

Propiedades psicométricas de las puntuaciones de los test más empleados en la evaluación de la regulación emocional

Pérez-Sánchez, Jennifer; R. Delgado, Ana; Prieto, Gerardo

Propiedades psicométricas de las puntuaciones de los test más empleados en la evaluación de la regulación emocional

Papeles del Psicólogo, vol. 41, núm. 2, 2020

Consejo General de Colegios Oficiales de Psicólogos, España

Disponible en: <https://www.redalyc.org/articulo.oa?id=77865485005>

DOI: <https://doi.org/10.23923/pap.psicol2020.2931>

Propiedades psicométricas de las puntuaciones de los test más empleados en la evaluación de la regulación emocional

Psychometric properties of the test scores most used in the assessment of emotion regulation

Jennifer Pérez-Sánchez
Universidad de Salamanca, España
jeps@usal.es

DOI: <https://doi.org/10.23923/pap.psicol2020.2931>
Redalyc: <https://www.redalyc.org/articulo.oa?id=77865485005>

Ana R. Delgado
Universidad de Salamanca, España

Gerardo Prieto
Universidad de Salamanca, España

Recepción: 06 Noviembre 2019
Aprobación: 04 Febrero 2020

RESUMEN:

La evaluación de la regulación emocional no solo requiere del conocimiento del constructo y de los test que lo evalúan sino también de la calidad de esos instrumentos. El objetivo de este artículo fue analizar, mediante el Cuestionario de Evaluación de Test Revisado, las propiedades psicométricas de los dos test más empleados en la evaluación de la regulación emocional, Difficulties in Emotion Regulation Scale (DERS) y Emotion Regulation Questionnaire (ERQ). Se analizaron 653 artículos que describían estudios en los que se emplearon estos test. La evidencia de validez de criterio difiere entre ambos tanto cualitativa como cuantitativamente. La fiabilidad expresada en coeficientes de consistencia interna puede valorarse positivamente en ambos test. Dado que la mayoría de las investigaciones emplearon una metodología basada en la Teoría Clásica de los Test se recomienda abordar el estudio de las propiedades de estos test desde modelos psicométricos avanzados.

PALABRAS CLAVE: DERS, ERQ, Propiedades psicométricas, Regulación emocional, Revisión sistemática..

ABSTRACT:

The assessment of emotion regulation requires not only knowledge of the construct and the tests that evaluate it, but also of the quality of those instruments. The objective of this article is to analyze, using the Revised Test Evaluation Questionnaire, the psychometric properties of the two most commonly used tests in the evaluation of emotion regulation, the Difficulties in Emotion Regulation Scale (DERS) and the Emotion Regulation Questionnaire (ERQ). We analyzed 653 articles describing studies in which these tests were used. Evidence of criterion validity differs between the two both qualitatively and quantitatively. The reliability expressed as internal consistency coefficients was positively evaluated in both tests. Given that most of the investigations used a methodology based on classical test theory, it is recommended to consider the study of the properties of these tests based on advanced psychometric models.

KEYWORDS: DERS, Emotion regulation, ERQ, Psychometric properties, Systematic review..

Muchas investigaciones tratan de explicar la psicopatología y/o el bienestar psicológico teniendo en cuenta el modo en el que somos capaces de manejar nuestras emociones (Chervonsky & Hunt, 2017; Ford, Gross & Gruber, en revisión; Sloan et al., 2017; Visted, Vollestad, Nielsen & Schanche, 2018). Existe un amplio consenso al entender la regulación emocional como un proceso en el que el sujeto es capaz de influir en el tipo de emoción, así como en el momento y en el modo en que se expresa (Ford, Gross & Gruber, en revisión; Sloan et al., 2017). El Modelo de Proceso (Gross, 1998) es el marco teórico habitual en la explicación de las estrategias de la regulación emocional. El Modelo, por un lado, propone cuatro fases principales en el proceso de la generación de las emociones: a) presentación de situaciones o estímulos, b) identificación y significado de los mismos, c) selección de técnicas para manejar esas valoraciones, y d) implementación de respuestas para promover el cambio esperado. Por otro lado, el modelo incluye diferentes categorías de estrategias de

regulación emocional en las diferentes fases del proceso: a) la selección de la situación y la modificación de la situación, b) el despliegue atencional, c) la reevaluación cognitiva, y d) la modificación en la respuesta. Es posible que las personas tiendan a utilizar estrategias (o fallar en su uso) correspondientes a la misma categoría como por ejemplo la distracción, rumiación y mindfulness; todas ellas centradas en el grupo de despliegue atencional (Naragon-Gainey, McMahon y Chacko, 2017).

Además del Modelo de Proceso, otra forma de conceptualizar las estrategias de regulación emocional consiste en tener en cuenta modelos basados en estrategias específicas y su relevancia en el desarrollo y/o mantenimiento de los trastornos psicopatológicos. En este sentido, encontramos tres categorías de clasificación. La magnitud de la relación entre la estrategia de regulación emocional y los síntomas patológicos permitiría la clasificación en estrategias adaptativas o no adaptativas (Aldao & Nolen-Noeksema, 2012). No obstante, no todas las estrategias clasificadas como no adaptativas tienen repercusión psicopatológica ya que el éxito de una estrategia puede depender de las situaciones o los objetivos concretos. Por ejemplo, aunque la evitación puede considerarse una estrategia no adaptativa, apartar de la conciencia ciertos pensamientos negativos puede resultar beneficioso para llevar a cabo tareas con una fuerte demanda cognitiva. La segunda clasificación corresponde a estrategias cognitivas (rumiación, aceptación) y conductuales (consumo, evitación conductual). Las estrategias conductuales no siempre tienen como objetivo la regulación emocional y, cuando lo tienen, suelen estar asociadas a trastornos psicopatológicos como el abuso de sustancias o los trastornos de la alimentación (Aldao & Dixon-Gordon, 2014; Dixon-Gordon, Aldao & De los Reyes, 2015). Por último, clasificar en función de la capacidad de disposición de estrategias alude a que la ausencia o mal uso de las mismas se asociaría a síntomas psicopatológicos como los de la depresión, la ansiedad o los trastornos de la personalidad, entre otros (Gratz & Roemer, 2004).

