Subjective well-being (four items), Problems/Symptoms (12 items), Functioning (12 items), and Risk (six items).[36,37] Eight of the items are positively cued (items 3, 4, 7, 12, 19, 21, 31, and 32). The focus is on the last 7 days, and items are scored in a five-point scale ranging from 0 (not at all) to 4 (most or all the time), where higher scores on all domains indicate more problems and high levels of psychological distress even for the Subjective well-being scale. The domains were named to designate their item content but never envisaged to be psychometric factors.[35,38] The Subjective well-being domain comprises four items capturing this aspect. The Problems/Symptom domain includes four items addressing anxiety, four for depression, and two each for physical problems and trauma. The Functioning domain includes four items covering general/work functioning, four addressing close relationships, and four for social functioning. The Risk domain has four items about risk to self and two about risk to others.

The CORE-OM was designed to be user-friendly for both clients and practitioners.[35,39] It takes 5–10 minutes to complete, and the total and domain scores are reported as means across items. Prorating, that is, using the item mean even with missing items, is recommended as long as $<10\%$ of the items in the score are missing.[36]

Psychometric properties were excellent in the original UK testing and in all subsequent explorations showing high internal consistency (Cronbach α between 0.75 and 0.94 for all scores, the lowest for Risk) and test–retest stability

Table 1 Demographic data

Sample	Total (missing data for sex)	Females (%)	Males (%)	Mean age (SD)	Age range (years)
Nonclinical sample	452 (15)	343 (75.9)	94 (20.8)	29.3 (14.4)	18–76
Students	310 (15)	250 (76.9)	60 (18.5)	23.2 (6.1)	18–69
Community sample	127 (0)	93 (73.2)	34 (26.8)	44.4 (17.6)	20–76
Clinical sample (outpatients)	192 (1)	130 (67.7)	61 (31.8)	41.3 (14.9)	18–78
Primary care	44 (0)	29 (65.9)	15 (34.1)	41.8 (12.7)	22–76
Secondary care	147 (1)	101 (68.2)	46 (31.1)	41.1 (15.5)	18–78
Test–retest sample	78 (0)	54 (69.2)	24 (30.8)	34.9 (18.8)	18–69
Students	32 (0)	26 (81.3)	6 (18.8)	20.7 (3.8)	18–34
Community sample	46 (0)	28 (60.9)	18 (39.1)	44.8 (18.8)	20–69

Abbreviation: SD, standard deviation.

of 0.91 (Spearman's ρ 0.91 for 1-week test–retest in a student sample). Discriminant validity showed large differences between clinical and nonclinical samples (Cohen's *d* from 0.71 Risk to 1.77 Problems/Symptoms) and high correlations with measures which are conceptually close, for example, Beck Depression Inventory-II (BDI-II) (ρ=0.85) and Symptom Checklist 90 Revised (SCL-90-R) (ρ=0.88). The CORE-OM is also sensitive to change in therapies.[15,36] As expected, the domains did not show neat factorial separation, but an oblique structure in which Risk items are clearly separated from other items and two strongly correlated main problem dimensions of the positively and the negatively cued items gave a moderate and just acceptable fit on confirmatory factor analysis.[38]

BDI-II is a 21-item self-administered inventory designed to measure the intensity of depressive symptoms in psychiatric and nonpsychiatric populations of both adults and adolescents.[40] Items are rated on a four-point scale (0–3), and total scores are obtained by tallying the ratings for all 21 items. Scores range from 0 to 63, with higher scores reflecting increased depressive severity. The BDI-II requires ~5–10 minutes to complete and may be administered to individuals 13–80 years of age. We used the Spanish-language version of the BDI-II.[41]

SCL-90-R is a 90-item self-report symptom inventory designed to screen for a broad range of psychological problems.[42] Each of the 90 items is rated on a five-point Likert scale of distress, ranging from "not at all" (0) to "extremely" (4). Subsequently, the answers are combined in nine primary symptom dimensions: Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Hostility, Depression, Anxiety, Paranoid Ideation, Phobic Anxiety, and Psychoticism. In addition, three global indices provide measures of overall psychological distress: the Global Severity Index, the Positive Symptom Total, and the Positive Symptom Distress Index. We used the Spanish-language version of the SCL-90-R.[43]

