



ORIGINAL

Outcomes Assessment in Psychological Treatment: Spanish adaptation of OQ-45 (Outcome Questionnaire)

Ioseba Iraurgi, Patricia Penas *

Department of Personality, Evaluation and Psychological Treatment, University of Deusto, Bilbao, Spain

Received 13 July 2020; accepted 8 March 2021

KEYWORDS

Outcome Questionnaire,
Spanish validation, OQ-45,
psychometric properties,
factorial structure,
functionality assessment

Abstract **Introduction:** The Lambert's Outcomes Questionnaire (OQ-45) is a 45-item self-administered instrument used to measure clinical outcomes in psychotherapy. It measures functioning through three areas: symptoms distress, interpersonal relations and social role. The objective of this paper is to assess the reliability and the validity of its Spanish version. **Method:** A sample of 639 subjects, non-clinical and clinical, completed the instrument. **Results:** The psychometric evidences of the questionnaire showed an adequate internal consistency (.97 and .91) and acceptable convergent validity with the BDI, STAI, PSS and SF-12. The Confirmatory Factor Analyses suggested a bifactor structure. The cut-off points have been established for the subscales and total score (54.5), considering the clinical criterion, which means, prioritizing slightly the sensibility over the specificity. The RCI was 3.80 and Minimum Change Score 17.56. **Conclusion:** The OQ-45 showed acceptable psychometric properties, providing support for using this version of the questionnaire to assess Spanish's functionality. Given these findings, this tool could help clinicians evaluate treatment efficacy and establish psychotherapy goals.

© 2021 Fundación Universitaria Konrad Lorenz. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Evaluación de resultados en el tratamiento psicológico: adaptación española del OQ-45 (Outcome Questionnaire)

PALABRAS CLAVE

Outcome Questionnaire,
validación española, OQ-45,
propiedades psicométricas,
estructura factorial,
evaluación de la funcionalidad

Resumen: **Introducción:** El instrumento *Outcome Questionnaire* (OQ-45) desarrollado por Lambert, se trata de una escala auto-administrada de 45-items que se utiliza para medir los resultados clínicos en psicoterapia. Esta herramienta evalúa el funcionamiento por medio de tres áreas: síntomas de malestar, relaciones interpersonales y rol social. El objetivo de este artículo es evaluar la fiabilidad y la validez de la versión española. **Método:** Una muestra de 639 personas, clínica y no clínica, cumplimentaron dicho instrumento. **Resultados:** Las evidencias psicométricas del cuestionario mostraron una consistencia interna adecuada (.97 y .91) y una validez convergente aceptable con las siguientes escalas: BDI, STAI, PSS y SF-12. Los análisis factoriales confirmatorios han sugerido una estructura bifactorial. Los puntos de corte se han establecido para las subescalas y la puntuación total (54.5), teniendo en cuenta el criterio clínico, lo que significa que se ha priorizado ligeramente la sensibilidad sobre la especificidad. El RCI ha sido de 3.80 y la puntuación mínima de cambio de 17.56. **Conclusión:** El OQ-45

* Corresponding author.

E-mail: patricia.penas@deusto.es

<https://doi.org/10.14349/rlp.2021.v53.7>

0120-0534/© 2021 Fundación Universitaria Konrad Lorenz. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ha mostrado propiedades psicométricas aceptables, que apoyan el uso de esta versión para valorar la funcionalidad de los españoles y, como consecuencia, podría ayudar a los clínicos a evaluar la eficacia del tratamiento y establecer objetivos en psicoterapia.

© 2021 Fundación Universitaria Konrad Lorenz. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Assessing the effectiveness of a treatment, and subsequently the patient's evolution is considered a breakthrough in the field of clinical practice. Traditionally, outcome assessment was based on efficacy research which evaluated discrete treatment interventions for specific groups of patients (McGlynn, 1996). Nevertheless, the focus has moved into the assessment of patients under real conditions, which is known as effectiveness research (Sederer et al., 1996).

Moreover, it has been a shift in the measures used to establish outcome in clinical practice. Researchers have used different tools to assess changes at symptomatologic levels in experimental trials. However, there has been a shift in the field and the patient's social capacity and adjustment are considered to be equally important to clinical change (Palomo-Vélez et al., 2020; Sederer et al., 1996). Thus, when measuring clinical assessment, the symptomatic and functional dimension should be considered given that the assessment of these different domains helps clinicians to have more positive outcomes in their daily practice, as it offers a useful methodology to know whether their efforts are consistent with the patient's psychotherapy process (Lambert, 2012; Sederer et al., 1996).

A benefit associated to measuring outcomes assessment is obtained by given the feedback, since it optimizes the treatment (Hawkins et al., 2004; Miller et al., 2013; Shimokawa et al., 2010). Lambert and his group suggested that regularly assessing the patient's psychotherapeutic process have beneficial effects on the treatment outcomes, since they proved that if the clinicians have feedback regarding their patients' process twice the clinically significant and reliable change rate and decrease both deterioration and the number of treatment sessions (Lambert et al., 2001). Additionally, it is especially helpful in the cases where a patient's treatment is not advancing well (Shimokawa et al., 2010). Many different researchers pointed out (Lambert, 2012; Walfish et al., 2012) that clinicians tend to have an optimistic view of their patients' processes, and the feedback allows them to notice those cases that are having a poor response to treatment, offering the therapist the chance to adjust the actual treatment to a more adequate one (Lambert, 2013). Likewise, the data obtained from the assessment is usually a more reliable option than the clinical judgment in order to recognise patients whose treatment is failing (Hannan et al., 2005).