El gran número de estrategias de regulación emocional existente permite utilizar más de una al mismo tiempo o para el mismo objetivo. En este sentido, se puede hacer referencia al término reciente “polirregulación” que alude a la capacidad de los sujetos para considerar múltiples objetivos y estrategias con el fin de dar respuestas diferentes en el proceso emocional (Ford, Gross & Gruber, en revisión). Esta capacidad requiere la comprensión e internalización del Modelo de Proceso y el uso de diferentes estrategias de regulación emocional para que no derive en una inadecuada selección de técnicas o una secuencia desorganizada y costosa para el sujeto.

Los estudios que han tratado de consolidar y definir la estructura de las estrategias de regulación emocional se han basado en los resultados obtenidos mediante instrumentos estandarizados (Naragon-Gainey, McMahon y Chacko, 2017; Seligowski & Orcutt, 2015). Los test son los instrumentos más empleados en los últimos años para la evaluación de la regulación emocional (Pérez-Sánchez, Delgado & Prieto, en revisión). Hacer balance de la calidad de esos instrumentos y realizar un análisis metodológico y exhaustivo de los mismos resulta esencial ya que en muchas ocasiones la redundancia empírica o la sobreestimación de factores puede derivar del instrumento de evaluación empleado. El estudio de las propiedades psicométricas de los test exige el análisis de los ítems, la fiabilidad de las puntuaciones, las evidencias de validez y la construcción de baremos (Muñiz, 2018).

La consistencia y precisión de las puntuaciones obtenidas en un test refleja la propiedad psicométrica denominada fiabilidad (Muñiz, 2018; Prieto y Delgado, 2010). En la Teórica Clásica de los Test (TCT) se evalúa la fiabilidad de las puntuaciones en un grupo de personas mediante diferentes estadísticos (coeficiente de fiabilidad y error típico de medida) que se estiman a partir de diversos procedimientos empíricos: equivalencia (formas paralelas), estabilidad (test-retest), consistencia entre las partes de una prueba (consistencia interna) y, en su caso, consistencia de las puntuaciones de distintos calificadores (inter-jueces) (Abad, Olea, Ponsoda & García, 2011; Prieto y Delgado, 2010). Los modelos psicométricos avanzados tales como el Modelo de Rasch hacen posible la estimación de la precisión del parámetro de cada persona y del parámetro de localización de cada ítem (error estándar). Además, permiten obtener estadísticos

de fiabilidad a nivel del grupo de personas y de ítems (Person Separation Reliability e Item Separation Reliability) que tienen una interpretación análoga a los índices de fiabilidad en la TCT.

La concepción más reciente de la validez implica: a) evidencias basadas en la relación entre las puntuaciones del test y otras variables (marcadores de constructos), evidencias basadas en la relación entre las puntuaciones del test y un criterio de interés, y evidencias de estructura interna del test (Abad, Olea, Ponsoda & García, 2011; Prieto y Delgado, 2010).

Existe un amplio consenso en el estudio y análisis de todos estos aspectos psicométricos para la construcción, análisis y estandarización de un test. La Comisión Internacional de Test (ITC), la EFPA Standing Committee on Tests and Testing (SCTT), el Buros Center for Testing (BUROS), el Test Review System of the Committee on Tests, el System for the Evaluation of Psychological Tests, el COTAN Evaluation system for test quality y el Cuestionario de Evaluación de Test (CET) son modelos metodológicos de referencia en la evaluación de la calidad de los test y de su uso adecuado. El CET (Prieto & Muñiz, 2000) fue el método estandarizado de evaluación de la calidad de cuestionarios diseñado por la Comisión de Test del Colegio Oficial de Psicólogos en España. En el año 2016 se publicó la versión revisada del modelo denominada Cuestionario de Evaluación de Test-Revisado (CET-R; Hernández et al., 2016) que incorpora avances y recomendaciones frente a la anterior.

En una revisión sistemática sobre la frecuencia de uso de los diversos instrumentos que evalúan la regulación emocional (Pérez-Sánchez, Delgado & Prieto, en revisión) se observó que el Difficulties in Emotion Regulation Scale (DERS, Gratz & Roemer, 2004) y el Emotion Regulation Questionnaire (ERQ, Gross & John, 2003) fueron los test más empleados en los últimos años. Lo cual justifica sobradamente nuestro objetivo de analizar, en el marco del CET-R, la calidad psicométrica de estos dos test.

METODOLOGÍA

Muestra

Para la selección de artículos se llevó a cabo una búsqueda en Scopus, una de las bases de datos de mayor cobertura en Ciencias de la Salud, Medicina, Tecnología y Ciencias Sociales permitiendo el acceso a resúmenes de artículos procedentes de más de 5000 editoriales. Se revisaron los artículos publicados desde el año 2013 al 2018 que contenían en el título, resumen o palabras clave el término emotion regulation, seleccionando los que incluían los test DERS y ERQ en el ámbito de la psicología. La búsqueda concluyó el 1 de Abril de 2019. Se revisaron todos los resúmenes para escoger los artículos empíricos que utilizaron el DERS y el ERQ para la evaluación de la regulación emocional. El número total de artículos analizados fue de 653, de los cuales 337 contenían el DERS y 316 utilizaron el ERQ.

