



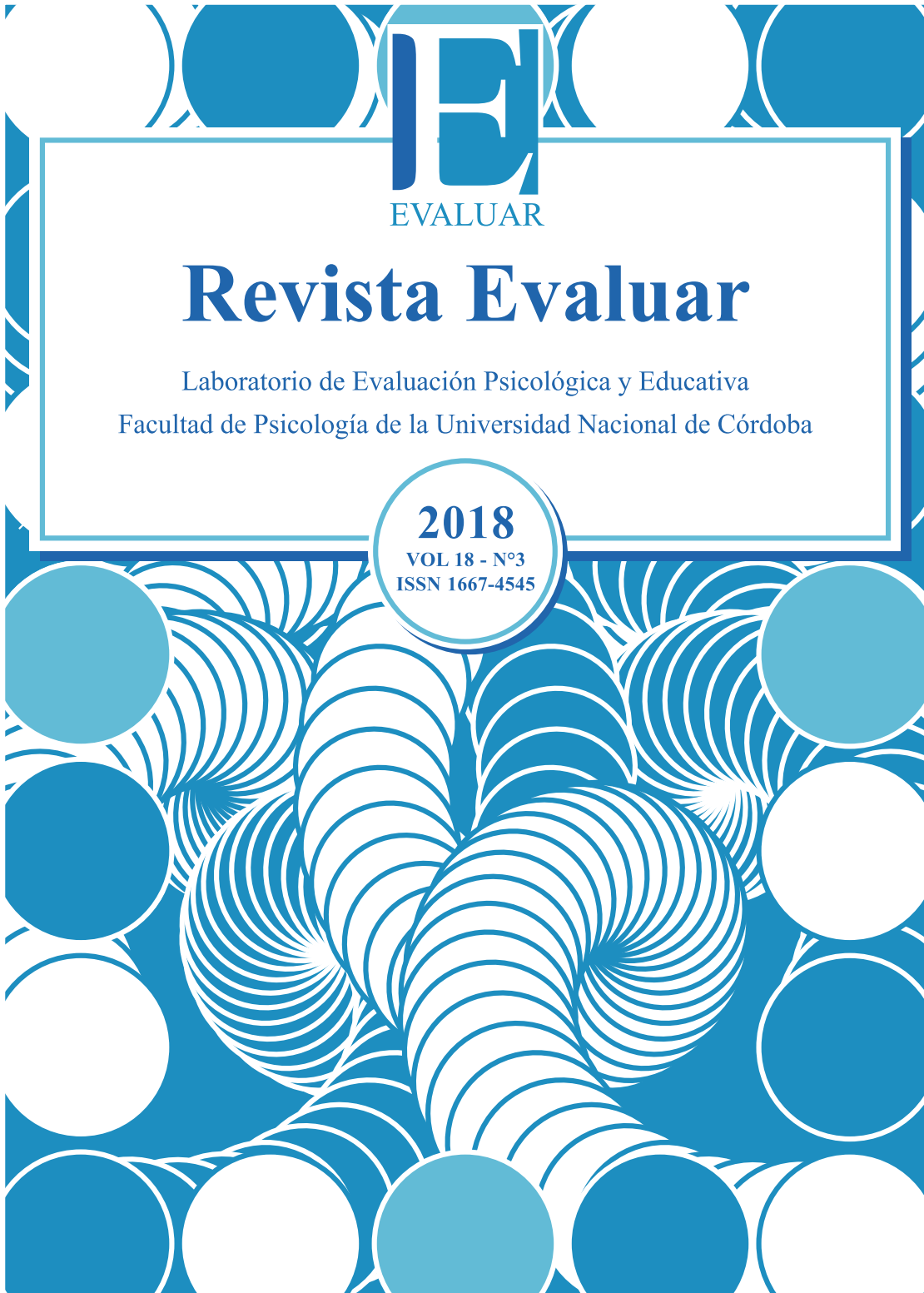
EVALUAR

Revista Evaluar

Laboratorio de Evaluación Psicológica y Educativa
Facultad de Psicología de la Universidad Nacional de Córdoba

2018

VOL 18 - Nº3
ISSN 1667-4545





Advantages of the Rasch Model for Analysis and Interpretation of Attitudes: the Case of the Benevolent Sexism Subscale

Ventajas del Modelo de Rasch para el análisis e interpretación de actitudes: El caso de la subescala de sexismo benevolente

José Andrey Zamora-Araya *^{1,3}, Vanessa Smith-Castro²,
Eiliana Montero-Rojas^{2,3}, Tania Elena Moreira-Mora⁴

1 - Escuela de Matemáticas, Universidad Nacional, Heredia, Costa Rica.

2 - Instituto de Investigaciones Psicológicas, Universidad de Costa Rica, San José, Costa Rica.

3 - Escuela de Estadística, Universidad de Costa Rica, San José, Costa Rica.

4 - Instituto Tecnológico de Costa Rica, Cartago, Costa Rica, Departamento de Orientación y Psicología.

Introduction
Method
Results
Discussion
Conclusion
References

Recibido: 24/03/2018 Revisado: 10/05/2018 Aceptado: 28/05/2018

Abstract

This paper describes how the use of Rasch Analysis (RA), compared with the Classical Test Theory (CTT) and other Item Response Theory (IRT) approaches, could enhance the study and interpretation of attitudinal scales. This is illustrated with data from 197 students from the University of Costa Rica who answered the Benevolent Sexism (BS) Scale (Glick & Fiske, 1996). Besides providing estimations of the measure's specific accuracy at different levels of the construct, the RA, thanks to the person versus item map, allowed us to generate respondents' profiles describing particular aspects of the construct and according to their estimated scores in the scale. The analysis indicated that construct categories for participants with scores between [-0.30, 0.5] in the logit scale are the most accurately represented, with more items covering this interval, and reflecting the three aspects of the scale described by the theory. On the other hand, results showed less measurement accuracy for a considerable number of respondents with lower scores, suggesting the need for the development of additional items for that level of the scale. These evidences are discussed in light of the benefits of using the RA for the understanding and interpretations of respondents' scores in attitudinal scales, according to the underlying theory.

Keywords: *Benevolent Sexism Scale, Classical Test Theory, Rasch Analysis, Extended Rasch Model, attitude scales, psychometric analysis*

Resumen

Este artículo describe cómo el uso del Análisis de Rasch (AR), comparado con la Teoría Clásica de los Tests (TCT) y otros modelos de Teoría de Respuesta a los Ítems (TRI), puede mejorar el estudio y la interpretación de escalas actitudinales. Esto se ilustra con datos de 197 estudiantes de la Universidad de Costa Rica que tomaron la escala de Sexismo Benevolente (SB; Glick & Fiske, 1996). Además de proveer estimaciones específicas de la precisión de la medición en diferentes niveles del constructo, el Análisis de Rasch, gracias al mapa de personas versus ítems, permitió generar perfiles de los participantes en términos de aspectos particulares del constructo y de acuerdo con sus puntajes estimados en la escala. El análisis indicó que las categorías del constructo para participantes con puntajes entre [-0.30, 0.5] en la escala logit son las que están mejor representadas, con más ítems que cubren este intervalo y reflejan los tres aspectos de la escala descritos en la teoría. Por otro lado, los resultados mostraron menos precisión en la medición para un considerable número de participantes con puntajes más bajos, lo que sugiere la necesidad de desarrollar ítems adicionales para este nivel de la escala. Estas evidencias son discutidas a la luz de los beneficios de usar el AR para el entendimiento e interpretación de puntajes en escalas de actitud, de acuerdo con la teoría subyacente.

Palabras Clave: *Escala de Sexismo Benevolente, Teoría Clásica de los Tests, Análisis de Rasch, Modelo Extendido de Rasch, escalas de actitud, análisis psicométrico*

How to cite: Zamora-Araya, J. A., Smith-Castro, V., Montero-Rojas, E., & Moreira-Mora, T. E. (2018). Advantages of the Rasch Model for analysis and interpretation of attitudes: The case of the Benevolent Sexism Subscale. *Revista Evaluar*, 18(3), 1-13. Available at <https://revistas.unc.edu.ar/index.php/revaluar>

* **Correspondence to:** José Andrey Zamora Araya, PO Box. 86-3000, Heredia, Costa Rica. Tel: (506) 2562-6029. E-mail: jzamo@una.ac.cr

Authors' note This research was supported by a grant of the Council of Rectors of Costa Rican Universities to the Project "Nueva formas de medir viejas ideologías: el caso del sexismo y sus implicaciones en el ámbito académico (New Forms of Measuring Old Ideologies: the Case of Sexism and its Implications in the Academic Domain)," of the University of Costa Rica, the National University and the Costa Rica Institute of Technology.

Introduction

For many years, researchers in the field of Social Psychology have commonly applied the Classical Test Theory (CTT) approach in order to gather evidence on the reliability of its measures. This approach is based on the assumption that the value of an attribute is represented by an observed score, which is the sum of a true score (error-free) and the measuring error. Although CTT can provide important evidence of the accuracy of measuring instruments, several new psychometric tools might complement or even replace this approach in order to collect more accurate evidence to support the inferences made about the meaning and interpretation of scores (Muñiz, 2017). According to Bond and Fox (2001), one of such tools is the Rasch Analysis (RA), through which trait levels (the probability of a correct response or the probability of endorsing any option on each item) are modeled as a mathematical function of the difference between the person and the item parameters (Prieto-Adanes & Dias-Velasco, 2003).

This study describes the results of applying both, CTT and RA, to test the measurement properties of the Benevolent Sexism Scale (BS), one of the two subscales of the Ambivalent Sexism Inventory (ASI) developed by Glick and Fiske (1996). Our goal is to illustrate, with this subscale, the benefits of using RA to attain a better understanding of the strengths and weaknesses of instruments in the affective domain.

The Rasch Model: Characteristics and Advantages over CTT and other IRT Models

As pointed out before, most psychometric tests have been analyzed using the CTT. This model assumes that X , the observed score in the test, is a linear combination of two quantities, the

true score (T) and the measurement error (E): $X = T + E$ (Muñiz, 2017).

One of the limitations of the CTT is the assumption that E is constant across true score values, i.e. the error associated to each examinee is the same, no matter what his/her X is. Even intuitively, this assumption seems empirically unlikely. If the test items are endorsed by the majority of the respondents, it is fair to conclude that scores in the higher level of the trait will be estimated with less precision (more error) than scores in the lower end of the trait. On the other hand, if the items are endorsed by only a few respondents, scores in the lower level of the trait will be estimated with less precision (more error) than the scores in the higher end of the trait. Therefore, when applying CTT, it is not possible to provide different precision estimates for the different levels of the construct being measured. However, researchers and practitioners often need to measure certain levels of the construct with more precision, depending on their particular purposes and applications.

Another fundamental shortcoming of CTT lays in the fact that the model does not allow for descriptive interpretations of the meaning of each particular score. This limitation was first noticed in the educational measurement community (Wilson, 2004), which traditionally criticized the CTT approach for not addressing the need of knowing what students can and cannot achieve according to their scores in the tests. Social psychologists could also benefit from the possibility of attaching specific meanings to each particular score in their scales in terms of the construct being measured.

To overcome the limitations of CTT, a family of models known as IRT (Item Response Theory) were proposed around the second half of the twentieth century (Hambleton & Swaminathan, 2013). These mathematical models attempt to describe the respondents' behavior based on their an-

swer to each item. In general, the logistic function is used to estimate the model, with three different formulations: 1PL is the One Parameter Logistic model, 2PL is the Two Parameter Logistic model, and, 3PL is the Three Parameter Logistic model. The difference between these models lies on the number of parameters needed for their definition. In the 1PL model only the item difficulty, b , is estimated, along with the examinee's ability; in the 2PL model the item discrimination, a , is also estimated; and in the 3PL model, a guessing parameter, c , is estimated as well. 1PL is obtained when the item discrimination is assumed constant for all the items and the guessing parameter is assumed to be zero; on the other hand, the 2PL is obtained when only the guessing parameter is set to zero. Thus, both models are special cases of the IRT 3PL model.

The 1PL model is also known as the Rasch Model in honor of the Danish mathematician Georg Rasch, who in the 1960's described the special properties that only this model possesses (Olsen, 2003), making it particularly useful, and very attractive for applied test users who are interested in knowing what their instruments allow them to infer in terms of substantive interpretations (Rasch, 1980). Its mathematical formula relates the probability of the outcome (response) to the level of the respondent in the construct under measurement, and, the item difficulty. *Difficulty* is a term also used for tests in the affective domain that describes how low or high is the mean score for a specific item. In this case it can be also described as *endorsability*.

The Rasch model, in its original form, for dichotomous items is written as follows:

(1)

$$P(X_{ij} = 1) = \frac{e^{(\theta - b_j)}}{1 + e^{(\theta - b_j)}}$$

Where,

$P(X_{ij} = 1)$: Probability that a specific person j answers correctly to the item i , and 0 for any other case.

θ : Parameter that describes a specific level of the trait for a person j .

b_i : Parameter that describes the difficulty (endorsability) of the item i .

θ and b can take any value in the real domain and they are both in the logit scale.

This initial formulation describes the Rasch Model, referring to the dichotomous items in the cognitive domain (1 correct, 0 incorrect). However, later developments have shown that it can be easily extended to data from rating scales for instruments in the affective domain, such as traits estimated through the Likert scale (Carvalho, Primi, & Meyer, 2012). For example, suppose that we have an item with $m+1$ response options. In this case, each of the m first options are described by the following expression:

(2)

$$P_{ik}(\theta) = \frac{e^{\sum_{k=0}^h (\theta - b_{ik})}}{\sum_{h=0}^m e^{\sum_{k=0}^h (\theta - b_{ik})}}$$

Where

$h = 0, 1, \dots, m$

$P_{ik}(\theta)$: indicates the probability of a subject with a specific θ score to endorse category k in item i .

b_{ik} is the endorsability parameter for item i in the k category.

$m+1$ is the total number of response categories.

Note that the probability for endorsement of the last category (i.e. reference category) is obtained when the examinee does not endorse any of the other m categories. In fact, these endorsement parameters estimated for affective scales are equivalent to the difficulty parameters estimated with dichotomous scales.

There is ample evidence that attribute the relative robustness of the Rasch Model to the deviation from the assumptions of equality of discrimination and zero guessing. In terms of robustness regarding these two specific assumptions, [Muñiz, Rogers and Swaminathan \(1989\)](#) found, by means of simulations, that estimations and fit indexes in the Rasch Model do not present great differences when there is guessing and variability in discrimination indexes.

Within the Rasch Model, as in the other IRT models, each particular estimated score has a specific estimation of its measurement error. Hence, it is possible to estimate how well the test's scores in the low, medium and high end of the scale might be. It also allows for the selection of the items that provide more precision (less errors) in pre-specified intervals of the trait under measurement. In other words, the measurement error is not the same for all examinees but it is a function of θ ([Muñiz, 1997](#)).

The specific advantage of the Rasch Model over other IRT models is that the estimated values for person and items are in the same scale of latent units (logits). This property is called *conjoint measurement*, which can be used to generate criterion-referenced interpretations in terms of qualitative descriptions of what the examinee can or cannot do (or what the examinee agrees or does not agree to do). This is possible thanks to the person-by-item map. Thus, the interpretation of scores in the Rasch Model is not based on group norms (as typically done in CTT), but it can be done in terms of item content and processes in which the examinee has a low or high probability of answering correctly (or has a low or high probability of endorsing). This trait provides the Rasch Model with a great diagnostic power.

Goodness of Fit Criteria

As [Bond and Fox \(2001\)](#) point out, before interpreting results in the Rasch Model it is necessary to check if the data adjusts reasonably to the model. There are several statistical measures of fit that can be used in this context, but one of the most widely used is called INFIT, which is an internal fit indicator corresponding to the residuals' weighted quadratic mean. Since items and persons are measured along the same scale, INFIT can be calculated for both of them.

The formula to obtain this measure is the following:

$$(3) \quad INFIT = \sum \frac{z_{vi}^2 W_{vi}}{N}$$

Where each observation, (item endorsability or person's level in the construct) is weighted by its individual variance.

INFIT gives more importance to the examinees or items whose trait level is located near the item difficulty or person ability. Thus, at the examinee level, the INFIT indicator will attach more weight to items with difficulties (agreeability or endorsability) near the examinee's score. Conversely, at the item level, INFIT will give more weight to persons' ability estimates that are near the item difficulty ([Bond & Fox, 2001](#)).

Smith, Schumaker and Bush (as cited by [Prieto-Adanes & Dias-Velasco, 2003](#)) recommend different intervals to evaluate INFIT depending on sample size. Thus, INFIT values higher than 1.3 indicate lack of fit in samples with less than 500 subjects, 1.2 is the threshold value for samples between 500 and 1000 subjects, and 1.1 is the threshold value for samples with more than 1000 subjects.

Rasch models have been employed to test psychometric properties of tests intended to measure performance, abilities and competences.

Their employment on instruments for measuring attitudes, motivations, interests, values, subjective appreciations or psychological traits (the so called affective domain) is less frequent. The present study illustrates, using the Benevolent Sexism Subscale, how the Rasch Model improves the analyses and interpretations of these types of scales, compared to the CTT.

Benevolent Sexism: Conceptualization

Over the past 30 years, research on sexism against women has provided compelling evidence of the pervasiveness of anti-female biases in our societies. *Sexism* has been traditionally defined as the endorsement of discriminatory or prejudicial beliefs and feelings based on sex, usually linked with stereotypical conceptions of the sexes and the adoption of a traditional gender-role ideology (Moya & Expósito, 2001).

Currently, considerable attention has been paid to contemporary forms of sexism against women in the light of two observations: First, in the current cultural climate it is unlikely that respondents will openly endorse prejudicial attitudes toward women (Campbell, Schellenberg, & Senn, 1997). Second, given the particular intimate relationship between men and women, sexism against women does not always reflect open hostility, but rather a profound ambivalence (Glick & Fiske, 1996).

In trying to capture the complexity of contemporary forms of sexism, several researchers have conceptualized it in different ways. For instance, Glick and Fiske (1996) describe sexism as a multidimensional construct that involves both, hostile and benevolent attitudes toward women. *Hostile sexism* is characterized as antipathy and derogatory attitudes, as in the classical definition of prejudice, while *benevolent sexism* is defined

as a set of subjectively positive attitudes that are sexist in terms of typecasting women in restricted roles.

Consequently, Glick and Fiske (1996) developed a scale attempting to measure this construct accurately and reliably: the Ambivalent Sexism Inventory (ASI), which comprises two subscales: the Hostile Sexism Scale (HS) and the Benevolent Sexism Scale (BS). A detailed description of the instrument is presented in the method section.

In the present study we focus specifically on BS because of the unique characteristics of the construct and its relevance for understanding contemporary forms of sexism. Despite of the subjectively positive content of the scale, it reflects sexism by justifying traditional gender roles and masculine dominance (e.g., the man as the provider and woman as his dependent). Interestingly, benevolent sexist attitudes are not always recognized as such by respondents, who tend to endorse BS items more strongly than HS items. Moreover, participants who endorse benevolently sexist beliefs are more likely to endorse other gender-traditional attitudes, including hostile sexism, unaware of the fact that they are endorsing two complementary aspects of the same sexist ideology.

Additionally, BS is harmful for women by itself, not only because of its relationship with HS. Data show that men who endorse BS are more likely to blame a female victim of rape if she has “infringed” traditional gender role expectations (Viki, Abrams, & Masser, 2004); and women who endorse BS are more likely to accept an ostensibly protective and restrictive male as a romantic partner, even if it implies a constraint to their career aspirations (Moya, Glick, Expósito, De Lemus, & Hart, 2007). In sum, as Glick and Fiske (1996) point out, the BS Scale measures an aspect of sexism with important consequences for women that many other instruments might overlook.

To the best of our knowledge, no study describing the adaptation or validation of this specific subscale using Rasch Analyses has been published so far. Therefore, analyzing it with this approach could be useful for illustrating the benefits of the Rasch Model when it comes to a deeper understanding of psychometric properties of scales in the affective domain.

Method

Participants

Analyses were run on a random cluster sample of 197 students from the University of Costa Rica, the National University of Costa Rica, and the Costa Rica Institute of Technology. These are the main State universities of the Country, located in the Metropolitan Area of the Central Valley of Costa Rica. One hundred and sixty six (84.3% of the sample) were women. The mean age was 21.69 years ($SD = 3.67$ years). Inclusion criteria were: a) being an active student of introductory Humanities and Math courses at these universities, and b) voluntarily participating in the study.

Instruments

The paper-pencil questionnaire contained a brief demographic section, along with several measures of attitudes toward women, including a Latin American adaptation of the ASI (Cárdenas, Lay, González, Calderón, & Alegría, 2010). The 22-Item ASI is made up of two subscales: HS, which basically matches the old sexism conceptualization, and BS, reflecting women as delicate creatures, confined to limited roles. Examples of HS items are *Women seek to gain power by getting control over men* and *Women exaggerate problems they have at work*. Examples of BS

items are *Many women have a quality of purity that few men possess* and *Women should be cherished and protected by men*. Items are rated in a 5-point Likert scale.

Glick and Fiske (1996) reported Cronbach's alpha coefficients for the overall scale ranging from .80 to .90. For the HS subscale, alphas have been ranged from .80 to .90, whereas for BS subscale alphas are lower, ranging from .70 to .85. Their validity studies yielded significant correlations between ASI, specially the HS, and other measures of sexism, racism and gender biases. Further analytic evidence supports the idea that the ASI scores show two correlated yet distinct primary dimensions: hostile and benevolent sexism (Glick & Fiske, 1996).

Other authors have also provided evidence of the reliability and validity of ASI-scores not only among adults (Becker & Wagner, 2009; Cárdenas et al., 2010), but also among adolescents (De Lemus, Moya, & Glick, 2010; Etchezahar & Ungaretti, 2014) in different social, cultural and linguistic contexts (Chen, Fiske, & Lee, 2009; Rodríguez-Castro, Lameiras-Fernández, & Carrera-Fernández, 2009; Rodríguez & Magalhães, 2013; Sakalli-Uğurlu & Glick, 2003). A general description of the psychometric properties of the measure from the CTT approach can be found in Fiske & North (2014). The complete scale and scoring instructions are available in Glick and Fiske (1996).

Procedures

Questionnaires were group administered to the students in their classrooms. Following the guidelines of the Institutional Revision Board (IRB) of the University of Costa Rica, respondents were informed about the purpose of the study, that their participation was voluntary, that

no reward would be given and that the personal information will remain confidential.

Analyses

To test the psychometric properties of BS from the perspective of the CTT, means, standard deviations, and item-total correlations were calculated for all items, as well as Cronbach's alpha coefficient and standard error of measurement for the total scale using SPSS 21 (IBM Corporation, 2012). RA comprised persons and items fit analyses, using INFIT statistics. INFIT values between 0.5 and 2.0 were considered acceptable for respondents' fit (Linacre, 2002), whereas values ranging from 0.7 to 1.3 were considered satisfactory for items' fit (Prieto & Delgado, 2003). Secondly, the Extended Rasch Model was estimated using joint maximum likelihood using the WINSTEPS 3.72.3, including respondents' scores and items' endorsabilities (difficulties), reliabilities, measurement errors, as well as the person-item map.

Data Preparation and Preliminary Analyses

In preparation for the main analyses, data was screened to detect major problems with asymmetrical distributions, missing values and outliers. Since diagnostic analyses revealed no major issues in this regard, all items were retained for further analyses. Only respondents who satisfactorily fitted the Extended Rasch Model ($N = 197$), were employed for comparison and contrast purposes.

Results

Table 1 shows some of the principal psychometric properties of BS obtained by means of

employing the CTT and RA. From the CTT perspective, data shows a Cronbach's alpha of .74, item means ranging from 1.49 to 2.67 (in a scale from 1 to 5), and item-total correlations from .30 to .51, with exception of item BS6, which shows an unacceptable item-total correlation of .04. Regarding RA, data revealed an average respondents' reliability of .68, which means that if the same group of participants were to answer to another set of items drawn from the same hypothetical item universe, the estimated correlation between the two estimations of the construct would be approximately .68. RA also showed item endorsability parameters ranging from -0.81 to 0.75, in the logit scale. Only one item, BS6, showed an unacceptable INFIT value of 2.13. Therefore, item BS6 was left out for the subsequent analyses. Standard errors of measurement ranged from .06 to .09, as shown in Figure 1, and in the persons vs. items map in Figure 2.

Finally, RA can be used to generate criterion-referenced interpretations about respondents' attitudes. For BS scale, the person vs. item-map allows creating categories of responses corresponding to different levels of endorsability. This is illustrated by the analysis of item content shown in Table 2.

Discussion

The purpose of the present study is to illustrate the benefits of using RA for the analysis and interpretation of attitudinal scales by applying RA and CTT procedures to the Benevolent Sexism Scale.

There are some similarities in the information provided by both approaches, since both models yielded statistics indicating poor psychometric quality for item BS6, in one case because it shows an item-total correlation of .04, in the

Table 1
Statistical properties for BS under CTT and RA.

Item	CTT				RA		
	Difficulty	Standard deviation	Standard error of measurement	Item-total correlation	Logit	Standard error of measurement	INFIT
BS1	2.05	1.324		.470	0.14	0.070	0.96
BS3	2.67	1.300		.304	-0.33	0.060	0.98
BS6	3.05	1.743		.038	-0.59	0.060	2.13
BS8	2.15	1.285		.508	0.05	0.070	0.80
BS9	3.37	1.451		.466	-0.81	0.060	0.92
BS12	2.21	1.352		.556	0.01	0.060	0.80
BS13	1.49	0.932		.509	0.75	0.090	0.84
BS17	1.92	1.254		.457	0.25	0.070	0.96
BS19	2.06	1.183		.459	0.13	0.070	0.79
BS20	1.60	1.025		.484	0.60	0.080	0.87
BS22	2.49	1.375		.407	-0.20	0.060	0.94
TOTAL	25.06	7.638	3.857				

Note. BS = Benevolent Sexism. Item numbering corresponds to Glick and Fiske (1996).

other because it presents an INFIT of 2.13. As highlighted before, both approaches emphasized the need of removing this item for any subsequent analyses.

In other aspects, however, both models offered different results. CTT assumes a constant measurement error, which in this case was equal to 3.857 (in a total scale ranging from 11 to 55), i.e., regardless of the construct level, all items are assumed to provide the same precision. On the other hand, RA relaxes this limiting assumption, estimating specific measurement errors for both respondents and items. Thus, in this case, measurement errors varied along different levels of the construct, being more accurate those near 0 in the logit scale, which is centered on the item average endorsability (see Figure 1).

CTT and RA also rendered different results regarding the total scores (i.e. the estimated construct level for each respondent). The Cronbach's alpha value of .745 suggests an acceptable internal consistency measure for research purposes. On the contrary, the estimated person reliability of 0.68 is clearly not satisfactory even for research purposes.

As it was previously mentioned, a very valuable feature of RA is its conjoint measurement property. It means that estimations for respondents' scores and items' means (i.e., endorsabilities) are calculated in the same logit scale. In this regard, Figure 2 depicts this useful property that CTT does not offer, showing that there are very few items measuring construct levels lower than 0. In addition, measurement accuracy increases as

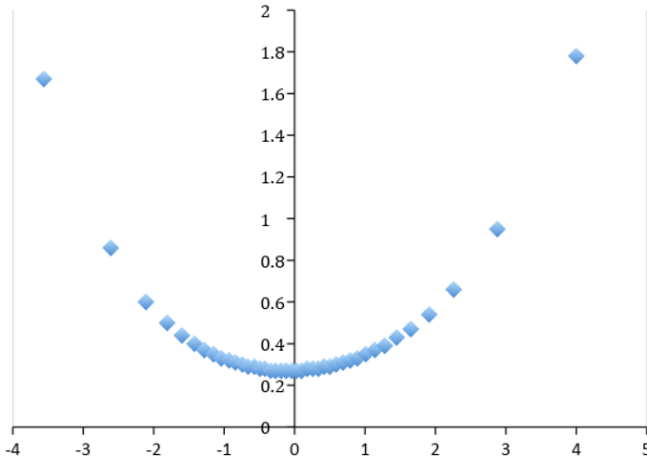


Figure 1
Measurement errors for BS in Rasch Analysis.

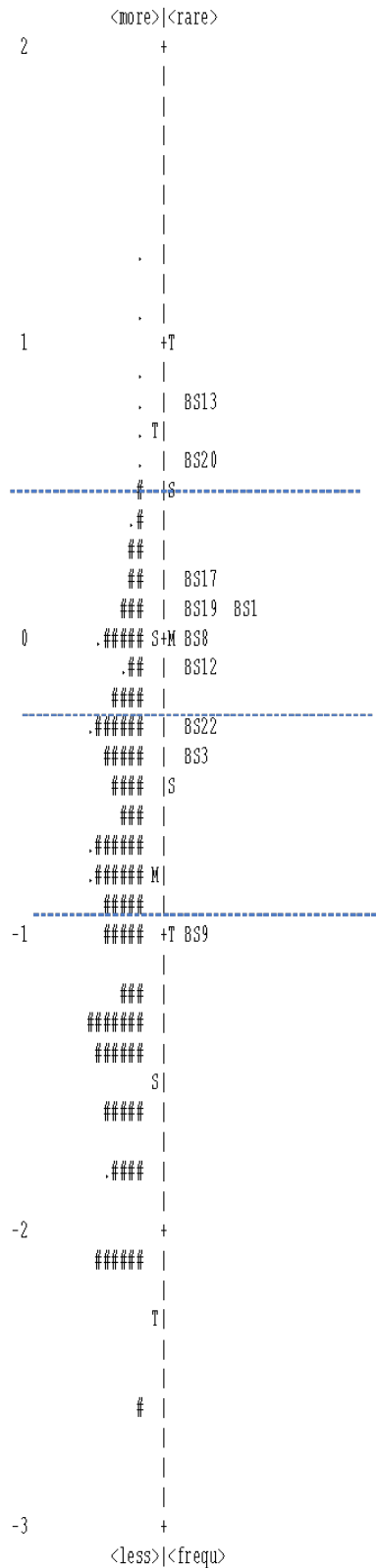
Note. Each dot represents the specific measurement error associated to each score in the construct. X axis: the level of the construct in logit units. Y axis: measurement error in logit units. 0 = Mean difficulty (endorsability) of the items.

the estimated scores reach values close to 0. This means that, for a considerable number of people, the construct cannot be accurately measured with the BS scale; in particular for those who are more likely to disagree with item content.

Taking into account the theoretical background of the BS, expert judgments and respondents' scores distribution, we noticed that only one item (BS9) represented the lower range of the BS scale [-2.61, -0.83] for this group. The content of this item reflects a mild kind of protective paternalism towards women, which might be seen as an inoffensive form of modern-day chivalry. Because these kind of benevolent sexist attitudes seem positive, participants might not recognize these beliefs as a form of gender-based prejudice, therefore this kind of items are more likely to be endorsed.

