CAMBRIDGE
UNIVERSITY PRESS

Psychometric Properties of the Outcome Rating Scale (ORS) in a Spanish Clinical Sample

Danilo Moggia[1], Noelia Niño-Robles[1], Scott D. Miller[2] and Guillem Feixas[1]

[1] *Universitat de Barcelona (Spain)*
[2] *The International Center for Clinical Excellence (USA)*

Abstract. The Outcome Rating Scale (ORS) is an ultra-brief measure of well-being designed to track outcome in psychotherapy. This research studied the psychometric properties of the ORS in a Spanish clinical sample. One-hundred and sixty-five adult participants from different primary care centers of the city of Barcelona were recruited. The psychometric properties of the ORS in the sample were explored and described, comparing them to the properties of other instruments already validated in Spain. Our results showed good reliability ($\alpha = .91$ [.88, .93]; $\alpha = .96$; test re-test correlations from .61 to .84), good validity (convergent validity correlations with distress and symptoms measures from $-.32$ to $-.76$), and good sensitivity to change (pre-post comparison through Wilcoxon signed-rank test, $Mdn_{pre} = 31.0$, $Mdn_{post} = 19.6$, $z = -7.38$, $p < .05$, $r = .42$). These results are consistent with previous findings in other countries. We conclude suggesting that the instrument can be applied to monitor outcome in psychotherapy and to test the effectiveness of treatments imparted with Spanish speaking clients. However, further research with the ORS in Spanish could provide more evidence of its psychometric properties.

Received 3 April 2017; Revised 25 June 2018; Accepted 29 June 2018

Keywords: Outcome Rating Scale, psychometrics, reliability, validity.

This research is aimed at studying the psychometric properties of the Spanish Outcome Rating Scale (ORS) (Harris, 2010; Miller, Duncan, Sparks, & Claud, 2003) in a Spanish clinical sample. The ORS is an instrument that assesses well-being, designed for routine outcome monitoring (ROM) in psychotherapy (Boswell, Kraus, Miller, & Lambert, 2015). It was developed as an ultra-brief alternative to the Outcome Questionnaire 45.2 (OQ–45.2) (Lambert et al., 1996). The OQ–45.2 is constituted by three scales: Individual, relational and social functioning, corresponding to the three main areas that psychotherapy improves as valid indicators of treatment outcome. These areas were incorporated into the ORS as three visual analogue scales to assess individual, relational and social well-being. A fourth item was added, general well-being.

Several studies have been conducted to study the psychometric properties of the ORS in different countries and cultures, been translated into several languages, such as Slovak (Biescad & Timulak, 2014), Dutch (Hafkenscheid, Duncan, & Miller, 2010; Janse, Boezenhilberdink, van Dijk, Verbraak, & Hutschemaekers, 2014), and Spanish (Harris, 2010; Donoso & Grez, 2006). The ORS has also been applied to several kind of populations, like students (Bringhurst, Watson, Miller, & Duncan, 2006), primary-care users (DeSantis, Jackson, Duncan, & Reese, 2017), rural clinical samples (Campbell & Hemsley, 2009), and in different formats, e.g. sign laguage (Munro & Rodwell, 2009). To date, all these studies of the ORS in English and other languages document similar psychometric properties.

Furthermore, the ORS has been utilized in different studies, mainly in those aimed at demonstrating the clinical usefulness, efficacy and effectiveness of feedback informed treatments and ROM (Anker, Duncan, & Sparks, 2009; Anker, Owen, Duncan, & Sparks, 2010; Miller, Duncan, & Sorrell, 2006; Overington & Ionita, 2012; Reese, Norsworthy, & Rowlands, 2009; Reese, Usher, et al., 2009).

Nowadays in Spain we count on just one instrument standardized and validated to assess psychotherapeutic outcome in routine care, the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) (Feixas et al., 2012; Trujillo et al., 2016). To count on new instruments aimed at providing clinicians with varied tools to evaluate systematically psychotherapeutic outcome, is required to guarantee the effectiveness of treatments imparted. Responding to these needs,

---

Correspondence concerning this article should be addressed to Danilo Moggia, Universitat de Barcelona, Departament de Psicologia Clínica i Psicobiologia. Pg. de la Vall d'Hebron,171, Edifici del Ponent, Facultat de Psicologia, despacho frente aula 2203, 08007 Barcelona (Spain). Telephone: +34–933125867.