Procedimiento

En cada artículo se examinó el año de publicación, los datos demográficos, el foco de estudio, los datos de fiabilidad y las evidencias de validez. También se tuvieron en cuenta las diferentes versiones utilizadas de ambos test. La información se organizó tomando en consideración tres apartados principales del CET-R (Hernández et al., 2016): a) descripción general del test, b) fiabilidad y c) validez. En la codificación de los datos, los infrecuentes casos dudosos se resolvieron mediante el consenso entre todos los investigadores.

RESULTADOS

Descripción de los test

El DERS es un instrumento destinado a la evaluación de las dificultades para la regulación emocional. Es un cuestionario de autoinforme en el que se indica la frecuencia con la que determinados comportamientos afectivos o estados emocionales corresponden a los del sujeto evaluado. El test consta de seis subescalas de las que, a continuación, se enumera la denominación en la escala original y el objeto de evaluación: a) nonacceptance (no aceptación de las respuestas emocionales), b) goals (dificultad para adoptar comportamientos dirigidos a objetivos), c) impulse (dificultad para el control de impulsos), d) awareness (falta de conciencia emocional), e) strategies (acceso limitado a estrategias de regulación emocional) y f) clarity (falta de claridad para identificar sus emociones). El DERS contiene 36 ítems con formato de respuesta tipo Likert y cinco categorías (desde 1=casi nunca, 2=algunas veces, 3=la mitad del tiempo, 4=la mayoría de las veces, 5=casi siempre). La calificación de este cuestionario se obtiene mediante las puntuaciones directas en las subescalas y en la escala total; las puntuaciones altas en la escala total indican dificultades en la capacidad para regular las emociones. Este instrumento incluye once ítems con codificación inversa (Tabla 1). El tiempo aproximado de aplicación es de cinco minutos y se administra en formato de papel y lápiz. La Tabla 2 refleja la descripción resumida del DERS. La adaptación al castellano se llevó a cabo por Hervás y Jódar (2008) con la denominación de Escala de Dificultades en la Regulación Emocional.

El ERQ es un cuestionario de autoinforme que evalúa la tendencia para regular las emociones teniendo en cuenta dos estrategias posibles: reevaluación cognitiva o supresión expresiva. El ERQ pretende reflejar información sobre la experiencia emocional subjetiva del sujeto o - cómo se siente interiormente - y la expresión emocional o - cómo se muestran las emociones mediante el habla, los gestos u otras conductas -. El test está compuesto por dos subescalas: reappraisal (reevaluación cognitiva) y suppression (supresión). La primera contiene 6 ítems mientras que la segunda 4. En conjunto, el ERQ contiene 10 ítems con formato de respuesta tipo Likert y siete categorías que van desde completamente en desacuerdo hasta completamente de acuerdo. La calificación de este cuestionario se obtiene mediante puntuaciones directas en las subescalas; cuanto más alta es la puntuación, mayor es el uso de esa estrategia de regulación de la emoción en particular, por el contrario, las puntuaciones más bajas representan un uso menos frecuente. El tiempo aproximado de aplicación es de dos minutos y se administra en formato de papel y lápiz. La Tabla 2 refleja la descripción resumida del ERQ. Este cuestionario fue adaptado al español por Cabello et al. (2013).

TABLA 1 ÍTEMS CON CODIFICACIÓN INVERSA DEL DERS	
Item	Subescala
1	Clarity
2	Awareness
6	Awareness
7	Clarity
8	Awareness
10	Awareness
17	Awareness
20	Goals
22	Strategies
24	Impulse
34	Awareness

TABLA 1
ÍTEMS CON CODIFICACIÓN INVERSA DEL DERS

TABLA 2 DESCRIPCIÓN GENERAL DEL DERS Y ERQ		
Características	Descripción DERS	Descripción ERQ
Nombre del test original	Difficulties in Emotion Regulation Scale	Emotion Regulation Questionnaire
Autor del test original	Kim L. Gratz y Elizabeth Roemer	James J. Gross y Oliver P. John
Fecha de publicación del test original	2004	2003
Variable que pretende medir	Dificultad en la regulación emocional	Tendencia para regular las emociones
Campo de aplicación	Personalidad y calidad de vida/bienestar	Personalidad y calidad de vida/bienestar
Población (objeto de medida)	Adultos	Adultos
Tipo de instrumento	Cuestionario de autoinforme	Cuestionario de autoinforme
Duración de aplicación	No consta	No consta
Formato	Papel y lápiz	Papel y lápiz
Número de escalas	Seis	Dos
Nombre de las escalas	Nonacceptance, goals, impulse, awareness, strategies y clarity	Cognitive Reappraisal y Suppression
Número total de ítems	36	10
Formato de respuesta	Tipo Likert	Tipo Likert
Número de categorías de respuesta	5	7
Ítems con puntuación inversa/ (n°)	Si/ (11)	No
Calificación	Puntuaciones directas	Puntuaciones directas

