Papeles del Psicólogo / Psychologist Papers, 2020 Vol. 41(2), pp. 116-124
https://doi.org/10.23923/pap.psicol2020.2931
http://www.papelesdelpsicologo.es
http://www.psychologistpapers.com

PSYCHOMETRIC PROPERTIES OF THE SCORES OF THE MOST COMMONLY USED TESTS IN THE EVALUATION OF EMOTION REGULATION

Jennifer Pérez-Sánchez, Ana R. Delgado and Gerardo Prieto
Universidad de Salamanca

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.*

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 Cuestionario de Evaluación de Test Revisado [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.
Key words: *DERS, Emotion regulation, ERQ, Psychometric properties, Systematic review.*

Many research investigations attempt to explain psychopathology and/or psychological well-being by looking at how we are able to manage our emotions (Chervonsky & Hunt, 2017; Ford, Gross, & Gruber, in review; Sloan et al., 2017; Visted, Vollestad, Nielsen, & Schanche, 2018). There is a broad consensus in understanding emotion regulation as a process in which the subject is able to have an influence on the type of emotion experienced, as well as the time and manner in which it is expressed (Ford, Gross, & Gruber, in review; Sloan et al., 2017). The process model (Gross, 1998) is the usual theoretical framework for explaining strategies of emotion regulation. The model, on the one hand, proposes four main phases in the process of generating emotions: a) presentation of situations or stimuli, b) identification and meaning of these, c) selection of techniques to manage these assessments, and d) implementation of responses to promote the expected change. On the other hand, the model includes different categories of emotion regulation strategies in the different phases of the process: a) situation selection and situation modulation, b) attentional deployment, c) cognitive reappraisal, and d) response modification. People may tend to use strategies (or fail to use them) that fall into the same category, such as distraction, rumination, and *mindfulness,* all of which focus on the attentional deployment group (Naragon-Gainey, McMahon, and Chacko, 2017).

In addition to the process model, another way of conceptualizing emotion regulation strategies is to take into account models based on specific strategies and their relevance in the development and/or maintenance of psychopathological disorders. In this sense, we find three categories of classification. The magnitude of the relationship between the emotion regulation strategy and the pathological symptoms allows the classification into adaptive or non-adaptive strategies (Aldao & Nolen-Noeksema, 2012). However, not all strategies classified as non-adaptive have a psychopathological impact as the success of a strategy may depend on specific situations or objectives. For example, although avoidance may be considered a non-adaptive strategy, removing certain negative thoughts from consciousness may be beneficial for carrying out high

Received: 6 November 2019 - Accepted: 4 February 2020
Correspondence: Jennifer Pérez-Sánchez. *Universidad de Salamanca. Avda. de la Merced, 109-131. 37005 Salamanca. España. E-mail: jeps@usal.es*

cognitive demand tasks. The second classification corresponds to cognitive (rumination, acceptance) and behavioral strategies (consumption, behavioral avoidance). Behavioral strategies do not always aim at emotion regulation, and when they do, they are often associated with psychopathological disorders such as substance abuse or eating disorders (Aldao & Dixon-Gordon, 2014; Dixon-Gordon, Aldao, & De los Reyes, 2015). Finally, classifying according to the ability to have strategies available alludes to the fact that the absence or misuse of these strategies is associated with psychopathological symptoms such as depression, anxiety, or personality disorders, among others (Gratz & Roemer, 2004).

The large number of existing emotion regulation strategies allows more than one to be used at the same time or for the same purpose. In this sense, reference can be made to the recent term "polyregulation" which refers to the capacity of subjects to consider multiple objectives and strategies in order to create different responses in the emotional process (Ford, Gross, & Gruber, in review). This capacity requires the understanding and internalization of the process model and the use of different strategies of emotion regulation so that it does not result in an inadequate selection of techniques or a disorganized and problematic sequence for the subject.