Analyses

To facilitate a comparison with the UK data, we followed the original study by assessing acceptability, internal consistency, test–retest reliability (with 15- to 30-day interval), influence of age and sex, correlations between domain scores, and discriminant validity against sample, reflected in the differences between clinical and nonclinical sample, along with the calculation of cutoff scores, and convergent validity in terms of the correlations between CORE-OM's scores and those on the BDI-II and SCL-90-R.[36] Following the UK study, most analyses were reported for each of the four content domains (Subjective well-being, Problems/Symptoms, Life/social functioning, and Risk) as well as for total scale, and for score of all items except those in the Risk domain. Internal reliability was reported as Cronbach's α for the subsample with no missing item data,[44] but results for domain scores were reported where a score could be computed by prorating up to 10% of missing items. To test the equality for the different coefficients in the samples and subsamples, a Felt's procedure was done.[45] Again following the UK validation study, nonparametric correlation coefficients (Spearman's ρ) and nonparametric tests of differences in central location of distributions (Wilcoxon test) were used as scores did not conform to Gaussian distributions. The BDI-II[40,41] and the SCL-90-R[42,43] were used to test convergent validity with other self-report measures. Clinically significant change was calculated according to the *c* criterion that uses a cutoff point based on the contrast between dysfunctional and general population samples.[46] Analyses were conducted using SPSS, version 20.0. As in the original paper, the methodology was mainly exploratory and descriptive rather than one of null hypothesis testing; wherever possible, 95% confidence intervals (CIs) were reported rather than *P*-values. This gave a test approximating to testing for $P<0.05$. Comparisons of parameters within this sample and against those reported in the UK data were generally informed in terms of overlap or not of CIs.[15]

This paper did not follow the original UK analysis in including a principal component analysis, as subsequent UK papers have shown that the CORE-OM, as its authors expected, has a complicated factor structure that would need larger clinical and nonclinical samples for the Spanish data than we have to date.[6] More psychometric exploration will be reported later when such significantly larger samples are available.

Results

Acceptability

All of the questionnaires have sufficiently few items missing to allow prorating for a usable overall score (ie, no participant omitted more than three items). One hundred and seventy-nine (93.2%) participants of the clinical and 432 (95.6%) of the nonclinical samples returned completed data. The overall omission rate was 0.17%. The items that were most often incomplete were items 3 (0.7%) and 25 (0.7%) in the nonclinical and items 21 (1%) and 32 (1%) in the clinical sample.

Table 2 Coefficient α (95% CI) denoting internal consistency for nonclinical and clinical samples

Domains	Nonclinical samples		Clinical samples		Pooled nonclinical samples (n=452)	Pooled clinical samples (n=192)
	Students (n=325)	Community (n=127)	Primary care (n=44)	Secondary care (n=148)		
Subjective well-being	0.80 (0.76, 0.83)	0.80 (0.74, 0.85)	0.79 (0.67, 0.88)	0.81 (0.75, 0.85)	0.80 (0.77, 0.83)	0.81 (0.76, 0.85)
Problems/Symptoms	0.88 (0.86, 0.90)	0.85 (0.81, 0.89)	0.86 (0.80, 0.91)[a]	0.90 (0.88, 0.92)[a]	0.88 (0.86, 0.90)	0.90 (0.87, 0.91)
Functioning	0.86 (0.84, 0.88)	0.84 (0.80, 0.88)	0.82 (0.73, 0.89)	0.86 (0.83, 0.89)	0.86 (0.84, 0.88)	0.85 (0.82, 0.88)
Risk	0.73 (0.68, 0.77)[b]	0.60 (0.48, 0.70)[b]	0.80 (0.68, 0.87)	0.76 (0.70, 0.82)	0.71 (0.66, 0.75)[c]	0.77 (0.71, 0.82)[c]
Nonrisk items	0.94 (0.93, 0.95)	0.93 (0.91, 0.95)	0.92 (0.88, 0.95)	0.85 (0.93, 0.96)	0.94 (0.93, 0.95)	0.94 (0.93, 0.95)
All items	0.94 (0.93, 0.95)[b]	0.92 (0.90, 0.94)[b]	0.93 (0.90, 0.95)	0.95 (0.93, 0.96)	0.94 (0.93, 0.95)	0.94 (0.93, 0.95)

Notes: [a]$P<0.05$ (significantly higher α in the secondary care sample in comparison with primary care sample). [b]$P<0.05$ (significantly higher α in the students sample in comparison with the community sample). [c]$P<0.05$ (significantly higher α in the clinical sample in comparison with the nonclinical sample).
Abbreviation: CI, confidence interval.

Internal consistency

To evaluate the internal reliability, we calculated Cronbach's α,[44] for all domains and the entire scale for the clinical and nonclinical groups. Furthermore, to test if the differences between these coefficients were statistically significant, we followed the procedure proposed by Feldt et al.[45] All domains showed an appropriate internal reliability in both samples. The levels were within the acceptable range, although being lower for the Risk domain (Table 2 and Figure 1).

In comparison with the UK referential data, the pooled clinical and nonclinical α values for all items and all nonrisk items showed tight 95% CIs covering the UK referential values, and when the clinical and nonclinical samples were pooled, the lower confidence limit (CL) was above that for the UK data. For Subjective well-being, the Spanish α was above the UK one; for Problems/Symptoms, the clinical sample α had a CI covering the UK one, and the nonclinical α was slightly lower than the UK nonclinical value with the upper CL below the UK value; for Functioning, the CIs included the UK referential values. The values for the Risk domain were lower than the UK ones (which were the same for clinical and nonclinical samples at 0.79), though the CI for the combined clinical sample included 0.79.