E-mail: danilomoggia@ub.edu

this study aims to explore the psychometric properties of the ORS in a Spanish clinical sample, comparing them with the properties of others instruments used in mental health settings already validated in Spain.

Method

Participants and procedures

One-hundred and sixty-five adult participants from different primary care centers of the city of Barcelona were recruited. These participants received psychotherapy in the context of an internship agreement between the University of Barcelona and the Catalan government, earmarked for students of the Master of Cognitive Social Therapy. The latter is a three-year training program that prepares students to do psychotherapy following the models of cognitive-constructivist psychotherapies and systemic family therapy. Participants were referred by their primary care physicians. All participants received an initial assessment before the start of the treatment. Exclusion criteria were to be received by another psychological treatment at the moment of intake, presence of psychotic symptoms, maniac or hypomanic episodes, or suicidal ideation.

The treatment offered is individual psychotherapy of a maximum of sixteen sessions one or two weeks spaced. All therapists were from the third year of the master's degree and were supervised by experienced teachers and senior therapists.

The approval to conduct this study was given by the primary care centers where the study was conducted. All participants referred were asked to give informed consent for the use of the information from the assessment of their treatment progress and for research. Receiving psychological treatment was not conditional on such consent.

Instruments and measures

Outcome Rating Scale (ORS)

Consists of a four-item scale aimed at measuring four areas of client functioning: Individual, interpersonal, social, and general well-being. The items are answered using a ten-centimeter visual analog scale (VAS) ranging from negative at the left pole, to positive at the right pole. To answer each item, the client has to make a vertical mark on the VAS. Using a ruler each mark is measured to obtain the item score. All scores are summed up to calculate the total score that can range from 0 to 40. High scores represent a good level of functioning and well-being; on the contrary, low scores represent a bad level of functioning and distress.

The Spanish version used for this study is the version translated by Rafael S. Harris, Jr., obtained through the website of Scott D. Miller on 2012 (Harris, 2010).

Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) (Core System Group, 1998)

It is an outcome questionnaire that assesses four factors distributed in thirty-four items: Well-being (four items), problems or symptoms (twelve items), general functioning (twelve items) and risk of self-harm and harm to others (six items). There are multiple versions of the instrument in different languages, developed for men and women. Shorter versions also exist (eighteen, ten or five items) for each gender (Feixas et al., 2012). The questionnaire is self-reported, the client has to answer each item on a five-option Likert scale that goes from *never* to *most or all the time.* The direct score varies from 0 to 136. To estimate the final score, the direct score is divided by the total number of items, obtaining a score that goes from 0 to 4. The same procedure is applied on the subscales to obtain the score for each factor. A total score without the risk scale can also be calculated (Evans et al., 2002). High scores represent high levels of psychological distress. The original English version and the Spanish version have good psychometric properties (Trujillo et al., 2016). In this study, for the initial assessment, the version of thirty-four items was applied, while for monitoring during psychotherapy sessions, the short version of eighteen items was applied (Short Form B, CORE-SFB).

Beck Depression Inventory–II (BDI–II) (Beck, Steer, & Brown, 1996; Sanz, Navarro, & Vázquez, 2003; Sanz, Perdigón, & Vázquez, 2003)

Self-reported questionnaire composed of twenty-one items that have to be answered on a Likert scale. The items have three answer options, except for the items sixteen and eighteen that have seven. The BDI–II assess two factors related to the DSM–IV diagnostic criteria for dysthymia and major depressive disorder: Somatic or motivational factor, and cognitive or affective factor. The total score goes from 0 to 63, and is calculated summing all items. According to the distribution of the scores in clinical samples, four groups of severity can be obtained: Minimal depression (0–13), mild depression (14–19), moderated depression (20–28), and severe depression (29–63) (Sanz, Navarro, et al., 2003; Sanz, Perdigón, et al., 2003).