The studies that have tried to consolidate and define the structure of emotion regulation strategies have been based on the results obtained through standardized instruments (Naragon-Gainey, McMahon & Chacko, 2017; Seligowski & Orcutt, 2015). Tests are the most widely used instruments in recent years for the assessment of emotion regulation (Pérez-Sánchez, Delgado, & Prieto, in review). Taking stock of the quality of these instruments and carrying out a methodological and exhaustive analysis of them is essential since often empirical redundancy or the overestimation of factors may derive from the assessment instrument used. The study of the psychometric properties of the tests requires the analysis of the items, the reliability of the scores, the validity evidence, and the construction of scales (Muñiz, 2018).

The consistency and accuracy of the scores obtained in a test reflects the psychometric property known as reliability (Muñiz, 2018; Prieto & Delgado, 2010). In classical test theory (CTT), the reliability of scores in a group of people is evaluated by means of different statistics (reliability coefficient and standard error of measurement) that are estimated from various empirical procedures: equivalence (parallel forms), stability (test-retest), consistency between the parts of a test (internal consistency) and, where appropriate, consistency of the scores of different markers (inter-rater) (Abad, Olea, Ponsoda, & García, 2011; Prieto & Delgado, 2010). Advanced psychometric models such as the Rasch Model make it possible to estimate the accuracy of each person's parameter and the location parameter of each item (standard error). Furthermore, these models allow us to obtain reliability statistics at the level of the group of people and of items (person separation reliability and item separation reliability) that have an interpretation which is analogous to the reliability indices in CTT.

The most recent conception of validity involves: a) evidence based on the relationship between test scores and other variables (construct markers), evidence based on the relationship between test scores and a criterion of interest, and evidence of internal test structure (Abad, Olea, Ponsoda, & García, 2011; Prieto & Delgado, 2010).

There is a broad consensus in the study and analysis of all these psychometric aspects for the construction, analysis and standardization of a test. The International Test Commission (ITC), the EFPA Standing Committee on Tests and Testing (SCTT), the Buros Center for Testing (BUROS), the Test Review System of the Committee on Tests, the System for the Evaluation of Psychological Tests, the COTAN Evaluation system for test quality and the *Cuestionario de Evaluación de Test* [Test Evaluation Questionnaire] (CET in Spanish) are methodological models of reference in the evaluation of the quality of tests and their appropriate use. The CET (Prieto & Muñiz, 2000) was the standardized method for evaluating the quality of questionnaires designed by the Test Commission of the Spanish Psychological Association. In 2016, the revised version of the model called the *Cuestionario de Evaluación de Test-Revisado* [Test Evaluation Questionnaire-Revised] (CET-R; Hernández et al., 2016) was published, incorporating advances and recommendations building on the previous one.

In a systematic review of the frequency of use of the various instruments that assess emotion regulation (Pérez-Sánchez, Delgado, & Prieto, in review), it was observed that the Difficulties in Emotion Regulation Scale (DERS, Gratz & Roemer, 2004) and the Emotion Regulation Questionnaire (ERQ, Gross & John, 2003) were the most widely used tests in recent years. This more than justifies our objective of analyzing, within the framework of the CET-R, the psychometric quality of these two tests.

METHODOLOGY

Sample

For the selection of articles, a search was carried out on Scopus, one of the databases with the greatest coverage in Health Sciences, Medicine, Technology, and the Social Sciences, allowing access to abstracts of articles from more than 5,000 publishers. Articles published from 2013 to 2018 that contained the term "emotion regulation" in the title, abstract, or keywords were reviewed, selecting those that included the DERS and ERQ tests in the field of psychology. The search concluded on April 1, 2019. All abstracts were reviewed to choose the empirical articles that used the DERS and ERQ for the assessment of emotion regulation. The total number of articles analyzed was 653, of which 337 involved the DERS and 316 used the ERQ.

Procedure

Each article examined the year of publication, demographics, focus of study, reliability data, and evidence of validity. The different versions used of both tests were also taken into account. The information was organized taking into consideration three main sections of the CET-R (Hernández et al., 2016): a) general description of the test, b) reliability, and c) validity. In the coding of the data, the infrequent undecided cases were resolved by consensus among all the researchers.