TABLA 2
DESCRIPCIÓN GENERAL DEL DERS Y ERQ

Perspectiva teórica

La relevancia clínica que tienen las dificultades en la regulación emocional fue el motivo principal para construir el DERS. Se entiende así la regulación emocional como a) la conciencia y comprensión de las emociones, b) aceptación de emociones, c) control de impulsos y d) flexibilidad en la modulación de respuestas emocionales (Gratz y Roemer, 2004). La ausencia de estas habilidades o el mal uso de las mismas indicarían dificultades en la regulación emocional que pueden derivar en comportamientos no adaptativos, alteraciones clínicas o trastornos de personalidad, entre otros. En cambio, el ERQ se desarrolló teniendo en cuenta el Modelo de Proceso (Gross, 1998); existen diferentes etapas en la generación de la emoción complementadas por varios procesos de regulación. Los autores del ERQ seleccionaron e incluyeron deliberadamente las dos estrategias de regulación del Modelo más empleadas en la vida diaria, más fáciles de manipular en situaciones de laboratorio y correspondientes a momentos específicos y diferentes del Modelo de Proceso. Se trata de la reevaluación cognitiva y de la supresión.

Validez

Evidencias basadas en las relaciones entre las puntuaciones del test y un criterio

El 67% de los artículos que utilizaron el DERS y el 68% de los que emplearon el ERQ tuvieron como foco de estudio la predicción de criterios. En ese caso, la regulación emocional actuaba como variable predictora o mediadora. La clasificación y la frecuencia de aparición de los criterios se realizaron teniendo en cuenta los datos correlacionales y los análisis de regresión en cada estudio. No obstante, para la evaluación e interpretación de los resultados se seleccionaron solo aquellos estudios que aportaban datos que permitían calcular los rangos y los promedios de las correlaciones entre los test de regulación emocional (la escala total del DERS y las dos subescalas del ERQ) y el criterio. Para interpretar los promedios se establecieron los siguientes rangos de valores basados en el CET-R (Hernández et al., 2016) y en Rosenthal y Rosnow (1984): bajo-medio ($r < .35$), medio ($.35 < r < .55$) y alto ($r > .55$).

En la Tabla 3 se observa que el DERS y el ERQ coincidieron en la predicción de la mayoría de los criterios a excepción del académico y del familiar. El DERS destacó en la predicción de 5 criterios (adicción, ansiedad, comportamientos, trastorno de la alimentación y trastorno postraumático) y el ERQ en la de 4 (ansiedad, comportamientos, emoción y social). La frecuencia de predicción del criterio ansiedad y comportamiento

fue similar en ambos test. En esos dos criterios el promedio de las correlaciones fue significativamente superior cuando se utilizó el DERS Total ($.35 > r < .55$) que con el ERQ ($r < .35$). El criterio depresión, funciones cognitivas, emoción y estrés presentaron los promedios de asociación más altos en el DERS Total ($r > .55$). Aunque el criterio afrontamiento y psicopatología presentaron los promedios más altos con la subescala reappraisal y la subescala suppression del ERQ, respectivamente, todos los promedios se incluían en la categoría baja-media ($r < .35$). Por último, como datos extremos cabe destacar que las correlaciones obtenidas en algunos estudios diferían notablemente de la mediana en el criterio; se trataba mayoritariamente de estudios en los que se apreciaba diferencia en las características o tamaño de las muestras o en los test que median el criterio.

Criterio	DERS				ERQ						
	N	k DERS	Rango	Mdn	N	k reap	Rango	Mdn	k sup	Rango	Mdn
Académico	-	-	-	-	2	2	.17 a .20	.19	1	.18	.18
Adicción	21	14	.03 a .58	.24	3	3	.06 a .25	.11	2	.10 a .11	.11
Afrontamiento	2	1	.38	.38	1	1	.25	.25	1	.17	.17
Agresión	16	11	.14 a .59	.28	5	3	.07 a .15	.12	2	.02 a .02	.02
Ansiedad	27	18	.16 a .82	.49	25	14	.03 a .35	.23	11	.03 a .37	.17
Calidad de vida	8	6	.30 a .65	.53	15	11	.00 a .42	.21	11	.01 a .39	.15
Comportamiento	23	16	.05 a .60	.42	22	15	.03 a .48	.24	15	.00 a .46	.15
Depresión	11	5	.53 a .75	.64	19	13	.02 a .34	.21	13	.02 a .31	.17
Funciones cerebrales	3	1	.35	.35	10	7	.04 a .45	.16	5	.07 a .42	.11
Funciones cognitivas	5	1	.57	.57	9	3	.01 a .14	.14	5	.02 a .25	.07
Emoción	7	3	.19 a .81	.60	23	13	.01 a .31	.13	13	.01 a .40	.05
Estrés	7	4	.51 a .76	.58	14	8	.04 a .36	.15	9	.10 a .30	.21
Familiar	3	3	.18 a .51	.28	-	-	-	-	-	-	-
Físico	5	5	.09 a .61	.42	12	4	.02 a .13	.09	4	.00 a .31	.24
Personalidad	8	6	.26 a .70	.35	2	1	.06	.06	1	.07	.07
Psicopatología	1	1	.47	.47	1	1	.22	.22	1	.25	.25
Sueño	3	2	.02 a .22	.12	1	1	.07	.07	1	.13	.13
Social	7	6	.03 a .75	.22	23	16	.01 a .27	.11	14	.03 a .35	.14
Suicidio	14	5	.25 a .39	.29	8	6	.02 a .34	.22	6	.00 a .19	.08
Trastorno de la alimentación	28	18	.05 a .49	.36	5	2	.02 a .05	.04	1	.13	.13
Trastorno postraumático	21	14	.25 a .64	.46	7	5	.09 a .43	.16	6	.01 a .43	.10
Trastornos (otros)	5	2	.01 a .60	.30	7	5	.06 a .20	.14	5	.06 a .26	.21

N=n° de artículos; k=n° indicadores de correlación; Mdn=mediana; reap=subescala de reappraisal; sup=subescala de suppression.