Depression, Anxiety and Stress Scale–21 (DASS–21) (Lovibond & Lovibond, 1995a; Lovibond & Lovibond, 1995b)

It is a shorter version of the original forty-two-items DASS (Lovibond & Lovibond, 1995b). DASS–21 is a self-reported questionnaire structured in three subscales: Depression, anxiety and stress. Each sub-scale is composed of seven items that have to be answered on a four-option Likert scale (from 0 to 3). The score of

each scale is obtained by adding the answers of the items of the scale, and multiplying it by two. The score of each scale goes from 0 to 42. High scores in each scale represent high levels of depression, anxiety and stress, respectively.

Data analyses

Traditional analyses to study psychometric properties were performed: Internal consistency, test-retest reliability, convergent validity, and sensitivity to change.

Internal reliability was reported as Cronbach's alpha considering the first administration ($n = 147$), and all sample administrations ($n = 1,875$) with no missing item data in both cases. Confidence intervals for Cronbach's alpha for the first administration were computed through the method proposed by Feldt (1965) (Feldt, Woodruff, & Salih, 1987).

Test-retest reliability was analyzed correlating the scores of one administration with the next, considering from the first to the fourth session. Convergent validity was analyzed between ORS and the other instruments at initial assessment and session by session. These two analysis were performed through nonparametric correlations (Spearman's rho) due to scores did not conform to normal distribution according to normality tests.

Sensitivity to change was estimated through Wilcoxon signed-rank test, considering the first and the last session of therapy. A non-parametric hypothesis testing procedure was chosen because, while it is true that the scores distribution did not show statistically significant heteroscedasticity, it was not Gaussian. Bootstrapped 95% confidence intervals (CI) for the difference in means and the effect sizes (as Pearsons's correlation coefficient r) were also computed. All the analysis were performed through the software IBM SPSS 24.

Results

Characteristics of the sample

This clinical sample ($N = 165$) was made up of 120 (72.7%) women, and 45 (27.3%) men. The age of the participants ranged from 18 to 81 years, with a mean of 43.57 ($SD = 13.3$), and they presented a variety of psychological problems (Table 1).

From the overall sample, two participants did not start the treatment after intake. One-hundred and sixty-three participants received at least one session of psychotherapy. The mean number of sessions was 12.2 ($SD = 5.0$). There were 62 therapist participants who saw a mean of 3 ($SD = 1.5$) clients each. The mean, standard deviation and confidence intervals for each scale and total score of the instruments administered at intake and first session of therapy are shown in Table 2.

Table 1. *Clinical Characteristics of the Sample*

	Frequency (%)
Somatoform disorders	1 (0.6)
Eating disorders	3 (1.8)
Substance abuse	3 (1.8)
Psychological problems due to medical conditions	6 (3.6)
Personality disorders	9 (5.5)
Adjustment disorder	13 (7.9)
Life distress	36 (21.8)
Mood disorders	40 (24.2)
Anxiety disorders	54 (32.7)

Internal consistency

Cronbach's alpha [95% CI] for the first administration ($n = 147$) and all administrations in the sample ($n = 1,875$), results in $\alpha = .91$ [.88, .93] for the first case, and $\alpha = .96$ for the second.

Test-retest reliability

It was estimated considering the scores at each administration from the first to the fourth session, correlating one score with the score at subsequent administration (Table 3).

Convergent validity

It was estimated through correlations between the instruments administered at intake and the ORS administered at the first session (Tables 4 and 5).

Furthermore, correlations between the ORS and CORE-SFB administered during all sessions were calculated (Table 6).

All correlations reported were statistically significant ($p < .05$).

Sensitivity to change

Through Wilcoxon signed-rank test, the total scores of the ORS and CORE-SFB from the first and last session of therapy were compared. Total scores of the ORS at last session ($Mdn = 31.0$) were significantly higher than total scores at first session ($Mdn = 19.6$), $z = -7.38$, $p < .05$, $r = .42$. For CORE-SFB, total scores at first session (all items, $Mdn = 1.55$; non-risk items, $Mdn = 1.73$) were significantly higher than total scores at last session (all items, $Mdn = 1.11$; non-risk items, $Mdn = 1.25$), for all items, $z = -4.94$, $p < .05$, $r = -.32$, and for non-risk items, $z = -5.04$, $p < .05$, $d = -.33$.