RESULTS

Description of the tests

The DERS is an instrument for the assessment of difficulties in emotion regulation. It is a self-report questionnaire that indicates the frequency with which certain affective behaviors or emotional states correspond to those of the subject being evaluated. The test consists of six subscales, of which the name on the original scale and the object of evaluation are listed below: a) nonacceptance (non-acceptance of emotional responses), b) goals (difficulty in adopting goal-oriented behaviors), c) impulse (difficulty in controlling impulses), d) awareness (lack of emotional awareness), e) strategies (limited access to emotion regulation strategies), and f) clarity (lack of clarity in identifying one's emotions). The DERS contains 36 items in a Likert response format and five categories (ranging from 1=almost never, 2=sometimes, 3=half the time, 4=most of the time, 5=almost always). The rating of this questionnaire is obtained by direct scores on the subscales and on the total scale; high scores on the total scale indicate difficulties in the skills of emotion regulation. This instrument includes eleven items with inverse coding (Table 1). The application time is approximately five minutes and it is administered in paper-and-pencil format. Table 2 reflects the summary description of the DERS. The adaptation to Spanish was carried out by Hervás and Jódar (2008) under the name *Escala de Dificultades en la Regulación Emocional*.

The ERQ is a self-report questionnaire that evaluates the tendency to regulate emotions by taking into account two possible strategies: cognitive reappraisal or expressive suppression. The ERQ aims to reflect information on the subject's subjective emotional experience or—how he or she feels inside—and emotional expression or - how emotions are shown through speech, gestures, or other behaviors. The test is composed of two subscales: cognitive reappraisal and suppression. The former contains 6 items while the latter contains 4. Overall, the ERQ contains 10 items with a Likert-type response format and seven categories ranging from *completely disagree* to *completely agree*. The rating of this questionnaire is obtained through direct scores on the subscales; the higher the score, the greater the use of that particular emotion regulation strategy, in contrast, lower scores represent less frequent use. The application time is approximately two minutes and it is administered in paper-and-pencil format. Table 2 reflects the summary description of the ERQ. This questionnaire was adapted into Spanish by Cabello et al. (2013).

TABLE 1
ITEMS WITH INVERSE CODIFICATION OF THE DERS

Item	Subscale
1	Clarity
2	Awareness
6	Awareness
7	Clarity
8	Awareness
10	Awareness
17	Awareness
20	Goals
22	Strategies
24	Impulse
34	Awareness

Theoretical perspective

The clinical relevance of difficulties in emotion regulation was the main reason for constructing the DERS. Emotion regulation is thus understood as follows: a) awareness and understanding of emotions, b) acceptance of emotions, c) impulse control, and d) flexibility in modulating emotional responses (Gratz & Roemer, 2004). The absence of these skills or their misuse indicate difficulties in emotion regulation that may lead to non-adaptive behaviors, clinical alterations, or personality disorders, among others. The ERQ, on the other hand, was developed taking into account the process model (Gross, 1998); there are different stages in the generation of emotion complemented by several regulation processes. The authors of the ERQ deliberately selected and included the model's two regulatory strategies that are most used in daily life, easiest to manipulate in laboratory situations, and correspond to specific and different moments of the process model. These are cognitive reappraisal and suppression.

Validity

Evidence based on the relationships between test scores and a criterion

Sixty-seven percent of the articles that used the DERS and 68% of those that used the ERQ had criteria prediction as a focus of study. In this case, emotion regulation acted as a predictive or mediating variable. The classification and frequency of appearance of the criteria were made based on the correlational data and regression analyses in each study. However, for the evaluation and interpretation of the results, only those studies were selected that provided data allowing the calculation of the ranges and averages of the correlations between the emotion regulation tests (the full scale of the DERS and the two subscales of the ERQ) and the criterion. To

interpret the averages, the following ranges of values were established based on the CET-R (Hernández et al., 2016) and on Rosenthal and Rosnow (1984): low-medium ($r < .35$), medium ($.35 \leq r < .55$), and high ($r \geq .55$).