TABLA 3
FRECUENCIA DE LOS CRITERIOS SELECCIONADOS

Evidencias basadas en la relación entre las puntuaciones del test y otras variables

En este apartado se optó por estudiar la relación entre el DERS y el ERQ, habida cuenta de que fueron contruidos para medir el mismo constructo. En este sentido, se calculó el promedio de las correlaciones entre las diferentes subescalas de ambos test como evidencias de validez convergente. Solo 16 artículos reflejaron datos que permitieron estos cálculos (Tabla 4).

De acuerdo con el CET-R, el promedio de correlaciones es por lo general inadecuado ($r < .35$) a excepción de la relación entre la subescala suppression del ERQ y awareness del DERS, que es adecuado pero con algunas carencias ($.35 < r < .50$). Cabe destacar que la relación entre la subescala suppression del ERQ y la de impulse del DERS es muy pequeña (.08). Además, la subescala reappraisal mantiene relaciones inversas con el DERS mientras que con la de suppression son asociaciones directas.

TABLA 4
PROMEDIO DE CORRELACIONES ENTRE DERS Y ERQ

ERQ	DERS						
	T	N	G	I	S	C	A
Reappraisal	-.33	-.22	-.22	-.33	-.33	-.31	-.33
Suppression	.27	.29	.12	.08	.21	.31	.40
T=Total; N=Nonacceptance; G=Goals; I=Impulse; S=Strategies; C=Clarity; A= Awareness							

TABLA 4
PROMEDIO DE CORRELACIONES ENTRE DERS Y ERQ

Evidencias basadas en la estructura interna del test

De todos los trabajos seleccionados que emplearon el DERS (n=337), solo cuatro tuvieron como foco de estudio la estructura interna del test original y nueve obtuvieron este tipo de evidencias a partir de versiones adaptadas o modificadas (Tabla 5). En la mayoría de los estudios se realizaron análisis factoriales exploratorios de los ítems para identificar el número y la composición de los factores. Casi el 50% de los artículos mantuvo la estructura original con 6 factores y 36 ítems. En aquellos estudios con 5 factores solía eliminarse la subescala awareness dado que las cargas factoriales de los ítems lo aconsejaban.

En el caso del ERQ, solo un estudio analizó la estructura interna del test original y siete se centraron en otras versiones del ERQ. Todos los estudios mantuvieron los dos factores propuestos en la versión original (reappraisal y suppression), sin embargo en varios estudios (Enebrink, Björnsdotter & Ghaden, 2013; Gómez-Ortiz et al., 2016; Westerlund, M., & Santtila, 2018) se eliminó el ítem 1 (I control my emotions by changing the way I think about the situation I'm in) y/o el ítem 5 (When I want to feel less negative emotion (such as sadness or anger), I change what I'm thinking about) ya que no ajustaban al modelo factorial propuesto por los autores de la versión original.

TABLA 5
ARTÍCULOS CUYO FOCO DE ESTUDIO FUE EL ANÁLISIS DE LA ESTRUCTURA INTERNA

1º Autor	Año	Versión	N	Muestra	Factores	Ítems	Método
DERS							
Bjoreberg	2016	DERS-16	180	Clínica y normativa	5	16	Correlaciones
Gómez-Simón	2014	DERS-Español	642	Adolescentes españoles	6	36	AFC y ESEM
Guzmán-González	2014	DERSE	1179	Población chilena	5	25	AFE y AFC
Kaufman	2016	DERS-SF	954	Adolescentes y adultos	6	18	AFE y AFC
Kökönyei	2014	DERS-Original	207	Pacientes de dolor crónico	6	36	AFC
Lavender	2017	DERS-5	484	Normativa de mujeres	4	28	AFE y AFC
Lee	2016	DERS-Original	840	Estudiantes universitarios	5	30	AFE y AFC
Miguel	2017	DERS-16	725	Población brasileña	5	16	AFC
Mitsopoulou	2013	DERS-Griego	780	Población griega	6	36	AFE y AFC
Osborne	2017	DERS-Original	344	Pacientes Tª Dialéctica	5	30	AFE y AFC
Ritschel	2015	DERS-Original	1050	Diversos grupos demográficos	6	36	AFC
Victor	2016	DERS-18	1624	Grupos diversos	6	18	AFE y AFC
Weiss	2015	DERS-Positivo	360	Estudiantes de psicología	3	13	AFE
ERQ							
Cobello	2013	ERQ-Español	866	Población española	2	10	AFC
Enebrink	2013	ERQ-Original	1433	Población sueca	2	8	AFC
Gómez-Ortiz	2016	ERQ-Español	2060	Estudiantes de ESO	2	8	AFE y AFC
Li	2018	ERQ-Taiwanés	909	Estudiantes taiwaneses	2	10	AFE y AFC
Pineda	2018	ERQ-Español	1980	Estudiantes universitarios	2	10	AFC
Spaapen	2014	ERQ-9	1033	Población australiana y británica	2	10	AFC
Teixeira	2015	ERQ-CA	809	Adolescentes	2	10	AFC
Westerlund	2018	ERQ-Finlandés	409	Finlandeses	2	9	AFE y AFC

*Bjoreberg (2016). El procedimiento consiste en la eliminación de las correlaciones entre ítems menores de $r=.50$

TABLA 5
ARTÍCULOS CUYO FOCO DE ESTUDIO FUE EL ANÁLISIS DE LA ESTRUCTURA INTERNA

Fiabilidad

El principal indicador de fiabilidad fue el coeficiente de consistencia interna α de Cronbach. La Tabla 6 muestra el número de coeficientes para cada escala o subescala de ambos test, el rango, el promedio y la valoración según el CET-R. Dado que el número de ítems está asociado positivamente con la magnitud del coeficiente, el valor promedio más alto correspondió a la escala total del DERS. Asimismo, el promedio de los coeficientes de la subescala reappraisal del ERQ es mayor que el de la subescala suppression, habida cuenta que la primera cuenta con un mayor número de ítems que la segunda. Con pocas excepciones, el tamaño de las muestras en los estudios de fiabilidad fue moderado (estudios con muestras grandes, $N > 500$, y muestras pequeñas, $N < 200$).