Bootstrapped 95% CI for the difference in means and for the effect sizes were also considered. Both instruments revealed a statistically significant improvement of the participants, with medium effect sizes (Table 7).

Table 2. *Mean, Standard Deviation, and 95% Confidence Interval of CORE-OM, DASS–21, and BDI–II at Intake, CORE-SFB and ORS at First Session*

	μ	σ	95% CI
Intake			
CORE-OM Subjective well-being	2.30	0.88	[2.02, 2.43]
CORE-OM Problem/Symptoms	2.03	0.82	[1.79, 2.20]
CORE-OM Functioning	1.61	0.75	[1.43, 1.80]
CORE-OM Risk	.34	0.51	[0.22, 0.45]
CORE-OM Non-risk items	1.87	0.76	[1.69, 2.05]
CORE-OM All ítems	1.65	0.69	[1.42, 1.79]
DASS–21 Depression	17.60	10.18	[14.98, 20.19]
DASS–21 Anxiety	14.27	10.01	[11.95, 17.07]
DASS–21 Stress	19.69	10.39	[16.54, 21.43]
BDI–II	23.71	12.10	[21.48, 27.22]
First Session			
CORE-SFB Subjective well-being	2.34	0.98	[2.11, 2.56]
CORE-SFB Problem/Symptoms	1.96	0.89	[1.75, 2.16]
CORE-SFB Functioning	1.50	0.90	[1.30, 1.71]
CORE-SFB Risk	.21	0.41	[0.12, 0.30]
CORE-SFB Non-risk items	1.88	0.84	[1.68, 2.07]
CORE-SFB All items	1.68	0.74	[1.51, 1.86]
ORS individual	4.02	2.70	[3.62, 4.50]
ORS interpersonal	5.22	2.85	[4.79, 5.72]
ORS social role	4.57	2.87	[4.10, 5.04]
ORS overall	4.42	2.56	[4.05, 4.88]
ORS total	18.18	9.91	[16.75, 19.97]

Note: μ = mean; σ = standard deviation; CI = confidence interval.

Table 3. *Test- Retest Correlations*

	1st–2nd	2nd–3rd	3rd–4th
ORS	.61	.77	.84
CORE-SFB All items	.81	.84	.83
CORE-SFB Non-risk items	.81	.84	.85

Note: Spearman rho correlation.

Table 4. *Correlation Coefficients between ORS Items, DASS–21 Subscales, and BDI–II*

	DASS–21 Depression	DASS–21 Anxiety	DASS–21 Stress	BDI–II
ORS individual	–.37	–.23	–.48	–.52
ORS interpersonal	–.32	–.30	–.35	–.44
ORS social role	–.41	–.24	–.40	–.49
ORS overall	–.45	–.34	–.52	–.49
ORS total	–.43	–.32	–.49	–.57

Note: Spearman rho correlation.

Discussion

The present research studied the psychometric properties of the ORS in a Spanish clinical sample, being the first study of this instrument in Spain. Their properties in the sample were described, comparing them to the properties of other instruments already standardized in the country.

Analyzing our results, the total score of the ORS at first administration are similar to those found in other clinical samples at the intake (Anker et al., 2009; Anker et al., 2010; Biescad & Timulak, 2014; Hafkenscheid et al., 2010; Janse et al., 2014; Miller et al., 2003; Reese, Norsworthy, et al., 2009). Total scores of the CORE-OM at intake are consistent with the scores of the clinical sample in the study of Trujillo et al. (2016) at the same point administration (Non-risk items,

$M = 1.86$ $SD = .78$ CI = .84, 1.05; All items, $M = 1.62$ $SD = .71$ CI = .75, .94).