Table 3 shows that the DERS and the ERQ coincided in predicting most of the criteria except for academic and family. The DERS was noteworthy in predicting 5 criteria (addiction, anxiety, behaviors, eating disorder, and posttraumatic disorder) and the ERQ was noteworthy in predicting 4 criteria (anxiety, behaviors, emotion, and social). The frequency of prediction of the criteria anxiety and behavior was similar in both tests. In these two criteria the average of the correlations was significantly higher when the Total DERS was used ($.35 \geq r < .55$) than with the ERQ ($r < .35$). The criteria depression, cognitive functions, emotion, and stress presented the highest average associations in the Total DERS ($r > .55$). Although the coping and psychopathology criteria presented the highest averages with the reappraisal subscale and the suppression subscale of the ERQ, respectively, all averages were included in the low-medium category ($r<.35$). Finally, as extreme data, it should be noted that the correlations obtained in some studies differed notably from the median in the criterion; these were mostly studies in which a difference was observed in the characteristics, the sample size, or in the tests that measure the criterion.

Evidence based on the relationship between the test scores and other variables

In this section, it was decided to study the relationship between the DERS and the ERQ, given that they were designed to measure the same construct. In this sense, the average of the correlations between the different subscales of the two tests was calculated as evidence of convergent validity. Only 16 articles reflected data that allowed these calculations (Table 4).

According to the CET-R, the average correlation was generally inadequate ($r<.35$) except for the relationship between the suppression subscale of the ERQ and awareness of the DERS, which was adequate but with some shortcomings ($.35 \leq r < .50$). It should be noted that the relationship between the suppression subscale of the ERQ and the impulse subscale of the DERS was very small (.08). In addition, the reappraisal subscale had inverse relationships with the DERS while it had direct associations with the suppression subscale.

Evidence based on the internal structure of the test

Of all the selected papers that used the DERS (n=337), only four had the internal structure of the original test as the focus of study and nine obtained this type of evidence from adapted or modified versions (Table 5). In most studies, exploratory factor analyses of the items were performed to identify the number and composition of factors. Almost 50% of the articles maintained the original structure with 6 factors and 36 items. In the studies with 5 factors, the awareness subscale was usually eliminated since the factor loads of the items made this advisable.

In the case of the ERQ, only one study analyzed the internal structure of the original test and seven focused on other versions of the ERQ. All studies maintained the two factors proposed in the original version (reappraisal and suppression), however in several studies (Enebrink, Björnsdotter, & Ghaden, 2013; Gómez-Ortiz et al., 2016; Westerlund, M., & Santtila, 2018) the following items were removed as they did not fit the factorial model proposed by the authors of the original version: item 1 (I control my emotions by changing the way I think about the situation I'm

TABLE 2
GENERAL DESCRIPTION OF THE DERS AND ERQ

Characteristics	Description DERS	Description ERQ
Name of the original test	Difficulties in Emotion Regulation Scale	Emotion Regulation Questionnaire
Author of the original test	Kim L. Gratz and Lizabeth Roemer	James J. Gross and Oliver P. John
Date of publication of the original test	2004	2003
Variable to be measured	Difficulty in Emotion Regulation	Tendency to regulate emotions
Field of application	Personality and quality of life/wellbeing	Personality and quality of life/well-being
Population (object of measurement)	Adults	Adults
Type of instrument	Self-report questionnaire	Self-report questionnaire
Duration of application	Not provided	Not provided
Format	Paper-and-pencil	Paper-and-pencil
Number of scales	Six	Two
Name of the scales	Nonacceptance, goals, impulse, awareness, strategies and clarity	Cognitive Reappraisal and Suppression
Total number of items	36	10
Response format	Likert Type	Likert type
Number of response categories	5	7
Items with inverse score / (n°)	Yes/ (11)	No
Scoring	Direct scores	Direct scores

in) and/or item 5 (When I want to feel less negative emotion (such as sadness or anger), I change what I'm thinking about).

