



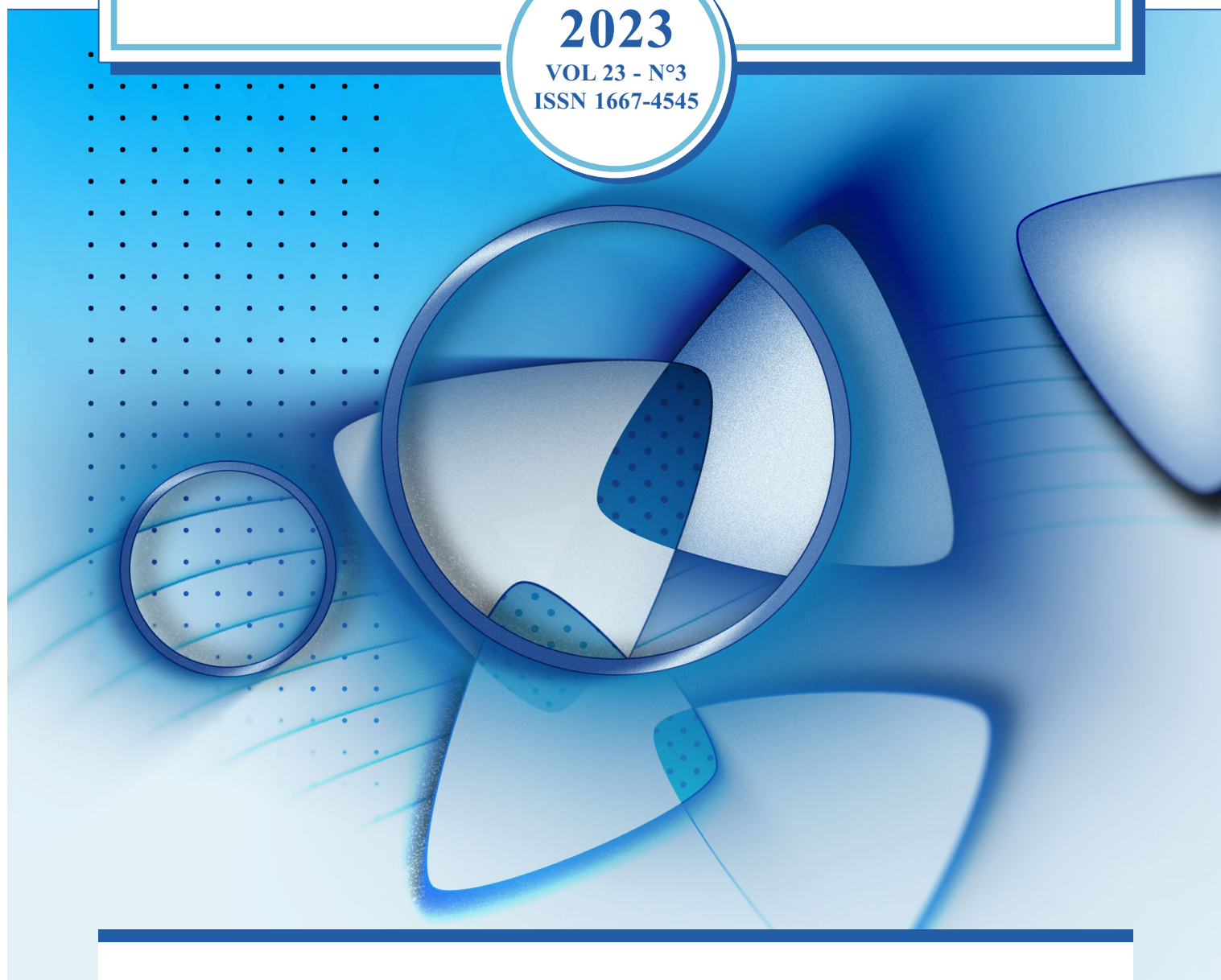
EVALUAR

Revista Evaluar

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

2023

VOL 23 - N°3
ISSN 1667-4545



Assessment of Cognitive Flexibility in the Elderly: Evidence for the Validity of the CAMBIOS Test

Evaluación de la Flexibilidad Cognitiva en la vejez: Evidencias de validez del Test CAMBIOS

Krzemien Deisy *¹, Barrio Daiana¹, Ferreyra Florencia¹, Richard's María¹

1 - Instituto de Psicología Básica, Aplicada y Tecnología (CONICET-UNMDP).

