

EVALUACIÓN EN PSICOTERAPIA: PROPIEDADES PSICOMÉTRICAS DE LA VERSIÓN ARGENTINA DE LA ESCALA DE CALIFICACIÓN DE RESULTADOS Y LA ESCALA DE CALIFICACIÓN DE SESIONES

PSYCHOTHERAPY ASSESSMENT: PSYCHOMETRIC PROPERTIES OF DE ARGENTINEAN VERSION OF THE OUTCOME RATING SCALE AND SESSION RATING SCALE

Pablo Rafael Santangelo

ORCID: <https://orcid.org/0000-0001-5515-2923>

Universidad Nacional de Mar del Plata. Argentina
CAPsi, Centro de Asistencia Psicológica Mar del Plata
Fundación Aiglé

Karina Conde

ORCID: <https://orcid.org/0000-0003-1913-8328>

Instituto de Psicología Básica, Aplicada y Tecnología (IPSIBAT),
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).
Universidad Nacional de Mar del Plata. Argentina

Helga Schupp

ORCID: <https://orcid.org/0000-0002-3487-5134>

Universidad Nacional de Mar del Plata
CAPsi, Centro de Asistencia Psicológica Mar del Plata. Argentina

Natalia Paoloni

ORCID: <https://orcid.org/0000-0002-1112-3052>

Universidad Nacional de Mar del Plata. Argentina

Cómo referenciar este artículo/How to reference this article:

Santangelo, P. R., Conde, K., Schupp, H. y Paoloni, N. (2021). Evaluación en psicoterapia: Propiedades psicométricas de la versión Argentina de la Escala de Calificación de Resultados y la Escala de Calificación de Sesiones. *Revista de Psicoterapia*, 32(119), 165-180. <https://doi.org/10.33898/rdp.v32i119.503>



Resumen

Los instrumentos que ponderan alianza terapéutica y resultados en psicoterapia son esenciales, ya que los profesionales tienden a presentar sesgos de juicio en sus evaluaciones. Sin embargo, en muchos países de habla hispana son escasos. En este estudio, se evaluó la Escala de calificación de resultados (ORS por sus siglas en inglés) y la Escala de calificación de sesiones (SRS 3.0 por sus siglas en inglés) en el contexto argentino. Las dos escalas, breves, contienen ítems que se responden usando una escala visual analógica de 10 centímetros, la ORS evalúa el progreso del paciente y la SRS la alianza terapéutica. Método: Se contó con una muestra clínica de $n = 167$ pacientes. Las escalas fueron aplicadas en cuatro momentos del proceso psicoterapéutico. Se analizó las puntuaciones de la ORS según sexo y edad, y se evaluó el Índice de Cambio Confiable (ICC). Se estimó la fiabilidad y validez concurrente en ambas escalas y la validez predictiva en SRS 3.0. Resultados: No se encontraron diferencias de sexo ni de edad en la ORS, con un ICC = 5.61. La consistencia interna fue mayor a $\alpha = .80$ en las dos escalas, la correlación test-retest fue moderada-alta y se encontró evidencia de validez concurrente. Las características que la ORS y la SRS 3.0 presentan para el contexto argentino parecen ser similares a las escalas originales, incluida el ICC, presentando potencial para aplicarse a nivel práctico.

Palabras claves: psicoterapia, ORS, SRS 3.0, psicometría, Argentina.

Abstract

Instruments that assess therapeutic alliance and the psychotherapy results are essential since professionals tend to present judgment biases in their evaluations. However, in many Spanish-speaking countries, they are rare. In this study, we evaluated the psychometric properties of the Outcome Rating Scale (ORS) and Session Rating Scale (SRS 3.0) in the Argentine context. Both brief scales have ten items with analog responses. The ORS assesses the patient's progress, and the SRS 3.0 the therapeutic alliance. Method: In a clinical sample of $n = 167$ patients, we applied the scales in four moments of psychotherapy. We analyzed the ORS scores according to gender and age, and we evaluated the Reliable Change Index (RCI). We estimated the reliability and concurrent validity in both scales and predictive validity in SRS. Results: No sex or age differences were found in the ORS scale, with an RCI = 5.61. The internal consistency was higher than $\alpha = .80$ in the scales, and the test-retest correlation was moderate-high. We found evidence of concurrent validity. Discussion: The characteristics that the ORS and the SRS 3.0 present for the Argentine context seem to be similar to the original scales, including the RCI and could be applied at a practical level.

Keywords: psychotherapy, ORS, SRS 3.0, psychometry, Argentina.

Es frecuente que los resultados de los procesos psicoterapéuticos se evalúen según el juicio de los profesionales. Sin embargo, los estudios indican que los terapeutas tienden a sobreestimar las tasas de mejora en sus pacientes y subestimar las tasas de deterioro (Chapman et al., 2012; Walfish et al., 2012). Por ello, se han creado distintos sistemas para evaluar de manera sistemática y objetiva al proceso terapéutico.

El Sistema de Monitoreo Rutinario de Resultados (ROM, por sus siglas en inglés) ha surgido como un método para abordar los problemas de terminación prematura, deterioro y falta de mejora en los pacientes (Miller et al., 2013; Lambert, 2015). Dentro de las distintas alternativas del ROM, el Outcome Questionnaire-45 (OQ-45) fue el primero que se implementó como sistema de feedback (Lambert et al., 2001). Otra de las alternativas ROM es el Sistema de Gestión de Resultados Socios para el Cambio (PCOMS, por sus siglas en inglés), un programa de retroalimentación para mejorar la efectividad de la psicoterapia en adultos y niños. El PCOMS fue diseñado para disminuir la terminación prematura en psicoterapia y para facilitar el logro de cambios confiables y clínicamente significativos. Está compuesto por dos escalas breves y globales, que se administran todas las sesiones evaluando predictores del éxito terapéutico, la ORS, que evalúa el progreso terapéutico del paciente (a través de las calificaciones del funcionamiento psicológico y el malestar) y el beneficio percibido en el tratamiento por parte del paciente; y la SRS 3.0, que evalúa la percepción del paciente respecto a la alianza (es decir, la calidad del vínculo y si comparten objetivos y medios entre paciente y psicoterapeuta). El psicoterapeuta administra el ORS al comienzo de la sesión de tratamiento, y el SRS 3.0 se administra hacia el final. Las puntuaciones de los pacientes para ambas medidas se analizan sesión por sesión para mantener su participación en el tratamiento, optimizar la alianza terapéutica y proporcionar un medio para que el paciente transmita sus preocupaciones. Si las calificaciones de los pacientes son muy bajas, el terapeuta puede optar por modificar el tipo y la intensidad del tratamiento (SAMHSA's National Registry of Evidence-based Programs and Practices, 2015). En este sentido, la utilización de las escalas ORS y el SRS favorece el diálogo, proporcionando al terapeuta información para luego profundizar sobre los temas que identifica el paciente (Bargmann y Robinson, 2012).