Reliability

The main indicator of reliability was Cronbach's internal consistency coefficient α. Table 6 shows the number of coefficients for each scale or subscale of the two tests, the range, the mean, and the rating according to the CET-R. Since the number of items was positively associated with the magnitude of the coefficient, the highest mean value corresponded to the full scale of the DERS. Likewise, the average of the coefficients of the reappraisal subscale of the *ERQ* was higher than that of the suppression subscale, given that the former has a higher number of items than the latter. With few exceptions, the sample size in the reliability studies was moderate (studies with large samples, N ≥ 500, and small samples, N < 200).

Eight studies that used the DERS provided coefficients of stability (test-retest), inter-rater agreement (intraclass correlation coefficient, ICC), and/or omega coefficient to express reliability data. It was not possible to establish a CET-R-based assessment of these latter reliability indicators due to their small number and because they were presented in range values. In addition, some studies only presented reliability data obtained in previous studies or in the original test study. Different versions and/or adaptations of the DERS were used in the selected studies. Six different versions of the DERS and twelve adaptations to different languages were recorded. Table 7 reflects the name of the version, author, and year of publication, the study sample, reliability and frequency of use of that version within our selection of articles. The full scale of the DERS in all its versions reflected excellent Cronbach coefficients (α ≥ .85) according to the CET-R. The highest coefficient (.97) was obtained with the DERS-M (Bardeen et al., 2016), a modified version of the DERS. On the contrary, the Cronbach's coefficient α of the subscale "strategies" of the Argentinean version (Medrano et al., 2014) was assessed as inadequate (α < .60) according to the CET-R.

TABLE 4
AVERAGE CORRELATION BETWEEN DERS AND ERQ

	DERS						
ERQ	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

TABLE 3
FREQUENCY OF THE SELECTED CRITERIA

	DERS				ERQ						
Criterion	N	k DERS	Range	Mdn	N	k reapp	Range	Mdn	k sup	Range	Mdn
Academic	-	-	-	-	2	2	.17 to .20	.19	1	.18	.18
Addiction	**21**	14	.03 to .58	.24	3	3	.06 to .25	.11	2	.10 to .11	.11
Coping	2	1	.38	.38	1	1	.25	.25	1	.17	.17
Aggression	16	11	.14 to .59	.28	5	3	.07 to .15	.12	2	.02 to .02	.02
Anxiety	**27**	18	.16 to .82	.49	**25**	14	.03 to .35	.23	11	.03 to .37	.17
Quality of life	8	6	.30 to .65	.53	15	11	.00 to .42	.21	11	.01 to .39	.15
Behavior	**23**	16	.05 to .60	.42	**22**	15	.03 to .48	.24	15	.00 to .46	.15
Depression	11	5	.53 to .75	.64	19	13	.02 to .34	.21	13	.02 to .31	.17
Brain functions	3	1	.35	.35	10	7	.04 to .45	.16	5	.07 to .42	.11
Cognitive functions	5	1	.57	.57	9	3	.01 to .14	.14	5	.02 to .25	.07
Emotion	7	3	.19 to .81	.60	**23**	13	.01 to .31	.13	13	.01 to .40	.05
Stress	7	4	.51 to .76	.58	14	8	.04 to .36	.15	9	.10 to .30	.21
Family	3	3	.18 to .51	.28	-	-	-	-	-	-	-
Physical	5	5	.09 to .61	.42	12	4	.02 to .13	.09	4	.00 to .31	.24
Personality	8	6	.26 to .70	.35	2	1	.06	.06	1	.07	.07
Psychopathology	1	1	.47	.47	1	1	.22	.22	1	.25	.25
Sleep	3	2	.02 to .22	.12	1	1	.07	.07	1	.13	.13
Social	7	6	.03 to .75	.22	**23**	16	.01 to .27	.11	14	.03 to .35	.14
Suicide	14	5	.25 to .39	.29	8	6	.02 to .34	.22	6	.00 to .19	.08
Eating disorder	**28**	18	.05 to .49	.36	5	2	.02 to .05	.04	1	.13	.13
Post traumatic disorder	**21**	14	.25 to .64	.46	7	5	.09 to .43	.16	6	.01 to .43	.10
Disorders (others)	5	2	.01 to .60	.30	7	5	.06 to .20	.14	5	.06 to .26	.21

N=n° of articles; k=n° of correlation indicators; Mdn=median; reapp= reappraisal subscale; sup= suppression subscale.