Test–retest stability

Test–retest correlations were strong within domains in the nonclinical data (Table 3). The stabilities for all domains were satisfactory (range: 0.76–0.87), except for the Risk domain (0.45) reflecting the high rate of zero responses in answering these items in the nonclinical group. Changes of mean values between first and second survey were not significant for all scores.

Convergent validity

Correlations between domain scores and the BDI-II and the SCL-90-R were calculated (Table 4). Across domain scores, correlations were highest against conceptually close measures showing an acceptable convergent validity. The pattern and the correlations were generally very similar to the UK findings,[36] although the Spanish correlations between the Risk scores and the BDI-II and SCL-90R were lower than the UK ones.

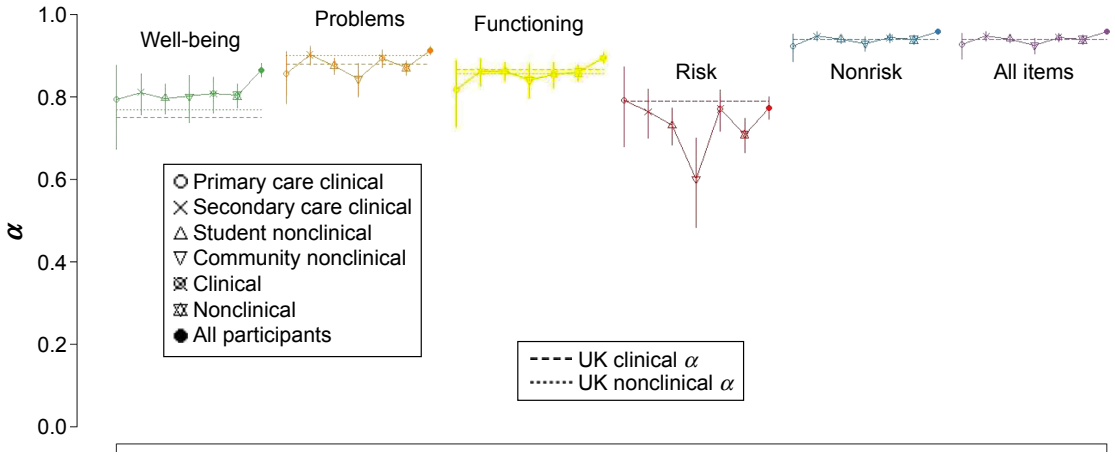

Figure 1 Forest plot showing comparison between Spanish scores and UK referential data.

Table 3 Test–retest stability and changes of mean values between first and second survey in a nonclinical sample (n=78)

Domains	Test–retest stability[a]	Change		
		Mean	95% CI	*P*-value[b]
Subjective well-being	0.76	–0.013	–0.13, 0.10	0.80
Problems/Symptoms	0.85	0.045	–0.03, 0.12	0.47
Functioning	0.79	0.045	–0.02, 0.12	0.16
Risk	0.45	0.008	–0.04, 0.04	0.90
Nonrisk items	0.87	0.037	–0.02, 0.10	0.36
All items	0.87	0.030	–0.02, 0.08	0.43

Notes: [a]Rho Spearman correlation. [b]Wilcoxon test.
Abbreviation: CI, confidence interval.

Differences between clinical and nonclinical samples

There were significant differences between clinical and nonclinical samples in all domains (Table 5) with higher scores for the clinical sample than the nonclinical one. With the exception of the Problem/Symptoms domain, the effect sizes of the differences were similar to the results of the UK study with CIs including the UK referential values.[36] At 1.4 (CI 1.22–1.59), the effect size for the Problem/Symptoms score is lower than the UK referential that was 1.7 but remains respectable as discriminant validity against the clinical/nonclinical distinction. As in the UK data, the effect size of the difference for the Risk score at 0.8 was smaller than for all the other scores, actually higher than that in the UK data (0.7) but with the CI including the UK value.

The box plot in Figure 2 shows no patients in the clinical sample scoring zero and a very few patients (outliers) in the nonclinical sample scoring very highly. The box for the one sample and the median line bisecting the box for the other sample do not overlap.

Table 4 Correlations with referential measures in clinical samples

Samples	n	Domains					
		W	P	F	R	–R	All
Primary care (present study)							
BDI-II	39	0.76	0.75	0.65	0.32	0.78	0.74
SCL-90-R	30	0.66	0.56	0.58	0.10	0.64	0.61
Secondary care (present study)							
BDI-II	123	0.80	0.82	0.77	0.55	0.85	0.85
SCL-90-R	125	0.70	0.81	0.75	0.51	0.82	0.82
Pooled clinical samples (present study)							
BDI-II	162	0.79	0.80	0.74	0.48	0.83	0.83
SCL-90-R	155	0.70	0.77	0.72	0.46	0.79	0.79
Clinical sample (Evans et al[36])							
BDI-II	29	0.79	0.74	0.78	0.32	0.83	0.81
SCL-90-R	34	0.68	0.87	0.79	0.83	0.85	0.88

Abbreviations: W, subjective well-being; P, problems/symptoms; F, functioning; R, risk; BDI-II, Beck Depression Inventory-II; SCL-90-R, Symptom Checklist-90-Revised.