As a consequence of the increasing demand for assessing outcomes in psychotherapy, during the 1900s *the Outcome Questionnaire* (OQ-45) was developed by Lambert and his group (Lambert et al., 1996). And nowadays, it has become one of the most used psychotherapy outcome instruments (Hatfield & Ogles, 2004). It is a 45-item self-report measure, designed for repeated administration during the course of a treatment. It is divided into three domains of patient functioning: symptoms and psychological distress (anxiety and depression), interpersonal relationships (problems related

to loneliness or conflict with others) and social role performance (daily activities or quality of life) (Lambert et al., 1996). The OQ-45 is a psychometrically appropriate instrument that has demonstrated to be sensitive to change in diverse populations over short intervals, and also, it remains relatively stable in untreated individuals (Vermeersch et al., 2000; Vermeersch et al., 2004). Therefore, the OQ-45 could be used by clinicians in their day-to-day work, to obtain the patients' perspective regarding the course of their treatment (Lambert, 2013). Due to its growing popularity, the OQ-45 has been translated into and validated in several languages, like: Japanese, Italian, German, Dutch and Norwegian, among others. Von Bergen and De la Parra (2002) have developed the Spanish version for the Chilean culture.

Given that the OQ-45 is not adapted for the Spanish population, and also, due to the high value that has been demonstrated in other countries, the objective of the present work is to adapt and validate the instrument for this culture. Moreover, with the aim of having an effective tool for assessing outcomes in clinical practice, a non-clinical and a clinical sample is going to be used in the procedure. The values of internal consistency are expected to be good and similar to the previous studies. In terms of the concurrent validity, it is expected that OQ-45 scores would correlate positively with clinical indicators and negatively with mental and physical health scores. Regarding the factorial structure, different models have been tested in previous researches (Lo Coco et al., 2008; Kim et al., 2010). So, three correlated factor, four-factor hierarchical and bi-factor models are going to be tested, with the expectation of obtaining appropriate adjustment indexes for a structure that would allow the explanation of the theoretical solution. Finally, through this study, criteria and norms to interpret the scores are going to be provided.

Method

Participants

The study included 639 adult participants, divided into two samples. The clinical sample was comprised of 139 patients who attended mental health centres with a demand for therapeutic assistance (most of them with affective symptomatology), 88 of which were female and 51 males who were aged between 19 and 61 years old (Mean - $M = 33.45$). The non-clinical sample ($n = 500$) included people from the community, between 17 and 82 years of age ($M = 44.45$), 53.4% were females and 46.6% males.

Instruments

The Outcome Questionnaire (OQ-45; Lambert et al., 1996) is a 45-item self-report measure, which is scored using

a five-point Likert Scale. The questionnaire generates a general dimension, taking into account all the items: OQ-45 total, but also, three subscales are generated. High scores indicate a perceived worse state. The psychometric properties of the scale have been widely investigated, resulting in an adequate internal consistency reliability coefficient ($\alpha = .93$) and concurrent validity with SCL-90-R and BDI with coefficients (persons' r) about .80 (Lambert et al., 2004). Although, there is a lack of support in its factor structure; users of the OQ-45 consider the total score and the scores obtained from the three subscales (Kim et al., 2010). The translation and preliminary analyses were developed by Iraurgi et al. (2009) and Penas et al. (2017), who found a high internal consistency (.88 to .91), an adequate concurrent validity, and an appropriate factorial structure.

Beck Depression Inventory (BDI; Beck et al., 1996) is 21-item self-report rating instruments that assess characteristics, attitudes and symptoms of depression. The Spanish version was developed by Sanz and Vázquez (1998) with an internal consistency of .87. The internal consistency of the current study was .90.

State-Trait Anxiety Inventory (STAI; Spielberg et al., 1970) is a self-report instrument that includes 20-items for assessing trait anxiety and 20 for state anxiety. The instrument was adapted to the Spanish population (Spielberg et al., 1982). In the present study an Alpha coefficient of .95 was obtained for the state dimension and .93 for the trait dimension.

The Perceived Stress Scale (PSS; Cohen et al., 1983) is 14-item self-report questionnaire which assesses the degree to which individuals consider situations in their lives to be stressful. The Spanish version was developed by Remor (2006) obtaining a reliability of .81. The Alpha the Cronbach of the current research has been .93.

Finally, the **Short-Form Health Survey (SF-12; Ware et al., 1996)** is a shortened version of the 36-items Short Form of a Health Survey (SF-36), which measures health related quality of life. Its Spanish version was developed by Vilagut et al. (2008). In the present study, the Alpha Cronbach obtained for the physical component was .78 and mental component .82.

Procedure

All the participants completed the OQ-45 in its Spanish version (Iraurgi et al., 2009). Apart from the OQ-45 the clinical sample completed the BDI, STAI and PSS, whereas, the non-clinical data filled out SF-12. The clinical sample was recruited from people attending outpatient Mental Health Services, whereas the non-clinical sample was composed of university students and their relatives. The characteristics and the purpose of the scale were explained to all of the participants. Informed consent was obtained, and participants were ensured that their responses would be confidential and anonymous. Additionally, all of the ethical requirements for conducting this type of the study were followed.