Participants in the next level [-0.83, -0.30], do not only endorsed paternalistic attitudes in

Figure 2
Persons vs. items map for the BS Scale.



Note. Each “#” is 2, each “.” is 1. BS = Benevolent Sexism.

Table 2
Subjects' BS profiles.

Score interval	Description	Content
[0.5, 1.29]	BS13. Women are incomplete without men (0.81)	Heterosexual intimacy
	BS20. Men should be willing to sacrifice their own well-being in order to provide financially for the women in their lives (0.64)	Protective paternalism.
[-0.30, 0.5]	BS17 A good woman should be set on a pedestal by her man (0.24)	Protective paternalism
	BS1 No matter how accomplished he is, a man is not truly complete as a person unless he has the love of a woman (0.09)	Heterosexual intimacy
	BS19 Women, compared to men, tend to have a superior moral sensibility (0.09)	Complementary gender differentiation
	BS8 Many women have a quality of purity that few men possess (0.00)	Complementary gender differentiation
	BS12 Every man ought to have a woman whom he adores (-0.06)	Heterosexual intimacy
[-0.83, -0.30]	BS22 Women, as compared to men, tend to have a more refined sense of culture and good taste (-0.32)	Complementary gender differentiation
	BS3 In a disaster, women ought to be rescued before men (-0.44)	Protective paternalism.
[-2.61, -0.83]	BS9 Women should be cherished and protected by men (-1.05)	Protective paternalism.

Note. BS = Benevolent Sexism. In the second column, values within parentheses are item difficulties.

form of chivalry, but also stereotypical complementary gender differentiation; that is, participants with this level of sexism tended to endorse the idea that women are delicate creatures, and that they therefore need to be protected.

Construct levels of those participants with scores between [-0.30, 0.5] are the most accurately represented with this instrument. There are five items covering this interval, reflecting the three aspects of BS described by the theory, i.e.

protective paternalism, complementary gender differentiation and heterosexual intimacy. At this level, participants not only endorsed mild forms of modern-day chivalry and the notion that women are delicate creatures in need of protection, but also the idea that men are incomplete without women, reflecting heterosexual intimacy, i.e. the belief that romantic intimacy is necessary to complete a man, but also that women are incomplete without men.

Finally, the most *difficult* items; i.e., those which are *more difficult* to be endorsed by participants, turned out to be a combination of an extreme form of protective paternalism (BS20), and a plain statement that a woman is incomplete without a man by her side (BS13). Participants with this level of benevolent sexism are more willing to accept that women are so defenseless that men should sacrifice themselves in order to protect them, reflecting not only the superiority of men over women, but also undermining the notion of women as competent and independent agents.

Notice that in both the lower and the higher levels of the construct, the measurement was less accurate; since not all three components of BS were present along the continuum and because of the reduced number of items, which resulted in a less precise measurement (see Figure 1).

Conclusion

Our data showed that the Extended Rasch Model is a useful tool for testing psychometric aspects of scales in the attitudinal domain such as the Benevolent Sexism Subscale. It also allows researchers and practitioners to generate meaningful interpretations about the construct being measured. CTT, although useful for some purposes, is more restrictive; presenting important shortcomings that RA helps to overcome. This paper illustrates several valuable features of RA, as fit statistics for both persons and items, and specific estimations for measurement error at different levels of the construct. More importantly, the conjoint measurement property provides the Extended Rasch Model with a particular advantage over other IRT models, allowing researchers to generate respondents' profiles and criterion-referenced interpretations.

This is particularly important for measures in the attitudinal domain. While CTT and other IRT models allow researchers to compute a score on BS, reflecting a global view of participants' BS (low or high) levels; RA allows researchers to understand which items are more probable to be endorsed by participants with different levels of BS, enhancing knowledge about the meaning of low and high scores in the measure. Using this tool, we can better understand how benevolent sexist attitudes toward women are constituted and organized, providing a deeper comprehension of contemporary sexism in our societies, which will also contribute to the development of educational programs and community interventions to foster social equity and justice.

References

- Becker, J. C., & Wagner, U. (2009). Doing gender differently: The interplay of strength of gender identification and content of gender identity in predicting women's endorsement of sexist beliefs. *European Journal of Social Psychology, 39*(4), 487-508. doi: [10.1002/ejsp.551](https://doi.org/10.1002/ejsp.551)
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Campbell, B., Schellenberg, E. G., & Senn, C. Y. (1997). Evaluating measures of contemporary sexism. *Psychology of Women Quarterly, 21*(1), 89-101. doi: [10.1111/j.1471-6402.1997.tb00102.x](https://doi.org/10.1111/j.1471-6402.1997.tb00102.x)
- Cárdenas, M., Lay, S. L., González, C., Calderón, C., & Alegría, I. (2010). Inventario de sexismo ambivalente: Adaptación, validación y relación con variables psicosociales. *Revista Salud y Sociedad, 1*(2), 125-135. doi: [10.22199/s07187475.2010.0002.00006](https://doi.org/10.22199/s07187475.2010.0002.00006)
- Carvalho, L., Primi, R., & Meyer, G. J. (2012). Aplicação do modelo de Rasch na medida de transtornos da personalidade. *Trends in Psychiatry and Psy-*

- chotherapy, 34(2), 101-109. doi: [10.1590/S2237-60892012000200009](https://doi.org/10.1590/S2237-60892012000200009)
- Chen, Z., Fiske, S. T., & Lee, T. L. (2009). Ambivalent sexism and power-related gender-role ideology in marriage. *Sex Roles, 60*(11-12), 765-778. doi: [10.1007/s11199-009-9585-9](https://doi.org/10.1007/s11199-009-9585-9)
- De Lemus, S., Moya, M., & Glick, P. (2010). When contact correlates with prejudice: Adolescents' romantic relationship experience predicts greater benevolent sexism in boys and hostile sexism in girls. *Sex Roles, 63*, 214-225. doi: [10.1007/s11199-010-9786-2](https://doi.org/10.1007/s11199-010-9786-2)
- Etchezahar, E., & Ungaretti, J. (2014). Woman stereotypes and ambivalent sexism in a sample of adolescents from Buenos Aires. *Journal of Behavior, Health & Social Issues, 6*(2), 87-94.
- Fiske, S. T., & North, M. S. (2014). Measures of stereotyping and prejudice: Barometers of bias. In G. Boyle & D. Saklofske (Eds.), *Measures of Personality & Social Psychological Constructs* (pp. 684-718). Boston, MA: Elsevier Academic Press.
- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*(3), 491-512. doi: [10.1037/0022-3514.70.3.491](https://doi.org/10.1037/0022-3514.70.3.491)
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Berlín: Springer Science & Business Media.
- IBM Corporation (2012). IBM SPSS Statistics for Windows (Version 21.0) [computer software]. Armonk, NY: IBM
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878. Retrieved from <https://www.rasch.org/rmt/index.htm>
- Moya, M., & Expósito, F. (2001). Nuevas formas, viejos intereses: Neosexismo en varones españoles. *Psicothema, 13*(4), 643-649. Retrieved from <http://www.psicothema.com>
- Moya, M., Glick, P., Expósito, F., De Lemus, S., & Hart, J. (2007). It's for your own good: Benevolent sexism and women's reactions to protectively justified restrictions. *Personality and Social Psychology Bulletin, 33*(10), 1421-1434. doi: [10.1177/0146167207304790](https://doi.org/10.1177/0146167207304790)
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid, España: Pirámide.
- Muñiz, J. (2017). *Teoría Clásica de los Tests*. Madrid, España: Pirámide.
- Muñiz, J., Rogers, J., & Swaminathan, H. (1989). Robustez de las estimaciones de modelo de Rasch en presencia de aciertos al azar y discriminación variable de los ítems. *Anuario de Psicología, 43*(4), 82-97. Retrieved from <https://www.raco.cat/index.php/AnuarioPsicologia>
- Olsen, L. W. (2003). Essays on Georg Rasch and his contributions to statistics (Unpublished doctoral dissertation, University of Copenhagen, Institute of Economics) Retrieved from <http://www.rasch.org/olsen.pdf>
- Prieto, G., & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema, 15*(1), 94-100. Retrieved from <http://www.psicothema.com>
- Prieto-Adanes, G., & Dias-Velasco, A. (2003). Uso del modelo de Rasch para poner en la misma escala las puntuaciones de distintos tests. *Actualidades en Psicología, 19*(106), 5-23. doi: [10.15517/ap.v19i106.43](https://doi.org/10.15517/ap.v19i106.43)
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Rodríguez-Castro, Y., Lameiras-Fernández, M., & Carreira-Fernández, M. V. (2009). Validación de la versión reducida de las Escalas ASI y AMI en una muestra de estudiantes españoles. *Psicogente, 12*(22), 284-295. Retrieved from <http://revistas.unisimon.edu.co/index.php/psicogente>
- Rodríguez, Y., & Magalhães, M. J. (2013). El sexismo moderno en estudiantes universitarios/as portugueses/as. *Revista Interdisciplinar de Ciencias Sociales y Humanas, 1*(2), 113-121. Retrieved from <http://independient.academia.edu/revistaagir>
- Sakalli-Uğurlu, N., & Glick, P. (2003). Ambivalent sexism and attitudes toward women who engage in premarital sex in Turkey. *The Journal of Sex Research, 40*(3),

296-302. doi: [10.1080/00224490309552194](https://doi.org/10.1080/00224490309552194)

Viki, G. T., Abrams, D., & Masser, B. (2004). Evaluating stranger and acquaintance rape: The role of benevolent sexism in perpetrator blame and recommended sentence length. *Law and Human Behavior*, 28(3), 295-303. doi: [10.1023/b:lahu.0000029140.72880.69](https://doi.org/10.1023/b:lahu.0000029140.72880.69)

Wilson, M. (2004). *Constructing measures: An item response modeling approach*. New York, NY: Routledge. doi: [10.4324/9781410611697](https://doi.org/10.4324/9781410611697)

Contextos y creatividad Variables sociodemográficas y datos normativos en el Test CREA

Contexts and Creativity Sociodemographic Variables and Normative Data in the CREA Test

Romina Elisondo *¹, Danilo Donolo¹

1 - Consejo Nacional de Investigaciones Científicas y Técnicas. Departamento de Ciencias de la Educación. Facultad de Ciencias Humanas. Universidad Nacional de Río Cuarto. Argentina.

Recibido: 29/04/2018 Revisado: 19/06/2018 Aceptado: 07/07/2018

Introducción
Método
Resultados
Discusión
Referencias

Resumen

El Test CREA es un instrumento de evaluación de la creatividad que valora la capacidad de las personas de formular interrogantes ante determinados estímulos. En el presente estudio se analizan desempeños en el Test CREA de niños, adolescentes y adultos que residen en la provincia de Córdoba (Argentina). Se utilizó el Test CREA para medir la creatividad por sus propiedades psicométricas. Al grupo de estudio lo constituyen 3237 niños, adolescentes y adultos. Las variables de incidencia analizadas fueron *género*, *edad* y *escolaridad*. Los resultados indican diferencias significativas según *edad* en los tres grupos estudiados, según *género* en adolescentes y según *escolaridad* en adultos. En el artículo se proponen datos normativos que atienden a las diferencias observadas. Los datos presentados son de relevancia para la interpretación de los resultados del Test CREA en grupos y contextos diversos.

Palabras clave: *creatividad, test CREA, normas, género, edad, escolaridad*

Abstract

The CREA Test is an instrument for evaluating creativity that assesses the ability of people to ask questions about certain stimuli. In this study we analyze the performances of the CREA Test of children, adolescents and adults residing in the province of Córdoba (Argentina). The CREA Test was used to measure creativity for its proven technical characteristics. The study group consists of 3237 children, adolescents and adults; the incidence variables analyzed were *gender*, *age* and *schooling*. The results indicate significant differences regarding *age* in the three groups studied, regarding *gender* in adolescents and regarding *schooling* in adults. The article proposes normative data that address the differences observed. The presented data are relevant for the interpretation of the results of the CREA Test in different groups and contexts.

Key words: *creativity, CREA test, norms, gender, age, education*

* **Correspondencia a:** Romina Elisondo. Departamento de Ciencias de la Educación. Facultad de Ciencias Humanas. Universidad Nacional de Río Cuarto. Ruta 36 Km 601. Río Cuarto CP 5800 Córdoba Argentina. relisondo@gmail.com 0054 358 4676473

Cómo citar este artículo: Elisondo, R., & Donolo, D. (2018). Contextos y creatividad. Variables sociodemográficas y datos normativos en el Test CREA. *Revista Evaluar*, 18(3), 14-29. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

En la bibliografía especializada, generalmente se acepta que todas las personas son creativas en algún sentido y en alguna actividad de su vida (Richards, 2007). En nuestros trabajos anteriores también hemos adherido a esta idea promisorio y alentadora para dar más espacio a la educación y al aprendizaje. Nuestro interés es analizar cómo se desempeña en una prueba de creatividad un grupo de sujetos en distintas etapas de su desarrollo y con diferentes niveles de *escolaridad*. El objetivo es contar con los elementos para sugerir acciones que tiendan a que la mayoría de las personas pueda encontrar un ambiente en el que expresar sus capacidades creativas.

La complejidad del constructo definido como *creatividad* dificulta contar con consensos amplios en cuanto a los alcances y los límites de su definición y a los acuerdos para su medición. Por eso, surgen inquietudes entre los investigadores respecto a la posibilidad de medir la creatividad, los aspectos o dominios que podrían evaluarse y los indicadores más apropiados.

Estas son algunas de las difíciles preguntas que intentan responder, sin pocas controversias y debates, los investigadores interesados en la medición de la creatividad (Baer, 2012; Batey, 2012; Barbot, Besançon, & Lubart, 2016). Desde los estudios psicométricos iniciales de Guilford (1962), hasta la actualidad, se han creado, perfeccionado y estandarizado diferentes instrumentos para evaluar la creatividad y sus diversas manifestaciones. La mayoría de los instrumentos estandarizados se orientan a valorar capacidades creativas a partir de tareas y procedimientos que evalúan el pensamiento divergente. El Test de Pensamiento Creativo de Torrance es uno de los instrumentos más utilizados en la intervención e investigación psicológica (Humble, Dixon, & Mporfu, 2017; Torrance, 1990). Se han realizado numerosos

estudios de validación y estandarización del instrumento en diversos lugares geográficos, como por ejemplo en Turquía, Corea, Portugal y Brasil (Primi, Nakano, Morais, Silva-Almeida, & Marques-David, 2013; Yarbrough, 2016; Yoon, 2017; Wechsler, 2006) y también en Argentina (Aranguren, 2014; Krumm, Aranguren, Arán-Filippetti, & Lemos, 2016). También se han desarrollado instrumentos estandarizados para evaluar logros creativos en diferentes campos del conocimiento como el Creative Behaviors Inventory (Hocevar, 1979), el Creative Achievement Questionnaire (Carson, Peterson, & Higgins, 2005) y el Biographical Inventory of Creative Behaviors (Batey, 2007). En Argentina se han confeccionado dos cuestionarios orientados a valorar comportamientos y acciones creativas en diferentes situaciones y áreas del conocimiento: la Escala de Comportamiento Creativo (Aranguren & Irrazabal, 2012) y el Cuestionario de Acciones Creativas (Elisondo & Donolo, 2016a).

Los estudios psicométricos son los más desarrollados en el campo de la creatividad, tal como lo afirman Long (2014) y Said-Metwaly, Van den Noortgate y Kyndt (2017) a partir del análisis de artículos de investigación publicados en revistas especializadas. Said-Metwaly et al. (2017) mencionan limitaciones en los instrumentos, específicamente aquellas vinculadas con la falta de validez y la valoración de aspectos individuales sin considerar variables contextuales. Asimismo, sugieren analizar las posibilidades de integración de diferentes instrumentos y ponen énfasis en la importancia de considerar la variabilidad de los perfiles de creatividad en contextos, dominios y etapas evolutivas diferentes.

Considerando la relevancia de disponer de datos normativos actualizados que atiendan a diferencias sociodemográficas, el presente artículo se dispone a analizar la incidencia del *género*, la *edad* y el nivel de *escolaridad* en las puntuaciones

en el Test CREA (Corbalán-Berná, Martínez-Zaragoza, Donolo, Alonso-Monreal, Tejerina-Areal y Limiñana-Gras, 2003) y presentar datos normativos para tres etapas etarias. El Test CREA evalúa la inteligencia creativa en niños, adolescentes y adultos a partir de una tarea general: la formulación de interrogantes ante estímulos visuales. El Test CREA se fundamenta teóricamente en las perspectivas teóricas de la creatividad que enfatizan el valor de los procesos de formulación de problemas. Desde estas perspectivas, se considera que la formulación de interrogantes es el proceso cognitivo básico de la creatividad. A partir de los procesos de formulación de problemas emergen ideas y productos creativos. En esta línea argumentativa, la creatividad es definida como la capacidad de formular cuestiones y resolver problemas de manera divergente (Corbalán-Berná, 2008).

El Test CREA ha sido utilizado en numerosas investigaciones en diversos países y con variadas poblaciones (Corbalán-Berná et al., 2003; Corbalán-Berná et al., 2014; López-Martínez & Navarro-Lozano, 2008). Se han realizado estudios de las propiedades psicométricas del instrumento en poblaciones española, argentina y estadounidense (Clapham & King, 2010; Martínez-Zaragoza, 2003), se hallaron índices de confiabilidad adecuados y se mostraron diversas evidencias de validez. Se han observado correlaciones entre capacidades creativas medidas con el CREA y las mediciones realizadas con la Batería de Guilford y el Test de Torrance (Corbalán-Berná et al., 2003; Clapham & King, 2010; López-Martínez & Navarro-Lozano, 2008). El Test CREA es un instrumento utilizado extensamente en la investigación de la creatividad, tal como se visualiza en investigaciones actuales con niños (Antoñanzas-Laborda, Toner-Pelay, Salavera-Bordás, Soler-Costa, & Usan-Supervía, 2015; Cárdenas-Avila, López-Fernández, & Arias-Cas-

tro, 2018), adolescentes (Castañeda-Rey, López, & Ezquerro-Cordón, 2017; Ramos-Moreno, López-Fernández, & Llamas-Salguero, 2017; Sánchez-Hernández, Méndez, & Garber, 2015) y adultos (Bogaert-García, 2017; Rodríguez-Suárez, Llamas-Salguero, & López-Fernández, 2015).

En Argentina se han realizado diferentes estudios del Test CREA: análisis de relaciones entre creatividad, inteligencia (Elisondo & Donolo, 2010) y personalidad (Elisondo, Donolo, & Corbalán-Berná, 2009), pruebas exploratorias con nuevos estímulos (Elisondo & Donolo, 2011), estudios de correlaciones entre creatividad verbal y gráfica (Freiberg-Hoffmann & Fernández-Liporace, 2017) e indagación de relaciones entre capacidades creativas y participación en actividades de ocio (Chiecher, Elisondo, Paoloni, & Donolo, 2018; Elisondo & Donolo, 2014; 2016a; 2016b). También se han realizado estudios exploratorios de diferentes procedimientos de puntuación de las preguntas (Elisondo, 2015) y análisis de la originalidad de los interrogantes (Elisondo, Donolo, & Limiñana-Gras, 2018).

El objetivo principal de esta investigación es analizar los desempeños en el CREA de una muestra amplia de niños, adolescentes y adultos argentinos de diferentes edades y niveles de escolaridad. En los tres grupos etarios, se analiza la incidencia del *género*, la *edad* y la *escolaridad* en el Test CREA y se presentan datos normativos que atienden a las particularidades de los grupos estudiados. Se considera de relevancia disponer de datos actualizados y referencias normativas sobre el instrumento que consideren particularidades sociodemográficas teniendo en cuenta que en estudios anteriores se han observado diferencias según género, edad y escolaridad de los participantes (Chacón-Araya & Moncada-Jiménez, 2006; Donolo & Elisondo, 2007; Elisondo & Donolo 2011; Limiñana-Gras, Bordoy, Juste-Ballesta, &

Corbalán-Berná, 2010). La relevancia del estudio presentado también se vincula con la amplitud de la muestra seleccionada ($N = 3237$) y la diversidad de casos incluidos (varones y mujeres de diferentes edades y niveles de escolaridad). La significatividad del estudio también se vincula con los desafíos actuales de la psicometría argentina y la necesidad de contar con estudios psicométricos y datos normativos para el uso de instrumentos de medición en contextos locales (Contini, 2013).

Método

Participantes

La muestra estuvo conformada por 3237 personas que residían al momento de la investigación en la provincia de Córdoba (Argentina). Participaron 192 niñas y 205 niños (entre 6 y 11 años), 421 adolescentes mujeres y 299 varones (entre 12 y 17 años) y 1352 mujeres y 768 hombres adultos (entre 18 y 82 años). Se incluyeron en la muestra jóvenes y adultos con diferentes experiencias educativas, según los tres niveles del Sistema Educativo Argentino: primario (8%), secundario (30%) y superior (62%). Las niñas y los niños de la muestra cursaban el nivel primario y los adolescentes el nivel secundario de la educación obligatoria argentina, en escuelas de gestión pública y privada. Todos los participantes eran de nacionalidad argentina y nivel socioeconómico medio.

Instrumentos

Se utilizaron como instrumentos de recolección de datos el Test CREA y un cuestionario breve sobre datos sociodemográficos. El cuestionario recoge datos acerca del género, la edad, los estudios en curso o el nivel de escolaridad alcanzado (en el caso de quienes ya no se encuentran

en el sistema educativo) y el lugar de residencia de los participantes.

El objetivo del Test CREA es medir la inteligencia creativa a partir de la evaluación de la capacidad de las personas de formular interrogantes ante estímulos visuales. Tres láminas (A, B y C) conforman el instrumento. La consigna del test es: *Se te va a presentar una ilustración. Tu tarea consiste en escribir brevemente todas las preguntas que puedas hacer sobre lo que representa la lámina. Trata de hacer el mayor número posible de preguntas en cuatro minutos.* La aplicación del test es escrita, individual o colectiva, a individuos mayores de 10 años. Entre los 6 y los 9 años la administración debe ser individual y oral. Las propiedades psicométricas del instrumento han sido documentadas a partir de investigaciones en diferentes contextos geográficos (Clapham & King, 2010; Corbalán-Berná et al., 2014; Corbalán-Berná et al., 2003; López-Martínez & Navarro-Lozano, 2008; López-Martínez & Brufau, 2010; Martínez-Zaragoza, 2003). En el manual del test se detallan estudios de fiabilidad y validez, se analizan formas paralelas ($.87$) y se presentan evidencias de validez convergente con la Batería de Guilford ($r = .61$; $r = .75$; $r = .59$; $r = .48$) y de validez discriminante con pruebas de inteligencia ($r = .25$; $r = .20$). Considerando criterios externos, también se observan correlaciones moderadas entre el CREA y las puntuaciones de expertos en una tarea creativa (Corbalán-Berná et al., 2003).

Análisis de datos

Se realizaron análisis mediante el programa especializado de estadísticas para ciencias sociales y humanas SPSS. Se realizaron estudios de frecuencia, media, desviación estándar, mínimo y máximo en las tres láminas y en los tres grupos.

Con el objetivo de analizar la incidencia de variables sociodemográficas (*género, edad y escolaridad*), se realizaron estudios de diferencia de media mediante prueba *t* y ANOVA. Para la construcción de datos normativos y la obtención de puntuaciones derivadas se calcularon percentiles para cada lámina y grupo. En el presente estudio, al igual que en el manual del test, se realizaron transformaciones no lineales para la construcción de baremos. De esta manera, es posible comparar los baremos disponibles en el manual del test con los datos normativos propuestos en la presente investigación. Los grupos considerados también son similares a los definidos en los baremos del manual del CREA.

Tabla 1

Media y prueba *t* en las tres láminas del CREA según género de los niños.

Género		CREA A	CREA B	CREA C
Niñas	M	10.94	11.19	8.38
	DE	5.35	7.35	4.99
Niños	M	10.80	11.74	7.65
	DE	5.70	7.07	5.50
		$t_{(1,313)} = .232$	$t_{(1,141)} = -.460$	$t_{(1,314)} = 1.241$
<i>p</i>		.817	.646	.215

La prueba ANOVA indica diferencias estadísticamente significativas en el CREA según la edad de los participantes (ver Tabla 2). A partir del análisis exploratorio de las medias se decidió conformar dos grupos que muestran diferencias en los desempeños en el CREA: el primero conformado por niños de 6 a 8 años y el segundo por niños de 9 a 11 años. Se observaron diferencias de medias estadísticamente significativas entre ambos grupos en el CREA (ver Tabla 3).

Resultados

Niños

Los análisis descriptivos del presente estudio (Lámina A: $M = 10.86$, $DE = 5.53$; Lámina B: $M = 11.48$, $DE = 7.18$; Lámina C: $M = 8.02$, $DE = 5.25$) indican puntajes medios superiores a los establecidos en el manual del Test CREA para la población argentina de 6 a 11 años (Lámina A: $M = 5.48$, $DE = 3.78$; Lámina C: $M = 4.22$, $DE = 3.47$). Con respecto a la incidencia del género en el CREA, no se observan diferencias estadísticamente significativas entre los puntajes de los niños y los de las niñas (ver Tabla 1).

Se realizaron contrastes *post hoc* y se observaron diferencias significativas en la Lámina A entre las medias de niños de 6,7 y 8 años y la medias de niños de los niños de 9, 10 y 11 años. Entre los niños de 9, 10 y 11 años no se observaron diferencias significativas de media en la Lámina A. En la Lámina B se observaron diferencias entre los niños de 11 años y los de 6, 7, 8, 9 y 10 años. En la Lámina C, se observaron diferencias entre los de 6 y los de 7, 8, 9, 10 y 11 años, entre los de 7, 9 y 10 años, entre los de 8, 9 y 10 años y entre los de 10 y 11 años.

Tabla 2

Media y prueba ANOVA en las tres láminas del CREA según edad de los niños.

Edad		CREA A	CREA B	CREA C
6 años (n = 44)	M	3.09	3.55	2.40
	DE	2.91	3.24	3.24
7 años (n = 50)	M	4.15	5.50	6.32
	DE	2.91	2.99	3.79
8 años (n = 27)	M	5.70	6.37	6.25
	DE	1.92	3.20	3.85
9 años (n = 82)	M	10.73	6.22	8.39
	DE	4.53	3.11	3.85
10 años (n = 87)	M	12.50	7.16	12.29
	DE	5.41	3.18	5.31
11 años (n = 107)	M	12.30	14.55	7.19
	DE	5.38	6.84	3.41
		$F_{(5,309)} = 18.65^{**}$	$F_{(5,137)} = 14.88^{**}$	$F_{(5,310)} = 36.66^{**}$

Nota. $** p < .01$ **Tabla 3**Media y prueba t en las tres láminas del CREA según dos grupos de edades: 6-8 años y 9-11 años.

Edad		CREA A	CREA B	CREA C
6-8 años	M	4.59	5.40	4.88
	DE	2.67	3.27	4.05
9-11 años	M	11.88	13.45	9.97
	DE	5.19	7.00	4.97
		$t_{(1,313)} = -9.11^{**}$	$t_{(1,141)} = -6.55^{**}$	$t_{(1,314)} = -9.47^{**}$

Nota. $** p < .01$

Conforme a las diferencias observadas según *edad*, se considera pertinente presentar datos normativos para cada uno de los grupos de niños (de 6 a 8 años y de 9 a 11 años). Se proponen a continuación percentiles para cada una de las láminas del CREA en los dos grupos de niños definidos según *edad* (ver Tabla 4).

Tabla 4
Percentiles para niños de 6 a 8 y de 9 a 11 años en las Láminas A, B y C del Test CREA.

Percentil	Niños de 6 a 8 años			Niños de 9 a 11 años		
	A	B	C	A	B	C
95	9	11	13	13	11	11
90	8	9	8	11	10	10
75	7	7	7	9	8	8
50	5	6	5	7	7	6
25	3	3	4	5	5	5
10	1	1	1	3	3	3
5				1	1	1
N	44	35	195	271	108	195
M	4.59	5.40	4.87	11.88	13.45	9.97
DE	2.67	3.27	4.05	5.19	7.00	4.97

Adolescentes

Los puntajes medios observados en el presente estudio (Lámina A: $M = 11$, $DE = 4.79$; Lámina B: $M = 10.84$, $DE = 5.10$; Lámina C: $M = 10.17$, $DE = 5.44$) son similares a los establecidos en el manual del Test CREA para la Lámina C ($M = 10.66$, $DE = 5.8$) e inferiores en la Lámina A ($M = 12.86$, $DE = 5.8$; ver Tabla 10). Al analizar las variaciones según *género*, se observan diferencias estadísticamente significativas a favor de las adolescentes en las Láminas A y B del CREA (ver Tabla 5).

El ANOVA mostró diferencias estadísticamente significativas en las tres láminas del CREA según la edad de los participantes (ver Tabla 6). Al igual que en la muestra de niños, se decidió conformar dos grupos según *edad*: de 12 a 14 años y de 15 a 17 años. Se observaron diferencias de medias estadísticamente significativas entre ambos grupos en las Láminas A y B del CREA (ver Tabla 7).

Tabla 5
Media y prueba t en las tres láminas del CREA según género en adolescentes.

Género	CREA A	CREA B	CREA C
M	11.32	11.45	10.34
DE	4.89	5.31	5.57
M	10.56	9.97	9.96
DE	4.636	4.67	5.29
	$t_{(2,718)} = 2.09$	$t_{(2,718)} = 3.87$	$t_{(2,618)} = .881$
p	.036	<.001	.379

Tabla 6

Media, desviación estándar y prueba ANOVA en las tres láminas del CREA según la edad de los adolescentes.

Edad		CREA A	CREA B	CREA C
12 años	M	8.29	8.05	8.47
(n = 71)	DE	2.96	3.03	3.59
13 años	M	9.58	9.01	10.67
(n = 53)	DE	4.34	4.27	5.57
14 años	M	10.39	10.77	10.64
(n = 89)	DE	4.63	5.23	6.18
15 años	M	11.24	11.26	10.26
(n = 124)	DE	4.47	4.912	4.75
16 años	M	11.38	11.40	10.17
(n = 119)	DE	5.12	5.13	6.24
17 años	M	11.94	11.52	10.36
(n = 264)	DE	5.00	5.42	5.51
		$F_{(5,714)} = 8.38$	$F_{(5,714)} = 7.27$	$F_{(5,614)} = 1.66$
<i>p</i>		<.001	<.001	.142

Se realizaron contrastes *post hoc* y se observaron diferencias significativas en la Lámina A entre las medias de los adolescentes de 12 años y los de 14, 15, 16 y 17; entre los de 13 y los de 15, 16, 17 y entre los de 14 y los de 17. En la Lámina B se observaron diferencias entre los adolescentes de 12 años y los de 13, 14, 15, 16 y 17 y entre los de 13 años y los de 14, 15, 16 y 17. En la Lámina C se observaron diferencias entre los de 12 y los de 13, 14, 15, 16 y 17 años.