Introduction
Methods
Results
Discussion
References

Recibido: 21/03/2023 Revisado: 12/05/2023 Aceptado: 18/05/2023

Abstract

This study aims to accomplish two objectives: (1) To analyze the performance of Cognitive Flexibility (CF) in order to establish normative values in older adults, and (2) to evaluate the external validity evidence of the CAMBIOS Test in the older population from the Rio de la Plata region. An intentional non-probabilistic sample of 200 participants aged 60 to 85 from Mar del Plata divided into two groups was evaluated: younger, on one hand, and middle-aged older adults, on the other hand. Concurrent validity evidence of the CAMBIOS Test with a computerized CF task and criterion-related validity evidence associated with age were presented. Performance differences in CF were observed based on age. The scores in the age-associated CF level were equivalent between both instruments. The CAMBIOS Test proved to be useful for assessing CF, enabling the development of normative values for the regional older adult population, discriminated by age subgroups in the aging process.

Keywords: *CAMBIOS Test, cognitive flexibility, older adults, validity*

Resumen

Este estudio propone dos objetivos: (1) analizar el desempeño de la Flexibilidad Cognitiva (FC) y establecer valores normativos en las personas mayores y (2) evaluar las evidencias de validez externa del Test CAMBIOS en la población rioplatense de personas mayores. Se evaluó una muestra no probabilística intencional de 200 participantes de 60 a 85 años de edad de Mar del Plata, distribuidos en dos grupos: mayores jóvenes y mayores intermedios. Se presentan evidencias de validez concurrente del Test CAMBIOS con una tarea informatizada de FC y evidencias de validez de criterio asociado a la edad. Según la edad, se evidenciaron diferencias en el desempeño en FC. Las puntuaciones en el nivel de FC, asociado a la edad, resultaron equivalentes entre ambos instrumentos. El Test CAMBIOS resultó útil para evaluar la FC, posibilitando la elaboración de valores normativos para la población regional de personas mayores, discriminado por subgrupos de edad en el envejecimiento.

Palabras clave: *test CAMBIOS, flexibilidad cognitiva, personas mayores, evidencias de validez*

*Correspondence to: deisykrzemien@gmail.com. Cuerpo V, Nivel III 3250, B7602AYJ, Deán Funes, B7602AYJ, Mar del Plata, Provincia de Buenos Aires, Argentina.

How to cite: Krzemien, D., Barrio, D., Ferreyra, F., & Richard's, M. (2023). Assessment of Cognitive Flexibility in the Elderly: Evidence for the Validity of the CAMBIOS Test. *Revista Evaluar*, 23(3), 1-15. Retrieved from <https://revistas.unc.edu.ar/index.php/revaluar>

Participaron en la edición de este artículo: Abigail Pérez, Andrea Suárez, Juan Cruz Balverdi Nieto, Florencia Ruiz, Benjamín Casanova, Jorge Bruera.

Introduction

Cognitive Flexibility (CF) is a component of executive function located in the prefrontal cortex (Uddin, 2021). CF is the ability to simultaneously switch between modes of thinking about multiple aspects of a complex object or situation (Diamond, 2013). It refers to the ability to process one set of stimuli and alternate with the simultaneous processing of another set (Seisededos, 2008). It also involves recognizing the intermittent change of rules or criteria in a task, having an appropriate tolerance for it, and not persisting with the previous criterion. As tasks demand change, the cognitive system can adapt itself by shifting the attention and selecting the specific information to provide a response and in so generating new activation states to provide feedback to the system. When these processes result in representations and actions that adapt to task changes, a person's behavior can be considered flexible (García-Coni & Vivas, 2014; Ionescu, 2017).

In the current context, maintaining or improving cognitive functions in older adults and reducing the prevalence of neurodegenerative diseases have become the focus of health care and research (Burke et al., 2019; Krzemien, 2013; Siagian et al., 2020; Mather, 2020). Given the role that executive functions play in maintaining independence and a good quality of life, they are considered a prerequisite for optimizing healthy aging (Betancourt-Zambrano et al., 2020; Lepe-Martínez et al., 2020; Stieger & Lachman, 2021). CF has been found to predict performance in daily life instrumental activities (Bell-McGinty et al., 2002; Vaughan & Giovanello, 2010).