Statistical Analyses

The following descriptive statistics of the OQ-45 items of both samples were calculated with the SPSS program: mean (M), standard deviation (SD), asymmetry (As), the value of the Cronbach's alpha if the item was removed from each

dimension of the scale (a). Furthermore, the differences between clinical and non-clinical scores were obtained through the t -student and its effect size by Cohen's d . Finally, the Felps test was used to compare the alpha of each dimension with the others. To address the concurrent validity, the correlation between the total scores and the three dimensions of OQ-45 with the scores obtained in the following instruments were measured: BDI, STAI-S and STAI-R and PSS for the clinical sample and SF-12 for the non-clinical one.

The factorial validity was analysed conducting a confirmatory factor analysis (CFA), using Mplus 7. In this process, the three-factor correlated model, four-factor hierarchical model and four-factor bi-level model were proved for both samples. The following indicators were used to assess the level of goodness of fit: the χ^2 chi-square, the degrees of freedom (df), the Akaike Information Criterion (AIC), the Adjusted Goodness-of-Fit Index (AGFI), the Normed Fit Index (NFI), the Comparative Fit Index (CFI), Standardized Root-Mean-Square Residual (SRMSR), the Root-Mean-Square Error of Approximation (RMSEA) and its confidence interval (90% CI).

Furthermore, for the cut-off points for the total score of the OQ-45 and each of the subscales the ROC curve (Receiver Operating Characteristics; Cerdá & Cifuentes, 2012) was used to estimate the four proposed cut-off points between the clinical and non-clinical sample. The procedure is developed by searching on the curve for the higher values of specificity and sensitivity. This point is determined by the Youden Index (Sensitivity + specificity -1) (Schisterman et al., 2005), which indicates the point where the sensitivity and specificity represents the highest value. Likewise, this type of methodology allows selecting cut-off points that are better suited to the objective of our study, knowing that the reduction or increase in sensitivity implies an increase or reduction of specificity, and vice versa.

The Reliable Change Index (RCI, Jacobson & Truax, 1991) and the clinically significant change were calculated following the procedure proposed by the aforementioned authors. The algorism is: $RCI = (\bar{x}_{\text{clinical}} - \bar{x}_{\text{non-clinical}}) / S_{\text{dif}}$, where S_{dif} is the standard error of the mean differences ($S_{\text{dif}} = DT \times \sqrt{1 - R_{xx}} \times \sqrt{2}$), DT is the pooled deviation of the clinical and non-clinical samples, and R_{xx} is the reliability of the OQ-45. Moreover, considering this algorism and the confidence interval of 95%, the Minimum Change Score was calculated ($MCS = S_{\text{dif}} \times 1.96$). The cut-off point (CP) has been calculated following this formula: $CP = (\bar{x}_{\text{clinical}} - SD_{\text{clinical}}) + (\bar{x}_{\text{non-clinical}} - SD_{\text{non-clinical}}) / (SD_{\text{clinical}} + SD_{\text{non-clinical}})$.

Results

Table 1 shows the differences between the clinical and the non-clinical scores in the global dimension and in the three subscales; all of the differences are statistically significant [$t_{(165,09)} = 12.65, p < .001$], with the clinical scores being higher. Also, there were differences in each item of the scale, with most of them being statistically significant. Moreover, the effect sizes are considerably high. Regarding internal consistency of the OQ-45, Cronbach's Alpha coefficients are calculated for the global score (.95) and also for the three dimensions. With the aim of comparing the Alpha Cronbach values, the Feldt test was calculated

Table 1 Differences between clinical and non-clinical samples and internal consistency