Específicamente la ORS permite una medida global del bienestar psicológico definido por cuatro áreas de funcionamiento: individual, interpersonal, social y general (Boswell et al., 2015), basados en las subescalas del OQ-45 (Chapman et al., 2017; Lambert et al., 1996). La SRS 3.0 es un instrumento que evalúa alianza terapéutica, creado con fines clínicos (Duncan et al., 2003). Tres de las dimensiones de la SRS 3.0 se basan en la conceptualización de la alianza de Bordin (1979), que incluyen la calidad del vínculo terapéutico, el acuerdo entre terapeuta y paciente de los objetivos y tareas a desarrollar; y una cuarta dimensión, que surge de los estudios de Hatcher y Barends (1996) evalúa como se sintió el paciente en la sesión en un sentido general.

Las dos escalas están diseñadas y normadas para adultos, adolescentes y niños (Bargmann y Robinson, 2012), han sido adaptadas a distintos contextos (Biescad y Timulak, 2014; Cazauvieilh et al., 2020; Hafkenschied et al., 2010; Janse et al., 2014; Moggia et al., 2018) y aplicadas a distintas poblaciones (Campbell y Hemsley, 2009; Bringhurst et al., 2006; DeSantis et al., 2017). Se ha demostrado que el ORS es sensible al cambio entre quienes reciben atención (Bargmann y Robinson, 2012). Numerosos estudios han documentado validez concurrente, discriminativa, relacionada con criterios y predictiva, confiabilidad test-retest y confiabilidad de consistencia interna (Anker et al., 2009; Campbell y Hemsley, 2009; Miller et al., 2006; Reese et al., 2010).

En síntesis, la información de escalas como la ORS y SRS 3.0 permite al psicoterapeuta hacer ajustes necesarios que beneficien el tratamiento como por ejemplo, intentar mejoras en la alianza terapéutica, en las intervenciones o intensidad del tratamiento, como también evaluar la posibilidad de una derivación en la medida que los cambios esperados no aparezcan. A estos beneficios se suman la brevedad y las propiedades psicométricas sólidas que han presentado en otros contextos, incluso en países de habla hispana con otra especificidad de dialecto (Moggia et al., 2018, 2020). Sin embargo, en Argentina no contamos con instrumentos breves validados que evalúen específicamente los resultados de tratamientos y la alianza terapéutica de una manera breve (Paz et al., 2021) y que puedan ser usados rutinariamente en el contexto clínico diario, los instrumentos con los que contamos son más extensos y se han creado con fines de investigación provocando que los psicoterapeutas sean reticentes a usarlos de manera rutinaria. Asimismo, estas escalas breves presentan características que las tornan flexibles a las diferentes realidades socio-económicas que atraviesan los países latinoamericanos (Paz et al., 2021).

Por ello, el objetivo de este trabajo es determinar las propiedades psicométricas del ORS y SRS 3.0 en el contexto argentino.

Método

Participantes

La muestra estuvo compuesta por 167 personas pertenecientes a dos grupos de tratamiento. Uno de ellos ($n = 128$), recibía atención en el marco del Programa de Formación y Entrenamiento en Psicoterapia, de la Facultad de Psicología, Universidad Nacional de Mar del Plata. El tratamiento que se brinda está basado en la evidencia (Santangelo, 2020) es gratuito y es para mayores de 18 años sin cobertura médica. El otro grupo de pacientes ($n = 39$) recibía atención privada, eran mayores de 18 años, pero el nivel socioeconómico fue más alto que el grupo anterior. Los criterios de exclusión, en adición a ser menores de 18 años, fueron tener algún trastorno severo como esquizofrenia, trastorno alimentario o consumo problemático de sustancias psicoactivas. Ninguno de los participantes se negó a responder los instrumentos. La edad de los pacientes osciló entre 18 y 74 años, con

una media de $M(DE) = 31.46$ (11.77), 59% mujeres. Recibieron un promedio de 11.97, $DE = 9.43$ sesiones, y los problemas clínicos por los que consultaron fueron muy variados. Los más preponderantes, en orden de frecuencia, fueron: conflictos interpersonales (20%), desarrollo personal (16%), ansiedad (15%), depresión (14%), derivación judicial (8%), y desarrollo de habilidades personales (4%). Los motivos de finalización del tratamiento, que derivaron en la mortalidad de la muestra, más frecuentes fueron: mudanza (20%), derivación a otro dispositivo (16%), abandono sin aviso (15%) y abandono con aviso (14%).

Los psicoterapeutas, 9 mujeres y 2 varones, no tenían más de 3 años de recibidos, con un promedio de edad de 28 años, 4 contaban con un año de experiencia en el ejercicio de la psicoterapia y 7 sin experiencia alguna. Tuvieron una formación de grado mayoritariamente psicoanalítica, pero muchos tomaron cursos de postgrado con orientación cognitiva y orientación sistémica antes de comenzar a trabajar en el proyecto de investigación. Todos los terapeutas formaban parte del Programa de Formación y Entrenamiento en Psicoterapia donde la psicoterapia que se enseña y se ejerce es una psicoterapia basada en la evidencia (Santangelo, 2020). Son los mismos psicoterapeutas que atienden en el programa y de manera privada, la diferencia es que los terapeutas que atendían pacientes pertenecientes al programa recibían supervisión vertical semanal. Los pacientes de los terapeutas que atendían de manera privada no eran supervisados semanalmente, ni de manera vertical. La media de pacientes atendidos por cada terapeuta fue de 5.93 ($DE = 6.55$).