On the other hand, six studies showed stability coefficients of the ERQ, with an adequate average (.65 ≤ r <.75) in the two subscales. Two studies presented omega coefficients and the Spearman's rank correlation to express the reliability of the test results. Eight studies only cited the reliability statistics of the original study. As in the DERS, different versions and adaptations of the ERQ were recorded (Table 8), specifically five new versions and fourteen adaptations. In general terms, the CET-R rating for the Cronbach α coefficients in each subscale of the ERQ was adequate (.70 < α < .80) in most of its versions or adaptations.

CONCLUSIONS

Firstly, the evidence from the criterion validation studies

TABLE 5
ARTICLES WHOSE FOCUS OF STUDY WAS THE ANALYSIS OF THE INTERNAL STRUCTURE

1st Author	Year	Version	N	Sample	Factors	Items	Method
DERS							
Bjureberg	2016	DERS-16	180	Clinical and normative	5	16	Correlations
Gómez-Simón	2014	DERS-Spanish	642	Spanish teenagers	6	36	CFA and ESEM
Guzmán-González	2014	DERS-E	1179	Chilean population	5	25	EFA and CFA
Kaufman	2016	DERS-SF	954	Adolescents and adults	6	18	EFA and CFA
KöKönyei	2014	DERS-Original	207	Chronic pain patients	6	36	CFA
Lavender	2017	DERS-S	484	Women normative	4	28	EFA and CFA
Lee	2016	DERS-Original	840	University students	5	30	EFA and CFA
Miguel	2017	DERS-16	725	Brazilian population	5	16	CFA
Mitsopoulou	2013	DERS-Greek	780	Greek population	6	36	EFA and CFA
Osborne	2017	DERS-Original	344	Dialectic therapy patients	5	30	EFA and CFA
Ritschel	2015	DERS-Original	1050	Various demographic groups	6	36	CFA
Victor	2016	DERS-18	1624	Various groups	6	18	EFA and CFA
Weiss	2015	DERS-Positive	360	Psychology students	3	13	EFA
ERQ							
Cabello	2013	ERQ-Spanish	866	Spanish population	2	10	CFA
Enebrink	2013	ERQ-Original	1433	Swedish population	2	8	CFA
Gómez-Ortiz	2016	ERQ-Spanish	2060	Middle school students	2	8	EFA and CFA
Li	2018	ERQ-Taiwanese	909	Taiwanese students	2	10	EFA and CFA
Pineda	2018	ERQ-Spanish	1980	University students	2	10	CFA
Spaapen	2014	ERQ-9	1033	Australian and British population	2	10	CFA
Teixeira	2015	ERQ-CA	809	Teenagers	2	10	CFA
Westerlund	2018	ERQ-Finnish	409	Finns	2	9	EFA and CFA

*Bjureberg (2016). The procedure consists of the elimination of correlations between items less than r=.50

TABLE 6
RELIABILITY IN CRONBACH'S OF THE DERS AND ERQ IN SELECTED STUDIES

Scale	k	Range	Mean	CET-R evaluation
DERS				
Total	199	.76 - .98	.93	Excellent
Nonacceptance	69	.70 - .95	.89	Excellent
Goals	65	.50 - .94	.87	Excellent
Impulse	65	.70 - .95	.86	Excellent
Strategies	80	.71 - .93	.89	Excellent
Clarity	65	.45 - .91	.80	Good
Awareness	74	.30 - .94	.79	Adequate
ERQ				
Reappraisal	240	.55 - .95	.83	Good
Suppression	225	.54 - .96	.76	Adequate