Ocho estudios que utilizaron el DERS proporcionaron coeficientes de estabilidad (test-retest), de acuerdo interjueces (ICC coeficiente de correlación intraclass) y/o coeficiente omega para expresar datos de fiabilidad. No fue posible establecer una valoración basada en el CET-R de estos últimos indicadores de fiabilidad debido a su escaso número y a que fueron presentados en valores de rango. Además, algunos estudios solo presentaron los datos de fiabilidad obtenidos en estudios anteriores o en el estudio original del test. En los estudios seleccionados se han utilizado diferentes versiones y/o adaptaciones del DERS. Se registraron seis versiones diferentes del DERS y doce adaptaciones a diferentes idiomas. La Tabla 7 refleja el nombre de la versión, autor y año de publicación, la muestra del estudio, la fiabilidad y la frecuencia de empleo de esa versión dentro de nuestra selección de artículos. La escala total del DERS en todas sus versiones reflejó coeficientes α de Cronbach excelentes ($\alpha > .85$) según el CET-R. Con el DERS-M (Bardeen et al., 2016), una versión modificada del DERS, se obtuvo el coeficiente más alto (.97). Por el contrario, el coeficiente α de Cronbach de la subescala strategies de la versión argentina (Medrano y Trógolo, 2014) puede ser valorado como inadecuado ($\alpha < .60$) según el CET-R.

Por otra parte, seis estudios mostraron coeficientes de estabilidad del ERQ, con un promedio adecuado ($.65 < r < .75$) en las dos subescalas. Dos estudios presentaron coeficientes omega y correlación r Spearman para expresar la fiabilidad de los resultados del test. Ocho estudios solo citaron los estadísticos de fiabilidad del estudio original. De la misma manera que en el DERS, se registraron diferentes versiones y adaptaciones del ERQ (Tabla 8), en concreto cinco nuevas versiones y catorce adaptaciones. En términos generales, la valoración del CET-R para los coeficientes α de Cronbach en cada subescala del ERQ fue adecuada ($.70 < \alpha < .80$) en la mayoría de sus versiones o adaptaciones.

TABLA 6 FIABILIDAD EN α DE CRONBACH DEL DERS Y ERQ EN LOS ESTUDIOS SELECCIONADOS				
Escala	k	Rango	Promedio	Valoración del CET-R
DERS				
Total	199	.76 - .98	.93	Excelente
Nonacceptance	69	.70 - .95	.89	Excelente
Goals	65	.50 - .94	.87	Excelente
Impulse	65	.70 - .95	.86	Excelente
Strategies	80	.71 - .93	.89	Excelente
Clarity	65	.45 - .91	.80	Buena
Awareness	74	.30 - .94	.79	Adecuada
ERQ				
Reappraisal	240	.55 - .95	.83	Buena
Suppression	225	.54 - .96	.76	Adecuada
k = n° de coeficientes α				

TABLA 6
FIABILIDAD EN α DE CRONBACH DEL DERS Y ERQ EN LOS ESTUDIOS SELECCIONADOS

TABLA 7 VERSIONES DEL DERS										
Versión	1º Autor	Año	N	Fiabilidad (α)						
				Rango	T	N	G	I	S	C
Alemana	Ehring	2008	84	.76 - .87						
Argentina	Medrano	2014	211			.84	.82	.87	.54	.70
Chilena	Guzmán-González	2014	2179		.92	.89	.87	.89	-	.68
DERS-16	Bjoreberg	2016	680		.92					
DERS-18	Victor	2016	427		.89	.87	.87	.89	.83	.81
DERS-M	Bardeen	2016	2399	.88 - .97	.97					
DERS-R	Bardeen	2012	145		.95					
DERS-S	Lavender	2017	484		.86	.92	.85			.65
DERS-SF	Kaufman	2016	1054		.90	.86	.91	.89	.85	.82
Española (a)	Hervás	2008	314		.93	.90	.87	.91	-	.78
Española (b)	Gómez-Simón	2014	642		.88	.84	.80	.81	.77	.71
Española (c)	Perez	2012	218			.90	.86	.92	.92	.81
Griega	Mitsapoulou	2013	708			.81	.79	.87	.85	.73
Hebrea	Kivity	2016	124		.91					
Inglesa	Grazt	2004	551		.93	.85	.89	.86	.88	.84
Italiana	Giromini	2012	422		.92	.83	.87	.86	.89	.83
Polaca	Czub	2012	1151			.78	.88	.92	.87	.66
Portuguesa	Coutinho	2010	324		.92	.86	.85	.80	.88	.75
Turca	Ruganci	2010	338	.75 - .90	.94	.83	.90	.90	.89	.82

T=Total; N=Nonacceptance; G=Goals; I=Impulse; S=Strategies; C= Clarity; A= Awareness; N=muestra en el estudios; k= estudios que emplean esa versión