Instrumentos

Escala de Calificación de Resultados (ORS). La ORS es una escala breve que evalúa el progreso del paciente de manera global a partir de cuatro áreas de funcionamiento: individual, interpersonal, social y general. Los ítems se responden a través de una escala visual analógica de 10 centímetros., donde el paciente tiene que colocar sólo una cruz o una marca en una línea de 10 centímetros para decir cuan mal o bien está en cada área. Cuanto más cerca del extremo izquierdo esta la marca representa mayor malestar. Brinda puntaje en cada área de 0 a 10 y un puntaje total a partir de su promedio. El estudio original se realizó en EE.UU, en idioma inglés, la muestra estuvo compuesta por una población clínica ($n = 435$) mayores de 18 años y por población no clínica ($n = 86$), entre 22 y 65 años; y género, nivel socioeconómico y étnico fueron mixtos. La consistencia interna fue de $\alpha = .93$, y la validez concurrente entre las puntuaciones totales del ORS y OQ-45.2 fue de $r = .59$ (Miller et al., 2003).

Escala de Calificación de Sesiones (SRS 3.0). La SRS 3.0 es un instrumento breve que evalúa la alianza terapéutica de manera global a partir de cuatro dimensiones: relación, objetivos y temas, enfoque o método y en general. Los ítems se responden a través de una escala visual analógica de 10 centímetros., donde el paciente tiene que colocar sólo una cruz o una marca en una línea de 10 centímetros para decir como ha sido su experiencia en la sesión. Cuanto más cerca del extremo

izquierdo esta la marca representa mayor disconformidad. Brinda puntaje en cada área de 0 a 10 y un puntaje total a partir de su promedio. El estudio original se realizó en EE.UU, en idioma inglés, la muestra estuvo compuesta por tres grupos seleccionados de distintas agencias de salud y programas de intervención. Grupo uno ($n = 81$), con un rango de edad entre 18 y 74 años. Grupo dos ($n = 100$), las edades oscilaron entre 18 y 83 años. Grupo tres ($n = 156$), no se especifica la edad. La consistencia interna fue de $\alpha = .88$, la confiabilidad test-retest fue de $.64$ y la validez concurrente entre las puntuaciones totales del SRS y la HAQ II fue de $r = .48$ (Duncan et al., 2003).

Una vez obtenido el permiso correspondiente para la traducción la ORS y SRS 3.0 fueron adaptadas siguiendo las Normas de la Comisión Internacional de Tests (Asociación Argentina de Estudio e Investigación en Psicodiagnóstico, 2000) y el método de traducción inversa a fin de generar dos versiones que se correspondan con el idioma local y los estilos lingüísticos propios del contexto en el que se aplicó. Las escalas traducidas fueron sometidas a examen y juicio crítico de un grupo de investigación conformado por psicólogos y un especialista en la construcción de instrumentos de evaluación psicológica; quienes estuvieron de acuerdo con la traducción, la cual no presentó modificaciones de la escala original. Por último, con el fin de evaluar el nivel de comprensión de los ítems, se realizó una prueba piloto con 10 psicólogos y durante la prueba se dio la posibilidad de realizar comentarios respecto de la versión traducida del ORS y SRS. Las escalas están publicadas en el sitio web de Scott D. Miller, <https://www.scottdmiller.com>.

Cuestionario de Resultados (OQ-45.2). El cuestionario OQ-45 fue diseñado específicamente para la evaluación de resultados y el monitoreo del progreso de los pacientes en psicoterapia. Es una prueba autoadministrable, de 45 ítems, con una escala tipo Likert de 5 puntos. Cuenta con tres subescalas: 1) Síntomas: (Symptom Distress, SD), que identifica los desórdenes mentales prevalentes en la población que requiere asistencia psicoterapéutica; 2) Relaciones Interpersonales (Interpersonal Relations, IR), que incluye ítems que evalúa tanto satisfacción como problemas en las relaciones interpersonales; 3) Rol Social (Social Role, SR), evalúa el grado de insatisfacción, conflictos ligados a las tareas relacionadas con el área laboral y el tiempo libre. Para cada una de estas dimensiones se obtiene un puntaje a modo de subescala y además el cuestionario brinda un puntaje total que evalúa el funcionamiento global del paciente. El nivel de correlación test re-test para la población general es $r = .864$, y la consistencia interna es de $\alpha = .923$. El estudio de validez concurrente indica un elevado nivel de correlación con el SCL-90-R y el Inventario de problemas interpersonales (IIP) y gran correspondencia con los estudios de la versión original (Maristany y Fernández-Álvarez, 2010). En este estudio se utilizó la versión validada en Chile por Von Bergen y De la Parra (2002).

Inventario de Alianza de Trabajo breve - versión paciente (WAI - SR - A). Es la adaptación argentina de la versión abreviada del inventario de alianza de trabajo (WAI, Hatcher y Gillaspy, 2006), construido a partir de la definición de

Alianza terapéutica de Bordin (1979). Está compuesto por 12 ítems distribuidos en tres sub-escalas: vínculo, tarea y objetivos. Con una escala tipo Likert de 7 puntos. La consistencia interna en el estudio original en la escala total fue de $\alpha = .91$, de la sub-escala vínculo $\alpha = .90$, de la sub-escala tareas $\alpha = .87$ y de la sub-escala objetivos $\alpha = .87$. En el estudio realizado en Argentina del cual se extrae la versión utilizada (Gómez Penedo et al., 2015) la consistencia interna de la escala total fue de $\alpha = .89$, de la sub-escala vínculo $\alpha = .77$, de la sub-escala tareas $\alpha = .84$ y de la sub-escala objetivos $\alpha = .67$.

Procedimiento

Se realizaron cuatro evaluaciones, de aproximadamente 10 minutos cada una: al inicio del tratamiento (S1, $n = 167$), en la tercera sesión (S3, $n = 92$), en la octava sesión (S8, $n = 50$) y en la sesión final (SF, $n = 33$). La pérdida de la muestra en las sucesivas administraciones corresponde a terminaciones prematuras de tratamiento, definidas como la terminación del tratamiento unilateralmente por parte del paciente (con o sin aviso) o a la conclusión de objetivos, es decir, cierre de tratamiento por logro de los mismos. Se les entregó información general sobre el estudio y datos de contacto de los investigadores. Se obtuvo consentimiento informado escrito, lo que no era un requisito para recibir el tratamiento. La investigación contó con el aval del Programa Temático Interdisciplinario en Bioética de la Universidad Nacional de Mar del Plata. Todos los sujetos participaron voluntariamente y no recibieron ningún tipo de compensación. Se resguardó el anonimato y la confidencialidad de las respuestas.