k = n° of coefficients alpha

TABLE 7
VERSIONS OF THE DERS

Version	1st Author	Year	N	Reliability (α)							k
				Range	T	N	G	I	S	C	A	
German	Ehring	2008	84	.76 - .87								1
Argentinian	Medrano	2014	211			.84	.82	.87	**.54**	.70	.71	1
Chilean	Guzmán-González	2014	2179		.92	.89	.87	.89	-	.68	.76	1
DERS-16	Bjureberg	2016	680		.92							8
DERS-18	Victor	2016	427		.89	.87	.87	.89	.83	.81	.79	2
DERS-M	Bardeen	2016	2399	.88 - .97	**.97**							1
DERS-R	Bardeen	2012	145		.95							1
DERS-S	Lavender	2017	484		.86	.92	.85			.65	.79	1
DERS-SF	Kaufman	2016	1054		.90	.86	.91	.89	.85	.82	.79	3
Spanish (a)	Hervás	2008	314		.93	.90	.87	.91	-	.78	.73	7
Spanish (b)	Gómez-Simón	2014	642		.88	.84	.80	.81	.77	.71	.62	3
Spanish (c)	Perez	2012	218			.90	.86	.92	.92	.81	.85	1
Greek	Mitsopoulou	2013	708			.81	.79	.87	.85	.73	.76	1
Hebrew	Kivity	2016	124		.91							1
English	Gratz	2004	551		.93	.85	.89	.86	.88	.84	.80	287
Italian	Giromini	2012	422		.92	.83	.87	.86	.89	.83	.77	8
Polish	Czub	2012	1151			.78	.88	.92	.87	.66	.65	4
Portuguese	Coutinho	2010	324		.92	.86	.85	.80	.88	.75	.74	2
Turkish	Ruganci	2010	338	.75 - .90	.94	.83	.90	.90	.89	.82	.75	4

T=Total; N=Nonacceptance; G=Goals; I=Impulse; S=Strategies; C= Clarity; A= Awareness; N=sample in the studies; k= studies that use this version

TABLE 8
VERSIONS OF THE ERQ

Version	1st Author	Year	N	Reliability (α)		k
				Reappraisal	Suppression	
German	Alber	2009	-	German	German	20
Chinese (a)	Deng	2011	-	Chinese	Chinese	1
Chinese (c)	Wang	2007	-	Chinese	Chinese	7
Chinese (d)	Chen	2011	-	Not available	Not available	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
Spanish	Cabello	2013	981	.79	.75	2
French	Christophe	2009	591	.76	.72	2
Greek	Kafetsios	2007	475	.84	**.70**	1
Hebrew	Carthy	2010	91	.87	.79	4
Dutch	Koole	2004	194	.74	.79	1
English	Gross	2004	1628	.79	.73	246
Italian	Balzarotti	2010	416	.84	.72	8
Japanese	Yoshizu	2013	-	Japanese	Japanese	3
Polish	Smeieja	2011	349	.77	.74	1
Rumanian	Heilman	2014	48	.74	.72	1
Taiwanese	Li	2018	909	.85	.72	1

corroborates what the conceptual analysis already pointed out: that the tests reviewed, those most used to assess emotion regulation, are based on different theoretical approaches (Gratz & Roemer, 2004; Gross, 1998). Although the DERS and ERQ were used with the same frequency to predict criteria such as "anxiety" and "behavior", most of the criteria predicted by each of the tests reflect the theoretical starting model. In other words, the research carried out with the DERS focuses on the functionality of emotion regulation and its psychopathological consequences, such as addictions or eating disorders. On the other hand, the research carried out with the ERQ focuses on the process of emotion generation, highlighting the two most used emotion regulation strategies in relation to criteria such as quality of life or emotion.