Sex and age differences

In the nonclinical sample, age was significantly and negatively related with all domain scores except Risk: Subjective well-being (ρ=–0.25, $P<0.001$), Problems/Symptoms (P=–0.23, $P<0.001$), and Functioning (ρ=–0.18, $P<0.001$); nevertheless, those relationships were weak. In the clinical sample only, the Functioning domain showed a significant correlation with age (ρ=–0.19, P=0.006), and again, this relationship was weak. Regarding sex, only the Subjective well-being domain showed a statistical difference between men and women in both samples with a small effect size (Table 6).

Correlations between domain scores

Table 7 shows, as expected, significant and generally strong correlations between all domains. However, correlations between Risk domain scores and the other scores were lower, especially in the nonclinical sample.

Clinically significant change

Values for clinical significant change were calculated for all domains following the *c* criterion which takes into account data from both clinical and nonclinical samples.[46] Cutoff scores (Table 8) separate typical clinical and nonclinical populations and will help to identify the extent to which change after treatment is clinically meaningful.

Discussion

To the extent that these psychometric analyses of these data from the Spanish version of the CORE-OM are good or acceptable, the translation is supported for use in Spanish-speaking populations.

Regarding acceptability, considered as the number of missing items and unusable measures, the results were excellent compared to those obtained in the original English-language test.[37] In our study, the percentage of complete item responses was higher for both the clinical and the nonclinical sample than in the initial UK testing, which could be taken as an evidence not only for the proper design of the questionnaire but also for the quality of the translation process carried out to adapt this instrument into Spanish.[33]

These results are consistent with other studies of validation such as the Italian, where the percentages of item response (96% for the clinical sample and 81% for nonclinical sample)[23] are comparable or lower than those observed in the current study (93.2% for the clinical and 95.6% for the nonclinical sample). Similarly, the results from Sweden have

Table 5 Mean and standard deviations for clinical and nonclinical samples

Domains	Present study						Evans et al*					
	Nonclinical (n=452)		Clinical (n=192)		95% CI		Nonclinical (n=1,084)		Clinical (n=863)		95% CI	
	Mean	SD	Mean	SD	Difference	*d*[a]	Mean	SD	Mean	SD	Difference	*d*[a,b]
Subjective well-being	1.18	0.76	2.41	0.95	1.08, 1.36	1.5 (1.31, 1.68)	0.91	0.83	2.37	0.96	1.38, 1.53	1.6 (1.54, 1.74)
Problems/Symptoms	0.99	0.62	1.98	0.87	0.86, 1.10	1.4 (1.22, 1.59)	0.90	0.72	2.31	0.88	1.33, 1.48	1.7 (1.67, 1.88)
Functioning	0.74	0.52	1.56	0.75	0.71, 0.92	1.3 (1.19, 1.55)	0.85	0.65	1.86	0.84	0.95, 1.09	1.3 (1.26, 1.46)
Risk	0.11	0.27	0.48	0.66	0.29, 0.44	0.8 (0.69, 1.04)	0.20	0.45	0.63	0.75	0.38, 0.49	0.7 (0.62, 0.81)
Nonrisk items	0.91	0.55	1.86	0.78	0.84, 1.05	1.5 (1.32, 1.70)	0.88	0.66	2.12	0.81	1.18, 1.31	1.7 (1.59, 1.80)
All items	0.77	0.48	1.62	0.71	0.75, 0.94	1.5 (1.33, 1.71)	0.76	0.59	1.86	0.75	1.04, 1.16	1.6 (1.55, 1.76)

Notes: [a]Cohen effect size parameter. [b]Cohen's *d* has been calculated with the data provided at UK study.[36] *Reproduced with permission from Evans C, Connell J, Barkham M, et al. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *Br J Psychiatry*. 2002;180:51–60.[36] Available from: http://bjp.rcpsych.org/content/180/1/51.long.

Abbreviations: CI, confidence interval; SD, standard deviation.

an omission rate of 0.44% of items,[25] compared with 0.17% in our study. There are no patterns regarding specific items in which omissions occurred, indicating that there appears to be no connection to any specific dimension.