TABLA 7
VERSIONES DEL DERS

TABLA 8 VERSIONES DEL ERQ						
Versión	1º Autor	Año	N	Fiabilidad (α)		k
				Reappraisal	Suppression	
Alemana	Alber	2009	-	Alemán	Alemán	20
China (a)	Deng	2011	-	Chino	Chino	1
China (c)	Wang	2007	-	Chino	Chino	7
China (d)	Chen	2011	-	No accesible	No accesible	1
ERQ-9	Spaapen	2014	1033	.79	.76	1
ERQ-CA	Gullone	2012	827	.84	.75	12
ERQ-CCA	Liu	2015	1381	.75	.72	1
ERQ-II	Remmes	2014	67	.90	.93	1
ERQ'	Melka	2011	1188	.79	.73	1
Española	Cabella	2013	981	.79	.75	2
Francesa	Christophe	2009	591	.76	.72	2
Griega	Kafetsios	2007	475	.84	.70	1
Hebrea	Carthy	2010	91	.87	.79	4
Holandesa	Koole	2004	194	.74	.79	1
Inglesa	Gross	2004	1628	.79	.73	246
Italiana	Balzarotti	2010	416	.84	.72	8
Japonesa	Yoshizu	2013	-	Japonés	Japonés	3
Polaca	Smeieja	2011	349	.77	.74	1
Rumana	Heilman	2014	48	.74	.72	1
Taiwanesa	Li	2018	909	.85	.72	1

TABLA 8
VERSIONES DEL ERQ

CONCLUSIONES

En primer lugar, la evidencia procedente de los estudios de validación de criterio corrobora lo que el análisis conceptual ya señalaba: que los test revisados, los más empleados para evaluar la regulación emocional, parten de distintos enfoques teóricos (Grazt & Roemer, 2004; Gross, 1998). Aunque el DERS y el ERQ se emplearon con la misma frecuencia para predecir criterios como los de “ansiedad” y “comportamiento”, la mayor parte de los criterios predichos por cada uno de los test refleja el modelo teórico de partida. Es decir, las investigaciones llevadas a cabo con el DERS se centran en la funcionalidad de la regulación emocional y en sus consecuencias psicopatológicas, tales como adicciones o trastornos de la alimentación; por su parte, las llevadas a cabo con el ERQ se enfocan al proceso de generación de las emociones, destacando las dos estrategias de regulación emocional más empleadas en relación con criterios como calidad de vida o emoción.

Prestando atención al marco conceptual, cabe destacar también las evidencias de validez derivadas de la relación entre ambos test. A diferencia de la revaluación, la supresión y la mayoría de las estrategias evaluadas por el DERS tratan de regular la experiencia emocional cuando esta ya ha tenido lugar, por lo que alivian a corto plazo. Esto explicaría las correlaciones directas entre la subescala suppression del ERQ y el DERS e inversas entre la subescala reappraisal y el DERS. Además, la correlación más elevada ($r=.40$) aparece entre la escala awareness del DERS y la subescala suppression del ERQ indicando que la supresión de emociones y el conocimiento de las mismas son procesos ligados.

Los promedios de los coeficientes de consistencia interna, en términos generales fueron superiores en el DERS que en el ERQ, en la escala total frente al resto de subescalas del DERS, así como en la subescala reappraisal del ERQ con respecto a la de suppression. Estos resultados pueden explicarse en parte por el número de ítems que es diferente en el DERS (36 ítems; 6 en cada subescala) y en el ERQ (10 ítems; 6 en revaluación y 4 en supresión). A pesar de contar con diferentes versiones y adaptaciones, la inglesa/original ha sido la más empleada en ambos test. Los coeficientes α de Cronbach de la mayoría de las versiones o adaptaciones fueron similares a los de la original.

Los análisis factoriales (exploratorio y confirmatorio) fueron por excelencia las técnicas más empleadas para analizar la estructura interna del DERS, el ERQ o sus versiones/adaptaciones. Estos estudios mantuvieron la dimensionalidad propuesta en las versiones originales de ambos test.

Las diferencias de validez y fiabilidad entre el DERS y el ERQ pueden verse determinadas también por el número de categorías de respuesta de los test, en ambos casos impares, así como por la recodificación de los ítems inversos. El estudio de las categorías de respuesta resulta tan importante como el análisis y desarrollo de los ítems. La investigación apoya que lo ideal serían entre cuatro y seis categorías de respuesta, ya que incluir o reducir opciones atenuaría la precisión psicométrica y sostiene que es preferible opciones pares dado que las respuestas intermedias pueden generar ambigüedad (Simms, Zelazny, Williams & Bersntein, 2019). La combinación de ítems con codificación inversa y directa influye en la precisión de medida, la dimensionalidad, la variabilidad y la influencia en las respuestas de los examinados (Suárez-Alvarez et al., 2018).

La metodología psicométrica utilizada en los artículos analizados corresponde a la Teoría Clásica de los Test (TCT). Resulta de interés proponer investigaciones que empleen modelos psicométricos avanzados como el de Escalas de Calificación (Andrich, 1988), que permitan estudiar, además de la calidad de las categorías de respuesta, otros aspectos destacados como el escalamiento conjunto de personas e ítems y el funcionamiento diferencial de los ítems y que ofrezcan resultados sobre la fiabilidad, validez, análisis de ítems y dimensionalidad comparables a los obtenidos hasta ahora a través de modelos clásicos.