Paying attention to the conceptual framework, the evidence of validity derived from the relationship between the two tests should also be highlighted. Unlike reappraisal, suppression and most of the strategies assessed by the DERS attempt to regulate the emotional experience when it has already taken place, thus alleviating it in the short term. This would explain the direct correlations between the suppression subscale of the ERQ and the DERS and the inverse correlations between the reappraisal subscale and the DERS. Furthermore, the highest correlation (r=.40) appears between the awareness scale of the DERS and the suppression subscale of the ERQ, indicating that the suppression and awareness of emotions are two linked processes.

The averages of the internal consistency coefficients were generally higher in the DERS than in the ERQ, in the full scale than in the rest of the subscales of the DERS, as well as in the reappraisal subscale of the ERQ than in the suppression scale. These results can be explained in part by the number of items which is different in the DERS (36 items; 6 in each subscale) and in the ERQ (10 items; 6 in reappraisal and 4 in suppression). In spite of having different versions and adaptations, the English/original version was the one most used for both tests. The Cronbach's α coefficients for most versions or adaptations were similar to those of the original.

Factor analyses (exploratory and confirmatory) were the most widely used techniques to analyze the internal structure of the DERS, the ERQ, and their versions/adaptations. These studies maintained the dimensionality proposed in the original versions of both tests.

The differences in validity and reliability between the DERS and the ERQ may also be determined by the number of response categories in the tests—in both cases an odd number—as well as by the recoding of the inverse items. The study of the response categories is as important as the analysis and development of the items. The research confirms that ideally between four and six response categories should be used, since including or reducing options would attenuate the psychometric accuracy, and it upholds that an even number of options is preferable since intermediate responses can generate ambiguity (Simms, Zelazny, Williams, & Bersntein, 2019). The combination of items with inverse and direct coding influences the precision of measurement, the dimensionality, the variability, and the influence on the responses of those examined (Suárez-Alvarez et al., 2018).

The psychometric methodology used in the articles analyzed corresponds to classical test theory (CTT). It is of interest to propose research that uses advanced psychometric models, such as the rating scale model (Andrich, 1988), that enable the study of, along with the quality of response categories, other outstanding aspects such as the joint scaling of persons and items, as well as the differential functioning of the items, which offer results on reliability, validity, item analysis, and dimensionality comparable to those obtained until now through classical models.

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

Abad, F.J., Olea, J., Ponsoda, V., & García, C. (Ed.) (2011). *Medición en ciencias sociales y de la salud [Measurement in social and health sciences]*. 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.

Boone, W.J., Staver, J.R., & Yale, Melissa, S. (2014). *Rasch analysis in the human sciences*. New York, NY: Springer.

Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences (2ª ed.)*. Mahwah, NJ: Lawrence Erlbaum.

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. doi: 10.1177/1754073919850314

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. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*, 1-26. doi: 10.1080/1047840X.2014.940781.

Hernández, A., Ponsoda, V., Muñiz, J., Prieto, G., & Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España [Assessing the quality of tests in Spain: Revision of the Spanish test review model]. *Papeles del Psicólogo, 37*, 192-197.

Hervás, G., & Jórdar, R. (2008). Adaptación al castellano de la Escala de Dificultades en la Regulación Emocional [Adaptation to Spanish of the Difficulties in Emotion Regulation Scale]. *Clínica y Salud, 19*, 139-156.

Muñiz, J. (Ed.) (2018). *Introducción a la Psicometría. Teoría Clásica y TRI [Introduction to Psychometry. Classical Theory and IRT]*. 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 [Analysis of the instruments used in the empirical investigation of emotion regulation]*. Article submitted for publication.

Rosenthal, R., & Rosnow, R.L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

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 options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*. Advance online publication. doi: 10.1037/pas0000648.

Sloan, E., Hall, K., Moulding, R., Bryce, S., Mildred, H., & 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.

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.

Victor, S.E., & Klonsky, E.D. (2016). Validation of a brief version of the Difficulties in Emotion Regulation Scale (DERS-18) in five samples. *Journal of Psychopathology and Behavioral Assessment, 38*, 582–589. doi: 10.1007/s10862-016-9547-9.

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.