In terms of internal consistency, the results show that is strong, with a high homogeneity and cohesion of their items. The latter is in line with the findings of other studies (Anker et al., 2010; Bringhurst et al., 2006; Hafkenscheid et al., 2010; Janse et al., 2014; Miller et al., 2003; Reese, Usher, et al., 2009), including previous Spanish translation (Donoso & Grez, 2006) and the phenomena observed with other ultra-brief scales (Boulet & Boss, 1991; Seidel, Andrews, Owen, Miller, & Buccino, 2017). Cronbach's alpha coefficients are similar

Table 5. *Correlation Coefficients between ORS Items and CORE–OM Subscales*

	CORE-OM	CORE-OM	CORE-OM	CORE-OM	CORE-OM	CORE-OM
	Subjective well-being	Problem/Symptoms	Functioning	Risk	Non-risk items	All items
ORS individual	−.63	−.60	−.59	−.42	−.64	−.63
ORS interpersonal	−.48	−.45	−.58	−.34	−.53	−.55
ORS social role	−.53	−.51	−.65	−.36	−.63	−.61
ORS overall	−.64	−.63	−.57	−.32	−.66	−.65
ORS total	−.64	−.63	−.68	−.40	−.70	−.70

Note: Spearman rho correlation.

Table 6. *Correlation coefficients between ORS items and CORE-SFB subscales*

	CORE-SFB	CORE-SFB	CORE-SFB	CORE-SFB	CORE-SFB	CORE-SFB
	Subjective well-being	Problem/Symptoms	Functioning	Risk	Non-risk items	All items
ORS individual	−.74	−.66	−.68	−.22	−.75	−.75
ORS interpersonal	−.67	−.61	−.68	−.19	−.70	−.70
ORS social role	−.66	−.60	−.67	−.16	−.70	−.70
ORS overall	−.73	−.67	−.69	−.24	−.75	−.75
ORS total	−.73	−.66	−.71	−.21	−.76	−.76

Notes: Spearman rho correlation.

Table 7. *Pre-post Bootstrapped Comparisons for ORS and CORE*

	μ pre-test (σ)	μ pre-test 95% CI	μ post-test (σ)	μ post-test 95% CI	μ difference 95% CI	Effect Size (Pearson *r*)	Effect Size 95% CI
ORS	19.12 (10.42)	[16.75, 21.64]	26.55 (11.61)	[23.69, 29.38]	[3.66, 11.20]	.32	[.16, .45]
CORE-SFB All items	1.65 (0.73)	[1.48, 1.82]	1.18 (0.69)	[1.02, 1.34]	[−0.71, −0.23]	−.31	[−.45, −.15]
CORE-SFB Non-risk items	1.84 (0.83)	[1.65, 2.03]	1.31 (0.77)	[1.14, 1.50]	[−0.80, −0.26]	−.31	[−.45, −.15]

Note: μ = mean; σ = standard deviation; CI = confidence interval.

to those found in the studies of Campbell & Hemsley (2009) (α = .90), and Donoso & Grez (2006) (α ranged from .91 to .96); and higher than what was found by Anker et al. (2009) (α = .83), Biescad & Timulak (2014) (α = .87), and Reese, Norsworthy, et al. (2009) (α = .88; .84).

In relation to test-retest reliability, coefficients are adequate and higher than reported in previous studies, e.g. Janse et al. (2014) (.64; .57; .69), but lower than observed in CORE-SFB. Test-retests correlations of the CORE-SFB are similar than those reported by Trujillo et al. (2016) for CORE-OM. This could be due to shorter measures nearly always having lower correlations than much longer measures such as CORE. In this sense, the CORE might be capturing more stable aspects of people's distress than ORS. However, in previous findings, ORS appears to be less sensitive

than CORE considering recovered and improved clients as separate groups, but more sensitive than CORE when the whole group is considered (Biescad & Timulak, 2014).

Regarding concurrent validity, the correlations between the total score of the ORS and the subscales of DASS–21 seem to be lower than what was reported by Campbell & Hemsley (2009) (−.71; −.46; −.60). The same is found if we compare the correlation with BDI–II in the study of Biescad & Timulak (2014) (−.73). This could be due to answering a questionnaire that asks about self-assessed well-being through a VAS, it may be capturing different areas of mental health functioning, rather than traditional measures that focus on symptoms and discomfort. Nevertheless, correlations between ORS and CORE are strong, showing that both instruments are measuring similar aspects of the same underlying construct (well-being and/or psychological distress).

Concerning sensitivity to change, the ORS was able to capture the improvement of the participants, being the effect size of the ORS similar to the effect sizes of CORE-SFB. However, in the study of Biescad & Timulak (2014) the effect size of the ORS seems to be lower ($d = .87$) than the effect sizes of the CORE-OM (Non-risk items, $d = .95$; All items, $d = .98$).