Tabla 7

Media, desviación estándar y prueba *t* en las tres láminas del CREA según dos grupos de edades: 12 - 14 años y 15 - 17 años.

Edad		CREA A	CREA B	CREA C
12-14 años	M	9.49	9.43	9.90
	DE	4.15	4.51	5.33
15-17 años	M	11.64	11.43	10.30
	DE	4.91	5.23	5.50
		$t_{(1,718)} = - 5,59$	$t_{(1,718)} = - 4,87$	$t_{(1,718)} = - ,86$
<i>p</i>		<.001.	<.001	.386

Considerando las diferencias observadas, es pertinente presentar baremos según género y edad para el grupo de adolescentes. Se proponen a continuación percentiles para cada una de las láminas del CREA en varones y mujeres de 12 a 14 años y de 15 a 17 años (ver Tablas 8 y 9).

Adultos

Los puntajes medios hallados en la presente investigación (Lámina A: $M = 11.40$, $DE = 5.11$; Lámina B: $M = 10.72$, $DE = 4.91$; Lámina C: $M = 10.08$, $DE = 5.00$) son similares a los establecidos en el manual del Test CREA para la población argentina adulta en la Lámina A ($M = 11.53$, $DE = 5.3$) y los valores superiores en la Lámina B ($M = 8.56$, $DE = 5.6$). Con respecto a la incidencia del género en el CREA, se observaron diferencias estadísticamente significativas a favor de las mujeres en las Láminas B y C del CREA (ver Tabla 10). También se observaron diferencias en el CREA según la *edad* (ver Tabla 11) y la *escolaridad* (ver Tabla 12). Los análisis de varianza en los que se mantienen fijos los demás factores indican diferencias significativas entre los puntajes medios según la edad y el nivel de escolaridad en las tres Láminas del CREA.

Tabla 8

Percentiles en las tres láminas del CREA para adolescentes mujeres de 12 a 14 y de 15 a 17 años.

Percentil	Adolescentes mujeres de 12 a 14			Adolescentes mujeres de 15 a 17		
	A	B	C	A	B	C
95	19	19	23	21	21	22
90	15	15	16	18	20	18
75	11	12	13	14	14	13
50	9	8	9	10	10	9
25	6	6	6	8	7	7

10	5	5	5	6	6	5
5	4	4	4	5	5	4
N	120	120	120	219	219	219
M	9.50	9.43	9.96	11.35	11.16	10.55
DE	4.51	4.72	5.58	4.85	5.14	5.56

Tabla 9

Percentiles en las tres láminas del CREA para adolescentes varones de 12 a 14 y de 15 a 17 años.

Percentil	Adolescentes varones de 12 a 14			Adolescentes varones de 15 a 17		
	A	B	C	A	B	C
95	16	18	20	20	21	21
90	15	15	16	18	20	18
75	12	12	12	15	14	13
50	9	9	9	11	11	9
25	7	6	6	8	8	6
10	5	5	5	6	6	4
5	4	4	4	5	5	3
N	93	93	85	288	288	196
M	9.47	9.43	9.81	11.85	11.63	10.02
DE	3.66	4.23	4.99	4.95	5.29	5.42

Discusión

Se han observado diferencias significativas en las puntuaciones medias del CREA según la variable *edad* en los tres grupos estudiados. En el caso de los adolescentes también se observaron diferencias según *género* y en los adultos según nivel de *escolaridad*. Se han propuesto datos normativos que atienden a la incidencia de las variables mencionadas. Los datos presentados pueden contribuir a una mejor interpretación de los resultados de grupos particulares. Asimismo, representan herramientas valiosas para el diagnóstico y el análisis de capacidades creativas individuales.

La *edad* de las personas es una variable importante para considerar en la evaluación de la

creatividad en general y en el Test CREA. Investigaciones previas han señalado la necesidad de tener en cuenta la edad como variable de relevancia en la medición de la creatividad (Wechsler, 2006; Hang-Wu, Cheng, Man-Ip, & McBride-Chang, 2005). También en el Test CREA se han observado diferencias según la edad de los participantes (Chacón-Araya & Moncada-Jiménez, 2006; Donolo & Elisondo, 2007; Elisondo & Donolo, 2011; Limiñana-Gras et al., 2010; López-Martínez & Navarro-Lozano, 2008).

En los análisis también se observaron diferencias según el nivel de *escolaridad* de los participantes. Estudios anteriores señalan diferencias según el nivel de escolaridad de los participantes en el CREA (Elisondo & Donolo, 2011) y en

otras pruebas de capacidades creativas (Matud, Rodríguez, & Grande, 2007). La consideración del nivel educativo en la evaluación de la creatividad, específicamente en el Test CREA resulta indispensable considerando las notorias diferencias observadas entre grupos de diferente nivel de escolaridad. El análisis de los datos permite observar claramente que los puntajes ascienden conforme aumenta el nivel de escolaridad alcanzada por los sujetos. En las interpretaciones de las puntuaciones individuales se sugiere considerar el nivel de escolaridad de las personas, los datos normativos propuestos permiten interpretaciones mejor adaptadas a las condiciones particulares de las personas que resuelven el CREA.

Tabla 10

Media, desviación típica y prueba *t* en las tres láminas del CREA según género para adultos.

Género		A	B	C
Mujeres	M	11.55	11.16	10.39
	DE	5.07	4.96	4.98
Varones	M	11.17	9.99	9.61
	DE	5.16	4.74	5.01
<i>p</i>		$t_{(1,1402)} = 1.35$.177	$t_{(1,2004)} = 5.22$ <.001	$t_{(1,1103)} = 2.60$.009

Tabla 11

Media, desviación típica y ANOVA en las tres láminas del CREA según edad de los adultos.

Edad		A	B	C
18 - 23 años (n = 1096)	M	12.20	11.59	11.02
	DE	5.23	4.96	4.98
23 - 40 años (n = 627)	M	11.42	10.70	9.79
	DE	4.85	4.92	4.88
41 - 55 años (n = 308)	M	9.69	9.00	7.84
	DE	4.65	4.09	4.41
56 años o más (n = 116)	M	8.41	8.00	6.57
	DE	4.89	4.18	4.07
F		$F_{(3, 1432)} = 23.02^{**}$	$F_{(3, 2032)} = 36.47^{**}$	$F_{(3, 1147)} = 25.14^{**}$

Nota. ** $p < .01$

Se realizaron contrastes *post hoc* y se observaron diferencias significativas en las tres láminas entre todos los grupos etarios, excepto entre 41-55 y más de 56 años.

Tabla 12

Media desviación típica y ANOVA en las tres láminas del CREA según nivel de escolaridad de los adultos.

Escolaridad		A	B	C
Primario	M	7.68	7.85	6.81
	DE	4.29	3.86	4.21
Secundario	M	9.77	9.11	8.43
	DE	4.46	4.12	4.44
Superior	M	12.44	11.90	10.91
	DE	5.07	5.02	4.98
		$F_{(2, 1435)} = 78.41^{**}$	$F_{(2, 2035)} = 108.26^{**}$	$F_{(2, 1150)} = 47.27^{**}$

Nota. ** $p < .01$

Se realizaron contrastes *post hoc* en las tres láminas y entre los tres grupos se observaron diferencias significativas en los puntajes medios considerando el nivel de *escolaridad* alcanzado.

Considerando las diferencias observadas en el grupo de adultos, es necesario construir baremos diferentes según edad y escolaridad. Se presentan a continuación datos normativos para adultos de diferentes grupos etarios y niveles de escolaridad alcanzados (ver Tablas 13 y 14).

Tabla 13

Percentiles para adultos (varones y mujeres) según diferentes rangos de edades para las láminas A, B y C del CREA.

	18 a 23			24-40			41-55			Más de 55		
	A	B	C	A	B	C	A	B	C	A	B	C
95	22	21	20	20	19	19	19	17	17	17	16	15
90	19	18	18	18	17	17	16	14	13	13	13	12
75	15	14	14	15	13	13	13	11	11	11	10	9
50	11	11	10	11	10	9	9	8	7	8	7	7
25	8	8	7	8	7	6	6	6	5	6	5	3
10	6	6	6	6	6	4	4	4	3	1	4	1
5	5	5	5	4	4	3	3	3	1	-	2	-
N	707	987	592	453	627	385	202	308	129	76	116	47
M	12.20	11.58	11.02	11.41	10.70	9.79	9.68	9.00	7.84	8.40	8.00	6.57
DE	5.19	4.96	4.98	4.85	4.92	4.88	4.65	4.09	4.40	4.89	4.18	4.07

Tabla 14

Percentiles para adultos (varones y mujeres) según los diferentes niveles de escolaridad para las láminas A, B y C del CREA.

Percentil	Nivel primario			Nivel secundario			Nivel superior		
	A	B	C	A	B	C	A	B	C
95	16	15	15	18	17	17	22	21	20
90	13	13	12	16	14	14	19	19	18
75	10	10	9	12	11	11	15	15	14
50	7	7	7	9	9	8	12	11	10
25	5	6	4	7	6	5	9	8	7
10	3	4	1	5	5	3	7	6	5
5	-	2	-	4	4	1	5	5	4
N	127	171	93	333	615	232	978	1252	828
M	7.67	7.85	6.80	9.77	9.11	8.42	12.43	11.89	10.90
DE	4.29	3.86	4.21	4.46	4.12	4.44	5.04	5.01	4.98

Con respecto a la incidencia de la variable *género*, se observaron algunas diferencias en los grupos de adolescentes y adultos. En el último grupo, cuando se analizó la incidencia del género en interrelación con la edad y la escolaridad no se observaron diferencias significativas. En estudios anteriores, se observaron resultados similares, es decir, inexistencia de diferencia según género en adultos (Chacón-Araya & Moncada-Jiménez, 2006; Clapham & King, 2010; Limiñana-Gras et al., 2010). Las investigaciones sobre creatividad y género muestran resultados diferentes según los grupos, los instrumentos que se utilizan y los contextos de análisis. Baer y Kaufman (2008) analizaron diferentes estudios sobre creatividad y género y concluyeron que los resultados de las investigaciones psicométricas referidas a diferencias de género son inconsistentes.

Los resultados hallados abren interesantes líneas para investigaciones futuras. Durante el proceso de recolección de datos, se han observado algunas dificultades en niños menores de 8 años para comprender la consigna del CREA. Es importante analizar, en futuros estudios, los procesos metacognitivos en los cuales se basa la

tarea propuesta. Asimismo, se sugiere en futuras investigaciones analizar con detalle el proceso de administración del test CREA en niños registrando expresiones, comportamientos y dificultades durante la resolución de la tarea. En el grupo de los adultos, la edad también es una variable importante a considerar, especialmente en personas mayores con ciertos déficits motrices, visuales y auditivos que dificultan la realización de las tareas propuestas por el CREA, tal como se ha observado en el proceso de recolección de datos. Se propone para futuros estudios ajustar las pautas de administración e interpretación considerando la situación de los participantes y sus dificultades para resolver las tareas. Desarrollar diferentes estrategias de validación según el proceso de respuesta (entrevistas, observaciones participantes, grupos focales, protocolos verbales) es una promisoriosa línea de estudio del Test CREA.

Las experiencias escolares de las personas inciden en las posibilidades de resolver la tarea verbal y de escritura que propone el CREA, se recomienda en investigaciones posteriores considerar otras técnicas no verbales u otras estrategias para la evaluación integral de la creatividad

en grupos con niveles bajos de escolaridad. En el caso de los adolescentes, las diferencias observadas según *género* abren interesantes líneas de investigaciones sobre creatividad en esta compleja etapa evolutiva. La realización de estudios mixtos con metodologías diversas puede ser de relevancia en la evaluación de la creatividad durante la adolescencia.

La conformación de la muestra y de los grupos según las variables consideradas es una limitante del estudio. En algunos casos, los datos normativos corresponden a grupos pequeños, como por ejemplo los de niños de 6 a 8 años en las Láminas A y B o los de adultos mayores de 55 años en la Lámina C. En futuras investigaciones es relevante ampliar el número de casos para estos grupos y construir datos normativos basados en muestras más numerosas y heterogéneas. La conformación de la muestra de manera no probabilística también representa una limitación del estudio y de los datos normativos presentados. Es necesario además, construir muestras representativas y ampliar los alcances de las mismas a otras regiones geográficas. Asimismo, en futuros estudios resulta de relevancia considerar variables como nivel socio-económico, contexto cultural, experiencias de ocio, etc. Sin embargo, se destaca como aporte la presentación de datos normativos que pueden contribuir en las interpretaciones de puntuaciones individuales y en el diagnóstico diferencial.

Referencias

- Antoñanzas-Laborda, J. L., Toner-Pelay, M., Salavera-Bordás, C., Soler-Costa, R., & Usan-Supervia, P. (2015). Creatividad y aprendizaje en niños de 4 y 5 años. *Revista de Psicología y Educación*, 10(1), 139-152. Recuperado de <http://www.revistadepsicologiayeducacion.es>
- Aranguren, M. (2014). Validez de constructo del Test de Pensamiento Creativo de Torrance en una muestra de jóvenes argentinos. *Anuario de Psicología/The UB Journal of Psychology*, 44(1), 55-70. Recuperado de <https://www.raco.cat/index.php/AnuarioPsicologia>
- Aranguren, M., & Irrazabal, N. (2012). Diseño de una escala para la evaluación del comportamiento creativo en diferentes dominios. *Ciencias Psicológicas*, VI(1), 29-41. Recuperado de <http://revistas.ucu.edu.uy/index.php/cienciaspsicologicas/index>
- Baer, J. (2012). Domain specificity and the limits of creativity theory. *Journal of Creative Behavior*, 46(1), 16-29. doi: 10.1002/jocb.002
- Baer, J., & Kaufman, J. (2008). Gender differences in creativity. *Journal of Creative Behavior*, 42(2), 75-106. doi: 10.1002/j.2162-6057.2008.tb01289.x
- Barbot, B., Besançon, M., & Lubart, T. (2016). The generality-specificity of creativity: Exploring the structure of creative potential with EPoC. *Learning and Individual Differences*, 52, 178-187. doi: 10.1016/j.lindif.2016.06.005
- Batey, M. (2007). *A psychometric investigation of everyday creativity* (Tesis doctoral inédita). University College, Londres.
- Batey, M. (2012). The measurement of creativity: From definitional consensus to the introduction of a new heuristic framework. *Creativity Research Journal*, 24(1), 55-65. doi: 10.1080/10400419.2012.649181
- Bogaert-García, H. (2017). La creatividad en los estudiantes universitarios. Una investigación psicométrica en la República Dominicana. *Ciencia y Sociedad*, 42(3), 51-68. doi: 10.22206/cys.2017.v42i3.pp51-68
- Cárdenas-Avila, N., López-Fernández, V., & Arias-Castro, C. (2018). Análisis de la relación entre creatividad, atención y rendimiento escolar en niños y niñas de más de 9 años en Colombia. *Psicogente*, 21(39), 75-87. doi: 10.17081/psico.21.39.2823
- Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. *Creativity*

- ty *Research Journal*, 17(1), 37-50. doi: [10.1207/s15326934crj1701_4](https://doi.org/10.1207/s15326934crj1701_4)
- Castañeda-Rey, E. L., López, V., & Ezquerro-Cordón, A. (2017). Creatividad, atención, rendimiento académico e interacción grupal en un aula de secundaria. *Ulu: Revista científica sobre la imaginación*, 2, 1-14. Recuperado de <http://digibug.ugr.es/handle/10481/39268>
- Chacón-Araya, Y., & Moncada-Jiménez, J. (2006). Relación entre personalidad y creatividad en estudiantes de educación física. *Actualidades Investigativas en Educación*, 6(1), 1-19. doi: [10.15517/aie.v6i1.9200](https://doi.org/10.15517/aie.v6i1.9200)
- Chiecher, A. C., Elisondo, R. C., Paoloni, P. V., & Donolo, D. S. (2018). Creatividad, género y rendimiento académico en ingresantes de ingeniería. *Revista Iberoamericana de Educación Superior*, 9(24), 138-151. Recuperado de <https://www.ries.universia.unam.mx>
- Clapham, M. M., & King, W. R. (2010). Psychometric characteristics of the CREA in an english speaking population. *Anales de Psicología*, 26(2), 206-211. Recuperado de <http://revistas.um.es/analesps>
- Contini, E. N. (2013). Valorización y desvalorización de los tests en el campo de la psicología argentina. Una disociación no superada. *PSIENCIA. Revista Latinoamericana de Ciencia Psicológica*, 5(2), 119- 122. Recuperado de <http://www.psiencia.org>
- Corbalán-Berná, J. (2008). ¿De qué se habla cuando hablamos de creatividad? *Cuadernos FHyCSUNJu*, 35, 11-21. Recuperado de <http://www.scielo.org.ar>
- Corbalán-Berná, J., Martín-Brufau, R., Donolo, D., Clapham, M., Limiñana, R., García-Peñas, V., ... King, R. (2014). CREA. A cross-cultural study. *Personality and Individual Differences*, 60, S54. doi: [10.1016/j.paid.2013.07.223](https://doi.org/10.1016/j.paid.2013.07.223)
- Corbalán-Berná, J., Martínez Zaragoza, F., Donolo, D. S., Alonso-Monreal, C., Tejerina-Arreal, M., & Limiñana-Gras, M. (2003). CREA. Inteligencia Creativa. *Una medida cognitiva de la creatividad*. Madrid, España: TEA Ediciones.
- Donolo, D. S., & Elisondo, R.C. (2007). Creativity for all. Consideration about a particular group. *Annals of Psychology*, 23(1), 148-151. Recuperado de <http://revistas.um.es/analesps>
- Elisondo, R. C. (2015). *Evaluación de la creatividad. Análisis de variables alternativas relacionadas con la forma y el contenido de las respuestas en el CREA* (Tesis doctoral). Universidad de Murcia, España. Recuperado de <https://www.tdx.cat>
- Elisondo, R. C., & Donolo, D. S. (2010). Creatividad o inteligencia. That is not the question. *Anales de Psicología*, 26(2), 220-225. Recuperado de <http://revistas.um.es/analesps>
- Elisondo, R. C., & Donolo, D. S. (2011). Los estímulos en un test de Creatividad. Incidencias según género, edad y escolaridad. *Boletín de Psicología*, 101, 51-65. Recuperado de <https://www.uv.es/seoane/boletin/boletin.html>
- Elisondo, R. C., & Donolo, D. S. (2014). Interculturalidad, apertura a experiencias y creatividad Aportes para una educación alternativa. *Red, Revista de Educación a Distancia*, 41, 1-19. Recuperado de <https://www.um.es/ead/red>
- Elisondo, R. C., & Donolo, D. S. (2016a). Construcción y análisis de las propiedades psicométricas del Cuestionario de Acciones Creativas en población argentina. *PSIENCIA, Revista Latinoamericana de Ciencia Psicológica*, 8(1), 1-21. Recuperado de <http://www.psiencia.org>
- Elisondo, R. C., & Donolo, D. S. (2016b). Creativity, involvement, and leisure time: The more, the better. *Business Creativity and the Creative Economy*, 2(1), 102-109. Recuperado de <http://www.icscpress.com/journals/bcce>
- Elisondo, R. C., Donolo, D. S., & Corbalán-Berná, F. J. (2009). Evaluación de la Creatividad ¿Relaciones con inteligencia y personalidad? *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 28(2), 67-79. Recuperado de <http://www.aidep.org/numeros-publicados>
- Elisondo, R. C, Donolo, D. S., & Limiñana-Gras, R. M.

- (2018). The measure of originality in CREA test responses. *Anales de Psicología/Annals of Psychology*, 34(1), 197-210. doi: [10.6018/analesps.34.1.286131](https://doi.org/10.6018/analesps.34.1.286131)
- Freiberg-Hoffmann, A., & Fernández-Liporace, M. (2017). Graphic Creativity Assessment: Psychometric properties in college students from Buenos Aires. *Psychological Thought*, 10(1), 60-79. doi: [10.5964/psyc.v10i1.192](https://doi.org/10.5964/psyc.v10i1.192)
- Guilford, J. P. (1962). Creativity: its measurement and development. En S. Parnes & H. Harding (Eds.), *A source book for creative thinking*. New York: Scribner's.
- Hang-Wu, C., Cheng, Y., Man-Ip, H., & McBride-Chang, C. (2005). Age differences in creativity: Task structure and knowledge base. *Creativity Research Journal*, 17(4), 321-326. doi: [10.1207/s15326934crj1704_3](https://doi.org/10.1207/s15326934crj1704_3)
- Hocevar, D. (abril, 1979). *The development of the Creative Behavior Inventory (CBI)*. Trabajo presentado en el encuentro anual Rocky Mountain Psychological Association. Recuperado de <https://eric.ed.gov/?id=ED170350>
- Humble, S., Dixon, P., & Mpofu, E. (2017). Factor structure of the Torrance Tests of Creative Thinking Figural Form A in Kiswahili speaking children: Multidimensionality and influences on creative behavior. *Thinking Skills and Creativity*, 27, 33-44. doi: [10.1016/j.tsc.2017.11.005](https://doi.org/10.1016/j.tsc.2017.11.005)
- Krumm, G., Aranguren, M., Arán-Filippetti, V., & Lemos, V. (2016). Factor structure of the Torrance Tests of Creative Thinking Verbal Form B in a Spanish speaking population. *Journal of Creative Behavior*, 50(2), 150-164. doi: [10.1002/jocb.76](https://doi.org/10.1002/jocb.76)
- Limiñana-Gras, R. M., Bordoy, M., Juste-Ballesta, G., & Corbalán-Berná, J. (2010). Creativity, intellectual abilities and response styles: Implications for academic performance in the secondary school. *Anales de Psicología*, 26(2), 212-219. Recuperado de <http://revistas.um.es/analesps>
- Long, H. (2014). An empirical review of research methodologies and methods in creativity studies (2003-2012). *Creativity Research Journal*, 26(4), 427-438. doi: [10.1080/10400419.2014.961781](https://doi.org/10.1080/10400419.2014.961781)
- López-Martínez, O., & Brufau, R. (2010). Creative intelligence and thinking styles. *Annals of Psychology*, 26(2), 254-258.
- López-Martínez, O., & Navarro-Lozano, J. (2008). Estudio comparativo entre medidas de creatividad: TTCT vs. CREA. *Anales de Psicología*, 24(1), 138-142. Recuperado de <http://revistas.um.es/analesps>
- Martínez-Zaragoza, F. A. (2003). Características psicométricas del CREA (inteligencia creativa). Un estudio con población española y argentina. *RIDEP*, 16(2), 71-83. Recuperado de <http://www.aidep.org>
- Matud, M., Rodríguez, C., & J. Grande (2007). Gender differences in creative thinking. *Personality and Individual Differences*, 43(5), 1137-1147. doi: [10.1016/j.paid.2007.03.006](https://doi.org/10.1016/j.paid.2007.03.006)
- Primi, R., Nakano, T. de C., Morais, M. de F., Silva-Almeida, L., & Marques-David, A. P. (2013). Factorial structure analysis of the Torrance Test with Portuguese students. *Estudos de Psicologia (Campinas)*, 30(1), 19-28. doi: [10.1590/s0103-166x2013000100003](https://doi.org/10.1590/s0103-166x2013000100003)
- Ramos-Moreno, A. M., López-Fernández, V., & Llamas-Salguero, F. (2017). Relación entre la creatividad, la memoria inmediata y lógica en relación con el rendimiento académico en la educación secundaria. *Revista Academia y Virtualidad*, 10(1), 123-130. doi: [10.18359/ravi.2674](https://doi.org/10.18359/ravi.2674)
- Richards, R. E. (2007). *Everyday creativity and new views of human nature: Psychological, social, and spiritual perspectives*. Washington, DC: American Psychological Association. doi: [10.1037/11595-000](https://doi.org/10.1037/11595-000)
- Rodríguez-Suárez, P. A., Llamas-Salguero, F., & López-Fernández, V. L. (2015). Relación entre la creatividad e inteligencia emocional en alumnado que cursa prácticas de Ingeniería en Colombia. *ReiDoCrea: Revista Electrónica de Investigación y Docencia Creativa*, 4, 192-199. Recuperado de <http://digibug.ugr.es/handle/10481/21881>
- Said-Metwaly, S., Van den Noortgate, W., & Kyndt, E. (2017). Methodological issues in measuring creativity: A systematic literature review. *Creativity. Theo-*

ries - Research - Applications, 4(2), 276-301. doi: [10.1515/ctra-2017-0014](https://doi.org/10.1515/ctra-2017-0014)

Sánchez-Hernández, O., Méndez, F. X., & Garber, J. (2015). Producción divergente explicativa: La relación entre resiliencia y creatividad. *Electronic Journal of Research in Educational Psychology*, 13(37), 551-568. doi: [10.14204/ejrep.37.14126](https://doi.org/10.14204/ejrep.37.14126)

Torrance, E. P. (1990). *The Torrance Tests of Creative Thinking: Norms Technical manual figural (streamlined) forms A & B*. Bensenville, IL: Scholastic Testing Service Inc.

Wechsler, S. (2006). Validity of the Torrance Tests of Creative Thinking to the Brazilian culture. *Creativity Research Journal*, 18(1), 15-25. doi: [10.1207/s15326934crj1801_3](https://doi.org/10.1207/s15326934crj1801_3)

Yarbrough, N. D. (2016). Assessment of creative thinking across cultures using the Torrance Tests of Creative Thinking (TTCT): Translation and validity issues. *Creativity Research Journal*, 28(2), 154-164. doi: [10.1080/10400419.2016.1162571](https://doi.org/10.1080/10400419.2016.1162571)

Yoon, C. H. (2017). A validation study of the Torrance Tests of Creative Thinking with a sample of Korean elementary school students. *Thinking Skills and Creativity*, 26, 38-50. doi: [10.1016/j.tsc.2017.05.004](https://doi.org/10.1016/j.tsc.2017.05.004)

Propiedades psicométricas de la Escala de Florecimiento en Puerto Rico

Psychometric Properties of the Flourishing Scale in Puerto Rico

Juan Aníbal González-Rivera * ¹

1 - Ponce Health Sciences University, San Juan University Center, Puerto Rico

Introducción
Método
Resultados
Discusión
Referencias

Recibido: 01/03/2018 **Revisado:** 20/06/2018 **Aceptado:** 07/07/2018

Resumen

En el presente estudio se analizan las propiedades psicométricas de la Escala de Florecimiento en Puerto Rico. Se realizaron tres estudios. En el primero, con 665 participantes, se analizaron la confiabilidad y la estructura factorial de la escala. En el segundo, con 821 participantes, se efectuó un análisis factorial confirmatorio. En el tercero, con 342 mujeres puertorriqueñas, se calculó la validez convergente y divergente de las escalas con otras medidas de bienestar y depresión. Los resultados de los estudios mostraron propiedades psicométricas adecuadas en la escala. El índice de confiabilidad Alfa de Cronbach de la escala fue de .92. La varianza explicada en el análisis factorial exploratorio fue de un 65.35%. Los resultados confirmaron que la escala posee una estructura unidimensional. Los ocho ítems de la escala cumplieron con los criterios de discriminación. Se realizaron análisis de validez convergente y discriminante mediante la evaluación de correlación entre el florecimiento, la satisfacción con la vida y la depresión. Se obtuvieron valores estadísticamente significativos. Estos resultados sugieren que la Escala de Florecimiento tiene el potencial para medir este constructo en adultos puertorriqueños.

Palabras clave: *bienestar, bienestar psicológico, florecimiento, propiedades psicométricas, validación*

Abstract

This article examines the psychometric properties of the Flourishing Scale in Puerto Rico. Three studies were carried out with this purpose. The first study, of 665 participants, analyzed the reliability and factorial structure of the scales. The second study, of 821 participants, carried out a confirmatory factorial analysis. The third study, of 342 Puerto Rican women, calculated the convergent and divergent validity of the scales with other measures of wellbeing and depression. The results of the studies showed adequate psychometric properties in the scale. The reliability index of the scale was of .95. The variance in the exploratory factorial analysis was 65.35%. The results confirmed that the scale has a one-dimensional structure. The eight items complied with the criteria of discrimination. Convergent and discriminant validity analyses were performed by evaluating correlation between flourishing, satisfaction with life and depression, obtaining statistically significant values. These results suggest that the Flourishing Scale has the potential to measure this construct among Puerto Rican adults.

Key words: *flourishing, psychological well-being, psychometric properties, validation, well-being*

* **Correspondencia a:** Dr. Juan A. González-Rivera, 500 West Main Suite 215, Bayamón, Puerto Rico, 00961. Tel.: 011 787 315 6034.
dr.juananibalgonzalez@outlook.com

Cómo citar este artículo: González-Rivera, J. A. (2018). Propiedades psicométricas de la Escala de Florecimiento en Puerto Rico. *Revista Evaluar*, 18(3), 30-43. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

Históricamente, la búsqueda de la felicidad y el bienestar humano han sido temas centrales en la literatura, la filosofía y la teología (González-Rivera, Quintero-Giménez, Veray-Alicea, & Rosario-Rodríguez, 2016). Por ejemplo, en la antigua Grecia, Aristóteles (2001) ya proponía el concepto de la suma felicidad como el fin último del ser humano. En Oriente, los discípulos de Confucio describieron la buena vida (una vida feliz) desde una sociedad ordenada en donde cada ciudadano tiene roles, deberes y responsabilidades claramente delimitados (Diener & Suh, 2000). Posturas filosóficas como estas inquietaron académicamente a los principales exponentes de la psicología positiva, quienes formalizaron científica y rigurosamente el estudio de la felicidad y el bienestar general en las ciencias sociales (Diener, 2000; Ryff, 1989; Seligman & Csikszentmihalyi, 2000; Seligman & Peterson, 2003). La intención principal de estos autores era entender los procesos psicológicos que subyacen en la búsqueda de la felicidad y que fomentan un mayor bienestar en las personas.

Con este propósito, los psicólogos positivos desarrollaron distintos constructos para operacionalizar la felicidad y el bienestar personal y facilitaron su estudio en los campos de la psicología y las ciencias sociales (p. ej. satisfacción con la vida, afecto positivo, afecto negativo, crecimiento y desarrollo personal, entre otros). Estos constructos suelen categorizarse en dos perspectivas distintas de investigación: la hedónica y la eudamónica. La perspectiva hedónica, caracterizada por el bienestar subjetivo, concibe el bienestar del individuo como la evaluación que las personas hacen sobre la satisfacción que experimentan en sus vidas y el balance entre los afectos posi-

vos y negativos (Diener, 2000; Keyes, Shmotkin, & Ryff, 2002).

Para medir el bienestar desde la perspectiva hedónica, se deben tener en consideración dos aspectos principales: el componente cognitivo –asociado al juicio valorativo-subjetivo que la persona hace sobre su vida– y el componente afectivo-emocional (Diener, 2000; Diener & Suh, 2000). Tomando como base estos elementos, se han desarrollado varios instrumentos para evaluar el bienestar desde la perspectiva hedónica: el Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985), el Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988) y el Scale of Positive and Negative Experience (Diener et al., 2010).