Considering reliability, the results are acceptable and consistent with the analysis made in other studies of adaptation and validation,[25–27] as well as with the original UK data. In all of these translations, including the present study, some differences in the internal consistency between clinical and nonclinical samples were identified; however, in all domains, the α value was between 0.7 and 0.9, which means that the reliability of the CORE-OM in Spanish has resulted as satisfactory as in other versions. α was lowest for the Risk domain, at 0.71 for the pooled nonclinical sample, lower than the observed value of 0.79 in the UK validation study (CI 0.77–0.81). It seems likely that this difference arises because the Risk items are tuned to catch mostly only quite significant levels of Risk, giving floor effects that curtail variance in nonclinical samples. It seems possible that both, the larger size of the UK nonclinical sample compared to that reported here and perhaps a higher rate of Risk to self in the UK populations, where, particularly in young adults, self-harm may be more prevalent than in some other countries probably including Spain,[47,48] may have led to more inter-item covariance appearing in this score due to floor effects rather than necessarily to much lower population covariances.

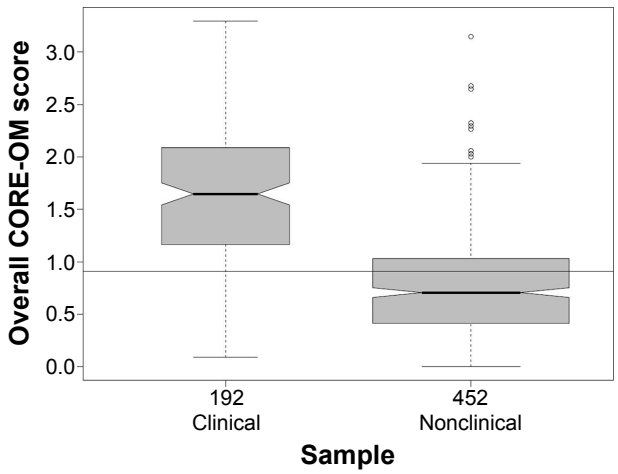

Figure 2 Box plot of mean item score for all items for clinical and nonclinical samples.

Abbreviation: CORE-OM, Clinical Outcomes in Routine Evaluation - Outcome Measure.

Test–retest stability in our study was good with the exception of the Risk domain score, which again is likely to be explained by its small length, floor effects, and the intrinsically impulsive (and thus unstable) nature of some of the phenomena addressed by these items. Stability correlations were strong but slightly lower than in the UK study,[36] which is consistent with other results such as the Icelandic data.[27]

Regarding convergent validity, correlations between the domain scores of the CORE-OM and the BDI-II and SCL-90-R were strong except again for the Risk scores, which is consistent with the original UK data.[36] In different studies, the CORE-OM has shown satisfactory convergent validity with other conceptually close measures which supports its value as a wider general measure for psychotherapy outcome assessment.[23,25,27,49]

Comparative analysis showed significant differences between clinical and nonclinical population in all domains, as in other validation studies, demonstrating discriminant validity across different countries and languages. The effect size (Cohen's *d*) values were large for all domains.

As in the original UK data, small but statistically significant correlations between scores and age were found in our study, more so in the nonclinical than the clinical samples. These seem likely to be genuine demographic associations, but the small effect size illustrates that age does not strongly and systematically contaminate scores. However,

Table 6 Sex differences in scores for clinical and nonclinical samples

Domains	Nonclinical						Clinical					
	Male (n=94)		Female (n=343)		95% CI		Male (n=61)		Female (n=130)		95% CI	
	Mean	SD	Mean	SD	Difference	*d*[a]	Mean	SD	Mean	SD	Difference	*d*[a]
W	0.95	0.77	1.23	0.76	−0.46, −0.11	−0.37 (−0.60, −0.14)	2.15	1.03	2.51	0.88	−0.65, −0.08	−0.39 (−0.69, −0.08)
P	0.96	0.68	0.99	0.60	−0.18, 0.10	−0.05 (−0.28, 0.18)	1.83	0.92	2.05	0.85	−0.48, 0.04	−0.25 (−0.56, 0.05)
F	0.70	0.57	0.74	0.50	−0.15, 0.08	−0.08 (−0.31, 0.15)	1.57	0.80	1.55	0.73	−0.21, 0.25	0.03 (−0.28, 0.33)
R	0.13	0.28	0.10	0.27	−0.03, 0.08	0.11 (−0.12, 0.34)	0.53	0.78	0.44	0.59	−0.11, 0.29	0.14 (−0.17, 0.44)
All – R	0.85	0.60	0.92	0.54	−0.19, 0.05	−0.13 (−0.35, 0.10)	1.77	0.83	1.90	0.75	−0.37, 0.10	−0.17 (−0.47, 0.14)
All	0.72	0.52	0.77	0.47	−0.16, 0.05	−0.10 (−0.33, 0.12)	1.56	0.78	1.64	0.68	−0.30, 0.13	−0.11 (−0.42, 0.19)

Note: [a]Cohen effect size parameter.
Abbreviations: CI, confidence interval; SD, standard deviation; W, Subjective well-being; P, Problems/Symptoms; F, Functioning; R, Risk.

the majority of participants in the nonclinical sample were students (72%) with a very different age mean and range from clinical population. Thus, larger replication studies with more diverse nonclinical samples are needed to ascertain the generalizability of these differences. Furthermore, a community sample of persons who exceed pensionable age, almost absent in these samples, would indicate whether specific norms are needed for older populations.[17]

In the analysis of sex differences in mean scores, only Subjective well-being domain showed a statistical difference between men and women in both samples, with a small effect size in the same direction as the results analyzed in the UK version.[15] According to the UK authors, sex should be considered in the interpretation of individual data regardless of clinical or nonclinical condition. In the Swedish and Italian studies, sex differences were very similar to those found in the UK study.[23,25,36] However, it seems highly plausible that there will be sex effects, which may be culture specific.