Análisis de los Datos

Los análisis preliminares incluyeron la determinación del tipo de distribución de los datos. Con excepción de la escala Rol Social, tanto para el ORS como para el OQ-45.2 las distribuciones fueron normales, prueba de Kolmogorov-Smirnov (K-S) $>.05$. Lo opuesto sucedió para la SRS 3.0 y la WAI, cuyas distribuciones no resultaron paramétricas (K-S $<.05$).

Dado que se evaluaron dos escalas analógicas de una escasa cantidad de ítems, el procedimiento de validación incluyó las medidas que se detallan a continuación, adaptando el procedimiento de Miller y Duncan (2003). En primer lugar, se compararon los puntajes del ORS para la S1 y SRS para S3 según el género y la edad. Para estos análisis se utilizó una prueba t bilateral para muestras independientes (género) y una correlación producto momento de Pearson (edad). Luego estimó la consistencia interna de ORS y SRS 3.0 mediante el coeficiente Alpha de Cronbach. Posteriormente, se estableció la validez concurrente de ambas escalas y el OQ-45.2 y la WAI respectivamente, mediante correlaciones bivariadas (Pearson o Spearman según distribución).

La correlación test-retest entre S1 y S3, S3 y S8, y S8 y SF se evaluó mediante correlaciones bivariadas (Pearson o Spearman). Asimismo, es esperable que el

instrumento sea capaz de estimar la mejoría de los pacientes. Se estimó entonces la sensibilidad al cambio, estableciendo diferencias entre S1 y SF (comparación pre y post-test, prueba t bilateral para muestras independientes). Para determinar el Índice de Cambio Confiable (ICC) entre el pre-test (S1) y post-test (S8) se utilizó la fórmula de Jacobson y Truax (1991). El ICC indica el cambio que es mayor que el producto de la casualidad, el error o la maduración del paciente, siendo de 6 en la versión original. El cambio que excede el ICC y cruza el corte clínico (25 en el estudio original) se considera clínicamente significativo (Miller et al., 2003).

Existen evidencias de que la calidad de la relación terapéutica influencia el resultado de los tratamientos (Flückiger et al., 2019). Por ello, se evaluó la validez predictiva de la SRS 3.0 mediante regresiones lineales entre la S3 y S8, y el puntaje total de SF del OQ-45.2.

Para los análisis se utilizó el SPSS versión 15, y para el cálculo del ICC se utilizó la aplicación de Morley y Dowzer (2014).

Resultados

Escala de Calificación de Resultados

Puntajes normativos. La media del puntaje total fue de 22.06, $DE = 9.88$, de la escala Individual 4.96, $DE = 2.59$, Interpersonal 5.66, $DE = 2.63$, Rol Social 5.97, $DE = 4.31$, General 4.49, $DE = 2.66$. No se hallaron diferencias respecto del género (mujeres: $n = 94$, $M(DE) = 21.99 (10.53)$, varones: $n = 67$, $M(DE) = 21.99 (8.86)$, $t(159) = -.004$, $p = .997$, $d = .01$ IC95% [-.31, .31]), ni relación entre los puntajes y la edad de los pacientes ($r = .09$, $p > .05$).

Índice Cambio Confiable. El ICC se estableció en 5.61, lo cual significa que los pacientes deben alcanzar al menos una mejoría de 5.61 puntos para que se trate de un cambio significativo.

Fiabilidad. La fiabilidad de la escala fue aceptable, encontrando una consistencia interna en S1 de $\alpha = .8$, en S3 de $\alpha = .89$, en S8 de $\alpha = .88$ y en SF de $\alpha = .92$. La correlación test-retest entre S1 y S3 fue moderada y significativa ($r = .55$, $p < .001$), y lo mismo ocurrió entre la S3 y la S8 ($r = .51$, $p < .001$), y la S8 y SF ($r = .69$, $p < .001$).

Validez Concurrente. La correlación entre los totales de la ORS y el OQ-45.2 para cada administración fue moderadamente fuerte en todas las administraciones (tabla 1). Los coeficientes fueron negativos debido a que menores puntajes en el OQ-45.2 indican mayor bienestar, y lo opuesto ocurre en la ORS. Los coeficientes de correlación entre las dimensiones y los totales de la ORS y el OQ-45.2 fueron en general altos (tabla 2).

Tabla 1
Coeficientes de correlación entre la ORS y el OQ-45, muestra clínica, Argentina, 2019

1 ^{era} Sesión (n=162)	3 ^{era} Sesión (n=91)	8 ^{ava} Sesión (n=50)	Sesión Final (n=32)
-.64**	-.66**	-.77**	-.71**

Nota: r de Pearson; ** $p < .01$; ORS= Outcome Rating Scale. OQ= Outcome Questionnaire.

Tabla 2
Coeficientes de correlación entre los totales y las dimensiones del ORS y el OQ-45, 1^{era} sesión, muestra clínica (n = 167), Argentina, 2019

	OQ-45 Malestar Subjetivo ^a	OQ-45 Interpersonal ^a	OQ-45 Rol Social ^b	OQ-45 Total ^a
ORS Individual ^a	-.53**	-.46**	-.39**	-.55**
ORS Interpersonal ^a	-.48**	-.54**	-.34**	-.53**
ORS Rol Social ^b	-.52**	-.51**	-.49**	-.57**
ORS General ^a	-.61**	-.56**	-.49**	-.65**
ORS Total ^a	-.60**	-.58**	-.50**	-.64**

Nota: ^a r de Pearson; ^b rho de Spearman; ** $p < .01$; * $p < .05$; ORS= Outcome Rating Scale; OQ= Outcome Questionnaire.

Sensibilidad al cambio. Por último, se hallaron diferencias significativas entre el pre-test (S1, $M(DE) = 22.94 (9.19)$ y el post-test (SF, $M(DE) = 26.19 (8.07)$, $t(198) = -4.53$, $p = .042$, $d = -.37$, IC95% [-.69, -.04]), indicando que la prueba es capaz de detectar cambios a lo largo del tiempo.

Escala de Calificación de Sesiones

Puntajes normativos. No se hallaron diferencias respecto del género (mujeres: $n = 52$, $M(DE) = 36.77(3.91)$, varones: $n = 36$, $M(DE) = 36.87(4.40)$, $t(86) = -.11$, $p = .913$, $d = -.02$ IC95% [-.45, .40]), y tampoco asociación entre los puntajes y la edad de los pacientes ($\rho = .09$, $p > .05$). La tabla 3 presenta los datos normativos de la escala, siendo en general los puntajes estables en las distintas administraciones.