	Clinical (<i>N</i> = 139)				Non-clinical (<i>N</i> = 500)				Comparison					
	<i>M</i>	<i>SD</i>	As	A	<i>M</i>	<i>SD</i>	Sk	α	<i>t</i>	d.f.	<i>p</i>	<i>d</i>	Feldt	<i>p</i>
2	1.27	1.10	0.31	.75	1.35	0.89	0.38	.88	-0.80	191.38	.430	0.09		
3	1.06	1.12	0.53	.75	1.18	0.84	0.61	.88	-1.17	183.88	.240	0.13		
5	2.22	1.04	0.22	.75	1.66	0.83	-0.06	.88	5.77	188.99	.001	0.64		
6	2.32	1.23	0.22	.75	1.53	0.76	0.02	.88	7.18	168.18	.001	0.90		
8	0.29	0.61	1.96	.76	0.15	0.49	4.23	.88	2.46	191.37	.015	0.27		
9	2.44	1.06	0.20	.75	1.13	0.83	0.39	.88	13.51	187.83	.001	1.48		
10	2.30	1.35	-0.07	.75	0.87	0.83	0.87	.88	11.94	168.06	.001	1.48		
11	0.34	0.87	2.40	.76	0.20	0.56	3.46	.88	1.82	170.42	.071	0.22		
13r	1.72	1.11	-0.23	.75	0.76	0.77	0.86	.88	9.53	176.96	.001	1.12		
15	2.12	1.46	-0.80	.75	0.45	0.72	1.76	.88	12.98	157.18	.001	1.80		
22	2.34	1.11	0.27	.75	1.42	0.98	0.64	.88	8.90	201.79	.001	0.91		
23	1.88	1.39	0.33	.75	1.24	1.01	0.48	.88	5.15	180.06	.001	0.58		
OQ-45 Symptoms	24r	2.71	1.16	-0.20	.75	0.83	0.84	0.93	.88	17.85	179.85	.001	2.05	
	25	2.29	1.14	0.24	.75	1.04	0.94	0.65	.87	11.94	193.25	.001	1.27	
	27	1.00	.83	0.00	.76	1.01	0.92	0.77	.88	-0.15	241.20	.880	0.01	
	29	1.59	1.07	-0.06	.75	0.83	0.94	1.03	.88	7.60	201.71	.001	0.79	
	31r	2.03	1.22	0.62	.75	0.79	0.86	1.00	.88	11.25	177.62	.001	1.31	
	33	2.42	1.16	0.91	.75	0.81	0.86	1.05	.88	15.21	182.43	.001	1.73	
	34	0.81	0.78	0.36	.76	1.38	1.00	0.47	.89	-7.20	276.94	.001	0.60	
	35	1.37	1.02	0.24	.75	0.27	0.65	2.76	.88	11.95	170.25	.001	1.48	
	36	2.45	1.02	0.07	.75	1.61	0.90	0.23	.88	8.86	201.38	.001	0.91	
	40	2.19	1.31	-0.14	.75	0.64	0.86	1.33	.88	13.24	172.30	.001	1.59	
	41	2.11	1.05	0.43	.75	1.22	0.98	0.53	.88	8.92	209.86	.001	0.90	
	42	2.72	1.16	-0.21	.75	1.10	0.83	0.43	.88	15.47	179.11	.001	1.78	
	45	0.65	0.81	0.79	.76	1.11	0.96	0.72	.88	-5.70	256.65	.001	0.50	
	Total	44.63	21.23	0.08	.76	24.56	10.89	0.61	.88	10.76	158.73	.001	1.45	2 .001
OQ-45 Inter- personal relationships	1r	0.84	0.86	0.86	.85	0.45	0.57	0.93	.74	5.06	173.25	.001	0.61	
	7	0.96	0.98	0.69	.85	0.87	1.03	1.29	.73	-0.85	637*	.400	0.09	
	16	1.44	1.02	0.11	.86	2.00	0.97	0.15	.76	-5.77	211.96	.001	0.57	
	17	1.73	1.03	-0.21	.86	1.15	1.19	0.90	.73	5.25	635*	.001	0.50	
	18	2.60	0.94	-0.02	.84	0.97	0.89	0.80	.71	18.33	212.26	.001	1.81	
	19	2.60	0.97	-0.11	.85	1.42	0.84	0.30	.74	13.10	199.42	.001	1.36	
	20r	1.13	1.00	0.44	.85	0.67	0.73	0.85	.72	5.08	180.75	.001	0.58	
	26	0.45	0.89	1.95	.88	0.34	0.73	2.49	.76	1.26	192.60	.210	0.14	
	30	1.19	1.08	0.53	.85	0.63	0.73	1.36	.74	5.84	174.92	.001	0.69	
	37r	2.36	1.00	0.24	.85	1.27	1.18	0.74	.72	-9.95	636*	.001	0.95	
	43r	2.37	1.03	0.20	.84	0.76	0.81	1.06	.72	17.01	188.38	.001	1.87	
	Total	17.68	7.05	0.09	.86	10.53	5.31	0.48	.75	11.10	183.90	.001	1.15	.56 .001
OQ-45 Social rol (functiona- lity)	4	1.94	0.98	0.65	.82	1.67	0.97	0.15	.52	2.80	635*	.005	0.28	
	12r	2.38	0.94	0.08	.82	1.13	0.98	0.83	.56	13.39	633*	.001	1.29	
	14	2.46	0.90	-0.03	.83	2.18	1.07	-0.11	.65	2.85	637*	.004	0.27	
	21r	2.39	0.97	0.20	.80	0.81	0.85	0.85	.59	17.30	200.30	.001	1.80	
	28	2.49	0.96	0.03	.82	1.19	0.96	0.67	.54	14.17	633*	.001	1.36	
	32	0.53	0.94	1.64	.88	0.10	0.34	3.77	.60	5.36	148.06	.001	0.81	
	38	2.27	0.95	0.31	.81	0.96	0.93	0.95	.54	14.68	634*	.001	1.40	
	39	1.59	0.95	0.09	.80	1.13	0.86	0.78	.53	5.12	205.10	.001	0.52	
	44	0.32	0.64	1.95	.83	0.26	0.59	2.52	.56	1.05	634*	.300	0.10	
	Total	16.37	5.49	0.05	.84	9.42	3.78	0.46	.60	14.01	176.36	.001	1.48	.40 .001
OQ-45 global	Total	78.69	30.44	-0.13	.97	44.48	17.72	.54	.91	12.65	165.09	.001	1.37	.33 .001

Note. r = the item has been recoded; *Levene's test of variances has not been significative; α =Alpha if element is eliminated for each item and the Alpha the Cronbach for the total of the scale.