Table 7 Correlations between Spearman's ρ values for clinical and nonclinical samples

Domains	W	P	F	R	All – R
Nonclinical (n=452)					
W					
P	0.79				
F	0.77	0.75			
R	0.33	0.39	0.40		
All – R	0.89	0.94	0.91	0.41	
All	0.88	0.94	0.91	0.45	0.99
Clinical (n=192)					
W					
P	0.85				
F	0.71	0.76			
R	0.51	0.56	0.57		
All – R	0.89	0.95	0.90	0.60	
All	0.88	0.94	0.90	0.67	0.99

Abbreviations: W, subjective well-being; P, problems/symptoms; F, functioning; R, Risk.

The strong and positive correlations between the domain scores are expected because the items of the CORE-OM are designed to evaluate related aspects of psychological distress, and the correlations found in this study are not dissimilar from those in all explorations to date with the only scale showing low correlations with respect to the others being Risk.[5,23,25–27] This corroborates the special characteristics of the Risk domain,[38] defined as an oblique factorial scale with fairly low positive correlation with the other items and domains illustrating that Risk issues are, generally, rather distinct from other aspects of psychological distress domains. The items were designed as much to provide flags of Risk more than to form a robust scale, while ensuring that the crucial issue of Risk would contribute to the overall score, in contrast to many measures which omit it. The findings in this study, in the UK, and of all other translations studied so far fit that design.

The cutoff scores obtained in our study are a little lower than those reported by the British and Lithuanian adaptations.[26,36] Our values seem more similar to those found in the Italian version,[23] with the exception of the Functioning domain, which again is lower in our data than the others. It seems entirely plausible that cutoff scores, which reflect

Table 8 Male and female cutoff scores between clinical and nonclinical populations

Domains	Present study		Evans et al*	
	Male	Female	Male	Female
Subjective well-being	1.46	1.82	1.37	1.77
Problems/Symptoms	1.33	1.43	1.44	1.62
Functioning	1.06	1.07	1.29	1.30
Risk	0.24	0.21	0.43	0.30
Nonrisk items	1.24	1.33	1.36	1.50
All items	1.06	1.13	1.19	1.29

Note: *Reproduced with permission from Evans C, Connell J, Barkham M, et al. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *Br J Psychiatry*. 2002;180:51–60.[36] Available from: http://bjp.rcpsych.org/content/180/1/51.long.

service provision (implicit in the separation of clinical and nonclinical populations), will show cultural/national variations. Currently, data about cutoff scores for reliable change are being collected, and we hope that the results will be published soon.

Overall, the results provide very reassuring information about the psychometric properties and the potential of the Spanish version of the CORE-OM. The limitations of the study are the nonrandom sample frames, the relatively limited sample sizes, lack of interview measures, and relatively limited number of convergent validity tests. However, these results clearly support the use of the measure and justify development of subsequent studies with the forms derived from this questionnaire. A Catalan translation has been completed, and psychometric exploration of it is currently in progress. Another translation into Spanish considered more suited for use in Argentina has been completed but not tested. Initial discussions with a few natives suggest that the Spanish version assessed in this article is considered acceptable for use in Chile, Mexico, and Colombia. Further exploration of its acceptability, and then its psychometric properties in other Spanish-speaking countries other than Spain, is encouraged.

In summary, this study presents the Spanish version of the CORE-OM showing that it is a reliable and valid instrument for assessing psychological distress in patients and providing feedback to their therapists about overall change and ongoing progress. An additional advantage of this instrument in all its versions, including Spanish, is that it can be used without payment of license fees, and this should facilitate generation of much more evidence about the efficacy and effectiveness of psychological therapies in Spain and at least some other Spanish-speaking populations.

Disclosure

The contract grant sponsor for this study was the Ministry of Economy and Competitiveness of Spain, contract grant number: PSI2011-23246. The authors report no other conflicts of interest in this work.