Tabla 3
Puntajes totales de la SRS para una muestra clínica, Argentina, 2019

	S1 (n=90)		S8 (n=50)		SF (n=28)	
	M(DE)	IC95%	M(DE)	IC95%	M(DE)	IC95%
Relación	9.36(.91)	9.15, 9.54	9.17(1.61)	8.65, 9.57	9.15(1.02)	8.75, 9.49
Objetivos	9.19(1.13)	8.93, 9.41	9.14(1.65)	8.61, 9.53	8.94(1.41)	8.38, 9.37
Enfoque	9.22(1.11)	8.97, 9.44	9.26(1.58)	8.74, 9.62	9.21(.94)	8.84, 9.54
General	9.06(1.38)	8.74, 9.32	8.81(2.14)	8.17, 9.32	9.15(1.08)	8.69, 9.51
Total	36.83(4.06)	35.93, 37.6	36.37(6.51)	34.34, 37.9	36.40(4.13)	34.79, 37.8

Nota: SRS 3.0= Session Rating Scale.

Fiabilidad. La fiabilidad de la escala presentó resultados óptimos, encontrando una consistencia interna en S3 de $\alpha = .91$, en S8 de $\alpha = .94$ y en SF de $\alpha = .93$. La correlación test-retest fue alta ($\rho = .71, p < .001$) entre la S3 y la S8, y no significativa entre la S8 y SF ($\rho = .39, p = .07$).

Validez Concurrente. La correlación entre los totales de la SRS 3.0 y la WAI para cada administración fue moderada a fuerte en todas las administraciones, (S3, $\rho = .48, p < .001$; S8, $\rho = .67, p < .001$; SF, $\rho = .59, p < .001$). Los coeficientes de correlación entre las dimensiones y los totales de ambas escalas en S3 fueron entre moderados y bajos en general, y todos significativos (tabla 4).

Tabla 4

Coefficientes de correlación entre los totales y las dimensiones de la SRS 3.0 y WAI, 1^{era} Sesión, muestra clínica (n=167), Argentina, 2019

	WAI Objetivos	WAI TAREA	WAI VINCULO	WAI TOTAL
Relación	.29**	.32**	.42**	.33**
Objetivos	.32**	.54**	.44**	.39**
Enfoque	.42**	.56**	.48**	.46**
General	.36**	.50**	.43**	.39**
Total	.45**	.55**	.50**	.48**

Nota: ρ de Spearman; ** $p < .01$; SRS 3.0= Session Rating Scale; WAI-SR-A= Inventario de Alianza de Trabajo breve - versión paciente.

Validez Predictiva. No se hallaron evidencias de validez predictiva de la escala, siendo los resultados de la regresión entre la SRS 3.0 S3 y el OQ-45.2 SF $F_{(1,29)} = 2.51, p = .124, R^2 = .08$, y SRS 3.0 S8 y el OQ-45.2 SF $F_{(1,24)} = .03, p = .862, R^2 = .001$.

Discusión

El objetivo general del trabajo fue adaptar las escalas ORS (Miller et al., 2003) y SRS 3.0 (Duncan et al., 2003), y proporcionar evidencias sobre el desempeño psicométrico de las mismas, siendo el primer estudio realizado en Argentina. Aunque no se puede esperar que medidas tan breves logren la misma precisión o profundidad de información que una medida más extensa como el OQ-45.2 y la WAI-SR-A en lo que refiere a la validez de constructo (Ziegler et al., 2014), este estudio encontró que ambas escalas tienen una validez adecuada, una confiabilidad sólida dada por su consistencia interna y correlación test-retest altas, y una alta factibilidad, por lo cual es recomendable su aplicación.

No se hallaron diferencias respecto del género, ni relación entre los puntajes y la edad de los pacientes, algo deseable en este tipo de medidas. Los puntajes en la primera administración en esta muestra fueron levemente inferiores a los reportados en una muestra española (Moggia et al., 2018) y similares a otros estudios (Biescad y Timulak, 2014; Cazauvieilh et al., 2020; Hafkenscheid et al., 2010).

El ICC, en el estudio original, fue de 6 y de 5.61 en el presente trabajo. Teniendo

en cuenta la clasificación de Jacobson y Truax (1991) los pacientes que están en tratamiento y que aumentan más de 5.61 su puntaje estarán *recuperados*. El ICC permite además establecer quienes no mejoran y quienes se deterioran.

La consistencia interna fue alta, considerando la cantidad de ítems, algo que se ha observado también en la literatura (Hafkenscheid et al., 2010; Janse et al., 2014; Miller et al., 2003; Moggia et al., 2018; Murphy et al., 2020). La correlación test-retest entre las administraciones S1/S3, S3/S8 y S8/SF fueron moderadas y significativas. Comparadas con el estudio realizado en España (Moggia et al., 2018) fueron menores. Una razón puede ser la diferencia en el tiempo de administración. Ya que en el estudio español las correlaciones fueron entre las sesiones 1-2 (.61); 2-3 (.77) y 3-4 (.84). Es decir, en el presente estudio al estar más espaciadas las sesiones se espera menor correlación entre ellas. Cabe tener en cuenta que al ser mediciones de personas que están en proceso psicoterapéutico, son personas que están cambiando, por lo tanto, no son mediciones test/re-test del todo fiables. Respecto a la validez, la correlación entre los totales del ORS y el OQ-45 para la primera administración fueron moderadas-altas, siendo similares o más altas que en otros estudios (Janse et al., 2014; Miller et al., 2003).

Los coeficientes de correlación entre los ítems y los totales del ORS y el OQ-45.2 fueron en general altos, con excepción de la dimensión Rol Social del OQ-45.2 y las dimensiones interpersonal e individual del ORS. Los ítems del ORS correlacionaron más con las puntuaciones totales del OQ-45.2 que con las subescalas del instrumento, similar al estudio original (Miller et al., 2003). En este sentido es aconsejable, al ser un instrumento de cuatro ítems, que la escala sea usada de manera global y no en función de las subescalas (Miller et al., 2003). La correlación general entre las puntuaciones totales de ORS y OQ-45.2 es mayor que en el estudio original, indicando presencia de validez concurrente. La brevedad del ORS sumado a su estructura analógica, pueden ser responsables de que los índices de correlación sean en ocasiones bajos (Miller et al., 2003). Sin embargo, la correlación es aceptable, y proporciona evidencia de que el ORS es una alternativa para evaluar de manera global el malestar subjetivo, con puntuaciones similares a la escala completa del OQ-45.2. Por último, se hallaron diferencias significativas entre el pre-test y el post-test, indicando que la prueba presentó sensibilidad al cambio como en otros estudios (Biescad y Timulak, 2014; Miller et al., 2003).