Table 2 Concurrent validity

OQ-45 →	Global measure	Symptoms	Interpersonal Relationships	Social Role
SF-12	Physical Function	-.51	-.49	-.37
	Physical Role	-.33	-.30	-.24
	Body Ache	-.31	-.29	-.22
	General Health	-.40	-.38	-.33
	Vitality	-.39	-.38	-.32
	Social Function	-.50	-.46	-.43
	Emotional Role	-.56	-.55	-.42
	Mental Health	-.61	-.61	-.49
	Physical Component	-.32	-.28	-.23
BDI	Mental Component	-.60	-.59	-.49
		.54	.54	.34
		.23	.18 ^s	.23
		.66	.68	.44
PSS		.79	.76	.62
				.66

Note. Pearson correlation coefficient -^s not significative, all the rest are statistically significative for $p < .001$. The SF-12 was administered to the non-clinical sample and the BDI, STAI and PSS to the clinical sample.

given that it is statistically significant for the global measure (Feldt = .33, $p < .001$) and for the three dimensions.

The data related to the concurrent validity is presented in the Table 2. For the clinical sample, all instruments correlated significantly, except the ones of the STAI-State with the symptomatology dimension of the OQ-45. In the non-clinical sample, the two components and the eight dimensions of the SF-12 correlated significantly in a negative way with the global score of the OQ-45, and also, with the three dimensions. All of these correlations are consistent with what would be expected, that is, higher scores for depression, anxiety and stress in the clinical sample, and higher OQ-45 scores. While the better the mental and physical health scores in the non-clinical sample, the lower the dysfunctionality.

In the Table 3, three different models were calculated through a CFA to support construct validity. For the two samples the bi-factorial structure suited more appropriately; only the AGFI index was slightly inferior than the criterion index, ($\chi^2_{(900)} = 2641.42$, $p < .001$, AGF = .88, CFI = .92, RMSEA = .049 [.037 to .062]) and ($\chi^2_{(900)} = 3930.47$, $p < .001$, AGF = .86, CFI = .91, RMSEA = .061 [.049 to .073]), respectively.

Figure 1 shows graphically the factor structure of the clinical sample. The factor loadings of three items (11, 26 and 32) were lower than .10. They assess problematic drinking or drug use, and as can be observed in the Table 1, they have skewed scores (the majority of the participants obtained 0). Nevertheless, each item is included in one of the three subscales. The factor structure obtained from the non-clinical sample is similar to the clinical one.

Table 3 Confirmatory Factor Analysis of the non-clinical sample ($n=500$) and the clinical sample ($n = 139$)

	Model	χ^2	d.f.	Goodness fit indexes					90% CI	
				AGFI	NFI	CFI	SRMSR	RMSEA	IL	SL
Non-clinical sample	Three correlated factors	3310.78	942	.82	.86	.86	.062	.058	.049	.067
	Four-Factor hierarchical model	3150.79	942	.82	.85	.85	.109	.088	.076	.101
	Bi-factorial model	2641.42	900	.88	.90	.92	.051	.049	.037	.062
Clinical sample	Three correlated factors	4639.24	942	.81	.85	.86	.089	.083	.072	.094
	Four-factor hierarchical model	4701.79	942	.80	.84	.85	.101	.090	.079	.101
	Bi-factorial model	3930.47	900	.86	.89	.91	.053	.061	.049	.073

Note. χ^2 : chi-square; df- degrees of freedom; AIC- Akaike Information Criterion; AGFI- Adjusted Goodness-of-Fit Index; NFI- Normed Fit Index; CFI- Comparative Fit Index; SRMSR- Standardized Root-Mean-Square Residual; RMSEA- Root-Mean-Square Error of Approximation and CI90%- Confidence Interval 90% RMSEA.

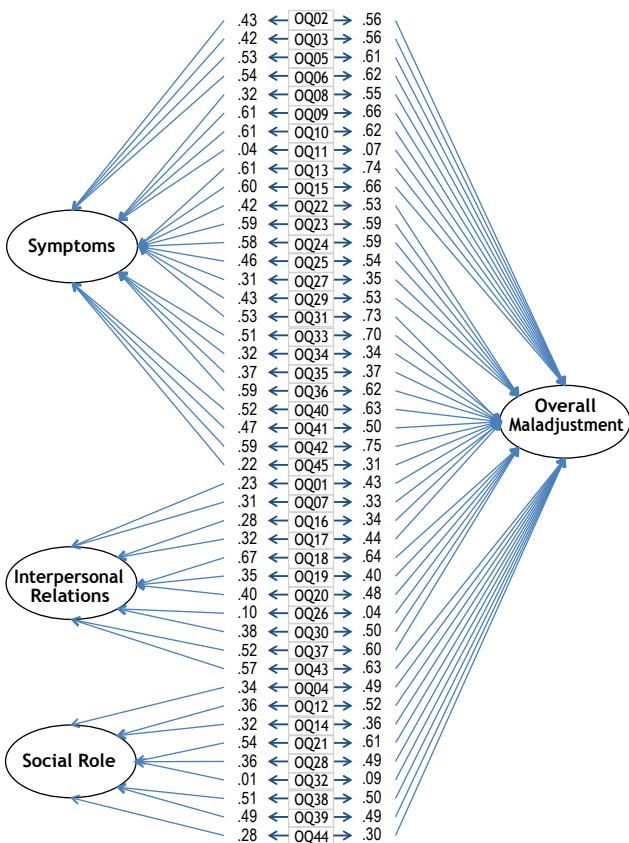


Figure 1. Bi-factor model of the clinical sample ($n = 139$).