Por su parte, la perspectiva eudamónica, representada por el bienestar psicológico, asocia el bienestar del individuo con el desarrollo del potencial humano. Los investigadores fieles a esta perspectiva de investigación explican que el bienestar humano está intrínsecamente ligado a la realización de actividades congruentes con los valores profundos del individuo que fomentan su desarrollo y crecimiento personal (Ryan & Deci, 2001). En síntesis, esta perspectiva concibe la felicidad como un estado de plenitud y armonía en el que el bienestar humano es entendido como la percepción del sentido y propósito en la vida, siendo el crecimiento y desarrollo personal, los principales factores del funcionamiento óptimo del ser humano (Chitgian-Urzúa, Urzúa, & Vera-Villaruel, 2013; González-Rivera et al., 2016).

Se han desarrollado varios instrumentos para medir el bienestar desde la perspectiva eudamónica; entre los más conocidos se encuentran: el Questionnaire for Eudaimonic Well-Being (Waterman et al., 2010) y el Scales of Psychological Well-Being (Ryff, 1995). Uno de los últimos instrumentos desarrollados en esta línea de investigación, que destaca por su brevedad y facilidad

de administración, es la Escala de Florecimiento (Flourishing Scale) de Diener et al. (2010). La Figura 1 presenta gráficamente el desarrollo teórico de los principales instrumentos que examinan el bienestar humano desde las perspectivas hedónica y eudaemónica.

La Escala de Florecimiento fue desarrollada por Diener et al. (2009; 2010) para medir la prosperidad psicosocial, tomando como fundamento conceptual la integración de las principales teorías psicológicas que tratan el florecimiento humano y la prosperidad personal. En primer lugar, en la elaboración de la escala utilizaron las necesidades psicológicas universales conceptua-

lizadas por Ryff (1989), estas son: necesidad de competencia, relaciones humanas (interpersonales) y autoaceptación. Asimismo, integraron otros elementos teóricos que, según los autores, fortalecen la perspectiva de las necesidades universales: la noción de capital social (Helliwell, Barrington-Leigh, Harris, & Huang, 2010), el flow y el compromiso (Csikszentmihalyi, 1990) y, por último, el optimismo, el propósito y el significado (Seligman, 2002). Todos estos elementos psicológicos están contemplados en la Escala de Florecimiento, dado que son beneficiosos para el funcionamiento óptimo del ser humano.

En cuanto a las propiedades psicométricas,

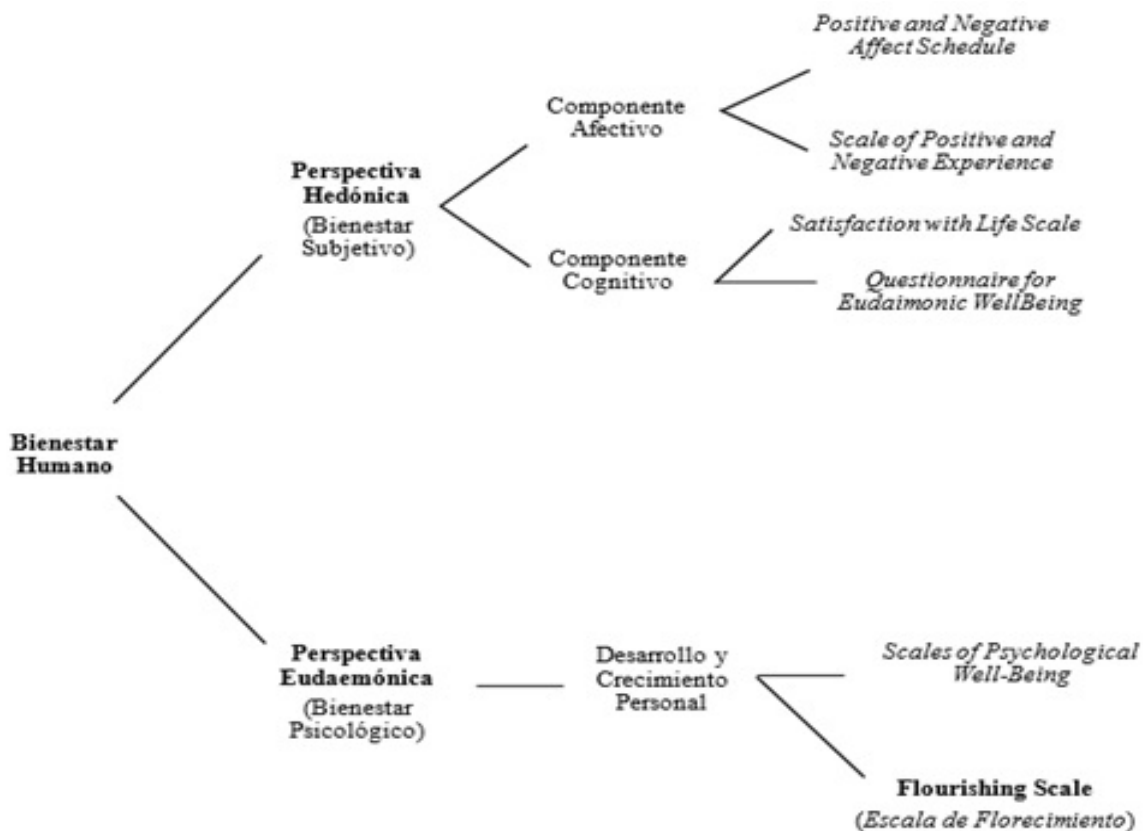


Figura 1

Desarrollo teórico de los principales instrumentos que examinan el bienestar humano desde las perspectivas hedónica y eudaemónica.

el instrumento ha demostrado una fuerte consistencia interna, con coeficientes alfa de Cronbach que oscilan entre .87 a .95 en diferentes países y culturas (Cassaretto-Bardales & Martínez-Uribe, 2017; Diener et al., 2010; Esch, Jose, Gimpel, Von Scheidt, & Michalsen, 2013; Hone, Jarden, & Schofield, 2014; Howell & Buro, 2015; Silva & Caetano, 2013; Sumi, 2013). Por su parte, cuando se validó la Escala de Florecimiento por primera vez en Estados Unidos, se evidenciaron índices de discriminación de los ítems que fluctuaron entre .57 y .71, y la escala reflejó una solución unifactorial que explicaba el 53% de la varianza de la escala (Diener et al., 2010). Estudios posteriores han confirmado la unidimensionalidad de la escala (Cassaretto-Bardales & Martínez-Uribe, 2017; Hone et al., 2014; Howell & Buro, 2015; Silva & Caetano, 2013).

Si bien la mayoría de los estudios que utilizan la Escala de Florecimiento se han llevado a cabo en Estados Unidos, se han registrado traducciones y validaciones en diferentes contextos internacionales, como Japón (Sumi, 2013), Portugal (Silva & Caetano, 2013), Canadá (Howell & Buro, 2015), Nueva Zelanda (Hone et al., 2014), Alemania (Esch et al., 2013) y Perú (Cassaretto-Bardales & Martínez-Uribe, 2017). Finalmente, cabe señalar que solo existe un estudio que examina las propiedades psicométricas del instrumento en el contexto latinoamericano (Perú), lo que hace necesario incrementar los estudios de carácter psicométricos que examinen la Escala de Florecimiento en otros países hispanohablantes como Puerto Rico.

El presente estudio se propuso verificar las propiedades psicométricas de la Escala de Florecimiento en la población puertorriqueña y comprobar si es un instrumento adecuado para examinar el bienestar psicológico en adultos puertorriqueños. En este sentido, nuestro estudio pretende validar y analizar las propiedades psicométricas

de la versión en español de la Escala de Florecimiento en adultos puertorriqueños. Para ello, se llevaron a cabo tres estudios independientes. En el Estudio 1 se analiza la estructura factorial de la escala mediante un análisis factorial exploratorio, se evalúa la confiabilidad del instrumento y se examina la capacidad de discriminación de los ítems. En el Estudio 2 se analiza la validez de constructo mediante un análisis factorial confirmatorio con ecuaciones estructurales y se evalúa la confiabilidad compuesta. En el Estudio 3 se evalúa la validez convergente y divergente de la escala con respecto a otras medidas de bienestar y una escala de depresión. En los tres estudios se utiliza un diseño de investigación instrumental (Montero & León, 2007).

Estudio 1

El objetivo de este estudio fue analizar la estructura factorial de la escala mediante un AFE de componentes principales, analizar los índices de discriminación de los ocho ítems y evaluar la confiabilidad de la escala mediante dos métodos: coeficiente alfa de Cronbach y división en mitades de Spearman-Brown.

Método

Participantes. Se trabajó con una muestra no probabilística constituida por 665 participantes: 473 mujeres (71.1%) y 192 hombres (28.9%). El rango de edad fue de entre los 21 y 65 años, con una edad promedio de 35.86 (DE = 10.22). En cuanto al estado civil, el 26.2% (n = 174) era soltero, el 44.8% (n = 298) era casado y el 29% n = 193 convivía con su pareja sin nupcias. En términos de preparación académica, el 40% (n = 266) poseía un bachillerato, el 24.5% (n = 163) poseía

un grado de maestría, el 18.8% (n = 125) tenía un grado técnico o grado asociado, el 6% (n = 40) completó únicamente la escuela superior, y el 10.7% (n = 71) contaba con un doctorado. Por último, la mayoría de la muestra (el 51.1%, n = 340) generaba ingresos menores que \$25 000, mientras que el 32.6% (n = 217) entre \$26 000 y \$50 000.

Instrumentos. La Escala de Florecimiento (Diener et al., 2010) está constituida por ocho ítems que examinan el bienestar psicológico desde una perspectiva eudaemónica. Cada ítem posee una escala de respuesta de siete puntos, cuyos límites lo constituyen las respuestas *Muy en desacuerdo* (1) y *Muy de acuerdo* (7). El rango posible es de 8 a 56 puntos donde a mayor puntuación, mayor grado de crecimiento o desarrollo psicológico.

Procedimientos generales. La recopilación de datos se llevó a cabo mediante el uso de cuestionarios a través de la plataforma PsychData utilizando como método de reclutamiento las redes sociales: Facebook, Twitter, Google+, WhatsApp, entre otras. Esta investigación fue aprobada por el Comité para la Ética en la Investigación (IRB) de

la Universidad Carlos Albizu en San Juan, Puerto Rico. Los datos fueron analizados en el programa estadístico IBM SPSS versión 24.0 (IBM Corporation, 2016).

Resultados

Análisis de componentes principales. En el primer estudio se realizó un análisis AFE utilizando el método de extracción de componentes principales con rotación oblicua para identificar las variables latentes que subyacen en los ítems. Los resultados mostraron una estructura unifactorial con un valor característico de 5.23, que explica el 65.35% de la varianza de los datos (KMO = .92; $\chi^2 = 3511.974_{(28)}, p < .001$). Las cargas factoriales de los ítems sobrepasan el .72, superando el mínimo de .50 recomendado por Stevens (2002). En la Tabla 1 se presentan las cargas factoriales obtenidas por los 8 ítems.

Análisis de Discriminación. Una vez que se analizó la estructura factorial de la escala, se llevó a cabo un análisis de reactivos para calcular los índices de discriminación de los ocho ítems de

Tabla 1

Análisis factorial exploratorio e índices de discriminación de los ítems.

Ítems	r_{bis}	1
1. Llevo una vida significativa y con propósito.	.78	.84
2. Mis relaciones sociales me apoyan y son reconfortantes.	.69	.76
3. Me intereso y me involucro en mis actividades diarias.	.79	.85
4. Contribuyo activamente a la felicidad y al bienestar de otros.	.74	.81
5. Soy competente y capaz en las actividades que son importantes para mí.	.74	.82
6. Soy una buena persona y vivo una buena vida.	.79	.86
7. Soy optimista acerca de mi futuro.	.75	.81
8. La gente me respeta.	.64	.72

Nota. r_{bis} = índice de discriminación del ítem; 1 = cargas factoriales (n = 665).

la Escala de Florecimiento, mediante el método de correlación ítem-total de la prueba (r_{bis}). Los resultados mostraron índices de discriminación que fluctuaban entre .64 y .79, de manera que todos los reactivos obtuvieron índices de discriminación mayores a .30, como recomienda Kline (2005). La Tabla 1 presenta los índices de discriminación de los ocho ítems de la escala.

Análisis de Confiabilidad. Los ocho ítems fueron sometidos a un análisis de consistencia interna para determinar el índice de confiabilidad de la escala. Para ello utilizamos dos métodos: el coeficiente Alfa de Cronbach y división en mitades de Spearman-Brown. Los resultados mostraron un coeficiente alfa de .92 y de .92 para la prueba Spearman-Brown. La Tabla 2 presenta el alfa de Cronbach, alfa de Cronbach estandarizado, el coeficiente de la prueba Spearman-Brown, la media y la desviación estándar de la Escala de Florecimiento.

Estudio 2

Los objetivos de este segundo estudio fueron analizar la validez de constructo mediante

un AFC con ecuaciones estructurales y evaluar la fiabilidad compuesta del instrumento.

Método

Participantes. Para este estudio se reclutó una muestra no probabilística constituida por 821 participantes: 410 mujeres (49.9%) y 411 hombres (50.1%). El rango de edad fue de entre los 19 y 75 años, con una edad promedio de 37.24 (DE = 13.30). En cuanto al estado civil, el 50.7% (n = 417) era soltero, el 33.6% (n = 275) era casado y el 15.7% (n = 129) convivía con su pareja sin nupcias. En términos de preparación académica, el 49% (n = 402) poseía un bachillerato, el 17.4% (n = 143) poseía un grado de maestría, el 16.9% (n = 139) tenía un grado técnico o grado asociado, el 8.9% (n = 73) cursó únicamente la escuela superior, y el 7.8% (n = 64) contaba con un doctorado. Por último, la mayoría de la muestra (el 58.2%, n = 478) generaba ingresos menores que \$25 000, mientras que el 29.4% (n = 241) entre \$26 000 y \$50 000.

Instrumento. En el Estudio 2 solo se utilizó la Escala de Florecimiento (véase descripción en el

Tabla 2

Medias, desviaciones estándar y coeficientes de confiabilidad.

Factor	M	DE	α	α_{est}	Spearman-Brown
Escala de Florecimiento	47.68	8.89	.92	.92	.90

Nota. M = media; DE = desviación estándar; α = alfa de Cronbach; α_{est} = alfa de Cronbach estandarizado. (n = 665).

Estudio 1) con el propósito de auscultar sus características psicométricas.

Procedimientos generales. La recopilación de datos se llevó a cabo de forma electrónica y presencial durante el 2017. Se reclutaron personas de forma presencial ($n = 348$) en dos centros universitarios de la zona metropolitana de Puerto Rico y en varias oficinas de empleo. A su vez, se reclutaron participantes electrónicamente ($n = 473$) mediante un anuncio pagado en Facebook y otras redes sociales. Esta investigación fue aprobada por el Comité para la Ética en la Investigación IRB de la Universidad Carlos Albizu en San Juan, Puerto Rico. Los programas estadísticos utilizados en esta fase fueron IBM SPSS Statistics versión 24 y STATA versión 14.1 (StataCorp., 2017).

Resultados

Análisis descriptivos de los ítems. En primer lugar, se calcularon las medias y las desviaciones estándar para cada ítem de la escala para analizar las propiedades de distribución de la escala. Las medias de los ítems fluctuaron entre 6.03 y 6.57, con desviaciones estándar que fluctuaron entre .95 y 1.22. Como muestra, la Tabla 3 y las pruebas de normalidad de Kolmogorov-Smirnov y Shapiro-Wilk evidencian que los datos no se distribuyen de forma normal ($p < .001$). Por la falta de normalidad en los datos, utilizamos las correcciones de Satorra y Bentler (2001) para calcular el ajuste de los modelos de ecuaciones estructurales, ya que la no normalidad de los datos altera los errores de estimación y el ajuste global del modelo.

Validez de Constructo. Se realizó un AFC con

ecuaciones estructurales utilizando el método de estimación de máxima verosimilitud. Para evaluar el ajuste del modelo se utilizaron los siguientes índices de bondad de ajuste y los criterios de Byrne (2010): chi-cuadrado (χ^2), error medio cuadrático de aproximación (RMSEA; los valores sugeridos deben ser $\leq .08$ para indicar un buen ajuste), índice Tucker-Lewis (TLI) y el índice de ajuste comparativo (CFI). Para que exista un buen ajuste del modelo, los valores sugeridos de CFI y TLI deben ser de $\geq .95$ (Byrne, 2010). Los resultados arrojaron un buen ajuste para el modelo conformado por un solo factor latente que recogía los ocho ítems de la escala (véase Figura 2), $\chi^2 = 181.315_{(20)}$, $p < .001$, RMSEA = .08, CFI = .97, TLI = .96, χ^2 corregido = $81.755_{(20)}$, $p < .001$, RMSEA corregido = .05, CFI corregido = .97, TLI corregido = .96. Estos índices cumplen con lo que Satorra y Bentler (2001) consideran como niveles aceptables. A su vez, se examinaron los coeficientes de regresión de cada ítem, esperando que cada uno de estos fuese mayor a .65, según recomiendan Schumaker y Lomax (2010). Los coeficientes de regresión fluctuaron entre .68 y .85 (véase Tabla 4). Luego, se analizó la confiabilidad compuesta del modelo unidimensional de la escala. Los análisis revelaron un índice de confiabilidad compuesta de .93, lo que supera el valor mínimo recomendado por la literatura (Raykov & Shrout, 2002).

Diferencias entre hombres y mujeres. Se realizó una prueba t independiente para analizar si existen diferencias estadísticamente significativas entre las medias de florecimiento de las mujeres ($M = 49.91$, $DE = 7.98$) y de los hombres ($M = 50.81$, $DE = 6.11$). Los resultados confirmaron que la diferencia a favor de los hombres no es estadísticamente significativa ($t_{(819)} = -1.815$, $p > .05$).

Tabla 3

Estadísticos de descriptivos y de distribución para los ítems de la escala.

Ítem	Media	Desviación Estándar	Asimetría	Curtosis	Kolmogorov-Smirnov	Shapiro-Wilk
1	6.33	1.14	-2.59	7.94	.32	.62
2	6.03	1.19	-1.74	3.55	.28	.76
3	6.29	1.07	-2.45	7.62	.29	.66
4	6.34	1.05	-2.50	7.99	.32	.64
5	6.57	.95	-3.70	16.68	.29	.49
6	6.49	.97	-3.20	13.17	.36	.57
7	6.20	1.19	-2.01	4.40	.29	.70
8	6.11	1.22	-1.94	4.24	.27	.72

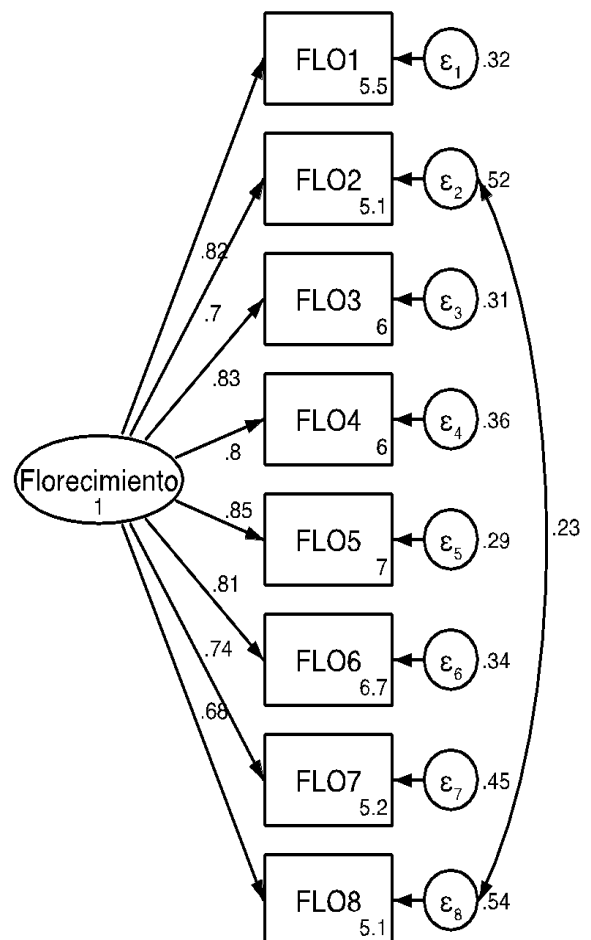
Nota. Error estándar de la asimetría = .09; Error estándar de la curtosis = .17. Grados de libertad Kolmogorov-Smirnov y Shapiro-Wilk = 821, todos los valores $p < .001$.

Estudio 3

Una vez confirmada la confiabilidad de la escala, su estructura factorial y que las puntuaciones de la escala no se alteran según el género, el objetivo de este tercer estudio fue determinar la validez convergente y divergente de las escalas con otras medidas de bienestar y una medida de sintomatología depresiva.

Método

Participantes. Este estudio contó con la participación de 342 mujeres adultas puertorriqueñas, seleccionadas por disponibilidad. El rango de edad fue de entre los 21 y 68 años, con una edad promedio de 37.66 (DE = 10.73). En términos de preparación académica, el 38.3% ($n = 131$) poseía un bachillerato, el 26.9% ($n = 92$) poseía un grado de maestría, el 20.7% ($n = 71$) tenía un grado técnico o grado asociado, el 5% ($n = 17$) completó únicamente la escuela superior, y el 9.1% ($n = 31$) contaba con un doctorado. Por último, la mayoría de la muestra (el 44.7%, $n = 153$) generaba ingre-

**Figura 2**

Modelo unidimensional de la Escala de Florecimiento.

Tabla 4

Estimados estandarizados de regresión y sus intervalos de confianza.

Ítems	β	p	I.C. 95%
1. Llevo una vida significativa y con propósito.	.82	<.001	[.77, .87]
2. Mis relaciones sociales me apoyan y son reconfortantes.	.70	<.001	[.63, .76]
3. Me intereso y me involucro en mis actividades diarias.	.83	<.001	[.79, .88]
4. Contribuyo activamente a la felicidad y al bienestar de otros.	.80	<.001	[.74, .86]
5. Soy competente y capaz en las actividades que son importantes para mí.	.85	<.001	[.80, .90]
6. Soy una buena persona y vivo una buena vida.	.81	<.001	[.75, .88]
7. Soy optimista acerca de mi futuro.	.74	<.001	[.69, .79]
8. La gente me respeta.	.68	<.001	[.60, .75]

Nota. β = estimados estandarizados de regresión; p = significación; I.C. 95% = intervalos de confianza de los coeficientes de regresión.

menos que \$25 000, mientras que el 35.7% ($n = 122$) entre \$26 000 y \$50 000.

Instrumentos. En el Estudio 3, además de utilizar la Escala de Florecimiento (véase descripción en el Estudio 1), se utilizaron la Escala de Satisfacción con la Vida (Diener et al., 1985), la Escala de Autoestima de Rosenberg (1965) y la Escala de Depresión del DASS-21 (Lovibond & Lovibond, 1995). La Escala de Satisfacción con la Vida está constituida por un total de cinco ítems (p. ej. *Estoy satisfecho/a con mi vida*) con una escala de respuesta de siete puntos que fluctúan de *Muy en desacuerdo* (1) a *Muy de acuerdo* (7). El puntaje más bajo que se puede obtener es 5 y el más alto es 35, donde puntuaciones elevadas sugieren alta presencia de satisfacción con la vida ($\alpha = .92$).

Por su parte, la Escala de Autoestima de Rosenberg está constituida por 10 ítems (p. ej. *En general, estoy satisfecha conmigo misma*) con una escala de respuesta de cuatro puntos que fluctúan de *Totalmente de acuerdo* (1) a *Totalmente en desacuerdo* (4). El rango posible es de 10 a 40 puntos, donde elevadas puntuaciones sugieren una alta valía personal y mucho respeto a sí misma ($\alpha = .90$). Por último, la Escala de Depresión

del DASS-21 está constituida por siete ítems (p. ej. *Me sentí triste y deprimido*) con una escala de respuesta de cuatro puntos que fluctúan de *Nada* (0) a *Mayor parte del tiempo* (3). El puntaje total se calcula con la suma de los ítems pertenecientes a esa escala y varía entre 0 y 21 puntos; a mayor puntuación, mayor sintomatología presenta la participante ($\alpha = .90$).

Procedimientos generales. La recopilación de datos se llevó a cabo mediante el uso de cuestionarios a través de la plataforma PsychData utilizando como método de reclutamiento las redes sociales: Facebook, Twitter, Google+, WhatsApp, entre otras. Esta investigación fue aprobada por el Comité para la Ética en la Investigación (IRB) de la Universidad Carlos Albizu en San Juan, Puerto Rico. Los datos fueron analizados en el programa estadístico IBM SPSS (versión 24).

Resultados

Evidencias de Validez Convergente y Divergente. Con el fin de evaluar la validez convergente y divergente de la Escala de Florecimiento, se

analizaron las correlaciones con otras medidas de bienestar (satisfacción con la vida y autoestima) y una medida de sintomatología asociada con la depresión. Como se puede observar en la Tabla 5, todas las correlaciones fueron significativas. Según lo esperado, las correlaciones entre florecimiento y las otras medidas de bienestar (satisfacción con la vida y autoestima) fueron positivas. En cambio, la correlación entre florecimiento y depresión fue negativa.

Discusión

El presente proyecto de investigación tuvo como propósito analizar las propiedades psicométricas de la Escala de Florecimiento de Diener et al. (2010) en una muestra de adultos en Puerto Rico. A partir de los resultados obtenidos, se puede establecer que la escala posee adecuadas características psicométricas para examinar el bienestar humano desde una perspectiva eudaimónica. Específicamente, la escala evalúa relaciones interpersonales positivas, sentimientos de competencia, sentido de propósito y optimismo. Además, los índices de confiabilidad obtenidos sugieren, según lo establece Kline (2000), que la Escala de Florecimiento cuenta con la suficiente consistencia interna para ser utilizada como instrumento de medición científica en investigacio-

nes en Puerto Rico y América Latina.

De forma general, el adiconamiento del AFE del Estudio 1 con el AFC del Estudio 2 demuestra que el modelo hipotetizado presenta un ajuste satisfactorio con los datos y confirma la estructura unifactorial del instrumento. Esto es congruente con investigaciones realizadas en otros países que aportan evidencia a la unidimensionalidad de la escala (Cassaretto-Bardales & Martínez-Urbe, 2017; Hone et al., 2014; Howell & Buro, 2015; Silva & Caetano, 2013). Según el AFE, la estructura factorial de la escala explica un 65% de la varianza total, lo que concuerda con los parámetros de varianza explicados en otras investigaciones internacionales que oscilan entre 47% y 73% (Cassaretto-Bardales & Martínez-Urbe, 2017; Diener et al., 2010; Hone et al., 2014; Silva & Caetano, 2013; Sumi, 2013).

En cuanto a la confiabilidad del instrumento, se obtuvieron índices superiores a los evidenciados por la escala en la mayoría de los estudios (Cassaretto-Bardales & Martínez-Urbe, 2017; Diener et al., 2010; Hone et al., 2014; Silva & Caetano, 2013), únicamente superados por el índice de consistencia interna obtenido en la validación de la escala en Japón (Sumi, 2013). Las correlaciones de cada ítem con la puntuación total manifiestan una consistencia interna notable y superan el rango de .61 - .71 informado en la literatura científica (Cassaretto-Bardales & Mar-

Tabla 5

Correlaciones entre las escalas, sus medias, desviaciones estándar y Alfa de Cronbach.

	α	M	DE	1	2	3
1. Florecimiento	.92	44.85	10.08			
2. Satisfacción con la Vida	.92	22.76	8.01	.73		
3. Autoestima	.90	33.49	6.18	.50	.50	
4. Depresión	.90	5.77	5.23	-.47	-.50	-.51

Nota. Todas las correlaciones significativas a .01. (n = 342); α = Alfa de Cronbach

tínez-Urbe, 2017; Diener et al., 2010).

Por su parte, la correlación moderadamente alta de la Escala de Florecimiento con la medida de satisfacción con la vida aporta evidencia sobre la validez convergente de la escala y concuerda con los patrones de correlaciones obtenidos en otras investigaciones, tanto en Puerto Rico como internacionalmente (González-Rivera, Pabellón-Lebrón, & Rosario-Rodríguez, 2017; Hone et al., 2014; Silva & Caetano, 2013). Esto demuestra que la Escala de Florecimiento es una medida confiable de bienestar eudaemónico y distinguible de las medidas de bienestar hedónico. Además, la relación obtenida con la presencia de sintomatología depresiva también es la esperada y la observada en estudios previos (Faulk, Gloria, & Steinhardt, 2013; Keyes, 2002; Keyes, 2006), lo que confirma que la escala posee validez discriminante.

En términos de las implicaciones teóricas del estudio, los hallazgos psicométricos demostraron que la escala está organizada de forma coherente con lo planteado por sus autores. A su vez, las similitudes de los resultados en todas las validaciones de la escala en distintos países sugieren que la escala no se ve afectada por variables y atributos culturales. Este detalle permite suponer cierta consistencia teórica sobre los procesos psicológicos que facilitan y desarrollan el florecimiento y el bienestar humano. En este sentido, nuestros resultados robustecen los fundamentos teóricos de la psicología del bienestar y la psicología positiva. Además, como señalan Diener et al. (2010) y Cassaretto-Bardales y Martínez-Urbe (2017), se confirmó con los análisis descriptivos la tendencia a experimentar altos niveles de bienestar, independientemente de la muestra que se utilice a nivel internacional.

En cuanto a las implicaciones prácticas del estudio, se demostró que la Escala de Florecimiento es útil en el desarrollo de nuevas investi-

gaciones en los campos de la psicología positiva y de la psicología de la salud en Puerto Rico y América Latina (Cassaretto-Bardales & Martínez-Urbe, 2017). En términos de conveniencia, las principales fortalezas de la escala son su lenguaje sencillo y su brevedad, lo que la consolida como un instrumento de fácil administración e interpretación. En este sentido, la Escala de Florecimiento se constituye como una herramienta práctica y efectiva en el quehacer investigativo y académico de la psicología latinoamericana.

Como parte de las limitaciones de esta investigación, se encuentra el tipo de muestreo no probabilístico por disponibilidad. Esto hace que la generalización de los resultados sea limitada. Sin embargo, se obtuvo una muestra heterogénea y de un tamaño razonable en los tres estudios para poder examinar las propiedades psicométricas del instrumento en tres muestras independientes. Además, el uso de una plataforma electrónica es una limitación porque limita la oportunidad de participación solo a personas que tengan acceso a internet, aún cuando se trata de una técnica en auge por facilitar una recolección ágil.