Neuropsychological studies (Dorbath et al., 2013; Ferguson et al., 2021) have shown that executive functions are affected by aging. There is evidence that aging is associated with a decline in the efficiency of certain cognitive processes,

such as cognitive flexibility, episodic memory, visuospatial functions and attention, among others, which are related to brain circuitry changes (Reuter-Lorenz & Park, 2014). However, other studies in older adults with pathologies, such as type 2 diabetes mellitus and hypertension found that the components of executive function, which were affected the most by aging, were working memory and planning, while mental flexibility and verbal fluency were more preserved (Aguilar-Mateu et al., 2014).

The developmental trajectory of CF has been found to show a progressive decline from adulthood to old age (Krzemien et al., 2018a; Richard's et al., 2023). Some authors report perseverative responses, an increase in the number of errors, and a decrease in the number of complete responses in older adults compared to young individuals (Herrera & Eizaguirre, 2020; Romo-Galindo et al., 2015; Roselli et al., 2008; Wecker et al., 2005). Among the proposed explanations, it has been suggested that older adults have difficulty generating new hypotheses about constantly changing rules (Rosselli et al., 2008).

On the other hand, several studies assert that neurocognitive changes associated with aging do not inevitably imply a cognitive decline in all performance domains, nor do they follow the same direction and sequence of slowing or deterioration (Baltes et al., 2007; Grandi & Tirapu-Ustároz, 2017). Instead, a variety of changes in neurocerebral activity takes place which are specific to normal aging. There is empirical evidence of brain plasticity and neuronal activation in old age (Clare et al., 2017; Hakun et al., 2015) which challenges the notion of frontal lobe vulnerability and the common assertion that processes mediated by this lobe deteriorate with age.

There are studies (Martins et al., 2015; Meunier et al., 2014; Ojeda et al., 2019) that explain preserved cognitive function in the context

of neuroanatomical and functional changes due to aging based on compensatory mechanisms as one of the main factors influencing cognitive performance. During aging, a greater number of brain areas are activated to support the control of a specific function at a given time (Wang et al., 2010). However, there is a cost in terms of reduced efficiency when multiple brain areas are activated in a generalized manner, making a more specialized system more efficient than a diffuse one (Van den Heuvel et al., 2009).

The method typically used to assess executive functions has generally been performance measures (Morales-Millán et al., 2021). A methodological issue that requires attention in the field of gerontological cognitive assessment is the validity of executive function performance tests of executive functions. Some of the instruments traditionally used to measure CF use a global index that assesses executive functions as a whole (e.g., Wisconsin Card Sorting Test, Trail Making Test, Rings Test), which hinders the evaluation, while the interpretation makes it difficult to assess and interpret each function independently (Miyake et al., 2000; Uddin, 2021). In particular, the CAMBIOS test allows the specific measurement of CF, since it can assess the alternation between different rules that the person must consider simultaneously and recognize when they change with increasing complexity.

On the other hand, the adaptation and development of regional normative values for executive function assessment techniques in older adults remain an open task. It is crucial to obtain standardization norms that are adapted to the characteristics of the specific population under study (Tornimbeni et al., 2004). The use of existing norms originally derived from other regions, cultures, or age groups is not recommended when studying the older adult population from a different regional context (Rosselli & Ardila, 2003).

This is especially important given the heterogeneity of cognitive performance within the aging process, where there are interindividual differences and characteristics based on age subgroups. Old age cannot be viewed as a homogeneous group with general patterns. Furthermore, given the cohort differences, it is advisable to have normative data specific to the age subgroups of older adults in the local context.

This study aims to analyze CF performance in old age and to provide evidence for the metric adaptation of the CAMBIOS test in the Rioplatense population of younger and middle-aged older adults.

To achieve this, evidence of the external validity (concurrent) of the CAMBIOS test is presented based on correlations with a computerized CF task, as well as evidence of age-related validity based on scores obtained on both instruments in older adults. This study proposes two specific objectives: first, to analyze CF performance based on educational level and gender and to develop preliminary normative values in a sample of Rioplatense older adults; and second, to obtain evidence of concurrent and age-related validity of the CAMBIOS test in a subsample of older adults.