Figure 2 shows the ROC curve obtained for OQ-45 scale, the continued curve represents the scale's capacity to discriminate between the global scores of the functional and non-functional populations. Likewise, the discontinued curves represent each of the dimensions of the scale: symptomatology (short lines), interpersonal relationships (medium lines) and social role (the longest lines). The diagonal of the

figure symbolizes the condition of zero discrimination, so any curve that stays away from such diagonal and covers an area towards the upper left corner would indicate a better diagnostic utility.

Table 4 presents the values obtained for the area under the curve (ADC), sensitivity, specificity, chi-square, Youden Index and the corresponding cut-off points of the global score of the OQ-45 and for the other three dimensions. For the global score of the OQ-45, the higher value of Youden Index stands at the value 57.5 and offers a sensitivity of .71 and specificity of .79. Likewise, the symptomatology dimension cut-off point is 32.5 the interpersonal relationships are 14.5, and the social role is 12.5. Furthermore, also it also showed the results obtained after taking into consideration the sensitivity prioritization criterion. For the global score, the value of the cut-off point is 54.5, and offers a sensitivity of .74 and specificity of .73. Likewise, the symptomatology dimension cut-off point is 29.5, the interpersonal relationships are 12.5 and the social role is 11.5.

Moreover, the Reliable Change Index has been calculated for the total score of the OQ-45 ($RCI = 3.80 > 1.96$), where the Minimum Change Score is 17.56, indicating an individual need to gain or lose 17.56 points to be considered a statistically significant improvement or deterioration. Finally, the cut-off point calculated considering the algorism proposed by Jacobson and Truax (1991) is established at 57.43.

Discussion

The Outcome Questionnaire should be considered reliable and valid to measure psychotherapeutic outcomes. No important differences between the original version, studies using other versions (Von Bergen & De la Parra, 2002), and the findings obtained in this research using the Spanish adaptation have been found.

The internal consistency was satisfactory in the total OQ-45 score both for the clinical and non-clinical populations: .97 and .91, respectively. Those Cronbach Alpha are similar to the research conducted by Lambert and his colleges (1996). The lower Cronbach Alpha for the non-clinical sample compared to the clinical sample could be due to its range. Furthermore, the internal consistency of the three dimensions was high and adequate. Additionally, significant differences were found in the OQ-45 total scores and in its three dimensions, between the clinical and non-clinical group. As was expected, higher scores were revealed in the clinical sample (patients who came to a mental health centres with a need for therapeutic assistance, affective symptomatology in most cases). Those differences confirm the instrument's ability to discriminate between clinical and non-clinical populations. Effects sizes reassured the size of those differences.

Regarding concurrent validity, the used instruments were different depending on the sample. Firstly, in the clinical population BDI, STAI and PSS correlated significantly with the three dimension and total scores of the scale, especially with those scales that evaluate the stress and the depressive symptomatology expressed by the individual. As was expected, the higher correlations were obtained in the symptomatology and global dimension of the scale, and instead the correlations were lower in interpersonal and social relationships areas, since these dimensions are measured less

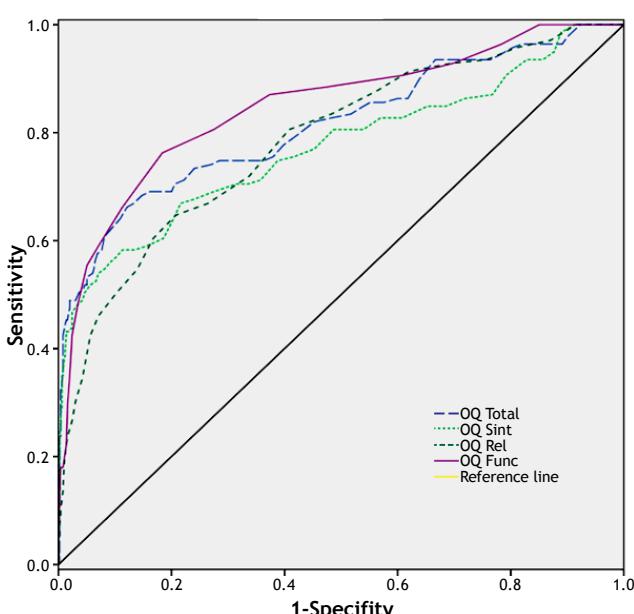


Figure 2. ROC Curve of the OQ-45.

Table 4 Area under the curve, sensitivity, specificity and chi-square values for each cut-off point considering Youden Index and sensitivity prioritization criterion

OQ-45		ADC	<i>p</i>	Sensitivity	Specificity	χ^2	Youden Index	Cut-off Point
Youden criterion	Total	.81	.001	.71	.79	125.13	.50	57.5
	Symptoms	.77	.001	.67	.78	103.22	.45	32.5
	Interpersonal	.79	.001	.65	.79	98.72	.44	14.5
	Social Role	.85	.001	.76	.82	169.73	.58	12.5
Sensitivity prioritization criterion	Total	.81	.001	.74	.73	102.54	.47	54.5
	Symptoms	.77	.001	.71	.68	68.27	.39	29.5
	Interpersonal	.79	.001	.72	.66	65.09	.38	12.5
	Social Role	.85	.001	.81	.73	128.49	.54	11.5