References

1. Margison FR, Barkham M, Evans C, et al. Measurement and psychotherapy evidence-based practice and practice-based evidence. *Br J Psychiatry*. 2000;177(2):123–130.
2. CORE System Trust [homepage on the Internet]. United Kingdom: CORE Information Management Systems. [updated 2015 Jun 1; cited 2015 Oct 10]. Available from: http://www.coresystemtrust.org.uk/copyright.pdf. Accessed March 18, 2016.
3. CORE System Trust [homepage on the Internet]. United Kingdom: CORE Information Management Systems. [updated 2015 Jun 1; cited 2015 Oct 10]. Available from: http://www.coresystemtrust.org.uk. Accessed March 18, 2016.
4. Spanish CORE-OM [homepage on the Internet]. Spain: CORE-OM Official Spanish Website. [updated 2015 Jun 1; cited 2015 Oct 10]. Available from: http://www.ub.edu/terdep/core. Accessed March 18, 2016.
5. Barkham M, Evans C, Margison F, et al. The rationale for developing and implementing core batteries in service settings and psychotherapy outcome research. *J Ment Health*. 1998;7(1):35–47.
6. Barkham M, Margison F, Leach C, et al. Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. *J Consult Clin Psychol*. 2001;69(2):184–196.
7. Barkham M, Gilbert N, Connell J, Marshall C, Twigg E. Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *Br J Psychiatry*. 2005;186:239–246.
8. Evans C, Connell J, Barkham M, Marshall C, Mellor-Clark J. Practice-based evidence: benchmarking NHS primary care counselling services at national and local levels. *Clin Psychol Psychother*. 2003;10(6):374–388.
9. Gilbert N, Barkham M, Richards A, Cameron I. The effectiveness of a primary care mental health service delivering brief psychological interventions: a benchmarking study using the CORE system. *Prim Care Ment Health*. 2005;3(4):241–251.
10. McHugh P, Gordon M, Byrne M. Evaluating brief cognitive behavioural therapy within primary care. *Ment Health Rev J*. 2014;19(3):196–206.
11. Hardy GE, Cahill J, Stiles WB, et al. Sudden gains in cognitive therapy for depression: a replication and extension. *J Consult Clin Psychol*. 2005;73(1):59–67.
12. Stiles WB, Leach C, Barkham M, et al. Early sudden gains in psychotherapy under routine clinic conditions: practice-based evidence. *J Consult Clin Psychol*. 2003;71(1):14–21.
13. Di Bona L, Saxon D, Barkham M, Dent-Brown K, Parry G. Predictors of patient non-attendance at improving access to psychological therapy services demonstration sites. *J Affect Disord*. 2014;169:157–164.
14. Royal College of Nursing. *Working Well? Results from the RCN Working Well Survey into the Wellbeing and Working Lives of Nurses*. London: Author; 2003.
15. Connell J, Barkham M, Stiles WB, et al. Distribution of CORE−OM scores in a general population, clinical cut-off points and comparison with the CIS-R. *Br J Psychiatry*. 2007;190(1):69–74.
16. Connell J, Barkham M, Mellor-Clark J. Mental health CORE-OM norms of students attending university counselling services benchmarked against an age-matched primary care sample. *Br J Guid Counc*. 2007;35(1):41–56.
17. Barkham M, Culverwell A, Spindler K, Twigg E. The CORE-OM in an older adult population: psychometric status, acceptability, and feasibility. *Aging Ment Health*. 2005;9(3):235–245.
18. Jenkins PE, Turner HM. An investigation into the psychometric properties of the CORE-OM in patients with eating disorders. *Couns Psychother Res*. 2014;14(2);102–110.
19. Richards DA, Lovell K, Gilbody S, et al. Collaborative care for depression in UK primary care: a randomized controlled trial. *Psychol Med*. 2003;38(02):279–287.
20. Feixas G, Bados A, García-Grau E, et al. A dilemma-focused intervention for depression: a multicenter, randomized controlled trial with a 3-month follow-up. *Depress Anxiety*. Epub 2016 Apr 22.
21. Nordgren LB, Hedman E, Etienne J, et al. Effectiveness and cost-effectiveness of individually tailored internet-delivered cognitive behavior therapy for anxiety disorders in a primary care population: a randomized controlled trial. *Behav Res Ther*. 2014;59:1–11.
22. CORE System Trust [homepage on the Internet]. United Kingdom: CORE Information Management Systems. [updated 2015 Jun 1; cited 2015 Oct 10]. Available from: https://www.coresystemtrust.org.uk/cst-translation-policy. Accessed March 18, 2016.
23. Palmieri G, Evans C, Hansen V, et al. Validation of the Italian version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM). *Clin Psychol Psychother*. 2009;16(5):444–449.