Para futuros estudios, se recomienda examinar la confiabilidad temporal a través de la técnica test-retest. A su vez, deben estudiarse las propiedades psicométricas de la escala en otras poblaciones como, por ejemplo, adolescentes y pacientes psiquiátricos. De igual forma, se deben llevar a cabo análisis de comparación según las variables demográficas, particularmente edad, estatus socioeconómico y preparación académica, para identificar diferencias entre los grupos. Por último, se sugiere replicar estos resultados en muestras más amplias y/o equilibradas por sexo.

Conclusiones generales

En síntesis, los resultados obtenidos en

cuanto a validez y confiabilidad indican que las propiedades psicométricas de la Escala de Florecimiento son adecuadas. El hecho de utilizar tres muestras independientes para validar y examinar la escala en Puerto Rico da un valor añadido al estudio. Se concluye que la Escala de Florecimiento es un excelente instrumento de medición para futuras investigaciones en Puerto Rico, así como para la práctica profesional de la psicoterapia y la consejería. Por último, existe suficiente evidencia psicométrica para considerar la Escala de Florecimiento como un índice transnacional para realizar estudios comparativos entre distintas culturas.

Referencias

- Aristóteles (2001). *Ética*. Madrid, España: Libsa.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York, NY: Psychology Press.
- Cassaretto-Bardales, M., & Martínez-Urbe, P. (2017). Validación de las escalas de bienestar, de florecimiento y afectividad. *Pensamiento Psicológico*, 15(1), 19-31.
- Chitgian-Urzúa, V., Urzúa, A., & Vera-Villaruel, P. (2013). Análisis preliminar de las Escalas de Bienestar Psicológico en población chilena. *Revista Argentina de Clínica Psicológica*, 22(1), 5-14.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper & Row.
- Diener, E. (2000). Subjective well-being: The science of happiness and proposal for a national index. *American Psychologist*, 55(1), 34-43. doi: 10.1037/0003-066X.55.1.34
- Diener, E., & Suh, E. (2000). *Culture and subjective well-being*. Cambridge: Massachusetts Institute of Technology.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). *The Satisfaction with Life Scale*. *Journal of Personality Assessment*, 49(1), 71-75. doi: 10.1207/s15327752jpa4901_13
- Diener, E., Wirtz, D., Biswas-Diener, R., Tov, W., Kim-Prieto, C., Choi, D.-w., & Oishi, S. (2009). New measures of well-being. En E. Diener (Ed.), *Social indicators research series: Vol. 39. Assessing well-being: The collected works of Ed Diener* (pp. 247-266). New York, NY: Springer Science + Business Media. doi: 10.1007/978-90-481-2354-4_12
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, 97(2), 143-156.
- Esch, T., Jose, G., Gimpel, C., Von Scheidt, C., & Michalsen, A. (2013). Die Flourishing Scale (FS) von Diener et al. liegt jetzt in einer autorisierten deutschen Fassung (FS-D) vor: Einsatz bei einer Mind-Body-medizinischen Fragestellung. *Forschende Komplementarmedizin*, 20(4), 267-275. doi: 10.1159/000354414
- Faulk, K. E., Gloria, C. T., & Steinhardt, M. A. (2013). Coping profiles characterize individual flourishing, languishing, and depression. *Anxiety, Stress, & Coping: An International Journal*, 26(4), 378-390. doi: 10.1080/10615806.2012.708736
- González-Rivera, J. A., Pabellón-Lebrón, S., & Rosario-Rodríguez, A. (2017). El rol mediador de la identificación ateísta en la relación entre discriminación y bienestar psicológico: Un estudio preliminar. *Revista Puertorriqueña de Psicología*, 28(2), 406-421. Recuperado de <http://www.ojs.repsasppr.net/index.php/reps/index>
- González-Rivera, J. A., Quintero-Giménez, N., Veray-Alicia, J., & Rosario-Rodríguez, A. (2016). Adaptación y validación de la escala de bienestar psicológico de Ryff en una muestra de adultos puertorriqueños. *Salud y Conducta Humana*, 3(1), 1-14. Recuperado de <http://rsych.com/>
- Helliwell, J. F., Barrington-Leigh, C., Harris, A., & Huang, H. (2010). International evidence on the social context of well-being. En E. Diener, D. Kahneman, & J. F. Helliwell (Eds.), *International differences in well-being*, (pp. 291-327). Oxford, UK:

- Oxford University Press. doi: [10.1093/acprof:oso/9780199732739.003.0010](https://doi.org/10.1093/acprof:oso/9780199732739.003.0010)
- Hone, L., Jarden, A., & Schofield, G. (2014). Psychometric properties of the Flourishing Scale in a New Zealand sample. *Social Indicators Research*, 119(2), 1031-1045. doi: [10.1007/s11205-013-0501-x](https://doi.org/10.1007/s11205-013-0501-x)^[1]_{SEP}
- Howell, A. J., & Buro, K. (2015). Measuring and predicting student well-being: Further evidence in support of the Flourishing Scale and the Scale of Positive and Negative Experiences. *Social Indicators Research*, 121(3), 903-915. doi: [10.1007/s11205-014-0663-1](https://doi.org/10.1007/s11205-014-0663-1)
- IBM Corporation (2016). IBM SPSS Statistics for Windows, Version 24.0 [Software de computación]. Armonk, NY: IBM.
- Keyes, C. L. M. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior*, 43(2), 207-222. doi: [10.2307/3090197](https://doi.org/10.2307/3090197)
- Keyes, C. L. M. (2006). Mental health in adolescence: Is America's youth flourishing? *American Journal of Orthopsychiatry*, 76(3), 395-402. doi: [10.1037/0002-9432.76.3.395](https://doi.org/10.1037/0002-9432.76.3.395)
- Keyes, C. L. M., Shmotkin, D., & Ryff, C. D. (2002). Optimizing well-being: The empirical encounter of two traditions. *Journal of Personality and Social Psychology*, 82(6), 1007-1022. doi: [10.1037/0022-3514.82.6.1007](https://doi.org/10.1037/0022-3514.82.6.1007)
- Kline, P. (2000). *The Handbook of Psychometric Testing*. New York: Routledge.
- Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks: Sage. doi: [10.4135/9781483385693](https://doi.org/10.4135/9781483385693)
- Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales* (2a Ed.). Sydney: Psychology Foundation.
- Montero, I., & León, O. G. (2007). A guide for naming research studies in psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847-862. Recuperado de <http://www.aepc.es/ijchp/busca.php?coid=Espa%F1ol>
- Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural equation modeling: A Multidisciplinary Journal*, 9(2), 195-212. doi: [10.1207/S15328007SEM0902_3](https://doi.org/10.1207/S15328007SEM0902_3)
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press. doi: [10.1515/9781400876136](https://doi.org/10.1515/9781400876136)
- Ryan, R., & Deci, E. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, 52(1), 141-166. doi: [10.1146/annurev.psych.52.1.141](https://doi.org/10.1146/annurev.psych.52.1.141)
- Ryff, C. (1989). Happiness is everything, or is it? Exploration on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57(6), 1069-1081. doi: [10.1037/0022-3514.57.6.1069](https://doi.org/10.1037/0022-3514.57.6.1069)
- Ryff, C. (1995). Psychological well-being in adult life. *Current Directions in Psychological Science*, 4(4), 99-104. doi: [10.1111/1467-8721.ep10772395](https://doi.org/10.1111/1467-8721.ep10772395)
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514. doi: [10.1007/BF02296192](https://doi.org/10.1007/BF02296192)
- Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3a ed.). Mahwah, NJ: Erlbaum. doi: [10.4324/9780203851319](https://doi.org/10.4324/9780203851319)
- Seligman, M. E. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. New York, NY: Free Press.
- Seligman, M. E., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, 55(1), 5-14. doi: [10.1037/0003-066X.55.1.5](https://doi.org/10.1037/0003-066X.55.1.5)
- Seligman, M. E., & Peterson, C. (2003). Positive clinical psychology. En L. G. Aspinwall & U.M. Staudinger (Eds.). *A Psychology of human strengths: Fundamental questions and future directions for a positive psychology* (pp. 305-317). Washington, DC: American Psychological Association.
- Silva, A. J., & Caetano, A. (2013). Validation of the Flourishing Scale and Scale of Positive and Negative Experience in Portugal. *Social Indicators Research*, 110(2), 469-478. doi: [10.1007/s11205-011-9938-y](https://doi.org/10.1007/s11205-011-9938-y)

- StataCorp. (2017). Stata: Release 15 [Software de computación]. College Station, TX: StataCorp LLC.
- Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4a ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sumi, K. (2013). Reliability and validity of Japanese versions of the Flourishing Scale and the Scale of Positive and Negative Experience. *Social Indicators Research*, 118(2), 601-615. doi: [10.1007/s11205-013-0432-6](https://doi.org/10.1007/s11205-013-0432-6)
- Waterman, A. S., Schwartz, S. J., Zamboanga, B. L., Ravert, R. D., Williams, M. K., Agocha, V. B., . . . Donnellan, M. B. (2010). The Questionnaire for Eudaimonic Well-Being: Psychometric properties, demographic comparisons, and evidence of validity. *The Journal of Positive Psychology*, 5(1), 41-61. doi: [10.1080/17439760903435208](https://doi.org/10.1080/17439760903435208)
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070. doi: [10.1037/0022-3514.54.6.1063](https://doi.org/10.1037/0022-3514.54.6.1063)
-



Best Practices in the Use of Bifactor Models: Conceptual Grounds, Fit Indices and Complementary Indicators

Buenas Prácticas en el uso de Modelos Bifactor: Teoría de Base, Índices de Ajuste e Indicadores Complementarios

Pablo Ezequiel Flores-Kanter *^{1,2}, Sergio Dominguez-Lara³,
Mario Alberto Trógolo^{1,2}, Leonardo Adrián Medrano^{1,2}

1 - Universidad Siglo 21, Córdoba, Argentina.

2 - Facultad de Psicología, Universidad Nacional de Córdoba, Argentina.

3 - Universidad de San Martín de Porres, Perú.

Introduction
Bifactor Model:
Basic Issues
Example: Factor
structure of
PANAS
Conclusion
References

Recibido: 18/05/2018 **Revisado:** 24/07/2018 **Aceptado:** 06/09/2018

Abstract

Bifactor models have gained increasing popularity in the literature concerned with personality, psychopathology and assessment. Empirical studies using bifactor analysis generally judge the estimated model using SEM model fit indices, which may lead to erroneous interpretations and conclusions. To address this problem, several researchers have proposed multiple criteria to assess bifactor models, such as a) conceptual grounds, b) overall model fit indices, and c) specific bifactor model indicators. In this article, we provide a brief summary of these criteria. An example using data gathered from a recently published research article is also provided to show how taking into account all criteria, rather than solely SEM model fit indices, may prevent researchers from drawing wrong conclusions.

Keywords: *Confirmatory factor analyses, bifactor models, PANAS, complementary statistical fit indices*

Resumen

En los últimos años, el uso de modelos bifactor ha tenido un creciente protagonismo en las literaturas de personalidad, de psicopatología y de evaluación psicológica. Los estudios empíricos que aplican este procedimiento generalmente hacen determinaciones sobre la estructura final utilizando únicamente índices de ajuste obtenidos a partir de modelos de ecuaciones estructurales, lo que puede derivar en conclusiones erróneas. Por este motivo, recientemente se han propuesto diferentes criterios a la hora de evaluar el ajuste de modelos bifactor que incluyen a) consideraciones teóricas, b) índices de ajuste global del modelo y c) indicadores específicos del modelo bifactor. En este artículo se repasan brevemente estos criterios. Asimismo, se ofrece un ejemplo utilizando datos provenientes de una publicación reciente con el fin de evidenciar cómo las decisiones basadas en los índices globales de ajuste, sin tener en cuenta las bases teóricas y los indicadores específicos del bifactor, pueden llevar a los investigadores a obtener conclusiones sesgadas.

Palabras clave: *Análisis factorial confirmatorio, modelos bifactor, PANAS, índices estadísticos complementarios*

Introduction

Despite the increasing use of bifactor models in psychology and related sciences, several researchers have noticed common errors in the bifactor analyses performed, particularly in the interpretation of model fit based solely in SEM (structural equation model) model fit indices (Gignac, 2016; Rodriguez, Reise, & Haviland, 2016). Based on literature review, the aim of the present article is to recommend multiple criteria when assessing bifactor models, including a) conceptual grounds, b) overall model fit indices, and c) additional, specific bifactor model indicators. Using data from a recent publication on factorial structure of the Positive and Negative Affect Schedule (Seib-Pfeifer, Pugnaghi, Beauducel, & Leue, 2017), we illustrate errors and misinterpretation that can result from exclusively relying on SEM model fit indices without taking into consideration certain theoretical basis and/or specific bifactor model fit indices.

Bifactor Model: Basic Issues

Within the framework of confirmatory factor analysis, albeit not exclusive to this approach, bifactor models (BM) represent one of the options available to the applied researcher for simultaneously testing the extent to which a particular set of items are explained by a general underlying factor and group-level factors. By doing so, BM constitutes a means to determine whether the construct being measured by a scale can be viewed primarily as either unidimensional or multidimensional. Bifactor models should be distinguished from hierarchical models with higher-order factors. Particularly, because in the latter the group-level factors represent dimensions of a general factor, whereas in BM, group-level factors are hypothe-

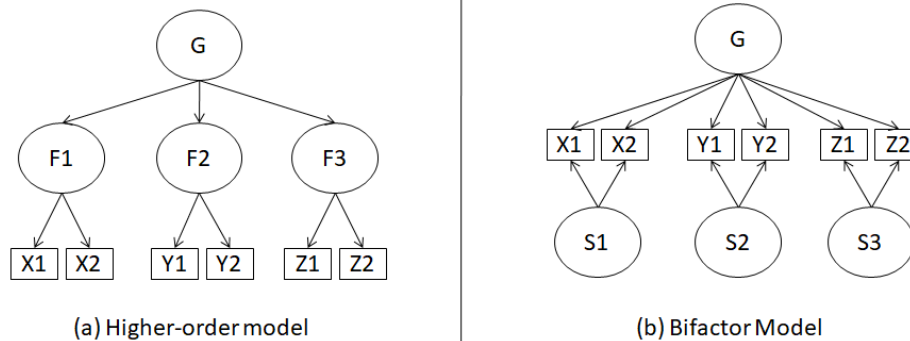
sized to be independent (i.e. orthogonal) and not lower-order factors of the general factor (see Figure 1). Accordingly, BM is especially useful for assessing the validity of an instrument intended to measure both the overall construct and its specific dimensions (Rodriguez et al., 2016).

Given that BM tends to outperform conventional confirmatory factorial analysis (CFA) models simply because of the way in which these are specified (Gignac, 2016; Morgan, Hodge, Wells, & Watkins, 2015), when a particular BM yields adequate model fit indices it is highly recommended to use complementary statistical indices for a more accurate interpretation. These additional, BM-specific fit indices are the Explained Common Variance (ECV), the Percentage of Uncontaminated Correlations (PUC), the hierarchical Omega coefficients (omega ω , omega subscale ω_s , omega hierarchical ω_h , and omega hierarchical subscale ω_{hs}) and the H coefficient (Rodriguez et al., 2016).

The ECV represents the proportion of the common variance attributable to the general factor. On the other hand, the ω_h reflects the proportion of the total variance explained by the general factor, while the ω_{hs} reflects the proportion of the total variance accounted for the specific factors after controlling for the influence of the general factor. High values on ECV ($> .60$) and ω_h ($> .70$) indicate that the variance of the indicators is substantially accounted for by the general factor and, therefore, the tenability of specific factors would be forced. Nonetheless, some authors (Bonifay, Lane, & Reise, 2017) have also pointed that if ω_{hs} explains a non-redundant amount of variance ($> .30$), then the specific factors could be retained along with the general factor, albeit considering the theoretical underpinnings of such decision. Also, there is an application of ECV at item level (ECV-I, Stucky, Thissen, & Edelen, 2013) with a similar interpretation to ECV.

Figure 1

Graphical representations of higher-order factor model and bifactor model.



Note. X1, X2, Y1, Y2, y Z1, Z3 = items; F1-F3 = factors; S1-S3 = group-level factors in bifactor model; G = general factor.

The PUC indicates the percentage of correlations not contaminated by multidimensionality (i.e., group-level factors) and moderates or supports the interpretation of ECV. According to [Rodriguez et al. \(2016\)](#), when ECV is $> .70$ and $PUC > .70$, the common variance can be regarded as essentially unidimensional, thus supporting the general factor. Finally, the H coefficient is a measurement of construct replicability and is defined as the extent to which a set of items represents a latent variable ([Hancock & Mueller, 2001](#)). This coefficient is calculated as the ratio of the percentage of variance explained by the latent variable to the percentage of variance unexplained by the latent variables. High H values ($> .80$) suggest that the latent variable is well-defined and adequately represented by the observed indicators, thus more likely to be replicated across studies.

Along with complementary model fit indices, theoretically-based reasoning should be applied when deciding in favor of a factor model other than overall model fit ([Morgan et al., 2015](#)), so that the interpretation of the resulting factors is conceptually well founded.

Example: Factor structure of PANAS

In their study, [Seib-Pfeifer et al. \(2017\)](#) concluded that the BM comprising positive affect (PA), negative affect (NA), and a general Affective Polarity factor had a better model fit compared to alternative two- and three-factor models. They made this assertion even when relevant BM statistical indices (e.g., ECV, ω_h , ω_{hs}) were not provided. Indeed, when BM statistical indices are calculated, conclusions differ markedly from [Seib-Pfeifer et al. \(2017\)](#). In particular, results show that the common variance accounted for by the general Affective Polarity factor is rather weak (ECV = 10.8 %), as it is also for the variance at item level (ECV- I_{mean} : .07 for PA and .16 for NA). The mean factor loading is moderate-to-high for both PA ($\lambda_{\text{mean}} = .52$) and NA ($\lambda_{\text{mean}} = .57$), and a significant amount of variance is also explained by the two specific factors (PA $\omega_{hs} = .78$; NA $\omega_{hs} = .72$). In contrast, both factor loadings and explained variance of the general Affective Polarity factor are close to zero ($\lambda_{\text{mean}} = .091$; $\omega_h = .049$). In addition, the PUC value is .53 and the H coeffi-

cient is acceptable for group-level factors ($H_{PA} = .80$; $H_{NA} = .77$) and low for the general factor ($H_G = .41$). All in all, these findings suggest that the conclusion drawn by Seib-Pfeifer et al. (2017) in their study, i.e that BM provides the most appropriate representation of the PANAS, is erroneous.

Conclusion

In a strict sense, a factor is a mathematical abstraction derived from the empirical covariance between a set of variables, which may (or may not) be interpreted as a common, substantive cause underlying a set of observable behaviors. The bifactor model is an alternative specification of the second order factor. In contrast to second-order models, the bifactor a) allows for the quantification of the direct effect of the general factor on observable variables without the need for such a relationship to be fully mediated by group factors, and b) facilitates independent evaluation of the merits of general and group factors. Underlying this model is the hypothesis that there is a general factor with causal influence on all items (Arias, Jenaro, & Ponce, 2018). Sometimes, the variance shared between different indicators may be artificial, the valence of the items (Bäckström & Björklund, 2016), or general features of those examined -not directly related to the construct- such as self-esteem (Davies, Connelly, Ones, & Birkland, 2015). These may increase the fit of the bifactor model, leading to the erroneous conclusion that there is a general, psychological, and psychometrically meaningful factor.

Bifactor models constitute a useful analytic method for the assessment of construct validity in psychological measuring. However, researchers should be wary when performing bifactor analysis and avoid using SEM model fit indices as the main or only criteria to judge the feasibility of a

BM, since conclusions based on such criteria can be misleading considering that complementary statistical fit indices provide researchers with relevant information concerning the salience of general and group-level factors. Such information, together with theoretical background, will allow for the selection of the most appropriate factor model and ensure the validity of the inferences drawn from the scale scores.

References

- Arias, V. B., Jenaro, C., & Ponce, F. P. (2018). Testing the generality of the general factor of personality: An exploratory bifactor approach. *Personality and Individual Differences, 129*, 17-23. doi: [10.1016/j.paid.2018.02.042](https://doi.org/10.1016/j.paid.2018.02.042)
- Bäckström, M., & Björklund, F. (2016). Is the general factor of personality based on evaluative responding? Experimental manipulation of item-popularity in personality inventories. *Personality and Individual Differences, 96*, 31-35. doi: [10.1016/j.paid.2016.02.058](https://doi.org/10.1016/j.paid.2016.02.058)
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science, 5*(1), 184-186. doi: [10.1177/2167702616657069](https://doi.org/10.1177/2167702616657069)
- Davies, S. E., Connelly, B. S., Ones, D. S., & Birkland, A. S. (2015). The general factor of personality: The “Big One,” a self-evaluative trait, or a methodological gnat that won’t go away? *Personality and Individual Differences, 81*, 13-22. doi: [10.1016/j.paid.2015.01.006](https://doi.org/10.1016/j.paid.2015.01.006)
- Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55*, 57-68. doi: [10.1016/j.intell.2016.01.006](https://doi.org/10.1016/j.intell.2016.01.006)
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit & D. Sorbom (Eds.), *Structural equation modeling: Present and future—A Fest-*

schrift in honor of Karl Joreskog (pp. 195-216). Lincolnwood, IL: Scientific Software International.

- Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo Simulations. *Journal of Intelligence*, 3(1), 2-20. doi: [10.3390/jintelligence3010002](https://doi.org/10.3390/jintelligence3010002)
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223-237. doi: [10.1080/00223891.2015.1089249](https://doi.org/10.1080/00223891.2015.1089249)
- Seib-Pfeifer, L.-E., Pugnaghi, G., Beauducel, A., & Leue, A. (2017). On the replication of factor structures of the Positive and Negative Affect Schedule (PANAS). *Personality and Individual Differences*, 107, 201-207. doi: [10.1016/j.paid.2016.11.053](https://doi.org/10.1016/j.paid.2016.11.053)
- Stucky, B. D., Thissen, D., & Edelen, M. O. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement*, 37(1), 41-57. doi: [10.1177/0146621612462759](https://doi.org/10.1177/0146621612462759)
-

Validación de la Escala de Búsqueda de Sensaciones Sexuales en mexicanos casados o cohabitantes

Validation of the Sexual Sensation Seeking Scale among married or cohabiting Mexicans

José Moral de la Rubia * ¹

1 - Facultad de Psicología, Universidad Autónoma de Nuevo León.

Introducción
Método
Resultados
Discusión
Referencias

Recibido: 28/06/2018 Revisado: 19/09/2018 Aceptado: 05/10/2018

Resumen

El objetivo del estudio fue validar la Escala de Búsqueda de Sensaciones Sexuales de 9 ítems (SSSS-9) en una población de adultos mexicanos casados o cohabitantes con pareja del sexo opuesto. Se recolectó una muestra de 807 participantes (53.4 % mujeres y 46.6 % hombres; 90.1 % casados y 9.9 % cohabitantes) por rutas aleatorias. Se aplicó la SSSS-9 y las escalas de conducta infiel y actitud hacia la exclusividad marital. La SSSS-9 presentó una consistencia interna muy alta. El número de factores fue uno según el análisis paralelo de Horn. El modelo unifactorial tuvo buen ajuste y resultó válido para ambos sexos por análisis factorial confirmatorio. La distribución de la SSSS-9 mostró asimetría positiva. Su promedio fue mayor en hombres que en mujeres. Tuvo correlación alta con conducta infiel y moderada con actitud hacia la exclusividad marital. Se concluye que las propiedades de SSSS-9 resultan mejores en esta población que las reportadas con SSSS-11 en otras poblaciones.

Palabras clave: *búsqueda de sensaciones sexuales, parejas concurrentes, actitud hacia la infidelidad marital, sexo, México*

Abstract

The objective of this study was to validate the 9-item Sexual Sensation Seeking Scale (SSSS-9) in a population of Mexican adults married or cohabiting with a partner of the opposite sex. A sample of 807 participants was collected (53.4% women and 46.6% men, 90.1% married and 9.9% cohabitants) from random sources. The SSSS-9, one scale of unfaithful behavior and other one of attitude toward marital exclusivity were applied. The SSSS-9 showed very high internal consistency. The number of factors was one through Horn's parallel analysis. The one-factor model had a good fit and was valid across both sexes through confirmatory factor analysis. The distribution of SSSS-9 showed positive asymmetry. Its average was higher among men than among women. SSSS-9 had a high correlation with unfaithful behavior, and moderate with attitude toward marital exclusivity. It is concluded that the properties of SSSS-9 are better in this population than those reported with SSSS-11 in other populations.

Key words: *sexual sensation seeking, concurrent partners, attitude toward marital infidelity, sex, Mexico*

* **Correspondencia a:** José Moral de la Rubia. c/Dr. Carlos Canseco 110. Col. Mitras Centro. C.P. 64460. Monterrey, Nuevo León, México. Tel. 00 52 81 83338233. Ext. 423. Fax. Ext. 103. Correo electrónico: jose_moral@hotmail.com

Cómo citar este artículo: Moral de la Rubia, J. (2018). Validación de la Escala de Búsqueda de Sensaciones Sexuales en mexicanos casados o cohabitantes. *Revista Evaluar*, 18(3), 49-72. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

Cuando una persona cree que su pareja le es infiel se siente traicionada y menos comprometida y usualmente la confirmación de la infidelidad conduce a la ruptura o alejamiento (Huang, Cassels, & Winer, 2015). Las causas de la búsqueda de parejas concurrentes son diversas, entre ellas se considera a la búsqueda de sensaciones sexuales como uno de sus factores de riesgo (Kogan, Cho, Barnum, & Brown, 2015). Más allá de la aplicación de este concepto a las conductas sexuales de riesgo (Xu, Zheng, Liu, & Zheng, 2016), la búsqueda de sensaciones sexuales puede ser una variable importante en el estudio de las parejas. Consecuentemente, poseer un instrumento adecuado para su evaluación adquiere relevancia.

La *búsqueda de sensaciones sexuales* fue definida por Kalichman et al. (1994) como la tendencia a obtener niveles más altos de excitación sexual mediante la implicación en experiencias sexuales novedosas. Para su evaluación, estos autores crearon la Escala de Búsqueda de Sensaciones Sexuales de nueve ítems (SSSS-9). En una muestra de 106 hombres homosexuales, encontraron un nivel aceptable de consistencia interna (α de Cronbach = .75) y de estabilidad temporal en un intervalo de dos semanas ($r_{[36]} = .78, p < .01$). Las correlaciones de los ítems con la escala variaron de .29 a .58 con una media de .43. Al calcular una función discriminante para diferenciar entre los participantes que no habían tenido sexo anal sin uso del condón en los últimos tres meses y los que sí, la SSSS-9 tuvo el coeficiente estructural más alto ($\beta = .56$).

Posteriormente, Kalichman y Rompa (1995) modificaron la escala para mejorar su consistencia interna. Se conservaron siete de sus nueve ítems (ítems 1, 4, 5, 6, 7, 8 y 9), se eliminaron dos (ítems 2 y 3) y se introdujeron cuatro nuevos: las sensaciones físicas son lo más importante en el

sexo; disfruto la sensación de mantener relaciones sexuales sin preservativo; mis parejas sexuales probablemente piensan que soy una persona que hace cosas arriesgadas; y cuando se trata de sexo, la atracción física es más importante que lo bien que yo conozca a la persona. De este modo la escala pasó a tener 11 ítems (SSSS-11). Esta nueva versión fue aplicada en dos muestras: una de 296 hombres que tienen sexo con hombres y otra de 60 hombres y 98 mujeres heterosexuales. En la primera muestra se encontró una consistencia interna aceptable (α de Cronbach = .79), una estabilidad temporal a los tres meses también aceptable ($r_{[195]} = .69, p < .01$) y correlaciones bajas con conductas sexuales de riesgo ($r_{[296]} = .22$ con número de parejas y .19 con coito sin condón). En la segunda, la consistencia interna fue alta (α de Cronbach = .81), al igual que la estabilidad temporal ($r_{[52]} = .73, p < .01$). Conforme con la expectativa de mayor libertad e impulso sexual en los hombres esgrimida por los autores, su media ($M = 2.34, DE = 0.67$) fue significativamente más alta ($t_{[143]} = 3.77, p < .001$) que la de las mujeres ($M = 1.95, DE = 0.57$). Finalmente, al comprobar la validez de constructo, se encontró que las correlaciones con conductas sexuales de riesgo fueron medianas.

Gaither y Sellbom (2003) elaboraron una versión de 10 ítems para su aplicación a estudiantes universitarios heterosexuales de ambos sexos (SSSS-10). En una muestra de 528 estudiantes universitarios estadounidenses, la consistencia interna de la escala fue alta en hombres (α de Cronbach = .83) y mujeres (α de Cronbach = .81), fue discriminante entre ambos sexos y correlacionó con permisividad sexual, siendo las correlaciones más altas en mujeres que en hombres (Gaither & Sellbom, 2003).

Kalichman et al. (1994) y Kalichman y Rompa (1995) asumieron que la escala presenta una estructura unidimensional. No obstante, es-

tudios factoriales recientes con la SSSS-11 han revelado una estructura de dos factores correlacionados (nuevas experiencias y nuevas sensaciones) que, tras eliminarse dos o tres de los ítems nuevos, presenta mejor ajuste a los datos (Ballester-Arnal, Ruiz-Palomino, Espada-Sánchez, Morell-Mengual, & Gil-Llario, 2018; De Oliveira-Santos, Soares-Ferreira, Carvalho-Duarte, & Ferreira, 2017; Santos-Iglesias, Moyano, Castro, Granados, & Sierra, 2018). La denominación dada a cada factor intenta reflejar el contenido dominante de sus ítems definidores. En el factor de nuevas sensaciones predominan los contenidos sobre sensaciones sin referencia explícita a lo externo, por ejemplo, estoy interesado en experimentar nuevas sensaciones sexuales o me gustan las experiencias y sensaciones sexuales nuevas y excitantes y en el factor de nuevas experiencias predominan los contenidos sobre situaciones con referencia explícita o dependientes de otras personas o estímulos externos, por ejemplo disfruto de la compañía de una persona sensual o disfruto ver películas pornográficas.

En México, no se ha validado la escala. No obstante, se tiene un estudio realizado en una muestra de 255 hombres que tienen sexo con hombres de Monterrey (Valdez, Moral, Onofre, Castillo, Benavides, & López, 2016). En esta investigación, se reportó una consistencia interna aceptable (α de Cronbach = .74) semejante a la obtenida con la SSSS-9 (Kalichman et al., 1994). La media en la puntuación total de la SSSS-11 ($M = 2.80$, IC 95 %: 2.74, 2.87) fue significativamente mayor que la obtenida entre 60 varones adultos estadounidenses usuarios de servicios socio-sanitarios ($M = 2.34$, IC 95 %: 2.17, 2.51; $t_{[240]} = 14.59$, $p < .001$) del estudio de Kalichman y Rompa (1995). Esta diferencia se puede atribuir a las distintas poblaciones de las cuales fueron extraídas las muestras, al ser la población de hombres que tienen sexo con hombres la que presenta

mayor riesgo de contraer enfermedades de transmisión sexual por sus prácticas sexuales (Iglesias & Reyes, 2017).