Methods

Design and participants

A correlational, nonexperimental and cross-sectional study was conducted (Montero & León, 2007). The total sample was selected using non-probability purposive sampling and consisted of 200 participants. Hypothesis testing sampling was also used (Padua, 1982), resulting in 100 women and 100 men. Within each gender group, two subgroups based on age were considered: Group 1 included younger older adults (60 to 69 years) and Group 2 included middle-aged

Table 1

Descriptive characteristics of the Study One sample discriminated by age groups (n = 200).

		Group 1 (60 to 69 years old)		Group 2 (70 to 85 years old)	
Age - M (SD)		64,06 (3,01)		74,26 (3,65)	
		Frequency	%	Frequency	%
Gender	Male	64	51	36	48,6
	Female	62	49	38	51,4
Educational level	Primary	14	11	12	16,2
	Secondary	47	37	28	37,8
	Tertiary / University	65	52	34	45,9
Retirement	Retired	82	65	71	95,9
	Not Retired	44	35	3	4,1
Occupation	Housewife	7	5,6	5	6,8
	Administrative employee	30	24	18	24,3
	Professional	37	29	20	27
	Technician / manager / trades	17	14	11	14,9
	Teacher	13	10	10	13,5
	Businessman	22	18	9	12,2
	Coexistence	Only	25	20	31
	With family member	101	80	43	58,1
Subtotal		126		74	

Note. Group 1 (group of young older people). Group 2 (group of middle-aged older adults).

older adults (70 to 85 years). Participants with at least complete primary education were included, while those residing in long-term care facilities and those with cognitive disorders were excluded. To rule out the presence of cognitive deficits, the ACE-III (Bruno et al., 2020) was used as a screening test, with an established cut-off point of 86 for individuals with basic education and a maximum score of 100. Group 1 had a mean score of 89.03 (2.90), and Group 2 had a mean score of 87.78 (3.51). The sample was drawn from various community institutions, including UPAMI (University for Integrated Older Adults) extension courses, the University Program for

Lifelong Learning, senior centers, day clubs, the Municipal Gerontological Unit, and individuals without institutional affiliation who were interviewed in their homes.

Based on the overall objectives, two studies were conducted, which are presented below:

Study 1: Analysis of the performance of CF in old age. Development of normative values in Rioplatense older adults. This study included a total sample of 200 participants with the characteristics described in the Methods section. The CAMBIOS test was administered to all participants, taking into account the gender and age group criteria (Table 1).

Table 2

Descriptive statistics of socio-occupational and educational variables for the subsample of Study 2, discriminated by age groups (n = 50).

		Group 1 (60 to 69 years old)		Group 2 (70 to 85 years old)	
Age - M (SD)		64,06 (3,01)		74,26 (3,65)	
		Frequency	%	Frequency	%
Gender	Male	64	51	36	48,6
	Female	62	49	38	51,4
Educational level	Primary	14	11	12	16,2
	Secondary	47	37	28	37,8
Retirement	Tertiary / University	65	52	34	45,9
	Retired	82	65	71	95,9
Occupation	Not Retired	44	35	3	4,1
	Housewife	7	5,6	5	6,8
Coexistence	Administrative employee	30	24	18	24,3
	Professional	37	29	20	27
Subtotal	Technician / manager / trades	17	14	11	14,9
	Teacher	13	10	10	13,5
	Businessman	22	18	9	12,2
	Only	25	20	31	41,9
	With family member	101	80	43	58,1
	Subtotal	126		74	

Note. Group 1 (group of young older people). Group 2 (group of middle-aged older adults).

Study 2: Evidence of the validity of the CAMBIOS test of cognitive flexibility. In order to obtain evidence of concurrent validity, a purposive sub-sample of 50 individuals (25 females) was purposefully selected from the total sample, while maintaining the same socio-occupational characteristics. These individuals were assessed using the CAMBIOS test and also completed the Fingers Task from the Cognitive Self-Regulation Battery (TAC), a computerized task. Age was also taken into account for these 50 individuals, using the same criteria as for the general sample: Group 1 consisted of younger older adults (60 to 69 years) and Group 2 consisted of middle-aged older adults (70 to 85 years) (see distribution in Table 2).