The limitations of this study are that it only assesses the psychometric properties in a clinical sample, so no-comparison with non-clinical subjects could be done. In this sense, nonrandom sample frame was applied, and the sample size is small. Another limitation is that the version of the ORS applied was the version translated to Spanish by the team of the original authors, without contrasting if this translation applies to Peninsula Spanish speakers, so generalizability has to be done with caution. Furthermore, the method to assess test-retest reliability is not totally accurate, because during the interval of each administration, the effect of therapy or external factors might be expected to produce change. Duncan et al. (2003) have argued that lower test-retest reliability can be obtained in instruments that are sensitive to change.

In summary, this article presents the first study of the ORS in Spain, showing that the instrument seems to be valid to assess well-being and psychotherapeutic outcome, being useful to obtain feedback about the progress of the client during the treatment. The ORS can be an instrument that clinicians can apply to monitor clients' outcome and to prove the effectiveness of treatments imparted with Spanish speaking clients. Further research is required to adapt and standardize the ORS to Spanish population.

References

Anker M. G., Duncan B. L., & Sparks J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology*, 77(4), 693–704. https://doi.org/10.1037/a0016062

Anker M. G., Owen J., Duncan B. L., & Sparks J. A. (2010). The alliance in couple therapy: Partner influence, early change, and alliance patterns in naturalistic sample. *Journal of Consulting and Clinical Psychology*, 78(5), 635–645. https://doi.org/10.1037/a0020051

Beck A. T., Steer R. A., & Brown G. K. (1996). *Beck Depression Inventory–II Manual (2nd Ed.)*. San Antonio, TX: The Psychological Corporation.

Biescad M., & Timulak L. (2014). Measuring psychotherapy outcomes in routine practice: Examining Slovak versions of three commonly used outcome instruments. *European Journal of Psychotherapy & Counselling*, 16(2), 140–162. https://doi.org/10.1080/13642537.2014.895772

Boswell J. F., Kraus D. R., Miller S. D., & Lambert M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, 25(1), 6–19. https://doi.org/10.1080/10503307.2013.817696

Boulet J., & Boss M. (1991). Reliability and validity of the Brief Symptom Inventory. *Journal of Consulting and Clinical Psychology*, 3, 433–437. https://doi.org/10.1037/1040-3590.3.3.433

Bringhurst M. D. L., Watson C. W., Miller S. D., & Duncan B. L. (2006). The reliability and validity of the Outcome Rating Scale: A replication study of a brief clinical measure. *Journal of Brief Therapy*, 5(1), 23–30.

Campbell A., & Hemsley S. (2009). Outcome Rating Scale and Session Rating Scale in psychological practice: Clinical utility of ultra-brief measures. *Clinical Psychologist*, 13(1), 1–9. https://doi.org/10.1080/13284200802676391

Core System Group (1998). *CORE System (information management) handbook.* Leeds, UK: Core System Group.

DeSantis B., Jackson M. J., Duncan B. L., & Reese R. J. (2017). Casting a wider net in behavioral health screening in primary care: A preliminary study of the Outcome Rating Scale. *Primary Health Care Research & Development*; 18, 188–193. https://doi.org/10.1017/S1463423616000311

Donoso C., & Grez M. A. (2006). *Chilean adaptation and validation of the Outcome Rating Scale (ORS)*. Santiago de Chile, Chile: Universidad Gabriela Mistral.

Duncan B. L., Miller S. D., Sparks J. A., Claud D. A., Reynolds L. R., Brown J., & Johnson L. D. (2003). The Session Rating Scale: Preliminary psychometric properties of a "working" alliance measure. *Journal of Brief Therapy*, 3(31), 3–12.

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. *The British Journal of Psychiatry*, 180(1), 51–60. https://doi.org/10.1192/bjp.180.1.51

Feixas G., Evans C., Trujillo A., Saúl L. A., Botella L., Corbella S., … López-González M. A. (2012). La versión española del CORE-OM: Clinical Outcomes in Routine Evaluation - Outcome Measure [Spanish version of CORE-OM: Clinical Outcomes in Routine evaluation - Outcome Measure]. *Revista de Psicoterapia*, 23(89), 109–135.