Hay investigaciones sobre la confiabilidad y validez de la SSSS en sus tres versiones, aunque estas presentan debilidades. La mayoría de ellas se han hecho con muestras incidentales, usualmente con hombres que tienen sexo con hombres y estudiantes universitarios o de media superior. Otra debilidad aparece en relación con las técnicas de análisis. Los ítems de la SSSS presentan un rango muy limitado de respuesta (cuatro valores ordinales), lo que provoca que el coeficiente alfa de Cronbach y la correlación producto-momento de Pearson subestimen su confiabilidad. En su lugar, sería más adecuado usar el coeficiente alfa ordinal y la correlación poliserial y policórica (Gadermann, Guhn, & Zumbo, 2014). Además, los estudios más recientes con la SSSS-11 sugieren reducir la escala a nueve u ocho ítems, eliminando dos o más de los ítems nuevos para definir un modelo bifactorial (Ballester-Arnal et al., 2018; De Oliveira-Santos et al., 2017; Santos-Iglesias et al., 2018), cuando se podría definir un modelo más parsimonioso de un factor con la escala original, en la que domina el contenido de situaciones interpersonales o dependientes de estímulos externos.

Este estudio tiene como propósito validar la SSSS-9 en población de personas mexicanas casadas o cohabitantes con pareja del sexo opuesto residentes en Monterrey. Con base en este propósito se definieron ocho objetivos: 1) comprobar la consistencia interna de los nueve ítems de la SSSS-9, así como su discriminabilidad y adecuación distribucional; 2) determinar el número de factores por criterios empíricos y en caso de dos o más factores explorar los mismos; 3) contrastar el modelo de un factor con nueve indicadores, el de dos factores propuesto para la SSSS-11 limitado a los nueve ítems de la SSSS-9 y posibles modelos

alternativos derivados del análisis factorial exploratorio; 4) contrastar la invarianza de los modelos entre ambos sexos; 5) una vez establecida la estructura factorial, calcular la consistencia interna de la escala; 6) describir la distribución de la escala y comparar su media entre ambos sexos; 7) comprobar la validez de constructo concurrente en relación con conducta infiel y actitud hacia la exclusividad marital comparando estas correlaciones entre ambos sexos y 8) observar si la deseabilidad social tiene efecto sobre la SSSS-9 y comprobar la sustantividad de las correlaciones de validez al parcializar la deseabilidad social.

En correspondencia con los ocho objetivos enunciados se esperan: 1) ítems consistentes y sin efecto suelo o techo ($\geq 80\%$ de los casos en uno de los polos), como ya se observó en el estudio original de Kalichman et al. (1994) y estudios previos (De Oliveira-Santos et al., 2017; Teva & Paz-Bermúdez, 2008), con la posibilidad de que los ítems de la SSSS-9 eliminados en la elaboración de la SSSS-11 (ítems 2 y 3) sean los menos consistentes; 2) un factor (Kalichman et al., 1994) o dos factores (nuevas sensaciones personales (ítems 1, 8 y 9) y nuevas experiencias interpersonales o dependientes de estímulos externos (ítems 2, 3, 4, 5, 6 y 7), a semejanza de los obtenidos con la SSSS-11 (Ballester-Arnal et al., 2018; De Oliveira-Santos et al., 2017; Santos-Iglesias et al., 2018); 3) mejor ajuste del modelo de dos factores que el de un factor, como en los estudios con la SSSS-11, aunque con esta versión original de la escala, los factores pueden presentar falta de validez discriminante o exceso de varianza compartida, indicando unidimensionalidad al ser predominante el contenido de situación interpersonal o dependiente de estímulo externo; 4) adecuadas propiedades de invarianza de los modelos factoriales entre ambos sexos, al hipotetizarse que las diferencias entre hombres y mujeres serán de magnitud (promedios), pero no de cualidad (con-

figuración y número de los factores subyacentes); 5) consistencia interna alta en la escala al ser estimada por un método más adecuado a la naturaleza ordinal de sus ítems (Gadermann et al., 2014); 6) asimetría positiva en la distribución (Teva & Paz-Bermúdez, 2008) por los casos de conducta sexual compulsiva en la cola derecha de la distribución y mayor promedio en los hombres que en las mujeres por la mayor impulsividad y búsqueda de sensaciones asociadas a estos primeros (Cross, Cyrenne, & Brown, 2013) y la mayor permisividad sexual que la cultura occidental concede al hombre (Ballester-Arnal et al., 2018; De Oliveira-Santos et al., 2017; Kalichman & Rompa, 1995; Santos-Iglesias et al., 2018); 7) correlación positiva de la puntuación total de la SSSS-9 con infidelidad marital y actitud hacia la infidelidad marital (Kogan et al., 2015), pudiendo ser las correlaciones más altas en hombres que en mujeres por una mayor permisividad cultural hacia los hombres en asuntos de infidelidad (Selterman & Koleva, 2015) y 8) fuerza de asociación trivial o pequeña entre la SSSS-9 y la deseabilidad social, lo que hace que las correlaciones de validez concurrente resulten sustantivas al parcializar la deseabilidad social, conforme con los resultados de otros estudios con escalas de conductas sexuales (Grubbs, Exline, Pargament, Hook, & Carlisle, 2015; Muise, Stanton, Kim, & Impett, 2016).

Método

Se realizó un estudio instrumental con un diseño ex post facto transversal y un muestreo por rutas al azar.

Participantes

Los criterios de inclusión fueron: tener al

Tabla 1

Distribución de frecuencias de las variables sociodemográficas.

Variable	Valor	n (p)	N (P)
Sexo	Mujer	431 (53.4 %)	
	Hombre	376 (46.6 %)	
Años de edad	[18, 30)	285 (35.3 %)	285 (35.3 %)
	[30, 40)	230 (28.5 %)	515 (63.8 %)
	[40, 50)	220 (27.3 %)	735 (91.1 %)
	[50, 60)	60 (7.4 %)	795 (98.5 %)
	[60, 94]	12 (1.5 %)	807 (100 %)
Escolaridad	Primaria	173 (21.4 %)	173 (21.4 %)
	Secundaria	231 (28.6 %)	404 (50.1 %)
	Media superior	218 (27 %)	622 (77.1 %)
	Licenciatura	164 (20.3 %)	786 (97.4 %)
	Posgrado	21 (2.6 %)	807 (100 %)
Estado civil	Casado	727 (90.1 %)	
	Cohabitante	80 (9.9 %)	
Años de matrimonio o cohabitación	[0, 7.17)	361(44.7 %)	361 (44.7 %)
	[7.17, 14.34)	146 (18.1 %)	507 (62.8 %)
	[14.34, 21.51)	171 (21.2 %)	678 (84 %)
	[21.51, 28.67)	91 (11.3 %)	769 (95.3 %)
	[28.67, 35.84)	27 (3.3 %)	796 (98.6 %)
[35.84, 43]	11 (1.4 %)	807 (100 %)	
Ingresos económicos familiares al mes en el último año en pesos mexicanos	< 3000	61 (7.6 %)	61 (7.6 %)
	[3,000, 6,000)	220 (27.3 %)	281 (34.8 %)
	[6,000, 12,000)	226 (28.0 %)	507 (62.8 %)
	[12,000, 24,000)	174 (21.6 %)	681 (84.4 %)
	[24,000, 48,000)	77 (9.5 %)	758 (93.9 %)
	[48,000, 92,000)	34 (4.2 %)	792 (98.1 %)
≥ 92,000	15 (1.9 %)	807 (100 %)	
Número total de hijos	0	130 (16.1 %)	130 (16.1 %)
	1	149 (18.5 %)	279 (34.6 %)
	2	237 (29.4 %)	516 (63.9 %)
	3	211 (26.1 %)	727 (90.1 %)
	4	60 (7.4 %)	787 (97.5 %)
	≥ 5	20 (2.5 %)	807 (100 %)
Afiliación religiosa	Cristiana católica	662 (82.0 %)	
	Cristiana no católica	76 (9.4 %)	
	Otra	13 (1.6 %)	
	Ninguna	56 (6.9 %)	

Nota. *n* = frecuencia absoluta simple, *p* = porcentaje simple, *N* = frecuencia absoluta acumulada, *P* = porcentaje acumulado.

menos 18 años, estar casado o cohabitar con una pareja del sexo opuesto, ser mexicano, residir en la ciudad de Monterrey en México, saber leer y escribir y dar el consentimiento informado. Los criterios de eliminación fueron: cuestionario incompleto e informar que la pareja estuvo presente en el momento de ser contestado, leyendo o preguntando por las respuestas.

Se completó una muestra de 807 participantes, de los cuales 431 (53.4 %) fueron mujeres y 376 (46.6 %) hombres, sin diferencia de frecuencia estadísticamente significativa entre ambos sexos por la prueba binomial ($p = .057$). Sus características sociodemográficas se presentan en la Tabla 1.

En la población de Monterrey, descontando al 31 % de la población menor de 18 años, hay 98 hombres por cada 100 mujeres, una media de 35.1 años de edad, una media de 9.5 años de escolaridad (entre primer semestre y segundo de educación media superior), una media de ingresos mensuales de 6,700 pesos mexicanos (unos 335 dólares estadounidenses), una media de 2 hijos y el 83 % de la población con afiliación religiosa cristiana católica, 7.4 % cristiana no católica, 2.5 % de otra religión y 6.9 % sin religión (Instituto Nacional de Estadística, Geografía e Informática [INEGI], 2016). Estos datos poblacionales son estadísticamente equivalentes a los de la muestra, como en la frecuencia de ambos sexos ($\chi^2_{[1, N=807]} = 1.87, p = .171$) y de la afiliación religiosa ($\chi^2_{[3, N=807]} = 7.13, p = .068$), así como en la media de edad ($t_{[806]} = 0.806, p = .421$) y de número de hijos ($t_{[806]} = -0.27, p = .787$), o muy semejantes, como en la media de escolaridad (Mdn = 4 que corresponde a media superior, incluyendo estudios truncados y terminados) e ingresos (Mdn = 3 que corresponde al intervalo [6,000, 11,999]). Si se reduce la población a adultos de la zona metropolitana, se obtienen 89 % casados y 11 % cohabitantes (INEGI, 2016), lo que es un porcentaje equiva-

lente al muestral ($\chi^2_{[1, N=807]} = 0.97, p = .324$). Por tanto, no se consideró necesario acudir a métodos de ponderación para corregir sesgos sociodemográficos

Instrumentos

Escala de búsqueda de sensaciones sexuales (SSSS-9) de Kalichman et al. (1994). Se compone de nueve ítems de codificación directa con cuatro categorías ordenadas de respuesta, de 1 = *Nada* a 4 = *Muchísimo*. La puntuación total se obtiene sumando los nueve ítems y dividiendo por nueve, con lo que esta varía en un continuo de 1 a 4. Mayor puntuación refleja mayor tendencia a la búsqueda de excitación, experiencias y novedades sexuales. Su consistencia interna es aceptable (α de Cronbach = .75) y se asume que a sus nueve ítems subyace un factor general (Kalichman et al., 1994). La versión en español que se aplicó en el presente estudio se obtuvo por traducción reversa (Streiner, Norman, & Cairney, 2015; véase Anexo).

Escala de conducta infiel de Romero-Palencia, Rivera-Aragón y Díaz-Loving (2007). Cuenta con 48 ítems directos con cinco categorías de respuesta, de 1 = *Nunca* a 5 = *Siempre*. A mayor puntuación, mayor frecuencia de deseos y situaciones con parejas concurrentes. Posee una consistencia interna muy alta (α de Cronbach = .98). Tiene una estructura de cuatro factores: infidelidad sexual con 21 ítems (α de Cronbach = .97), deseo de infidelidad emocional con 14 ítems (α de Cronbach = .96), deseo de infidelidad sexual con ocho ítems (α de Cronbach = .96), e infidelidad emocional con cinco ítems (α de Cronbach = .87; (Romero et al., 2007).

Escala de actitud hacia la exclusividad marital.

Fue creada por Weis y Felton (1987). Consta de siete ítems directos con cinco categorías ordenadas de respuesta que se puntúan de 1 = *Total rechazo* a 5 = *Total aceptación*. Su consistencia interna es alta ($\alpha = .87$) y presenta una estructura unidimensional (Weis & Felton, 1987). Las puntuaciones en la escala se obtienen por suma simple de los ítems. A mayor puntuación, mayor aceptación de la infidelidad a la pareja. La versión en español que se aplicó en el presente estudio se obtuvo por traducción reversa.

Inventario balanceado de deseabilidad social (BIDR) con la adaptación a México de Moral, García y Antona (2012). Se usó la versión de 20 ítems directos con siete categorías ordenadas de respuesta, de 1 = *Nada de acuerdo* a 7 = *Totalmente de acuerdo*. Esta versión quedó constituida por los 10 ítems directos de autoengaño (α de Cronbach = .76) y los 10 ítems directos de mane-

jo de la impresión (α de Cronbach = .71), con los que se logra mejorar la consistencia interna total (α de Cronbach = .77). El modelo de dos factores correlacionados en que se fundamenta la escala mostró buen ajuste a los datos por mínimos cuadrados generalizados ($\chi^2/gl = 1.74$, RMSEA = .03, GFI = .95 y AGFI = .94). Las puntuaciones en la escala y los dos factores se obtienen por suma simple. A mayor puntuación, mayor tendencia a inflar el autoinforme en un sentido socialmente deseable (Moral et al., 2012).

En la presente muestra, los valores de consistencia interna por el coeficiente α ordinal y el coeficiente α de Cronbach variaron de muy altos a altos (Tabla 2).

Procedimiento

El muestreo de rutas empleado consistió en seleccionar aleatoriamente 81 calles a partir de la

Tabla 2

Consistencia interna en la muestra conjunta, de mujeres y de hombres.

Escalas y factores	Conjunta (N = 807)		Mujeres (N = 431)		Hombres (N = 376)	
	α	α	α	α	α	α
	ordinal	Cronbach	ordinal	Cronbach	ordinal	Cronbach
Escala de búsqueda de sensaciones sexuales (SSSS-9)						
Total	.994	.972	.994	.972	.993	.969
Infidelidad sexual	.993	.985	.994	.984	.992	.985
Deseo inf. emoc.	.985	.974	.985	.971	.984	.975
Deseo inf. sexual	.989	.978	.991	.977	.987	.977
Infidelidad emoc.	.972	.941	.972	.935	.961	.942
Escala de actitud hacia la exclusividad marital						
Total	.968	.938	.974	.940	.961	.934
Inventario balanceado de deseabilidad social						
Total	.883	.865	.888	.869	.879	.861
Autoengaño	.849	.823	.852	.828	.847	.819
M. de impresión	.853	.827	.853	.825	.856	.832

Guía Roji. De cada calle se recolectaron 10 unidades (con consentimiento informado y datos completos). Se pretendía que fueran cinco mujeres y cinco hombres independientes (uno por casa). Si la persona deseaba participar voluntariamente y satisfacía los criterios de inclusión, se dejaba un cuestionario para recogerlo media hora después o cuando lo indicase la persona.

Para determinar el tamaño de la muestra se usó la fórmula para la estimación bilateral de una media con población infinita, al ser el tamaño poblacional mayor que 100,000. Bajo un modelo de curva normal, si se desea estimar la media de la SSSS-9 con una desviación estándar esperada de 0.518 (Kalichman et al., 1994), un error absoluto de estimación de .036 y un intervalo de confianza del 95 %, se requeriría una muestra mínima de 796 participantes.

El estudio fue aprobado en sus aspectos éticos por el Comité de Doctorado, autoridad concerniente dentro de la institución académica, al ser sometido de forma extraordinaria al juicio de éste para cumplimentar requisitos del programa de financiamiento. Fue financiado por el Programa de Apoyo a la Investigación Científica y Tecnológica. Se solicitó el consentimiento expreso informado de los participantes en la primera hoja del cuestionario. En la hoja de consentimiento se garantizó el anonimato de las respuestas y se informó sobre la identidad del responsable del estudio, con quien se podía contactar por correo electrónico por cualquier cuestión suscitada por la investigación. No se solicitó ningún dato de identificación personal. Así, se respetaron las normas éticas de investigación de la [Asociación Americana de Psicología \(2017\)](#).

Análisis de datos

En relación con el primer objetivo, la con-

sistencia interna de los ítems se estableció por la correlación poliserial (método de máxima verosimilitud de dos pasos) del ítem con el resto de la escala y la disminución del coeficiente alfa ordinal de la escala al ser calculado sin incluir al ítem. Si la correlación era menor que .30 y el valor del coeficiente alfa ordinal se incrementaba, se consideró que el ítem presentaba problemas de consistencia interna. La discriminabilidad se comprobó por la diferencia significativa de tendencia central entre el grupo de puntuaciones bajas (menores o iguales al cuartil inferior) y el de puntuaciones altas (mayores o iguales al cuartil superior) usando la prueba U de Mann-Whitney.

En relación con el segundo objetivo de determinar el número de factores, dicha determinación se hizo a través de la convergencia del análisis paralelo de Horn, coordenadas óptimas, factor de aceleración, promedio mínimo de las correlaciones parciales al cuadrado (criterio de Velicer) y a cuarta potencia (criterio de Velicer revisado) y autovalores de la matriz de correlaciones mayores que uno (criterio de Kaiser). Todos estos criterios se calcularon desde la matriz de correlaciones policóricas.

El método de análisis paralelo se implementó generando 1000 matrices de datos aleatorios a través de la permutación de los datos empíricos, por lo que se conservan los 807 casos y las nueve variables. A partir de las mismas se obtuvieron 1000 matrices de correlación y se calcularon los nueve autovalores de cada matriz de correlación. De este modo, se obtuvo una distribución para cada autovalor. Se calculó el percentil 95 de la distribución de cada autovalor. Los autovalores empíricos (de la matriz de correlación policórica original) que quedaron por encima de estos percentiles se consideran sustantivos y aquellos que quedaron por debajo se consideran espurios. El análisis de coordenadas óptimas compara las pendientes de los autovalores retenidos por el análisis

paralelo de Horn. Requiere que estos gradientes o pendientes sean descendentes. En caso de que un gradiente sea ascendente, indica un cambio anómalo de curvatura y se retienen sólo los autovalores previos con gradientes descendentes. El factor de aceleración, desde los gradientes de cada autovalor, busca el cambio más brusco de pendiente y retiene los autovalores que quedan por encima de ese punto de cambio brusco de curvatura. Estos dos últimos métodos son variantes matematizadas del test de Cattell, el cual se basa en la búsqueda visual del punto de cambio de curvatura en la gráfica de sedimentación de los autovalores de la matriz de correlación de los datos empíricos. En la prueba de Velicer, los factores sustantivos se buscan parcializando los factores que fueron extraídos por análisis de componentes principales; se calculan sucesivas matrices de correlación parcial y el promedio de las correlaciones fuera de la diagonal principal elevadas al cuadrado o a la cuarta potencia. Un promedio mayor que el obtenido en el cálculo anterior indica que no queda varianza compartida y que se está extrayendo un factor espurio (Courtney, 2013).

En relación con los objetivos tercero y cuarto de comprobar el ajuste del modelo a los datos y su invarianza entre ambos sexos, ambos contrastes se hicieron por mínimos cuadrados no ponderados desde la matriz de correlaciones policóricas. Los errores estándar de las estimaciones se calcularon por el método percentiles corregidos de sesgo (con la extracción de 2,000 muestras aleatorias). Se estableció la validez convergente de cada factor (varianza de los indicadores explicada por su factor mayor que la explicada por otros factores no atribuibles o azar) por la varianza media extraída (AVE) y el coeficiente omega (ω) de McDonald. Valores de AVE > .50 y $\omega \geq .70$ muestran validez convergente. A su vez, se estableció la validez discriminante entre dos factores por una varianza compartida menor que dos ter-

cios y menor que la AVE de cada factor. Se consideraron nueve índices de ajuste: χ^2/gl = chi-cuadrada relativa o cociente entre el valor mínimo de la función de discrepancia y los grados de libertad del modelo, GFI = índice de bondad de ajuste de Jöreskog y Sörbom, AGFI = índice de bondad de ajuste corregido de Jöreskog y Sörbom, CFI = índice comparativo de ajuste de Bentler, NFI = índice de ajuste normado de Bentler y Bonett, NNFI = índice de ajuste no normado de Bentler y Bonett, RFI = índice de ajuste relativo de Bollen, RMSEA = error de aproximación cuadrático medio de Steiger y Lindt y SRMR = error estandarizado cuadrático medio. A su vez, se estimó la parsimonia del modelo por la razón de parsimonia (PR) de James, Mulaik y Brett. Siguiendo a Byrne (2016), se estipularon como valores de buen ajuste: $\chi^2/gl \leq 2$, GFI, CFI, NFI, NNFI y RFI $\geq .95$, AGFI $\geq .90$, y RMSEA y SRMR $\leq .05$; y aceptable: $\chi^2/gl \leq 3$, GFI, CFI, NFI; NNFI y RFI $\geq .90$, AGFI $\geq .85$, RMSEA < .08 y SRMR < .10. Una PR de .50 a .749 se interpretó como parsimonia alta y de .75 a 1 muy alta.

En relación con el quinto objetivo de calcular la consistencia interna de la escala y sus factores, esta se calculó a través del coeficiente alfa ordinal; valores mayores que .70 se consideraron aceptables y mayores que .80 altos.

En relación con el sexto objetivo de describir la distribución y su diferencia de tendencia central entre ambos sexos, el ajuste de la distribución a una curva normal se contrastó por la prueba de Kolmogorov-Smirnov-Lilliefors. Las medias se compararon por la prueba *t* de Student usando el método repetitivo (con la extracción de 1000 muestras aleatorias) para estimar el error estándar y el intervalo de confianza de la diferencia media por el incumplimiento del supuesto de distribución normal. El tamaño del efecto del sexo sobre la escala se determinó por la *g* de Hedges corregida de sesgo. Valores de *g* menores que .20

se interpretaron como un tamaño del efecto trivial, de .20 a .49 pequeño, de .50 a .79 mediano y mayor o igual que .80 grande (Téllez, García, & Corral-Verdugo, 2015).

En relación con los objetivos séptimo y octavo de comprobar la validez de constructo concurrente y el sesgo por deseabilidad social, las correlaciones con las escalas de conducta infiel, actitud hacia la exclusividad marital y deseabilidad social se calcularon por la correlación producto-momento de Pearson. Debido al incumplimiento de normalidad se usó el método de muestreo repetitivo (con la extracción de 1000 muestras aleatorias) para estimar los errores estándar e intervalos de confianza. Se compararon las correlaciones entre ambos sexos por la prueba Z de Fisher. También se usó el método de muestreo repetitivo (con la extracción de 1000 muestras) para estimar los errores estándar e intervalos de confianza de las diferencias medias.

Tabla 3

Distribución, discriminabilidad y consistencia interna de los ítems.

Ítem	Porcentaje				Q_1	Q_2	Q_3	R_{SIQ}	A_{IQ}	Discr.	Consistencia		
	1	2	3	4							Z_U	$CPS_{i,t}$	$CPS_{i,t-i}$
1	38.0	34.6	16.5	10.9	1	2	3	1	0	-16.27***	.75	.66	.925
2	78.6	15.2	4.7	1.5	1	1	1	0	I	-11.71***	.71	.65	.923
3	67.2	18.6	7.2	7.1	1	1	2	0.5	1	-14.15***	.74	.66	.923
4	31.5	27.8	22.9	17.8	1	2	3	1	0	-17.07***	.80	.71	.921
5	54.5	29.9	9.5	6.1	1	1	2	0.5	1	-15.07***	.74	.66	.923
6	77.9	14.4	4.3	3.3	1	1	1	0	I	-11.89***	.73	.66	.922
7	45.5	28.1	14.7	11.6	1	2	3	1	0	-17.09***	.83	.75	.918
8	29.1	38.0	18.1	14.7	1	2	3	1	0	-16.44***	.81	.73	.922
9	25.9	32.6	21.3	20.2	1	2	3	1	0	-17.68***	.86	.79	.918

Nota. N = 807. Valores de respuesta de los ítems: 1 = nada, 2 = algo, 3 = bastante y 4 = muchísimo.

Cuartiles: primero (Q^1), segundo o mediana (Q^2) y tercero (Q^3), R_{SIQ} = rango semiintercuartílico, A_{IQ} = asimetría intercuartílica. Discriminabilidad: Estadístico estandarizado de la prueba U de Mann-Whitney *** $p < .0001$ (probabilidad de la hipótesis nula de igualdad de rango medio entre los grupos de puntuaciones altas y bajas en la escala). Consistencia interna: $CPS_{i,t}$ = correlación poliserial del ítem con la escala, $CPS_{i,t-i}$ = correlación poliserial del ítem con el resto de la escala, Ord. α_{t-i} = alfa ordinal eliminado el ítem, siendo la Ord. $\alpha_t = .930$.

Los contrastes fueron bilaterales con un nivel de significación de .05. Los cálculos se hicieron con SPSS22, módulo R4.2 para SPSS22, AMOS16 y Excel 2013.

Resultados

Distribución, discriminabilidad y consistencia interna de los ítems

En la muestra total, los nueve ítems fueron discriminantes. El efecto de este poder discriminativo fue grande (r^2 de Rosenthal de .25 a .39) en siete ítems y mediano en los ítems 2 ($r^2 = .17$) y 6 ($r^2 = .18$). A su vez presentaron consistencia interna, como reflejo de las correlaciones de cada ítem con el resto de la escala que variaron de altas ($CPS_{i,t-i} = .65$) a muy altas ($CPS_{i,t-i} = .79$), con una media de .70 y un descenso en el valor de la consistencia interna de la escala al ser eliminado el ítem (Tabla 3).

Determinación del número de factores

El análisis paralelo de Horn, el de coordenadas óptimas y el del factor de aceleración convergieron en uno tanto en la muestra total como en las de mujeres y hombres, lo que concuerda con la hipótesis de unidimensionalidad. El criterio de Kaiser indicó un factor en la muestra total, pero dos en las muestras de mujeres y hombres. El número de factores fue de dos por los criterios de Velicer en las tres muestras.

Exploración de la estructura de dos factores

Al extraer dos factores en la muestra total, se explicó el 62.3 % de la varianza total. El primer factor quedó configurado por tres indicadores (ítems 2, 5 y 6) sobre conductas deshonestas (mentir) y con baja aceptación social (ver pornografía), las cuales requerirían una mayor sinceridad para su reconocimiento. Tuvo consistencia interna alta (α ordinal = .88) y validez convergente (AVE = .74 y ω = .89). El segundo factor quedó conformado por seis indicadores (ítems 1, 3, 4, 7,

Tabla 4

Correlación de los ítems de la SSSS_9 con deseabilidad social.

SSSS-9	BIDR			MI			AE		
	CPS	Wald		CPS	Wald		CPS	Wald	
	(CI 95 %)	W	p	(CI 95 %)	W	p	(CI 95 %)	W	p
1	-.09 (-.16, -.01)	5.49	.019	-.20 (-.27, -.13)	28.93	<.001	.07 (-.01, .15)	3.41	.065
2	-.28 (-.36, -.20)	43.81	<.001	-.34 (-.42, -.26)	73.10	<.001	-.12 (-.21, -.03)	6.76	.009
3	-.18 (-.26, -.10)	19.92	<.001	-.24 (-.32, -.16)	35.40	<.001	-.06 (-.14, .03)	1.82	.177
4	-.08 (-.15, -.01)	4.11	.043	-.21 (-.28, -.14)	34.03	<.001	.11 (.03, .18)	8.08	.004
5	-.27 (-.34, -.20)	52.46	<.001	-.33 (-.39, -.26)	86.22	<.001	-.11 (-.19, -.04)	8.40	.004
6	-.32 (-.40, -.24)	64.40	<.001	-.36 (-.44, -.28)	81.45	<.001	-.18 (-.26, -.09)	16.00	<.001
7	-.17 (-.24, -.09)	19.78	<.001	-.26 (-.33, -.19)	50.96	<.001	-.01 (-.09, .07)	0.07	.798
8	-.08 (-.15, -.01)	4.91	.027	-.20 (-.27, -.13)	32.11	<.001	.09 (.02, .17)	5.99	.014
9	-.10 (-.18, -.03)	7.06	.008	-.22 (-.29, -.15)	36.00	<.001	.07 (-.01, .15)	3.49	.062

Nota. CPS = correlación poliserial, IC 95 % = intervalo de confianza al 95 %, W = estadístico W de Wald, p = probabilidad exacta de W desde una distribución chi-cuadrado con 1 grado de libertad en un contraste unilateral derecho para la hipótesis nula de $r_{ps} = 0$. SSSS-9 = escala de búsqueda de sensaciones sexuales, BIDR = inventario balanceado de deseabilidad social, MI = manejo de la impresión, y AE = autoengaño.

8 y 9) sobre la búsqueda de nuevas experiencias y sensaciones sexuales. Su consistencia interna fue muy alta (α ordinal = .91) y mostró validez convergente (AVE = .64 y ω = .86). La varianza compartida entre los dos factores fue muy alta, 55.4 %, aunque fue menor que la varianza media extraída de cada uno de ellos, por lo que los dos factores presentaron validez discriminante. El resultado fue muy semejante al extraer dos factores en la muestra de mujeres y hombres, con varianzas compartidas entre los factores de 53.3 % y 51.8 %, respectivamente.

En apoyo de la interpretación dada al primer factor, los ítems 6, 2 y 5 presentaron las correlaciones más altas con manejo de la impresión y la puntuación total del BIDR y fueron los únicos ítems con correlaciones significativas y negativas con autoengaño, lo que concuerda con la afirmación de una mayor sinceridad para reconocer esas conductas deshonestas y con baja aceptación so-

cial (Tabla 4). A su vez, el factor conformado por estos tres ítems tuvo correlaciones más altas con deseabilidad social que el otro factor. Sus correlaciones fueron significativas y negativas con una fuerza de asociación moderada con manejo de la impresión, $r_{[807]} = -.32$, IC del 95 % (-.38, -.25), y baja con la puntuación total del BIDR, $r_{[807]} = -.27$, IC del 95 % (-.34, -.21), y autoengaño, $r_{[807]} = -.13$, IC del 95 % (-.20, -.06). El otro factor de seis ítems tuvo correlaciones significativas, negativas y con una fuerza de asociación baja con manejo de la impresión, $r_{[807]} = -.25$, IC del 95 % (-.31, -.18) y la puntuación total del BIDR, $r_{[807]} = -.12$, IC del 95 % (-.19, -.06), pero fue independiente del autoengaño, $r_{[807]} = .06$, IC del 95 % (-.01, .13).

Figura 1

Modelo de un factor en la muestra total.