Note. ADC = Area under the curve; *p* = probability value; χ^2 = Chi square test.

by those additional instruments (BDI, STAI and PSS). Nonetheless, the state subscale of the STAI questionnaire obtains different results, where it has a non-significative correlation with the symptomatology dimension, and also, their correlations, despite being low, are higher with the interpersonal relationships and social role dimensions, indicating that this subscale of the STAI is more related to these areas of functioning. Secondly, in the non-clinical sample, the concurrent validity was measured by the SF-12 questionnaire, all the subscales correlated negatively which means that an increase in the OQ-45 scores is associated with a decrease in the scores of the SF-12. The mental component of the SF-12 is higher related to the OQ-45, indicating that the OQ-45 is related to the mental dimension in all of its subscales.

As most of the non-clinical sample was collected in a community setting, it was considered more appropriate to use questionnaires related to quality of life. Whereas, for the clinical sample scales related to stress, anxiety and depression symptomology were used since they are routinely used in the clinical setting, and also because they provided additional information about the patient's need and problem. Thus, as a consequence of the different instruments used in order of being adequate with the sample, the appropriate concurrent validity between the OQ-45 and the other instruments that measure diverse areas of functioning has been confirmed.

The results of the confirmatory factor analyses for both samples indicated that the OQ-45 is a bi-factorial scale composed of one general factor and three subscales. In this structure each item loaded on one of the three subscales, but also, in a general factor of distress. This model offers an appropriate description of the OQ-45 structure, since it fits with the interpretation process given by the authors (Lambert et al., 1996). This structure was also provided by Lo Coco et al. (2008) with an Italian sample.

In spite of considering the cut-off points that maximize the sensitivity and specificity (Youden's higher value), the ones that prioritize the sensitivity in a slight way have been established. The objective of the current study is to adapt and validate a scale that is going to be used with clinical patients, and consequently, it is important to capture all the possible cases in order not to leave them unattended. In other words, the clinical criterion has been chosen over the

statistical criterion, as the cut-off point suggested through the Youden Index prioritizes the specificity over the sensitivity. The cut-off point (CP) obtained for the OQ-45 in the present study, 54.5, is lower than the result of the original research (CP = 63) (Lambert et al., 1996). The resulting values of sensitivity and specificity after the calibration are also smaller in the Spanish adaptation, .74 for sensitivity and .73 for specificity. Timman et al. (2017) also reported a smaller cut-off point for the Dutch version (CP = 55) in comparison to the original one.

Considering the procedure for calculating the cut-off points proposed by Jacobson and Truax (1991), 57.43, which is 3 points superior to the one that resulted from maximizing the sensitivity. Moreover, the amount of points in OQ-45 that an individual needs to improve are 17.56, whereas in the American sample 14 points are necessary (Beckstead et al., 2003). Both indexes obtained from this research would allow the clinician to classify the patients as followed: recovered, improved, no change, or deteriorated. Moreover, they will also be able to choose the most convenient cut-off point based on their judgment, while also taking into account the person that they are working with.

Due to the differences found in terms of cut-off points, the current study has some possible limitations in the collected sample, especially in the size of the clinical sample. As a consequence, it would be adequate to increase the number of participants in the clinical sample. For future research it would be interesting to track the patients' progress and prove the instrument's sensitivity to change.

Based on the data presented, the Spanish adaptation of the OQ-45 has appropriate psychometric properties in order to be considered a useful instrument. Moreover, it could be an adequate scale for assessing the functionality of Spanish patients, and consequently, it could help clinicians to evaluate treatment efficacy and establish psychotherapy goals.

Conflict of interest

There is no conflict of interest.