24. Dias Sales MC, de Matos Moleiro CM, Evans C, Gomes Alves PC. The Portuguese version of the CORE-OM: translation, adaptation, and preliminary study of psychometric properties. *Rev Psiquiatria Clín*. 2011;39(2):54–59.
25. Elfström ML, Evans C, Lundgren J, Johansson B, Hakeberg M, Carlsson SG. Validation of the Swedish Version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM). *Clin Psychol Psychother*. 2013;20(5):447–455. doi:10.1002/cpp.1788.
26. Viliūnienė R, Evans C, Hilbig J, et al. Translating the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM) into Lithuanian. *Nord J Psychiatry*. 2012;67(5):305–311.
27. Kristjánsdóttir H, Sigurðsson BH, Salkovskis P, et al. 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. *Clin Psychol Psychother*. 2015;22(1):64–74. doi:10.1002/epp.1874.
28. Jokić-Begić N, Lauri Korajlija A, Jurin T, Evans C. Faktorska struktura, psihometrijske karakteristike i kritična vrijednost hrvatskoga prijevoda CORE-OM upitnika [Factor structure, psychometric properties and cut-off scores of Croatian version of clinical outcomes in routines evaluation – outcome measure (CORE-OM)]. *Psihologijske Teme*. 2014;23(2):265–288. Slovenian.
29. Stiles WB, Barkham M, Wheeler S. Effect of duration of psychological therapy on recovery and improvement rates: evidence from UK routine practice. *Br J Psychiatry*. 2015;207:115–122.
30. Biescad M, Timulak L. Measuring psychotherapy outcomes in routine practice: examining Slovak versions of three commonly used outcome instruments. *Eur J Psychother Couns*. 2014;16(2):140–162.
31. Mechler J, Holmqvist R. Deteriorated and unchanged patients in psychological treatment in Swedish primary care and psychiatry. *Nord J Psychiatry*. 2016;70(1):16–23.
32. Sturludottir K, Gestsdottir S, Rafnsson RH, Johannsson E. Áhrif hreyfiihlutunar á einkenni geðklofa, andlega líðan og líkamssamsetningu hjá ungu fólki [The effects of physical activity intervention on symptoms in schizophrenia, mental well-being and body composition in young adults]. *Laeknabladid*. 2015;101(11):519. Icelandic.
33. Feixas G, Evans C, Trujillo A, et al. La versión española del CORE-OM: Clinical Outcomes in Routine Evaluation – Outcome Measure [Spanish version of the CORE-OM: Clinical Outcomes in Routine Evaluation – Outcome Measure]. *Rev de Psicot*. 2012;89:109–135. Spanish.
34. International Test Commission. International Test Commission guidelines for translating and adapting tests. 2010. Available from: http://www.intestcom.org/files/statement_using_tests_for_research.pdf. Accesed March 18, 2016.
35. Evans C. The CORE-OM (Clinical Outcomes in Routine Evaluation) and its derivatives. *Integra Sci Pract*. 2012;2(2):12–14.
36. Evans C, Connell J, Barkham M, et al. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *Br J Psychiatry*. 2002;180:51–60.
37. Evans C, Mellor-Clark J, Margison F, et al. CORE: clinical outcome in routine evaluation. *J Ment Health*. 2000;9:247–255.
38. Lyne KJ, Barrett P, Evans C, Barkham M. Dimensions of variation on the CORE-OM. *Br J Clin Psychol*. 2006;45(2):185–203.
39. Barkham M, Mellor-Clark J, Connell J, Cahill J. A core approach to practice-based evidence: a brief history of the origins and applications of the CORE-OM and CORE System. *Couns Psychother Res*. 2006;6(1):3–15.
40. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory – Second Edition (BDI-II)*. San Antonio, TX: Psychological Corporation; 1996.
41. Sanz J, Perdigón AL, Vázquez C. Adaptación Española del inventario para la depresión de Beck-II (BDI-II): Propiedades psicométricas en población general [Spanish adaptation for depression inventory Beck-II (BDI-II): psychometric properties in general population]. *Clín Salud*. 2003;14:249–280. Spanish.
42. Derogatis LR. *SCL-90-R: Administration, Scoring and Procedures: Manual II*. Towson, MD: Clinical Psychometric Research; 1983.
43. González de Rivera JL, De las Cuevas C, Rodríguez M, Rodríguez F. *Cuestionario de 90 síntomas SCL-90-R de Derogatis, L. Adaptación española* [*90 Symptoms Questionnaire SCL-90-R of Derogatis, L. Spanish Adaptation*]. Madrid, Spain: Pearson; 2002. Spanish.
44. Cronbach LJ. Coefficient alpha and the internal structure of test. *Psychometrika*. 1951;16(3):297–334.
45. Feldt LS, Woodruff DJ, Salih FA. Statistical inference for coefficient alpha. *Appl Psychol Meas*. 1987;11(1):93–103.
46. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59(1):12–19.
47. Bernal M, Haro JM, Bernert S, et al. Risk factors for suicidality in Europe: results from the ESEMED study. *J Affect Disord*. 2007;101(1–3):27–34.
48. Casey PR, Dunn G, Kelly BD, et al. Factors associated with suicidal ideation in the general population five-centre analysis from the ODIN study. *Br J Psychiatry*. 2006;189(5):410–415.
49. Cahill J, Barkham M, Stiles WB, Twigg W, Rees A, Hardy GE. Convergent validity of the CORE measures with measures of depression for clients in cognitive therapy for depression. *J Couns Psychol*. 2006;53(2):253–259.