La confiabilidad de la SRS 3.0 fue óptima en las distintas administraciones siendo mayor que en el estudio original, igual en la tercera sesión que el estudio realizado en España (Moggia et al., 2020) y similar a otras investigaciones (Campbell y Hemsley, 2009; Fisher et al., 2016; Kuhlman et al., 2013; Mikeal et al., 2016; Owen et al., 2014). La correlación test-retest entre la S3 y S8 fue alta, siendo mayor al estudio original y a la encontrada entre la primera y segunda sesión (Reese et al., 2009). Sin embargo, la correlación entre S8 y SF fue baja. Esto último podría deberse a que dada la naturaleza de la confiabilidad test-retest, es difícil comparar estas estimaciones con interpretaciones estándar ya que la estabilidad de la puntua-

ción de la prueba no es necesariamente una expectativa de una medida de alianza terapéutica en el tiempo. Además, la SRS 3.0 está diseñada para medir la alianza terapéutica en la sesión y no de varias sesiones o en todo el proceso psicoterapéutico, es decir, tiene un objetivo clínico y no de investigación.

La validez concurrente con la WAI fue de moderada a fuerte similar a la encontrada en otro estudio (Reese et al., 2013). En este trabajo no se hallaron evidencias de validez predictiva de la escala teniendo en cuenta los resultados obtenidos en el OQ-45.2, una manera de explicar este resultado es que al ser una escala breve de tan solo 4 ítems tiene mayor utilidad como recurso de diálogo terapéutico, de instrumento clínico en vez de predictor de resultados. También cabe aclarar que la administración de las escalas fue pasiva, los terapeutas, en la mayoría de los pacientes, no utilizaban esa información para el trabajo clínico, Quizás esta es una de las razones de la falta de validez predictiva.

Es necesario considerar que el presente estudio presenta limitaciones. En las terminaciones prematuras no se pudo identificar en nivel de malestar de los pacientes. Si bien todos los tratamientos se realizaron en función de objetivos consensuados no todos tenían un tiempo estipulado. Además, presenta limitaciones relacionadas con las características de la muestra en el sentido que la mayoría de los pacientes son de escasos recursos económicos. No se realizaron estimaciones de validez divergente; y la administración de las escalas fue implementada de manera pasiva, es decir, las escalas eran administradas para hacer seguimiento de los pacientes, pero luego no se trabajaba con esa información en las sesiones, salvo contadas excepciones. Sin embargo, al ser el primer estudio realizado en Argentina con estas escalas la intención era evaluar su funcionamiento comparativamente con el OQ-45.2 y la WAI; para en posteriores estudios utilizarlas en contexto de feedback, es decir, rastrear el progreso de los pacientes y la alianza terapéutica para luego trabajarlo en las sesiones.

No obstante, estas limitaciones, la presente adaptación permitiría la utilización de la ORS y la SRS 3.0 en nuestro contexto (Argentina) como dos instrumentos breves, que se responden en un minuto, para evaluar el progreso de la psicoterapia y la relación terapéutica. Una de las características de las escalas es la factibilidad para su uso, permitiéndole a los psicoterapeutas tener una evaluación global de sus pacientes, tanto respecto al progreso como de la alianza terapéutica en muy poco tiempo, indagando sobre las dificultades que van apareciendo en el proceso. Además, como la puntuación se lleva a cabo en la sesión, la retroalimentación es inmediata. Las escalas permitirían un feedback fluido en temas de interés para el paciente y el psicoterapeuta, facilitándole al profesional contar con información para hacer los ajustes necesarios en función de los comentarios de los pacientes. Además, tienen la ventaja de involucrar directamente tanto al clínico como al paciente en el proceso de medir y discutir el progreso del proceso psicoterapéutico. Su utilización aumentaría las probabilidades de una menor terminación prematura de los procesos psicoterapéuticos, como de realizar un proceso exitoso (Miller et

al., 2013), lo que debería ser evaluado en futuras investigaciones.