Feldt L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika*, 30, 357–370. https://doi.org/10.1007/BF02289499

Feldt L. S., Woodruff D. J., & Salih F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11(1), 93–103. https://doi.org/10.1177/014662168701100107

Hafkenscheid A., Duncan B. L., & Miller S. D. (2010). The Outcome and Session Rating Scales: A cross-cultural examination of the psychometric properties of the Dutch translation. *Journal of Brief Therapy*, 7, 1–12.

Harris R. S., Jr., (2010). *Encuesta de Resultado* [Outcome Rating Scale] [Measurement instrument]. Retrieved from https://betteroutcomesnow.com/wp-content/uploads/2018/01/spanish.docx

Janse P., Boezen-Hilberdink L., van Dijk M. K., Verbraak M. J. P. M., & Hutschemaekers G. J. M. (2014). Measuring feedback from clients. The psychometric properties of the Dutch Outcome Rating Scale and Session Rating Scale. *European Journal of Psychological Assessment*, 30(2), 86–92. https://doi.org/10.1027/1015-5759/a000172

Lambert M. J., Hansen N. B., Umphress V., Lunnen K., Okiishi J., Burlingame G. M., … Reisigner G. (1996). *Administration and scoring manual for the Outcome Questionnaire (OQ–45.2)*. Stevenson, MD: American Professional Credentialing Services.

Lovibond P. F., & Lovibond S. H. (1995a). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*, 335–343. https://doi.org/10.1016/0005-7967(94)00075-U

Lovibond S. H., & Lovibond P. F. (1995b). *Manual for Depression, Anxiety, Stress Scales (2nd ed.)*. Sydney, Australia: Psychology Foundation.

Miller S. D., Duncan B. L., & Sorrell R. (2006). Using formal client feedback to improve retention and outcome: Making ongoing, real-time assessment feasible. *Journal of Brief Therapy*, *5*(1), 5–22.

Miller S. D., Duncan B. L., Sparks J. A., & Claud D. A. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy*, *2*(2), 91–100.

Munro L., & Rodwell J. (2009). Validation of an Australian sign language instrument of outcome measurement for adults in mental health settings. *Australian and New Zealand Journal of Psychiatry*, *43*(4), 332–339. https://doi.org/10.1080/00 048670902721111

Overington L., & Ionita G. (2012). Progress monitoring measures: A brief guide. *Canadian Psychology*, *53*(2), 82–92. https://doi.org/10.1037/a0028017

Reese R. J., Norsworthy L. A, & Rowlands S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy*, *46*(4), 418–431. https://doi.org/10.1037/a0017901

Reese R. J., Usher E. L., Bowman D. C., Norsworthy L. A., Halstead J. L., Rowlands S. R., & Chisholm R. R. (2009). Using client feedback in psychotherapy training: An analysis of its influence on supervision and counselor self-efficacy. *Training and Education in Professional Psychology*, *3*, 157–168. http://doi.org/10.1037/a0015673

Sanz J., Navarro M., & Vázquez C. (2003). Adaptación española del Inventario para la Depresión de Beck–II (BDI–II): 1. Propiedades psicométricas en estudiantes universitarios [Spanish adaptation of the Beck Depression Inventory–II (BDI–II): 1. Psychometric properties in university students]. *Análisis y Modificación de Conducta*, *29*, 239–288.

Sanz J., Perdigón A., & Vázquez C. (2003). Adaptación española del Inventario para la Depresión de Beck–II (BDI–II): 2. Propiedades psicométricas en población general [Spanish adaptation of the Beck Depression Inventory–II (BDI–II): 2. Psychometric properties in general population]. *Clínica y Salud*, *14*, 249–280.

Seidel J. A., Andrews W. P., Owen J., Miller S. D., & Buccino D. L. (2017). Preliminary validation of the Rating of Outcome Scale and equivalence of ultra-brief measures of well-being. *Psychological Assessment*, *29*(1), 65–75. https://doi.org/10.1037/pas0000311

Trujillo A., Feixas G., Bados A., García-Grau E., Salla M., Medina J. C., … 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