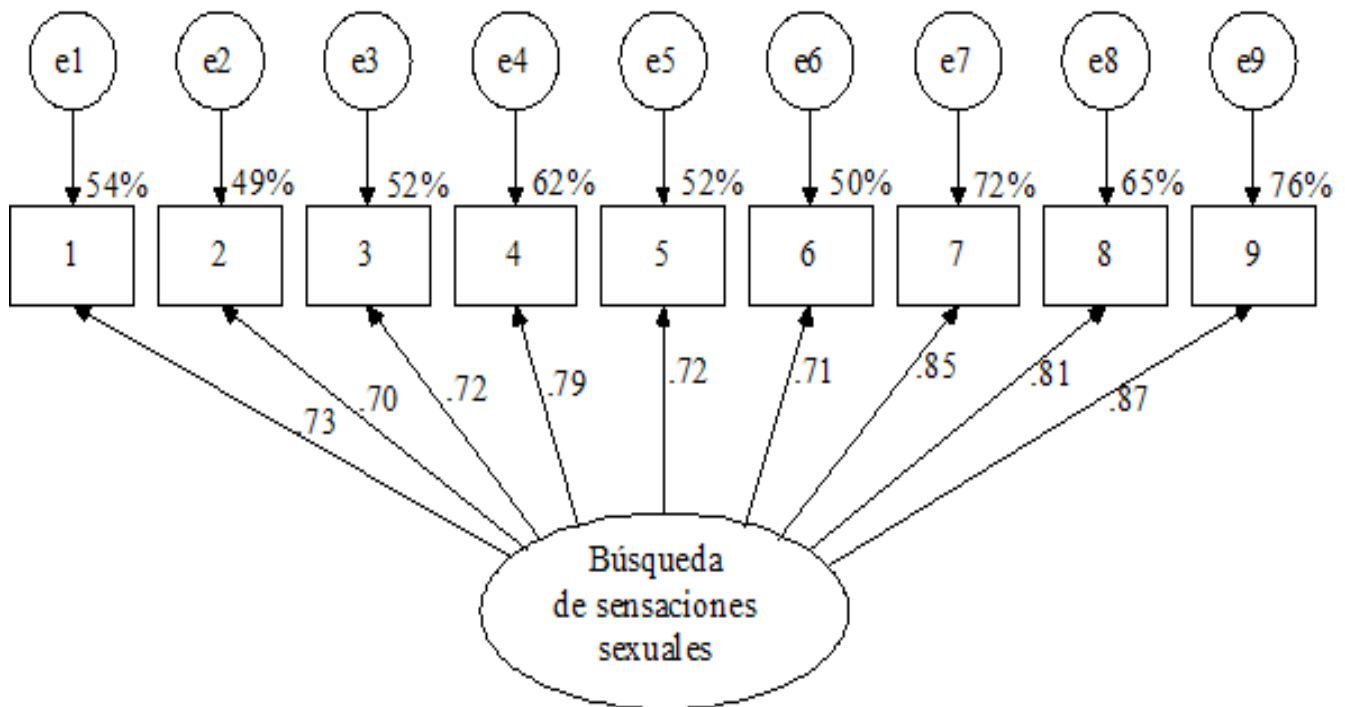


Tabla 5

Índices de ajuste del modelo de un factor y dos factores correlacionados en el contraste unigrupo (muestra total) y multigrupo (sexos).

Índices	Unigrupo				Multigrupo de 1F		
	1F	2F_AFE	2F_EXP	SR	CPM	CVE	CRM
χ^2	76.415	48.640	69.573	81.947	157.351	446.411	562.125
gl	27	26	26	54	62	63	72
χ^2/gl	2.830	1.871	2.676	1.518	2.538	7.086	7.807
GFI	.993	.996	.994	.992	.985	.957	.946
AGFI	.989	.993	.989	.987	.978	.938	.932
CFI	.994	.997	.995	.996	.987	.948	.933
NFI	.991	.994	.992	.989	.979	.940	.924
NNFI	.992	.996	.993	.995	.985	.940	.933
RFI	.988	.992	.988	.985	.975	.931	.924
RMSEA	.048	.033	.046	.025	.044	.087	.092
SRMR	.075	.054	.071	.088	.099	.174	.091
PR	.750	.722	.722				

Nota. Método: Mínimos cuadrados no ponderados. Datos de entrada: Matriz de correlaciones policóricas. Modelos: 1F = un factor con nueve indicadores, 2F_AFE = dos factores correlacionados desde los resultados del análisis factorial exploratorio: búsqueda de experiencias y sensaciones sexuales (ítems 1, 3, 4, 7, 8 y 9) y conductas censurables (ítems 2, 5 y 6), y 2F_EXP = dos factores correlacionados esperados desde los propuestos para la SSSS-11: búsqueda de sensaciones personales (ítems 1, 8 y 9) y búsqueda de experiencias dependientes de otras personas o estímulos externos (ítems 2, 3, 4, 5, 6 y 7). Contraste de un grupo: en la muestra total de 807 participantes. Contraste multigrupo: separando la muestra de 376 hombres y 431 mujeres y usando modelos anidados: SR = sin restricciones, CPM = con restricciones en los pesos de medida, CVE = con restricciones en la varianzas-covarianzas estructurales, y CRM = con restricciones en los residuos de medida. Índices de ajuste: χ^2 = valor mínimo de la función de discrepancia o chi-cuadrada, gl = grados de libertad de modelo, χ^2/gl = chi-cuadrada relativa, GFI = índice de bondad de ajuste de Jöreskog y Sörbom, AGFI = índice de bondad de ajuste corregido de Jöreskog y Sörbom, CFI = índice comparativo de ajuste de Bentler, NFI = índice de ajuste normado de Bentler y Bonett, NNFI = índice de ajuste no normado de Bentler y Bonett, RFI = índice de ajuste relativo de Bollen, RMSEA = error de aproximación cuadrático medio de Steiger y Lindt, SRMR = error estandarizado cuadrático medio y PR = razón de parsimonia de James, Mulaik y Brett.

Contraste de los modelos por análisis factorial confirmatorio

Se especificaron tres modelos recursivos con todos sus residuos independientes, uno de un factor con nueve indicadores, y dos de dos factores correlacionados, uno especificado a partir del análisis factorial exploratorio y otro a partir de expectativa conforme a los resultados con la SSSS-11.

En el modelo de un factor, todos los parámetros fueron significativos al ser estimados con un intervalo de confianza bilateral al 95 %. Los pesos de medida fueron mayores que .70 (Figura 1). La varianza media extraída fue de .59 y la confiabilidad compuesta por el coeficiente omega de McDonald fue .93, lo que indica que el

factor presentó validez convergente. El ajuste del modelo fue bueno por siete índices y aceptable por dos índices (χ^2/gl y SRMR), y su parsimonia fue muy alta (Tabla 5).

En el modelo de factores derivado del análisis factorial exploratorio, todos los parámetros fueron significativos. El factor de tres indicadores (AVE = .68 y ω = .90) y el de seis (AVE = .64 y ω = .92) mostraron validez convergente, pero carecieron de validez discriminante, al ser la varianza compartida mayor que dos tercios (r^2 = .67) y mayor que la varianza media extraída del factor de seis indicadores. El ajuste del modelo fue bueno por ocho índices y aceptable por uno (SRMR), y su parsimonia fue alta (Tabla 5 y Figura 2).

Igual ocurrió con el modelo especificado a partir de los dos factores correlacionados pro-

puestos para la SSSS-11 y reducido a los nueve ítems de la SSSS-9: búsqueda de sensaciones personales (ítems 1, 8 y 9) y búsqueda de experiencias dependientes de otras personas o estímulos externos (ítems 2, 3, 4, 5, 6 y 7). Sus parámetros fueron significativos, su ajuste fue bueno por siete índices y adecuado por dos (χ^2/gl y SRMR) y su parsimonia alta (Tabla 5). Ambos factores tuvieron validez convergente ($AVE = .685$ y $\omega = .866$ para el factor de tres ítems y $AVE = .644$ y $\omega = .906$ para el factor de seis ítems), pero care-

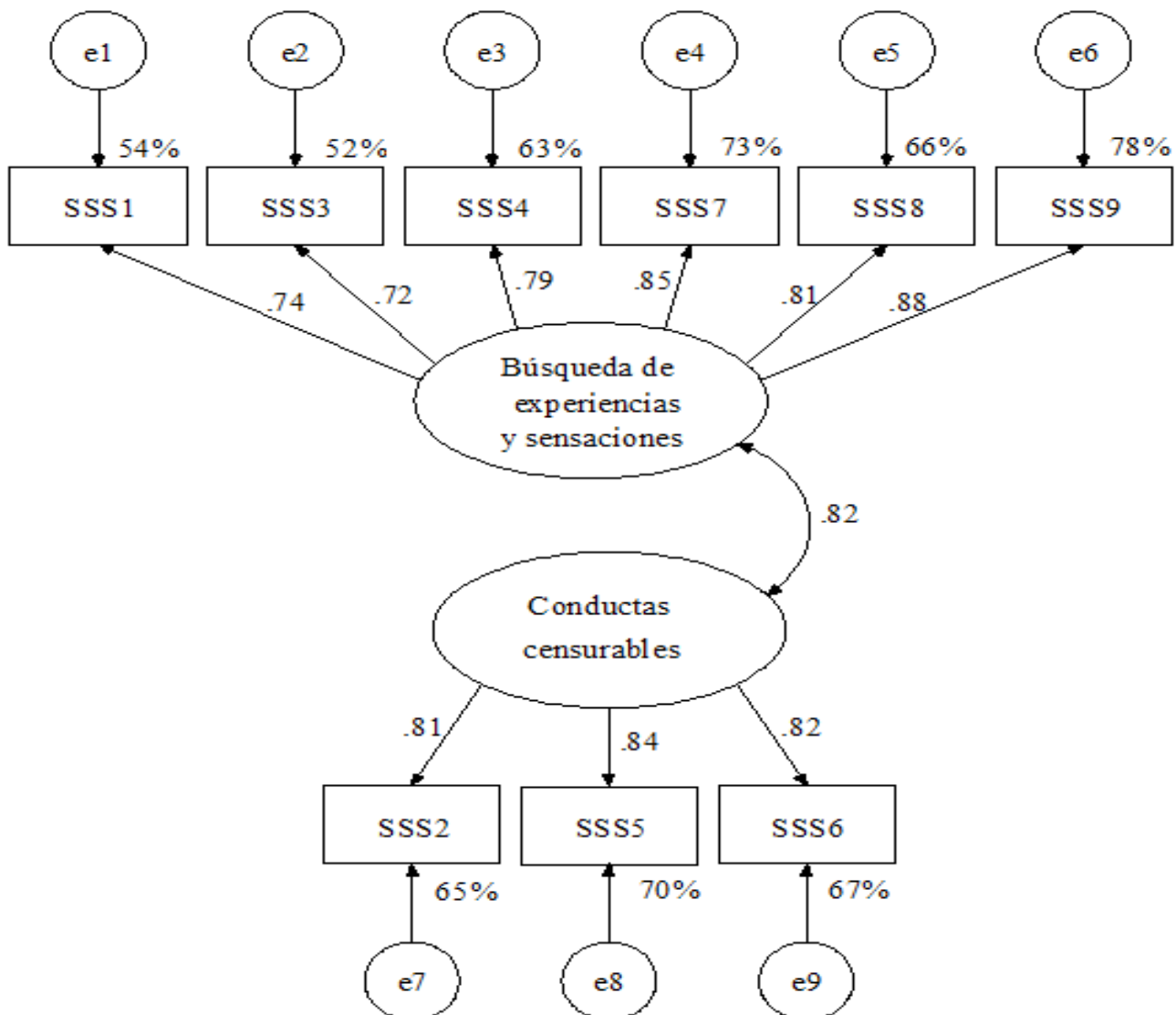
cieron totalmente de validez discriminante, al tener una varianza compartida del 94.1 %.

Contraste de la invarianza entre ambos sexos.

Al contrastar la invarianza del modelo unifactorial entre ambos sexos, los parámetros fueron significativos en los cuatro modelos anidados en las dos muestras, salvo la varianza del residuo de medida del ítem 6 en el modelo con restric-

Figura 2

Modelo de dos factores correlacionados, derivado del análisis factorial exploratorio, en la muestra total.



ciones en la varianza estructural en mujeres ($S^2_{\epsilon_6} = 03$, IC del 95 %: [-.01, .67], $p = .259$), al quedar este ítem totalmente explicado por el factor ($\lambda = .95$, IC del 95 % [.85 1.05], $p = .001$). En el modelo sin restricciones, los parámetros fueron equivalentes entre ambas muestras, salvo las varianzas de los residuos de los ítems 1, 2, 3 y 6, que fueron más altas en hombres; el ajuste fue bueno por todos los índices, salvo por SRMR, en el que fue aceptable. Al restringir los pesos de medida, las mujeres tuvieron valores más altos en los pesos de los ítems 3 y 6 y en la varianza del residuo del ítem 9; los hombres en los pesos de los ítems 4 y 9, la varianza del factor y las varianzas de los residuos de los ítems 2, 3, 5 y 6; empeoró la bondad de ajuste con respecto al modelo anterior ($\Delta\chi^2/\Delta gl = 9.426$, $\Delta CFI = .009$, $\Delta NFI = .010$, $\Delta NNFI = .010$ y $\Delta RFI = .010$), siendo esta buena por siete índices y aceptable por dos (χ^2/gl y SRMR). Al restringir adicionalmente la varianza estructural, los pesos de medida de los ítems 2, 3, 5, 6 y 7 fueron más altos en mujeres y las varianzas de los residuos 2, 3, 5, 6, 7 y 8 fueron más altas en hombres; empeoró la bondad de ajuste con respecto al anterior ($\Delta\chi^2/\Delta gl = 289.059$, $\Delta CFI = .039$, $\Delta NFI = .039$, $\Delta NNFI = .045$ y $\Delta RFI = .044$), siendo esta mala por tres índices (χ^2/gl , SRMR y RMSEA), adecuada por cuatro (NFI, NNFI, CFI y RFI) y buena por dos (GFI y AGFI). Al restringir adicionalmente las varianzas de los residuos de medida, esto es, al hacer los parámetros iguales entre ambas muestras, el ajuste fue malo por dos índices (χ^2/gl y RMSEA), adecuado por cinco (GFI, NFI, NNFI, CFI y RFI) y bueno por uno (AGFI). La bondad de ajuste empeoró con respecto al modelo anterior ($\Delta\chi^2/\Delta gl = 12.857$, $\Delta CFI = .015$, $\Delta NFI = .016$, $\Delta NNFI = .007$ y $\Delta RFI = .007$; Tabla 5).

Consistencia interna de la escala

Al carecer el modelo de dos factores de validez discriminante, se optó por el modelo de un factor, cuyas propiedades de validez convergente y ajuste a los datos fueron buenas y que mostró invarianza en sus pesos de medida entre ambos sexos en el modelo sin restricciones del contraste multigrupo.

La escala integrada por sus nueve ítems presentó una consistencia interna muy alta (α ordinal = .93 en la muestra total, .92 en mujeres y .93 en hombres). No obstante, si se calcula por el coeficiente alfa de Cronbach, la consistencia interna sería alta ($\alpha = .89$ en la muestra total y de hombres y .86 en la de mujeres).

Distribución de la SSSS-9 y diferencia de medias entre ambos sexos

La distribución de la puntuación total de la SSSS-9 mostró asimetría positiva y no se ajustó a una distribución normal en la muestra total ni en las muestras de mujeres y hombres. La media quedó por encima de la mediana, con mayor concentración de puntuaciones por debajo de la media que por encima, esto es, mayor concentración en puntuaciones bajas (Tabla 6).

La media de los hombres ($M = 2.05$, IC del 95 %: 1.97, 2.12) fue significativamente mayor ($t_{[705.11]} = 8.21$, $p < .001$; sin asumir igualdad de varianza con base en la prueba de Levene: $F_{[1, 805]} = 35.79$, $p < .001$) que la de las mujeres (1.66, IC del 95 %: 1.61, 1.72) con una diferencia media de 0.39 (IC del 95 %: 0.29, 0.48). El tamaño del efecto del sexo sobre la búsqueda de sensaciones sexuales fue mediano ($g = 0.59$, IC del 95 %: 0.45, 0.73).

Tabla 6

Distribución de la SSSS-9 en la muestra total, de mujeres y de hombres.

Estadísticos	Mujeres	Hombres	Total
N	431	376	807
Mínimo	1	1	1
Máximo	3.89	4	4
Media	1.662	2.048	1.842
	[1.605, 1.716]	[1.972, 2.123]	[1.796, 1.892]
Desviación estándar	0.575	0.737	0.683
Asimetría	1.009	0.491	0.804
	[0.777, 1.224]	[0.328, 0.663]	[0.672, 0.943]
Curtosis	0.751	-0.559	0.011
	[0.001, 1.534]	[-0.841, -0.194]	[-0.327, 0.390]
Percentil 10	1	1.111	1
Percentil 20	1.111	1.333	1.222
Percentil 25	1.222	1.444	1.333
Percentil 30	1.333	1.556	1.333
Percentil 40	1.422	1.778	1.556
Percentil 50	1.556	2	1.667
Percentil 60	1.667	2.111	1.889
Percentil 70	1.889	2.444	2.111
Percentil 75	2	2.667	2.222
Percentil 80	2.111	2.778	2.444
Percentil 90	2.444	3.111	2.889
$ D_{\max} $	0.137***	0.103***	0.121***

Nota. Intervalos de confianza del 95 %, creados por el procedimiento de remuestreo con la extracción de 1000 muestras aleatorias. $|D_{\max}|$ = Estadístico de contraste de la prueba de Kolmogorov-Smirnov. La probabilidad fue calculada con las tablas de Lilliefors. *** $p < .001$ (a una cola).

Validez de constructo concurrente y efecto de la discapacidad social

En la muestra conjunta, la correlación de la SSSS-9 fue significativa, positiva y alta con la puntuación total de la escala de conducta infiel y sus factores de infidelidad sexual, deseo de infidelidad emocional y deseo de infidelidad sexual. Fue moderada con el factor de infidelidad emo-

cional y la escala de actitud hacia la exclusividad marital. Estas seis correlaciones fueron estadísticamente equivalentes entre ambos sexos por la prueba de Fisher (Tabla 7).

En la muestra conjunta, la correlación de la SSSS-9 fue significativa, negativa y baja con la puntuación total de la escala de discapacidad social y su factor de manejo de la impresión, y fue independiente del factor de autoengaño. Estas

Tabla 7

Correlaciones de la SSSS-9 con las escalas de actitud hacia la exclusividad marital, conducta infiel y deseabilidad social.

Escalas	C	SSSS-9			Fisher		
		Mujeres (n = 431)	Hombres (n = 376)	Total (n = 807)	DM	Z	p
Actitud hacia exclusividad marital (infidelidad)	r	.413*** (.316, .501)	.512*** (.430, .590)	.491*** (.432, .550)	-0.126 (-.265, .013)	-1.782	.075
	r_p	.373*** (.279, .473)	.474*** (.378, .560)	.451*** (.387, .513)			
Conducta infidel	r	.538*** (.456, .613)	.558*** (.471, .630)	.578*** (.521, .630)	-.029 (-.167, .110)	-0.404	.687
	r_p	.506*** (.424, .594)	.522*** (.428, .607)	.544*** (.486, .598)			
Infidelidad sexual	r	.472*** (.379, .560)	.516*** (.427, .593)	.528*** (.461, .582)	-.058 (-.197, .081)	-0.822	.411
	r_p	.440*** (.346, .540)	.485*** (.385, .569)	.496*** (.431, .556)			
Deseo de infidelidad emocional	r	.540*** (.459, .618)	.561*** (.476, .631)	.578*** (.523, .626)	-.030 (-.169, .109)	-0.425	.671
	r_p	.506*** (.425, .588)	.521*** (.425, .605)	.541*** (.483, .599)			
Deseo de infidelidad sexual	r	.517*** (.420, .600)	.542*** (.449, .615)	.562*** (.500, .616)	-.035 (-.174, .104)	-0.491	.624
	r_p	.490*** (.405, .588)	.503 (.401, .586)	.529*** (.469, .589)			
Infidelidad emocional	r	.396*** (.277, .501)	.448*** (.364, .530)	.453*** (.387, .514)	-.063 (-.202, .076)	-0.894	.372
	r_p	.367*** (.259, .476)	.418*** (.371, .501)	.423*** (.350, .488)			
Deseabilidad social	r	-.150** (-.238, -.060)	-.198*** (-.295, -.103)	-.179*** (-.248, -.113)	.050 (-.089, .188)	0.699	.485
	r	.054 ^{ns} (-.037, .143)	-.032 ^{ns} (-.137, .070)	.010 ^{ns} (-.063, .085)	.086 (-.053, .225)	1.215	.224
Manejo Impresión	r	-.283*** (-.367, -.192)	-.279*** (-.370, -.185)	-.289*** (-.350, -.226)	-.004 (-.143, .134)	-0.061	.951

Nota. C = coeficiente de correlación: r = coeficiente de correlación producto-momento de Pearson y r_p = coeficiente de correlación parcial de Fisher, parcializando el manejo de la impresión. DM = diferencia media. Intervalos de confianza estimados con la extracción de 1,000 muestras aleatorias. Significación en un contraste bilateral por la prueba Z de Fisher: ns = $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$.

tres correlaciones fueron estadísticamente equivalentes entre ambos sexos (Tabla 7).

Al parcializar el manejo de la impresión, las correlaciones de SSSS-9 con la escala de conducta infiel y sus cuatro factores y la escala de actitud hacia la exclusividad marital siguieron siendo significativas con fuerzas de asociación de altas a moderadas, disminuyendo en cuatro o menos centésimas (Tabla 7).

Discusión

Como primer objetivo se planteó comprobar la consistencia interna, discriminabilidad y variabilidad de los ítems. Como en estudios previos (De Oliveira-Santos et al., 2017; Kalichman et al., 1994; Teva & Paz-Bermúdez, 2008), los ítems mostraron consistencia interna. Asimismo, tuvieron capacidad para diferenciar al grupo de puntuaciones altas y bajas en la escala con un poder discriminativo grande en siete de los nueve ítems y mediano en dos, lo que anteriormente no se había comprobado. Además, ninguno presentó efecto suelo (80 % o más de los casos en el primer valor) o techo (80 % o más de los casos en el último valor) y la variabilidad fue alta en la mayoría de los ítems. Conforme con la expectativa (Gadermann et al., 2014), los indicadores de consistencia interna fueron más altos que en estudios previos por el uso de un método más adecuado a la naturaleza ordinal de los ítems. A su vez, esta mejora en los índices se puede atribuir a la unidimensionalidad que presenta la versión original de la escala frente al modelo de dos factores de la versión revisada (Ballester-Arnal et al., 2018; De Oliveira-Santos et al., 2017; Santos-Iglesias et al., 2018). Los ítems 2 y 3, eliminados en la revisión de la escala (Kalichman & Rompa, 1995), no resultaron débiles en su consistencia interna. Su nivel fue semejante a los ítems 1, 5 y 6, siendo

los más fuertes en esta propiedad los cuatro ítems restantes, los ítems 4, 7, 8 y 9.

Se enunció como segundo objetivo determinar el número de factores. Conforme con la expectativa de Kalichman et al. (1994), los tres criterios empíricos más válidos en los estudios de simulación convergieron en un factor, el cual quedó definido por pesos de medida mayores que .70, lo que proporciona garantías de que es el verdadero número de factores (Courtney, 2013).

No obstante, el criterio de Velicer, considerado también válido, junto con el criterio de Kaiser, considerado el menos válido, convergieron en dos, conforme al número de factores que presenta la versión revisada de la escala (Ballester-Arnal et al., 2018; De Oliveira-Santos et al., 2017; Santos-Iglesias et al., 2018). Al explorar el modelo bidimensional, se configuraron dos factores consistentes y con validez convergente, aunque con una varianza compartida muy alta. Estos dos factores difieren en contenido de los obtenidos con la SSSS-11. No separan a los ítems sobre sensaciones con más acento en lo personal de los ítems sobre experiencias dependientes de otras personas o estímulos externos, sino que separan a los ítems con más sesgo de deseabilidad social por su contenido (mentir y ver pornografía) de los ítems con menos potencial de rechazo social, como muestra la relación de los ítems con el BIDR y sus factores de manejo de la impresión y autoengaño.

El tercer objetivo propuesto fue contrastar los modelos y el cuarto comprobar la invarianza de los mismos entre mujeres y hombres. El ajuste a los datos del modelo unidimensional fue bueno por mínimos cuadrados no ponderados, el cual es un método más adecuado para variables ordinales que el de máxima verosimilitud (Byrne, 2016), como se emplea en otros estudios (Ballester-Arnal et al., 2018; De Oliveira-Santos et al., 2017; Santos-Iglesias et al., 2018). El modelo de un factor fue válido para mujeres y hombres por

la significación de las estimaciones, el tamaño del efecto muy grande del factor sobre los ítems y el buen ajuste del modelo sin restricciones y con restricciones en los pesos de medida (Byrne, 2016). No obstante, el modelo no fue estrictamente invariante. Para poder hablar de invarianza, la estimación por intervalo (al 95 %) de cada parámetro (peso de medida, error de medida y varianza del factor) debería solaparse en la muestra de mujeres y hombres en los cuatro modelos anidados en restricciones, las diferencias entre los valores mínimos de la función de discrepancia entre los cuatro modelos anidados deberían ser estadísticamente equivalentes con un nivel de significación unilateral de .05, el cociente entre la diferencia de los valores mínimos de la función de discrepancia y la diferencia de sus grados de libertad ($\Delta\chi^2/\Delta gl$) debería ser menor que 2, y las diferencias entre los índices de ajuste relativos (ΔGFI , $\Delta AGFI$, ΔNFI y ΔRFI) menores que .01 (Byrne, 2016).

También se contrastó el modelo de dos factores revelado por el análisis factorial exploratorio. Aunque su ajuste e invarianza entre ambos sexos fue mejor que el modelo de un factor, el análisis factorial confirmatorio evidenció claramente el problema de validez discriminante, o exceso de varianza compartida entre los factores, que ya parecía entreverse en el análisis factorial exploratorio. Un modelo más complejo, como el modelo de dos factores correlacionados, en comparación con un modelo más simple y parsimonioso como es el de un factor, suele mostrar mejores índices de ajuste, por lo que es muy importante comprobar la validez discriminante entre los factores (Byrne, 2016; Kóbor, Takács, & Urbán, 2013), lo que falló con los nueve ítems de la SSSS-9, de ahí que se descartó.

En el caso del modelo de dos factores esperado desde las propuestas para la SSSS-11 (Ballester-Arnal et al., 2018; De Oliveira-Santos et al., 2017; Santos-Iglesias et al., 2018), la falta

de validez discriminante fue mucho más fuerte, aparte de no ser revelado por el análisis factorial exploratorio, lo que indica que es totalmente forzado para los nueve ítems de la SSSS-9, de ahí que también se descartó.

Se formuló como quinto objetivo comprobar la consistencia interna de la SSSS-9. En la presente muestra de adultos mexicanos casados o cohabitantes con parejas del sexo contrario, la consistencia interna de los nueve ítems de la SSSS-9 fue muy alta, al ser calculada por el coeficiente alfa ordinal, tanto en la muestra conjunta como al separar mujeres y hombres. Una consistencia interna muy alta apunta hacia un único factor subyacente (Sijtsma, 2015), como revelaron el análisis paralelo de Horn, coordenadas óptimas y factor de aceleración, por lo que este resultado apoya a la conclusión de unidimensionalidad. No obstante, con esta misma versión en una muestra de 106 hombres con orientación homosexual, Kalichman et al. (1994) obtuvieron una consistencia interna aceptable. Cabe señalar que la consistencia interna de la SSSS, calculada por el alfa de Cronbach y con independencia de la versión usada, es menor en las muestras de hombres que tienen sexo con hombres (Kalichman et al., 1994; primer estudio de Kalichman & Rompa, 1995; Valdez et al., 2016), lo que resulta en valores aceptables; por el contrario, es mayor en población general (segundo estudio de Kalichman & Rompa, 1995) y estudiantes universitarios (De Oliveira-Santos et al., 2017; Santos-Iglesias et al., 2018). En estas últimas poblaciones se alcanzan valores altos, como en este estudio con personas heterosexuales casadas o que cohabitan.

Se enunció como sexto objetivo describir la distribución de la SSSS-9. La distribución esperada para rasgos de personalidad es la distribución normal (Lyon, 2012). Este no fue el caso de la SSSS-9. Su distribución se desvió de la normalidad debido a la presencia de asimetría positiva.

Las colas de la distribución fueron asimétricas y la cola larga apareció hacia la derecha, lo que provocó que haya más casos por debajo de la media aritmética que por encima. Esto indica que hay un grupo pequeño de personas con un rasgo de búsqueda de sensaciones sexuales muy definido o alejado del promedio que cae en lo que podría calificarse como conducta sexual compulsiva (Kalichman et al., 1994). Si se observa el perfil del histograma, este no fue acampanado, sino que tuvo forma de escalera descendiente, aproximándose más al perfil de las escalas que evalúan rasgos psicopatológicos, como las escalas clínicas del Inventario Multifásico de Personalidad de Minnesota (Ben-Porath & Tellegen, 2016). Por tanto, es mejor realizar la baremación de la escala a través de medidas de posición o percentil en lugar de usar puntuaciones T basadas en la media aritmética y desviación estándar (Ben-Porath & Tellegen, 2016). Dentro de esta población de adultos mexicanos casados o cohabitantes con parejas del sexo contrario que residen en Monterrey, una puntuación mayor o igual de 2.11 en mujeres y 2.78 en hombres (percentil 80) indica que el rasgo de la búsqueda de sensaciones sexuales está muy definido o destacado.

Un rasgo de búsqueda de sensaciones más definido en hombres que en mujeres es un hecho bien establecido (Cross et al., 2013), lo que aplica al constructo de la búsqueda de sensaciones sexuales (Kalichman & Rompa, 1995) y se acentúa por la mayor libertad sexual que goza el hombre en la mayoría de las culturas, incluida la occidental (Ballester-Arnal et al., 2018; De Oliveira-Santos et al., 2017; Kalichman & Rompa, 1995; Santos-Iglesias et al., 2018). Los presentes datos con la versión de nueve ítems confirmaron esta expectativa. Además, mostraron que el tamaño del efecto del sexo sobre la búsqueda de sensaciones sexuales es mediano, como también se ha observado en las versiones de 10 ítems (Gaither

& Sellbom, 2003) y de 11 ítems (Ballester-Arnal et al., 2018; De Oliveira-Santos et al., 2017; Kalichman & Rompa, 1995; Santos-Iglesias et al., 2018; Teva & Paz-Bermúdez, 2008).