References

- Beck, A. T., Steer, R.A., & Brown, G. K. (1996). *BDI-II. Beck Depression Inventory-Second Edition Manual*. The Psychological Corporation.
- Beckstead, D. J., Hatch, A. L., Lambert, M. J., Eggett, D. L., Goates, M. K., & Vermeersch, D. A. (2003). Clinical significance of the Outcome Questionnaire (OQ-45.2). *The Behaviour Analyst Today*, 4, 86-90. <https://doi.org/10.1037/h0100015>
- Cerda, J., & Cifuentes, L. (2012). Uso de curvas ROC en investigación clínica: Aspectos teórico-prácticos. *Revista Chilena de Infectología*, 29(2), 138-141. <https://doi.org/10.4067/S0716-10182012000200003>
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behaviour*, 24, 385-396.
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, 61, 155-163. <https://doi.org/10.1002/jclp.20108>
- Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice*, 35, 485-491.
- Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K., & Tuttle, K. (2004). The effects of providing patient progress information to therapist and patient. *Psychotherapy Research*, 14, 308-327.
- Iraurgi, I., Hormaechea, A., Jiménez-Lerma, J. M., & Landabaso, M. (2009). Valoración de resultados en intervenciones psicosociales: datos preliminares de la adaptación y características psicométricas del OQ-45. *XIII Congreso Nacional de Psiquiatría, Madrid*.
- 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, 12-19. <https://doi.org/10.1037/0022-006X.59.1.12>
- Kim, S. H., Beretvas, S. N., & Sherry, A. R. (2010). A validation of the factor structure of OQ-45 scores using factor mixture modelling. *Measurement and Evaluation in Counselling and Development*, 42(4), 275-295. <https://doi.org/10.1177/0748175609354616>
- Lambert, M. J., Hansen, N. B., Umpress, V., Lunnen, K., Okiishi, J., & Burlingame, G. M. (1996). *Administration and Scoring Manual for the OQ-45.2*. American Professional Credentialing Services LLC
- Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Lars, S., & Hawkins, E. J. (2001). The effects of providing therapist with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research*, 11(1), 49-68. <https://doi.org/10.1080/713663852>
- Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., Shimokawa, K., Christopherson, C., & Burlingame, G. M. (2004). *Administration and Scoring Manual for the Outcome Questionnaire (OQ-45.2)*. American Professional Credentialing Service.
- Lambert, J. M. (2012). Helping clinicians to use and learn from research-based systems: The OQ-analyst. *Psychotherapy (Chicago, Ill), Training*, 49, 109-114. <https://doi.org/10.1037/a0027110>
- Lambert, J. M. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy*, 50(1), 42-51. <https://doi.org/10.1037/a0030682>
- Lo-Coco, G., Chiappelli, M., Bensi, L., Gullo, S., Prestano, C., & Lambert, J. M. (2008). The factorial structure of the Outcome Questionnaire-45: A study with an Italian sample. *Clinical Psychology and Psychotherapy*, 15, 418-423. <https://doi.org/10.1002/cpp.601>
- McGlynn, E. A. (1996). Domains of Study and Methodological Challenges. In L. L. Sederer & B. Dickey. (Ed.), *Outcomes Assessment in Clinical Practice* (pp. 20-24). Williams & Wilkins.
- Miller, S. D., Hubble, M. A., Choe, D. L., & Seidel, J. A. (2013). The outcome of psychotherapy: Yesterday, today and tomorrow. *Psychotherapy*, 50(1), 88-97. <https://doi.org/10.1037/a0031097>
- Palomo-Vélez, G., García, F., Arauna, D., Muñoz-Mendoza, C. L., Fuentes, E., & Palomo, I. (2020). Effect of frailty status on happiness and life satisfaction: The mediating role of self-perceived health. *Revista Latinoamericana de Psicología*, 52, 95-103. <https://doi.org/10.14349/rlp.2020.v52.10>
- Penas, P., Iraurgi, I., Gorbeña, S., Matellanes, B., & Montero, M. (2017). Outcomes assessment: Psychometric properties of the Spanish adaptation of the outcome questionnaire (OQ-45). *European Psychiatry*, 41, S781. <https://doi.org/10.1016/j.europsy.2017.01.1486>
- Remor, E. (2006). Psychometric properties of a European Spanish version of the Perceived Stress Scale (PSS). *The Spanish Journal of Psychology*, 9(1), 86-93.
- Sanz, J., & Vázquez, C. (1998). Fiabilidad, validez y datos normativos del inventario para la depresión de Beck. *Psychothema*, 10, 303-318.
- Schisterman, E. F., Perkins, N. J., Liu, A., & Bondell, H. (2005). Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*, 17, 73-81.
- Sederer, L. L., Dickey, B., & Hermann, R. C. (1996). The imperative of outcomes assessment in psychiatry. In L. L. Sederer & B. Dickey (Eds.), *Outcomes Assessment in Clinical Practice* (pp. 1-7). Williams & Wilkins.
- Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, 78(3), 298-311. <https://doi.org/10.1037/a0019247>
- Spielberg, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State/Trait Anxiety Inventory*. Consulting Psychologist Press.
- Spielberg, C. D., Gorsuch, R. L., & Lushene, R. E. (1982). *Cuestionario de Ansiedad Estado/Rasgo*. TEA.
- Timman, R., De Jong, K., & De Neve-Enthoven, N. (2017). Cut-off scores and Clinical Change Indices for the Dutch Outcome Questionnaire (OQ-45) in a large sample of normal and several psychotherapeutic populations. *Clinical Psychology and Psychotherapy*, 24, 72-81. <https://doi.org/10.1002/cpp.1979>
- Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000). Outcome Questionnaire: Item sensitivity to change. *Journal of Personality Assessment*, 74, 242-261. https://doi.org/10.1207/S15327752JPA7402_6
- Vermeersch, D. A., Whipple, J. L., Lambert, M. J., Hawkins, E. J., Burchfield, C. M., & Okiishi, J. C. (2004). Outcome Questionnaire: Is it sensitive to changes in counselling centre clients? *Journal of Counselling Psychology*, 51, 38-49. <https://doi.org/10.1037/0022-0167.51.1.38>
- Vilagut, G., Valders, J. M., Ferrer, M., Garin, O., López-García, E., & Alonso, J. (2008). Interpretación de los cuestionarios de salud SF-36 y SF-12 en España: Componentes físico y mental. *Medicina Clínica*, 130(19), 726-735.
- Von Bergen, A., & De la Parra, G. (2002). OQ-45.2, An Outcome Questionnaire for monitoring change in psychotherapy: Adaptation, validation and indications for its application and interpretation. *Terapia Psicológica*, 20(38), 161-176.
- Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports*, 110, 639-644. <https://doi.org/10.2466/02.07.17.PR0.110.2.639-644>
- Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-item Short-Form Health Survey: Construction of scales and preliminary test reliability and validity. *Medical Care*, 34, 220-233.