Referencias

- Asociación Argentina de Estudio e Investigación en Psicodiagnóstico. (2000). *Pautas Internacionales para el Uso de los Tests. Versión Argentina. ADEIP*. http://www.adeip.org.ar/pautas_min_test.htm
- Anker, M. G., Duncan, B. L. y Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting [Uso de los comentarios de los clientes para mejorar los resultados de la terapia de pareja: un ensayo clínico aleatorio en un entorno naturalista]. *Journal of Consulting and Clinical Psychology, 77*(4), 693-704. <https://doi.org/10.1037/a0016062>
- Bargmann, S. y Robinson, B. (2012). *Manual 2: Feedback-informed clinical work: The basics [Manual 2: Trabajo clínico basado en retroalimentación: conceptos básicos]*. International Center for Clinical Excellence.
- Biescad, M. y Timulak, L. (2014). Measuring psychotherapy outcomes in routine practice: Examining Slovak versions of three commonly used outcome instruments [Medición de los resultados de la psicoterapia en la práctica habitual: examen de los resultados de tres instrumentos eslovacos de uso común]. *European Journal of Psychotherapy & Counselling, 16*(2), 140-162. <https://doi.org/10.1080/13642537.2014.895772>
- Bringhurst, M. D. L., Watson, C. W., Miller, S. D. y Duncan, B. L. (2006). The reliability and validity of the Outcome Rating Scale: A replication study of a brief clinical measure [La fiabilidad y validez de la escala de calificación de resultados: un estudio de replicación de una breve medida clínica]. *Journal of Brief Therapy, 5*(1), 23-30.
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working Alliance [La generalización del concepto psicoanalítico de la alianza terapéutica]. *Psychotherapy: Theory, Research & Practice, 16*(3), 252-260. <https://doi.org/10.1037/h0085885>
- Boswell, J. F., Kraus, D. R., Miller, S. D. y Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions [Implementación de la monitorización de resultados en la rutina de la práctica clínica: beneficios, desafíos y soluciones]. *Psychotherapy Research, 25*(1), 6-19. <https://doi.org/10.1080/10503307.2013.817696>
- Campbell, A. y Hemsley, S. (2009). Outcome Rating Scale and Session Rating Scale in psychological practice: Clinical utility of ultra-brief measures [Escala de calificación de resultados y escala de calificación de sesión en la práctica psicológica: utilidad clínica de las medidas ultra breves]. *Clinical Psychologist, 13*(1), 1-9. <https://doi.org/10.1080/13284200802676391>
- Cazauvieilh, C., Gana, K., Miller, S. D. y Quintard, B. (2020). Validation of the French versions of two brief, clinician-friendly outcome monitoring tools: The ORS and SRS [Validación de dos versiones francesas breves y clínicamente amistosas sobre las herramientas de monitorización de resultados: La ORS y SRS]. *Current Psychology, 1*-13. <https://doi.org/10.1007/s12144-020-00992-x>
- Chapman, N. A., Black, S. W., Drinane, J. M., Bach, N., Kuo, P. y Owen, J. J. (2017) Quantitative performance systems: Feedback-informed treatment [Sistemas de ejecución cuantitativo: tratamiento informado por retroalimentación]. En T Rousmaniere, R. K. Goodyear, S. D. Miller y B. E. Wampold (Eds.), *The cycle of excellence: Using deliberate practice to improve supervision and training* (pp.123-143). John Wiley. <https://doi.org/10.1002/9781119165590.ch6>
- Chapman, C. L., Burlingame, G. M., Gleave, R., Rees, F., Beecher, M. y Porter, G. S. (2012). Clinical prediction in group psychotherapy [Predicciones clínicas en psicoterapia grupal]. *Psychotherapy Research, 22*(6), 1-9. <https://doi.org/10.1080/10503307.2012.702512>
- DeSantis, B., Jackson, M. J., Duncan, B. L. y Reese, R. J. (2017). Casting a wider net in behavioral health screening in primary care: A preliminary study of the Outcome Rating Scale [Proyectando una red más amplia en los exámenes de detección de salud conductual en la atención primaria: un estudio preliminar de la escala de calificación de resultados]. *Primary Health Care Research & Development, 18*(2), 188-193. <https://doi.org/10.1017/S1463423616000311>
- Duncan, B. L., Miller, S. D., Sparks, J. A., Claud, D. A., Reynolds, L. R., Brown, J. y Johnson, L. D. (2003). The Session Rating Scale: Preliminary psychometric properties of a "working" alliance measure [La escala de calificación de la sesión: propiedades psicométricas preliminares de una medida de alianza "funcional"]. *Journal of brief Therapy, 3*(1), 3-12.

- Fisher, H., Atzil-Slonim, D., Bar-Kalifa, E., Rafaeli, E. y Peri, T. (2016). Emotional experience and alliance contribute to therapeutic change in psychodynamic therapy [La experiencia emocional y la alianza contribuyen al cambio terapéutico en la terapia psicodinámica]. *Psychotherapy*, 53(1), 105–116. <https://doi.org/10.1037/pst0000041>
- Flückiger, C., Del Re, A. C., Wampold, B. E. y Horvath, A. O. (2019) Alliance in adult psychotherapy [Alianza en psicoterapia adulta]. En J. C. Norcross y M. Lambert (Eds.), *Psychotherapy relationships that work: Volume 1: Evidence-based therapist contributions* (pp. 86–132). Oxford University Press. <https://doi.org/10.1093/med-psych/9780190843953.003.0002>
- Gómez Penedo, J. M., Waizmann, V. y Roussos, A. J. (2015). Propiedades psicométricas de la adaptación argentina del Inventario de Alianza de Trabajo breve-versión paciente (WAI-SR-A): un análisis factorial confirmatorio. *Investigaciones en Psicología*, 20(1), 49–62.
- Hafkenscheid, A., Duncan, B. L. y Miller, S. D. (2010). The Outcome and Session Rating Scales: A cross-cultural examination of the psychometric properties of the Dutch translation [Las escalas de calificación de resultados y sesiones: un examen intercultural de las propiedades psicométricas de la traducción al holandés]. *Journal of Brief Therapy*, 7(1-2), 1–12.
- Hatcher, R. L. y Barends, A. W. (1996). Patients' view of the alliance in psychotherapy: Exploratory factor analysis of three alliance measures [Visión de los pacientes sobre la alianza en psicoterapia: análisis factorial exploratorio de tres medidas de alianza]. *Journal of Consulting and Clinical Psychology*, 64(6), 1326. <https://doi.org/10.1037/0022-006X.64.6.1326>
- Jacobson, N. S. y Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research [Significación clínica: un enfoque estadístico para definir un cambio significativo en la investigación en psicoterapia]. *Journal of Consulting and Clinical Psychology*, 59(1), 12–19. <https://doi.org/10.1037/0022-006X.59.1.12>
- Janse, P., Boezen-Hilberdink, L., van Dijk, M. K., Verbraak, M. J. P. M. y Hutschemaekers, G. J. M. (2014). Measuring feedback from clients. The psychometric properties of the Dutch Outcome Rating Scale and Session Rating Scale [Medir la retroalimentación de los clientes. Las propiedades psicométricas de la escala holandesa de calificación de resultados y la escala de calificación de la sesión]. *European Journal of Psychological Assessment*, 30(2), 86–92. <https://doi.org/10.1027/1015-5759/a000172>
- Kuhlman, I., Tolvanen, A. y Seikkula, J. (2013). The therapeutic alliance in couple therapy for depression: Predicting therapy progress and outcome from assessments of the alliance by the patient, the spouse, and the therapists [La alianza terapéutica en la terapia de pareja para la depresión: predecir el progreso y el resultado de la terapia a partir de las evaluaciones de la alianza por parte del paciente, el cónyuge y los terapeutas]. *Contemporary Family Therapy*, 35(1), 1–13. <https://doi.org/10.1007/s10591-012-9215-5>
- Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future [Comentarios sobre el progreso y el sistema OQ: el pasado y el futuro]. *Psychotherapy*, 52(4), 381–390. <https://doi.org/10.1037/pst0000027>
- Lambert, M. J., Hansen, N.B., Umpruss, V., Lunnen, K., Okiishi, J. y Burlingame, G. M. (1996). *Administration and Scoring Manual for the OQ45.2* [Administración y manual de puntuaciones para el OQ45.2]. American Professional Credentialing.
- Lambert, M. J., Whipples, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L. y Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcome enhance? [Efectos de proporcionar feedback al terapeuta sobre el progreso del paciente durante la psicoterapia. ¿Mejoran los resultados?] *Psychotherapy Research*, 11(1), 49–68. <https://doi.org/10.1080/713663852>
- Maristany, M. y Fernández-Álvarez, H. (2010). Evaluación de resultados en psicoterapia: Estudio de las propiedades psicométricas del OQ-45.2, en Argentina (Artículo no publicado).
- Mikeal, C. W., Gillaspay Jr, J. A., Scoles, M. T. y Murphy, J. J. (2016). A dismantling study of the partners for change outcome management system [Un estudio de desmantelamiento de los socios para el sistema de gestión de resultados del cambio]. *Journal of Counseling Psychology*, 63(6), 704–709. <https://doi.org/10.1037/cou0000168>
- Miller, S. D. (2021). Performance Metrics – License for the ORS and SRS. <https://scott-d-miller-ph-d.myshopify.com/collections/performance-metrics/products/performance-metrics-licenses-for-the-ors-and-srs>
- Miller, S. D., Duncan, B. L., Brown, J., Sorrell, R. y Chalk, M. B. (2006). Using formal client feedback to improve retention and outcome: Making ongoing, real-time assessment feasible [Utilizando el feedback del cliente para mejorar la retención y el resultado: Hacer factible la evaluación continua en tiempo real]. *Journal of Brief Therapy*, 5(1), 5–22.

- Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A. y Claud, D. A. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure [El resultado de rango de escala: Un estudio preliminar de fiabilidad, validez y viabilidad de una breve medida visual analógica]. *Journal of brief Therapy*, 2(2), 91-100.
- Miller, S. D., Hubble, M. A., Chow, D. L. y Seidel, J. A. (2013). The outcome of psychotherapy: Yesterday, today, and tomorrow [El resultado de psicoterapia: Ayer, hoy y mañana]. *Psychotherapy*, 50(1), 88-97. <https://doi.org/10.1037/a0031097>
- Moggia, D., Niño-Robles, N., Miller, S. D. y Feixas, G. (2018). Psychometric Properties of the Outcome Rating Scale (ORS) in a Spanish Clinical Sample [Propiedades psicométricas de la escala de valoración de resultados (ORS) en una muestra clínica española]. *The Spanish Journal of Psychology*, 21, e30. <https://doi.org/10.1017/sjp.2018.32>.
- Moggia, D., Niño-Robles, N., Miller, S. D. y Feixas, G. (2020). Psychometric properties of the Session Rating Scale 3.0 in a Spanish clinical sample [Propiedades psicométricas de la escala de calificación de la sesión 3.0 en una muestra clínica Española]. *British Journal of Guidance & Counselling*, 1-12. <https://doi.org/10.1080/03069885.2020.1778635>
- Morley, S. y Dowzer, C. N. (2014). *Manual for the Leeds Reliable Change Indicator: Simple Excel(tm) applications for the analysis of individual patient and group data* [Manual para el indicador de cambio fiable de Leeds: aplicaciones simples de Excel (tm) para el análisis de datos individuales de pacientes y grupos]. University of Leeds.
- Murphy, M. G., Rakes, S. y Harris, R. M. (2020). The psychometric properties of the Session Rating Scale: A Narrative Review [Las propiedades psicométricas de la escala de calificación de session]. *Journal of Evidence-Based Social Work*, 17(3), 279-299. <https://doi.org/10.1080/26408066.2020.1729281>
- Owen, J., Duncan, B., Reese, R. J., Anker, M. y Sparks, J. (2014). Accounting for therapist variability in couple therapy outcomes: what really matters? [Tener en cuenta la variabilidad del terapeuta en los resultados de la terapia de pareja: ¿qué importa realmente?]. *Journal of Sex and Marital Therapy*, 40(6), 488-502. <https://doi.org/10.1080/0092623X.2013.772552>
- Paz, C., Mascialino, G., Proaño, K. y Evans, C. (2021). Psychological intervention change measurement in Latin America: Where from? Where to? [Medida del cambio de intervención psicológica en Latinoamérica: ¿de dónde? ¿a dónde?]. *Psychotherapy Research*, 31(1), 132-141. <https://doi.org/10.1080/10503307.2020.1750728>
- Reese, R. J., Gillaspay, Jr., J. A., Owen, J. J., Flora, K. L., Cunningham, L. C., Archie, D. y Marsden, T. (2013). The influence of demand characteristics and social desirability on clients' ratings of the therapeutic alliance [La influencia de las características de la demanda y la deseabilidad social en las calificaciones de los clientes sobre la alianza terapéutica]. *Journal of Clinical Psychology*, 69(7), 696-709. <https://doi.org/10.1002/jclp.21946>
- Reese, R. J., Norsworthy, L. A. y Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? [¿El feedback continuo mejora el resultado de psicoterapia?]. *Psychotherapy: Theory, Research, Practice, Training*, 46(4), 418-431. <https://doi.org/10.1037/a0017901>
- Reese, R. J., Toland, M. D., Slone, N. C. y Norsworthy, L. A. (2010). Effect of client feedback on couple psychotherapy outcomes [Efecto del feedback del cliente sobre el resultado de la psicoterapia de pareja]. *Psychotherapy: Theory, Research, Practice, Training*, 47(4), 616-630. <https://doi.org/10.1037/a0021182>
- SAMHSA's National Registry of Evidence-based Programs and Practices. (2015). *Partners for Change Outcome Management System (PCOMS)* [Sistema de Gestión de Resultados Socios para el Cambio].
- Santangelo, P. R. (2020). Programa de formación y entrenamiento en psicoterapia: Fundamentos teóricos e implementación. *Revista de Psicoterapia*, 31(117), 331-346. <https://doi.org/10.33898/rdp.v31i117.377>
- Von Bergen, A. y de la Parra, G. (2002). OQ-45.2, Cuestionario para evaluación de resultados y evolución en psicoterapia: Adaptación, validación e indicaciones para su aplicación e interpretación. *Terapia psicológica*, 20(2), 161-176.
- Walfish, S., McAlister, B., O'Donnell, P. y Lambert, M. J. (2012). An investigation of self assessment bias in mental health providers [Una investigación del sesgo de autoevaluación en los proveedores de salud mental]. *Psychological Reports*, 110(2), 639-644. <https://doi.org/10.2466%2F02.07.17.PR0.110.2.639-644>
- Ziegler, M., Kemper, C. J. y Kruey, P. M. (2014). Short Scales-Five misunderstandings and ways to overcome them [Escala cortas: cinco malentendidos y formas de superarlos]. *Journal of Individual Differences*, 35(4), 185-189. <https://doi.org/10.1027/1614-0001/a000148>