Se propuso como séptimo objetivo comprobar la validez de constructo concurrente en relación con la conducta infiel y actitud hacia la exclusividad marital. Se confirmaron las expectativas de validez de constructo concurrente (Kogan et al., 2015). Una mayor búsqueda de sensaciones actuó como un factor de riesgo con un tamaño del efecto grande sobre desear y tener parejas sexuales concurrentes y desear emocionalmente otras parejas. Su tamaño del efecto fue moderado con conductas de implicación emocional con otras parejas y actitud favorable hacia la concurrencia de parejas. Aunque estas correlaciones aparentemente fueron más altas en hombres que en mujeres, al contrastar la diferencia con la prueba Z de Fisher para datos emparejados de dos muestras independientes y calcular el intervalo de confianza de la diferencia entre las correlaciones por un procedimiento paramétrico de remuestreo, las diferencias resultaron nulas, es decir, no significativas. La búsqueda de sensaciones sexuales es un rasgo más definido en hombres, pero su fuerza de asociación con la actitud hacia la exclusividad marital y conducta infiel es equivalente en hombres y mujeres, como previamente habían reportado Gaither y Sellbom (2003) con permisividad sexual.

El octavo objetivo fue observar el efecto de la deseabilidad social sobre la SSSS-9. Esta variable de sesgo tuvo un efecto pequeño sobre la búsqueda de sensaciones sexuales debido a su factor de manejo de la impresión, sin resultar este sesgo diferencial entre ambos sexos. La varianza compartida entre el manejo de la impresión y la búsqueda de sensaciones sexuales fue menor del 9 %. Al parcializar el efecto del manejo de la impresión en las correlaciones para establecer la

validez de constructo concurrente, estas siguieron siendo significativas y descendieron en un máximo de cuatro centésimas, lo que indica la sustantividad de las mismas y fortalece la prueba de validez de constructo concurrente.

Una primera limitación del presente estudio es que las inferencias sólo son aplicables a adultos mexicanos casados o cohabitantes con una pareja del sexo contrario que residen en Monterrey. Con la debida precaución, se podrían generalizar o usar como hipótesis para poblaciones afines, esto es, urbanas de cultura latina. Una segunda limitación es que los datos no permiten inferencias causales, ni estimar la confiabilidad temporal de las puntuaciones, ni la estabilidad temporal de la estructura factorial, ya que se usó un diseño *ex post facto* transversal. Una tercera limitación es la naturaleza de autorreporte de los datos, por lo que estos pueden diferir de los obtenidos con técnicas proyectivas, observacionales o procedimientos para evaluar actitudes implícitas.

Se concluye que, en adultos mexicanos casados o cohabitantes con una pareja del sexo contrario residentes en Monterrey, la escala SSSS-9 presenta una consistencia interna global muy alta por el coeficiente alfa ordinal y alta por el coeficiente alfa de Cronbach. Sus nueve ítems cumplen con las propiedades de consistencia interna y discriminabilidad y no muestran anomalías distribucionales del tipo efecto techo, efecto suelo o bimodalidad en valores extremos. La estructura de la escala es unidimensional como indican la convergencia del análisis paralelo de Horn y coordenadas óptimas en un factor, el tamaño del efecto muy grande del factor sobre sus nueve indicadores, la validez convergente y consistencia interna del factor único y el buen ajuste a los datos del modelo unifactorial. El modelo de un factor es válido para ambos sexos, aunque no estrictamente invariante. La distribución de la escala presenta sesgo hacia la cola derecha, con más

valores por debajo que por encima de la media aritmética, por lo que no se ajusta a una distribución normal y por ende su estandarización puede realizarse a través de puntuaciones de percentil. La media de los hombres es mayor que la de las mujeres, el tamaño del efecto del sexo sobre la escala es mediano, por lo que se requieren baremos diferenciales para mujeres y hombres. La SSSS-9 presenta validez de constructo concurrente en relación con conducta infiel y actitud hacia la exclusividad marital. La deseabilidad social impacta sobre la SSSS-9 con un tamaño del efecto pequeño y debido a su factor de manejo de la impresión.

Se sugiere el uso de la SSSS-9 como una medición unidimensional de la búsqueda de sensaciones sexuales en el estudio de las parejas casadas o cohabitantes, ya que constituye un factor de riesgo de infidelidad con un tamaño del efecto grande. Se sugiere estudiar la estabilidad de sus puntuaciones y de la estructura unifactorial. En futuros estudios con la SSSS-9 sería interesante comprobar si el modelo de dos factores (búsqueda de sensaciones y experiencias sexuales y conductas censurables) posee validez discriminante en otras poblaciones.

Referencias

- American Psychological Association (2017). *Ethical principles of psychologists and code of conduct. With the 2016 amendment to standard 3.04*. Washington, DC: American Psychological Association.
- Ballester-Arnal, R., Ruiz-Palomino, E., Espada-Sánchez, J. P., Morell-Mengual, V., & Gil-Llario, M. D. (2018). Psychometric properties and validation of the Sexual Sensation Seeking Scale in Spanish adolescents: Brief screening method for use in research and clinical practice. *Personality and Individual Differences*, 122, 47-54. doi: 10.1016/j.paid.2017.10.006
- Ben-Porath, Y. S., & Tellegen, A. (2016). *Part 3: restandar-*

- dizing the MMPI. Minneapolis, MN: University of Minnesota Press.
- Byrne, B. M. (2016). *Structural equation modeling with AMOS: basic concepts, applications, and programming* (3a ed.). New York, NY: Routledge.
- Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-menu v2.0 to make more judicious estimations. *Practical Assessment*, 18, 1-14 (artículo 8). doi: [10.2147/JHL.S35483](https://doi.org/10.2147/JHL.S35483)
- Cross, C. P., Cyrenne, D. L. M., & Brown, G. R. (2013). Sex differences in sensation-seeking: A meta-analysis. *Scientific Reports*, 3 (artículo 2486). doi: [10.1038/srep02486](https://doi.org/10.1038/srep02486)
- De Oliveira-Santos, M. J., Soares-Ferreira, E. M., Carvalho-Duarte, J. C., & Ferreira, M. M. (2017). Adaptação portuguesa e validação da Sexual Sensation Seeking Scale para estudantes do ensino superior. *Revista de Enfermagem Referência*, 4(15), 11-20. doi: [10.12707/riv17054](https://doi.org/10.12707/riv17054)
- Gadermann, A., Guhn, M., & Zumbo, B. D. (2014). Ordinal alpha. En A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 4513-4515). Dordrecht: Springer Netherlands. doi: [10.1007/978-94-007-0753-5_2025](https://doi.org/10.1007/978-94-007-0753-5_2025)
- Gaither, G. A., & Sellbom, M. (2003). The Sexual Sensation Seeking Scale: Reliability and validity within a heterosexual college student sample. *Journal of Personality Assessment*, 81(2), 157-167. doi: [10.1207/s15327752jpa8102_07](https://doi.org/10.1207/s15327752jpa8102_07)
- Grubbs, J. B., Exline, J. J., Pargament, K. I., Hook, J. N., & Carlisle, R. D. (2015). Transgression as addiction: Religiosity and moral disapproval as predictors of perceived addiction to pornography. *Archives of Sexual Behavior*, 44(1), 125-136. doi: [10.1007/s10508-013-0257-z](https://doi.org/10.1007/s10508-013-0257-z)
- Huang, C. E., Cassels, S. L., & Winer, R. L. (2015). Self-reported sex partner dates for use in measuring concurrent sexual partnerships: Correspondence between two assessment methods. *Archives of Sexual Behavior*, 44(4), 873-883. doi: [10.1007/s10508-014-0414-z](https://doi.org/10.1007/s10508-014-0414-z)
- Iglesias, M. C., & Reyes, G. (2017). 30 años del VIH/SIDA. *Perspectiva desde México*. Ciudad de México: Centro de Investigación en Enfermedades Infecciosas y Fundación México Vivo.
- Instituto Nacional de Estadística, Geografía e Informática (2016). Panorama sociodemográfico de Nuevo León 2015. Ciudad de México: INEGI. Recuperado de <http://www.beta.inegi.org.mx>
- Kalichman, S. C., Adair, V., Rompa, D., Multhauf, K., Johnson, J., & Kelly, J. (1994). Sexual Sensation-Seeking Scale: Development and predicting AIDS-risk behavior among homosexually active men. *Journal of Personality Assessment*, 62(3), 385-397. doi: [10.1207/s15327752jpa6203_1](https://doi.org/10.1207/s15327752jpa6203_1)
- Kalichman, S. C., & Rompa, D. (1995). Sexual Sensation Seeking and Sexual Compulsivity Scales: Reliability, validity, and predicting HIV risk behaviors. *Journal of Personality Assessment*, 65(3), 586-601. doi: [10.1207/s15327752jpa6503_16](https://doi.org/10.1207/s15327752jpa6503_16)
- Kóbor, A., Takács, A., & Urbán, R. (2013). The bifactor model of the Strengths and Difficulties Questionnaire. *European Journal of Psychological Assessment*, 29(4), 299-307. doi: [10.1027/1015-5759/a000160](https://doi.org/10.1027/1015-5759/a000160)
- Kogan, S. M., Cho, J., Barnum, S. C., & Brown, G. L. (2015). Correlates of concurrent sexual partnerships among young, rural African American men. *Public Health Reports*, 130(4), 392-399. doi: [10.1177/003335491513000418](https://doi.org/10.1177/003335491513000418)
- Lyon, A. (2012). Mathematical explanations of empirical facts, and mathematical realism. *Australasian Journal of Philosophy*, 90(3), 559-578. doi: [10.1080/00048402.2011.596216](https://doi.org/10.1080/00048402.2011.596216)
- Moral, J., García, C. H., & Antona, C. J. (2012). Traducción y validación del Inventario Balanceado de Deseabilidad Social al Responder en una muestra probabilística de estudiantes universitarios mexicanos. *Revista de Psicología GEPU*, 3(2), 54-72. Recuperado de <https://revistadepsicologiagepu.es.tl>
- Muise, A., Stanton, S. C. E., Kim, J. J., & Impett, E. A. (2016). Not in the mood? Men under- (not over-)

- perceive their partner's sexual desire in established intimate relationships. *Journal of Personality and Social Psychology*, 110(5), 725-742. doi: 10.1037/pspi0000046
- Romero-Palencia, A., Rivera-Aragón, S., & Díaz-Loving, R. (2007). Desarrollo del Inventario Multidimensional de Infidelidad (IMIN). *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 23(1), 121-147. Recuperado de <http://www.aidep.org/pt/pdf/4596/459645446008.pdf>
- Santos-Iglesias, P., Moyano, N., Castro, A., Granados, M. R., & Sierra, J. C. (2018). Validation of the Sexual Sensation Seeking Scale in Spanish samples. *Journal of Personality Assessment*, 100(4), 421-430. doi: 10.1080/00223891.2017.1328424
- Selterman, D., & Koleva, S. (2015). Moral judgment of close relationship behaviors. *Journal of Social and Personal Relationships*, 32(7), 922-945. doi: 10.1177/0265407514554513
- Sijtsma, K. (2015). Delimiting coefficient α from internal consistency and unidimensionality. *Educational Measurement: Issues and Practice*, 34(4), 10-13. doi: 10.1111/emip.12099
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5a ed.). New York, NY: Oxford University.
- Téllez, A., García, C. H., & Corral-Verdugo, V. (2015). Effect size, confidence intervals and statistical power in psychological research. *Psychology in Russia: State of the Art*, 8(3), 27-46. doi: 10.11621/pir.2015.0303
- Teva, I., & Paz-Bermúdez, M. (2008). Adaptación castellana y propiedades psicométricas de la Escala de Búsqueda de Sensaciones Sexuales en adolescentes españoles. *Revista Mexicana de Psicología*, 25(1), 129-137. Recuperado de <http://www.redalyc.org/articulo.oa?id=243016300010>
- Valdez, C., Moral, J., Onofre, D. J., Castillo, L., Benavides, R., & López, F. (2016). Búsqueda de sensaciones sexuales en hombres que tienen sexo con hombres. *Medicina Universitaria*, 18(Supl. 1), 87. Recuperado de <http://www.elsevier.es/en-revista-medicina-universitaria-304-seccion-xxviii-congreso-nacional-investigacion-medicina-66641>
- Weis, D. L., & Felton, J. R. (1987). Marital exclusivity and the potencial for future marital conflict. *Social Work*, 32(1), 45-49. doi: 10.1093/sw/32.1.45
- Xu, W., Zheng, L., Liu, Y., & Zheng, Y. (2016). Sexual sensation seeking, sexual compulsivity, and high-risk sexual behaviours among gay/bisexual men in Southwest China. *AIDS Care*, 28(9), 1138-1144. doi: 10.1080/09540121.2016.1153587

Anexo**Escala de búsqueda de sensaciones sexuales (SSSS-9).**

Señale qué tanto le caracterizan las siguientes afirmaciones en la siguiente escala.

	1 Nada	2 Algo	3 Bastante	4 Muchísimo
1. Me gustan los encuentros sexuales desinhibidos y salvajes			1	2 3 4
2. He hecho promesas sin la intención de mantenerlas para que otra persona tenga sexo conmigo			1	2 3 4
3. Siento curiosidad en tener sexo anal sin condón			1	2 3 4
4. Disfruto de la compañía de una persona sensual			1	2 3 4
5. Disfruto ver películas pornográficas			1	2 3 4
6. He dicho cosas que no son exactamente ciertas para lograr que una persona tenga sexo conmigo			1	2 3 4
7. Estoy interesada/o en probar nuevas experiencias sexuales			1	2 3 4
8. Me gusta explorar mi sexualidad			1	2 3 4
9. Me gustan las experiencias y sensaciones sexuales nuevas y excitantes			1	2 3 4

Otras formas de entender la d de Cohen

Other ways of understanding Cohen's d

José Ventura-León * ¹

1 - Universidad Privada del Norte, Lima, Perú.

Introducción
Referencias

Recibido: 27/02/2018 Revisado: 18/05/2018 Aceptado: 28/05/2018

Resumen

La d de Cohen (d) es una medida del tamaño del efecto bastante utilizada y su reporte es una condición necesaria para los análisis estadísticos. No obstante, los investigadores reportan que la diferencia entre dos distribuciones es pequeña ($d > .20$). Sin embargo, la interpretación de ese coeficiente no es clara en estudios de psicología. En ese sentido, es necesario convertir la d en una medida de probabilidad y de esa forma facilitar la interpretación de las distribuciones que son objeto de comparación. Dentro de las medidas más frecuentes se encuentran: U_3 de Cohen, el coeficiente de superposición (OVL), la probabilidad de superioridad (PS) y el número necesario para tratar (NNT) que pueden ser considerados como medidas alternativas de la magnitud de una diferencia. Para tales fines se proporcionan códigos en R que los investigadores pueden usar fácilmente, además de una tabla que evidencia las modificaciones de las medidas alternativas ante el incremento del tamaño del efecto.

Palabras clave: d de Cohen, tamaño del efecto, medidas alternativas, probabilidad de superioridad, Coeficiente de superposición

Abstract

Cohen's d (d) is quite a used measure of the size of the effect and its report is compulsory necessary in statistical analyzes. Nevertheless, researchers report that the difference between two distributions is small ($d > .20$). However, the interpretation of this coefficient is not clear in psychology studies. In this sense, it is necessary to convert the d into a probability measure to facilitate the interpretation of the distributions that are object of comparison. Among the most frequent measures are: Cohen's U_3 , the superposition coefficient (OVL), the probability of superiority (PS) and the number needed to treat (NNT), which can be considered as alternative measures of the magnitude of a difference. For such purposes, R codes that can be easily used by the researchers are provided, as well as a table showing the modifications of the alternative measures before the increase in the size of the effect.

Key words: Cohen's d , effect size, alternative measures, probability of superiority, superposition coefficients

* Correspondencia a: José Ventura-León. Av. Tingo María 1122, Breña, Lima. E-mail: jose.ventura@upn.pe

Cómo citar este artículo: Ventura-León, J. (2018). Otras formas de entender la d de Cohen. *Revista Evaluar*, 18(3), 73-78. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Introducción

El *tamaño del efecto* (TE) puede ser entendido como “el grado en que el fenómeno está presente en la población” (Cohen, 1988, p. 9) y es clasificado en dos grandes grupos de familia: (a) familia *r*, que implica las medidas de asociación y (b) familia *d*, que se refiere a las medidas de diferencias de grupos (Ellis, 2010). De acuerdo con revisiones sistemáticas, estos coeficientes aparecen entre el 9 % y el 12 % en revistas psicológicas (Castillo, 2014; García, Ortega, & De la Fuente, 2008) y se incluyen entre las recomendaciones para el análisis de datos (American Psychology Association [APA], 2010). Tanto ha sido su impacto que el artículo A Power Primer, publicado por Cohen en 1992, cuenta con 6839 citas, lo que lo convierte en el artículo más citado en la historia de la psicología (Ho & Hartley, 2016).

Pese a la existencia de numerosos coeficientes acerca del TE, la *d* continúa siendo el más utilizado para examinar la diferencia de medias estandarizadas entre dos grupos independientes. Matemáticamente se expresa como $d = (y'1 - y'2) / S_p$, donde *y'1* e *y'2* son las medias de las poblaciones y *S_p* es la desviación estándar combinada; de ese modo, el coeficiente resultante puede ser considerado pequeño ($d > 0.20$), mediano ($d > 0.50$) o grande ($d > 0.80$; Cohen, 1988). No obstante, estos valores fueron establecidos de forma arbitraria, y el propio Cohen (1988) advertía que “esta es una operación cargada de muchos peligros” (p. 12). El principal riesgo es que estos valores sean tomados como reglas rígidas que conviertan el tamaño del efecto en una especie de *p*-valor $< .05$ (Gurnsey, 2017). Por esto, existe un consenso creciente de que estos puntos de corte pueden ser engañosos y deben evitarse (Baguley, 2009), porque no pueden realizarse independientemente del contexto de investigación y de los conceptos asociados a este como la potencia,

significación y tamaño de la muestra (Gurnsey, 2017). Dado que el TE representa el grado en que la hipótesis nula es falsa, un TE alto implica un incremento en la potencia estadística (Cohen, 1988, 1992), y cuanto más pequeña sea la significancia, menor será esta potencia (Cárdenas-Castro & Arancibia, 2014).

En ese contexto, es importante transformar la *d* en una medida de probabilidad para facilitar la interpretación de la magnitud de la diferencia entre dos medias a partir de medidas alternativas. En este sentido, Magnusson (2014) expone los siguientes coeficientes, que pueden ser utilizados si se satisfacen los supuestos de normalidad, igualdad de varianza y tamaños de muestra equivalentes (Cohen, 1988): U3 de Cohen, el coeficiente de superposición (OVL), la probabilidad de superioridad (PS) y el número necesario para tratar (NNT), cuya utilización sólo es posible si se satisfacen los supuestos de normalidad, igualdad de varianza y tamaños de muestra equivalentes (Cohen, 1988).

Previo a la explicación de las medidas alternativas, se presenta la Tabla 1 que recoge a modo de resumen todas las fórmulas y comandos de R del presente artículo. Si bien los comandos de la U3 y OVL pueden resultar innecesarios, son útiles para buscar un valor preciso basado en una *d* específica (por ejemplo, $d = 0.623$). Si el usuario decide no utilizarlos, puede recurrir a la Tabla 2 en la que se presentan los diferentes valores de las medidas alternativas para cada incremento de una *d* previamente establecida. Asimismo, se advierte que la PS de la Tabla 2 asume igualdad muestral y el NNT está basado en una tasa de evento de grupo control de .20. Si se desea cambiar las condiciones, será necesario recurrir a las fórmulas y comandos de la Tabla 1.

Tabla 1

Coeficiente, fórmulas y códigos en R de las diferentes medidas alternativas.

Coeficiente	Fórmula	Códigos en R
U3	$U_3 = \Phi(\delta)$	d<-.69 U3<-pnorm(d) U3
OVL	$OVL = 2\Phi\left(\frac{- \delta }{2}\right)$	d<-.69 OVL<-pnorm(-(d)/(2))*2 OVL
PS	Grupos Iguales	d<-.69 CL1<-pnorm(d/sqrt(2)) CL1
	$PS = \Phi\left(\frac{d}{\sqrt{2}}\right)$	
	Grupos Desiguales	p1<-.40 p2<-.60 DE1<-(4.55)^2 DE2<-(9.11)^2 <-pnorm(d*(sqrt(p1*DE1)+(p2*DE2)/(DE1+DE2))) Pse
	$PS = \Phi\left(d\sqrt{\frac{p_1s_1^2 + p_2s_2^2}{s_1^2 + s_2^2}}\right)$	
NNT*	$NNT = \frac{1}{\Phi(d - \Psi(CER)) - CER}$	d<-.69 CER<-.2 NNT<-1/(pnorm(d + qnorm(CER)) - CER) NNT

Nota. Es la función de distribución acumulativa para la distribución normal y es la d de Cohen de la población; s^2 corresponde a la desviación estándar y p es la proporción de sujetos; es la inversa; CER es la tasa de eventos del grupo control. * = Los códigos en R son de Magnusson (2014).

U3 de Cohen

El TE puede ser expresado en términos de superposición de una distribución, y en este caso se denomina medida U (Cohen, 1988). La medida U3 resulta de la multiplicación de la d por una función de distribución acumulativa para la distribución normal (véase Tabla 1), lo que convierte a la d en la probabilidad de que ese valor esté por debajo o encima de la media de la segunda distribución.

Supóngase que un grupo de estudiantes de psicología que asisten a clase puntualmente supera a los estudiantes que llegan tarde; siendo él TE $d = 0.30$, la U es aproximadamente de 62. Esto significa que, si un estudiante del grupo de los puntuales dejara de ser puntual y pasara a formar parte del grupo de los impuntuales, entonces pasaría del percentil 50 al 62.

Coeficiente de Superposición (OVL). El OVL es un coeficiente que revela el porcentaje en que las

distribuciones de ambos grupos se superponen (Reiser & Faraggi, 1999), y de esa forma permite probar la similitud o cercanía entre las dos distribuciones (Al-Saleh & Samawi, 2007). Este coeficiente resulta de multiplicar dos veces la función de la distribución acumulativa para la distribución normal por el negativo de la mitad de la d (véase Tabla 1).

Supóngase que un grupo de niños con TDAH (trastorno por déficit de atención con hiperactividad) es superado por otro grupo de niños sin TDAH en lectura ($d = 0.75$); para este TE, la OVL es aproximadamente de .71. Esto significa que el 71% de las puntuaciones de lectura de niños con y sin TDAH se superponen.

Probabilidad de Superioridad (PS). La PS es un coeficiente que indica la probabilidad de que un individuo extraído al azar de un grupo sea superior en su puntaje con respecto a una variable, comparado con un individuo extraído al azar del otro grupo (McGraw & Wong, 1992). No obstante, es importante advertir que existen diferencias en las ecuaciones según se trate con grupos iguales o desiguales. En el caso de la igualdad de grupos, el coeficiente resulta de la multiplicación de la función de distribución acumulativa por la d entre la raíz cuadrada de dos y, en el caso de grupos desiguales, se considera la proporción de sujetos de cada grupo con sus respectivas desviaciones estándar (véase Tabla 1).

Supóngase que un psicólogo observa que la diferencia entre un grupo de 20 personas diagnosticadas con depresión ($p_1 = .54$ [proporción de sujetos del grupo 1]; $DE = 4.55$) y un grupo de 17 personas sin depresión ($p_2 = .46$ [proporción de sujetos del grupo 2]; $DE = 9.11$) es moderada con una $d = 0.55$; la PS es aproximadamente .98. Esto indica que hay una probabilidad del 98 % de que una persona escogida al azar del grupo con

depresión obtenga una puntuación mayor que una persona escogida al azar del grupo sin depresión.

Número necesario para tratar (NNT). El NNT es un cálculo que permite conocer la cantidad de pacientes necesarios para que un grupo experimental tenga un éxito más (o uno menos) en comparación con el grupo de control (Furukawa & Leucht, 2011). Para calcular este coeficiente es necesario contar con el CER (por sus siglas en inglés), que es el porcentaje de veces que ocurre el evento en el grupo de control. En caso de desconocerlo, se puede establecer un valor por defecto de .20 (en dicho caso los resultados deben tomarse con precaución). Además, el NNT refleja la importancia clínica relacionada con la reducción del riesgo absoluto (McGough & Faraone, 2009). Pese a ello, hay investigadores que indican que existe una pérdida de información en la conversión (Scholten, de Beurs, & Bouter, 1999).

Supóngase que un psicólogo desea conocer la efectividad de un programa de intervención en agresión para niños; por tanto, cuenta con un grupo de control de 20 niños que no reciben el programa y un grupo experimental en el que se efectúan sesiones de intervención. La magnitud de la diferencia entre los grupos es grande ($d = 0.95$) y, por ende, su NNT es 2.91. Esto indica que será necesario tratar a 3 pacientes con el programa de intervención para obtener una mejor respuesta que la del grupo de control.

Tabla 2

Valores para todas las medidas alternativas según diferentes tamaños de la d de Cohen.

d	U3	OVL	PS	NNT
0.00	50.00	100.00	50.00	Inf.
0.10	53.98	96.01	52.82	34.30
0.20	57.93	92.03	55.62	16.51

0.30	61.79	88.08	58.40	10.63
0.40	65.54	84.15	61.14	7.73
0.50	69.15	80.26	63.82	6.01
0.60	72.57	76.42	66.43	4.89
0.70	75.80	72.63	68.97	4.10
0.80	78.81	68.92	71.42	3.53
0.90	81.59	65.27	73.77	3.09
1.00	84.13	61.71	76.02	2.76
1.10	86.43	58.23	78.17	2.49
1.20	88.49	54.85	80.19	2.27
1.30	90.32	51.57	82.10	2.10
1.40	91.92	48.39	83.89	1.95
1.50	93.32	45.33	85.56	1.84
1.60	94.52	42.37	87.11	1.74
1.70	95.54	39.53	88.53	1.65
1.80	96.41	36.81	89.85	1.58
1.90	97.13	34.21	91.04	1.53
2.00	97.72	31.73	92.14	1.48
2.20	98.61	27.13	94.01	1.40
2.40	99.18	23.01	95.52	1.35
2.60	99.53	19.36	96.70	1.31
2.80	99.74	16.15	97.61	1.29
3.00	99.87	13.36	98.31	1.27
3.20	99.93	10.96	98.82	1.26
3.40	99.97	8.91	99.19	1.26
3.60	99.98	7.19	99.45	1.25
3.80	99.99	5.74	99.64	1.25
4.00	100.00	4.55	99.77	1.25

Nota. $d = d$ de Cohen; el NNT es calculado según un CER de .20; el PS es calculado para muestral igual. Estos valores fueron creados mediante códigos ad hoc en el programa de acceso libre R. Inf. = un valor infinito

En conclusión, transformar la d en valores de probabilidad resulta de mucha utilidad, porque facilita la interpretación de la magnitud de la diferencia y el investigador puede conocer mejor lo que significa un coeficiente d . Aunque cabe advertir que la relevancia clínica no puede solo deducirse de la interpretación de d de Cohen (Scholten et al., 1999), resulta útil explorar e interpretar de forma más sencilla las distribuciones de los dos

grupos a partir de los coeficientes aquí brindados. Además, el hecho de brindar códigos en R facilitará los cálculos de investigadores inexpertos en estadística, que podrán reproducir los códigos, modificando el valor de la d y otros índices.

Referencias

- Al-Saleh, M. F., & Samawi, H. M. (2007). Interference on overlapping coefficients in two exponential populations. *Journal of Modern Applied Statistical Methods*, 6(2), 503-516. doi: [10.22237/jmasm/1193890440](https://doi.org/10.22237/jmasm/1193890440)
- American Psychological Association (2010). *Publication Manual of the American Psychological Association*. Washington, DC: American Psychological Association.
- Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, 100(3), 603-617. doi: [10.1348/000712608X377117](https://doi.org/10.1348/000712608X377117)
- Castillo, R. (2014). *Reporte del tamaño del efecto en los artículos de tres revistas de psicología peruanas en los años 2008 al 2012* (Tesis de pregrado). Universidad Mayor de San Marcos. Lima, Perú.
- Cárdenas-Castro, J. M., & Arancibia, H. (2014). Potencia estadística y cálculo del tamaño del efecto en G* Power: Complementos a las pruebas de significación estadística y su aplicación en psicología. *Salud & Sociedad*, 5(2), 210-244. doi: [10.22199/s07187475.2014.0002.00006](https://doi.org/10.22199/s07187475.2014.0002.00006)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2a ed.). Hillsdale, NJ: Erlbaum. doi: [10.4324/9780203771587](https://doi.org/10.4324/9780203771587)
- Cohen, J. (1992). Cosas que he aprendido (hasta ahora). *Anales de Psicología*, 8(1-2), 3-17. Recuperado de <http://revistas.um.es/analesps>
- Ellis, P. D. (2010). *The essential guide to effect sizes. Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, UK: Cambridge University. doi: [10.1017/cbo9780511761676](https://doi.org/10.1017/cbo9780511761676)

- Furukawa, T. A., & Leucht, S. (2011). How to obtain NNT from Cohen's d: Comparison of two methods. *PloS One*, 6(4), 1-5. doi: [10.1371/journal.pone.0019070](https://doi.org/10.1371/journal.pone.0019070)
- García, J., Ortega, E., & De la Fuente, L. (2008). Tamaño del efecto en las revistas de Psicología indizadas en Redalyc. *Informes Psicológicos*, 10(11), 173-188.
- Gurnsey, R. (2017). *Statistics for research in Psychology: A modern approach using estimation*. Thousand Oaks, CA: SAGE.
- Ho, Y. S., & Hartley, J. (2016). Classic articles in psychology in the Science Citation Index Expanded: A bibliometric analysis. *British Journal of Psychology*, 107(4), 768-780. doi: [10.1111/bjop.12163](https://doi.org/10.1111/bjop.12163)
- Magnusson, K. (3 de febrero de 2014). *Interpreting Cohen's d effect size: An interactive visualization*. Recuperado de <http://rpsychologist.com/d3/cohend>
- McGough, J. J., & Faraone, S. V. (2009). Estimating the size of treatment effects: Moving beyond P values. *Psychiatry (Edgmont)*, 6(10), 21-29. Recuperado de <https://www.ncbi.nlm.nih.gov/pmc/journals/901>
- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, 111(2), 361-365. doi: [10.1037/0033-2909.111.2.361](https://doi.org/10.1037/0033-2909.111.2.361)
- Reiser, B., & Faraggi, D. (1999). Confidence intervals for the overlapping coefficient: The normal equal variance case. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 48(3), 413-418. doi: [10.1111/1467-9884.00199](https://doi.org/10.1111/1467-9884.00199)
- Scholten, R. J., de Beurs, E., & Bouter, L. M. (1999). From effect size into number needed to treat. *The Lancet*, 354(9178), 598. doi: [10.1016/s0140-6736\(05\)77952-6](https://doi.org/10.1016/s0140-6736(05)77952-6)