CONFLICTO DE INTERESES

No existe conflicto de intereses

REFERENCIAS

- Abad, F.J., Olea, J., Ponsoda, V., & García, C. (Ed.) (2011). *Medición en ciencias sociales y de la salud*. Madrid: Editorial Síntesis.
- Aldao, A., & Dixon-Gordon, K.L. (2014). Broadening the scope of research on emotion regulation strategies and psychopathology. *Cognitive Behaviour Therapy*, 43, 22-33. doi: 10.1080/16506073.2013.816769.
- Aldao, A., & Nolen-Hoeksema, S. (2012). The influence of context on the implementation of adaptive emotion regulation strategies. *Behaviour Research and Therapy*, 50, 493-501. doi:10.1016/j.brat.2012.04.004.
- Andrich, D. (1988). *Rasch models for measurement*. Beverly Hills, CA: Sage.

- Bardeen, J.R., Fergus, T.A., Hannan, S.M., & Orcutt, H.K. (2016). Addressing psychometric limitations of the difficulties in Emotion Regulation Scale through item modification. *Journal of Personality Assessment*, 98, 298-309. doi: 10.1080/00223891.2015.1091774.
- Cabello, R., Salguero, J.M., Fernández-Berrocal, P., & Gross, J.J. (2013). A Spanish adaptation of the Emotion Regulation Questionnaire. *European Journal of Psychological Assessment*, 29, 234-240. doi: 10.1027/1015-5759/a000150.
- Chervonsky, E., & Hunt, C. (2017). Suppression and expression of emotion in social and interpersonal outcomes: A meta-analysis. *Emotion*, 17, 1-15. doi:10.1037/emo0000270.
- Dixon-Gordon, K.L., Aldao, A., & De los Reyes, A. (2015). Emotion regulation in context: Examining the spontaneous use of strategies across emotional intensity and type of emotion. *Personality and Individual Differences*, 86, 271-276. doi: 10.1016/j.paid.2015.06.011.
- Enebrink, P., Björnsdotter, A., & Ghaden, A. (2013). The Emotion Regulation Questionnaire: Psychometric properties and norms for Swedish parents of children aged 10-13 years. *Europe's Journal of Psychology*, 9, 289-303. doi: 10.5964/ejop.v9i2.535.
- Ford, B.Q., Gross, J.J., Gruber, J. (in press). Broadening our field of view: The role of emotion polyregulation. *Emotion Review*, 1-12.
- Gómez-Ortiz, O., Romera, E.M., Ortega-Ruiz, R., Cabello, R., & Fernández-Berrocal, P. (2016). Analysis of emotion regulation in Spanish adolescents: Validation of the Emotion Regulation Questionnaire. *Frontiers in Psychology*, 6, 1-11. doi: 10.3389/fpsyg.2015.01959.
- Gratz, K.L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment*, 26, 41-54. doi: 10.1007/s10862-008-9102-4.
- Gross, J.J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, 2, 271-299.
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85, 348-362. doi: 10.1037/0022-3514.85.2.348
- Hernández, A., Ponsoda, V., Muñoz, J., Prieto, G., & Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 37, 192-197.
- Hervás, G., & Jórdar, R. (2008). Adaptación al castellano de la Escala de Dificultades en la Regulación Emocional. *Clínica y Salud*, 19, 139-156.
- Medrano, L.A., & Trógolo, M. (2014). Validación de la Escala de Dificultades en la Regulación Emocional en la población universitaria de Córdoba, Argentina. *Universitas Psychologica*, 13, 1345-1356. doi: 10.11144/Javeriana.UPSY13-4.vedr
- Muñoz, J. (Ed.) (2018). *Introducción a la Psicometría. Teoría Clásica y TRI*. Madrid: Pirámide.
- Naragon-Gainey, K., McMahon, T.P., Chacko, T.P. (2017). The structure of common emotion regulation strategies: A meta-analytic examination. *Psychological Bulletin*, 143, 384-427.
- Pérez-Sánchez, J., Delgado, A.R., & Prieto, G. (2019). *Análisis de los instrumentos empleados en la investigación empírica de la regulación emocional*. Artículo entregado para la publicación.
- Prieto, G. y Delgado, A.R. (2010). Fiabilidad y Validez. *Papeles del Psicólogo*, 31, 67-74.
- Prieto, G. y Muñoz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-71.
- Rosenthal, R., & Rosnow, R.L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.
- Sloan, E., Hall, K., Moulding, R., Bryce, S., Mildred, H., y Staiger, P.K. (2017). Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: A systematic review. *Clinical Psychology Review*, 57, 141-163.

- Seligowski, A. V., & Orcutt, H. K. (2015). Examining the structure of emotion regulation: A factor-analytic approach. *Journal of Clinical Psychology*, 71, 1004–1022. doi: 10.1002/jclp.22197.
- Simms, L.J., Zelazny, K., Williams, T.F., & Bernstein, L. (2019). Does the number of response option matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*. Advance online publication. doi: 10.1037/pas0000648.
- Suárez-Alvarez, J., Pedrosa, I., Lozano, L.M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30, 149-158. doi: 10.7334/psicothema2018.33.
- Visted, E., Vollestad, J., Nielsen, M.B., & Schanche, E. (2018). Emotion regulation in current and remitted depression: A systematic review and meta-analysis. *Frontiers in Psychology*, 9, 1-20. doi: 10.3389/fpsyg.2018.00756.
- Westerlund, M., & Santtila, P. (2018). A Finnish adaptation of the emotion regulation questionnaire (ERQ) and the difficulties in emotion regulation scale (DERS-16). *Nordic Psychology*, 70, 304-323. doi: 10.1080/19012276.2018.1443279