Instruments

Socio-Educational Data Questionnaire. This is a short instrument with easy administration, developed ad hoc and consisting of closed-ended questions about participants' socio-occupational data and educational level. This questionnaire has been used in several studies conducted in our context and with samples of older adults (Krzemien, et al., 2018a). Estimated time of administration: five minutes.

Addenbrooke's Cognitive Examination-III (ACE-III, Argentine adaptation by Bruno et al., 2020). This brief test detects dementia and assesses five cognitive domains: attention, memory, verbal fluency, language and visuospatial abili-

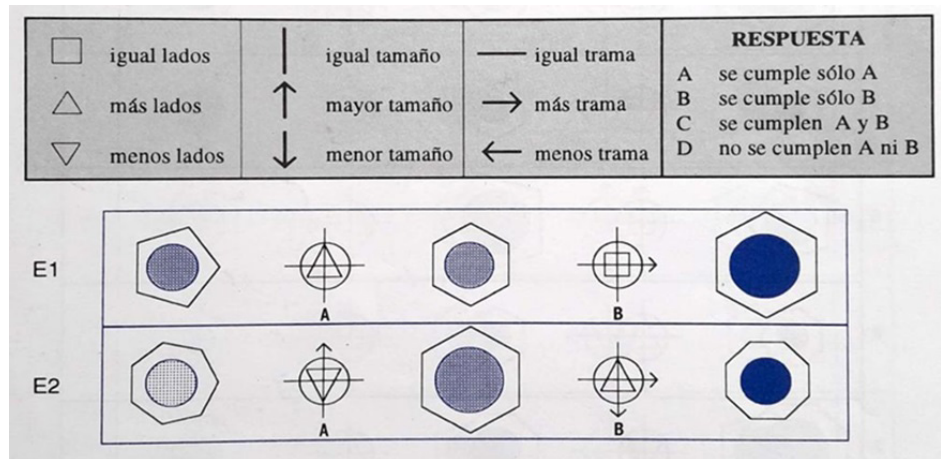


Figure 1

Description of the CAMBIOS Test command.

Source: Seisdedos, N. Cognitive Flexibility Test (CAMBIOS). TORCH. 2008 (Image taken from the original CAMBIOS test protocol, pages 11-13 of the manual).

ties. Individuals with basic education must score above the cutoff of 86, with a maximum score of 100. The Argentine version of the ACE-III has good internal consistency (Cronbach's alpha = .87), and significant differences in total ACE-III scores were found between the control and dementia groups ($p < .05$). Estimated time of administration: 20 minutes.

CAMBIOS Cognitive Flexibility Test (Seisdedos, 2008). It is a visual test free of verbal influence that assesses the use of a flexible and efficient action strategy when faced with simple tasks. It is designed to measure the logical processes necessary to comply with change conditions. This test measures organized and systematic behavior that responds quickly to classification stimuli. It assesses the ability to concentrate while attending to various changing conditions in the stimulus and the flexibility to analyze whether or not different requested changes are fulfilled and when they have ceased to be fulfilled. The test consists of 27 elements that contain simple geometric figures (polygons from five to nine sides with an interior or color pattern), on which three simple types of changes can be requested: the increase or decrease of one or several of its characteristics, that is, the

number of sides of the polygon, the size of the figure and the intensity of the pattern. The types of errors that the examinee could commit are perseverance or redundancy in the answer, omission, and non-recognition of the type of rule that is requested. The final score is only analyzed from the hits achieved within the predetermined time limit. In its original Spanish version, the test has adequate psychometric properties (the reliability coefficient for the two-half method obtained a correlation of .92). It has a minimum average execution time of seven minutes for the Spanish youth population. In previous studies in groups of older people (Krzemien, et al., 2018b), the execution time of the complete test (from one to 27 exercises) was 25 minutes, therefore, a maximum time limit of 14 minutes was determined for older people from the Rio de la Plata context. This limit as a cut-off point makes it possible to evaluate the examinee's performance (in terms of the number of correctly completed exercises as hits) in a time appropriate to the age group of older people (twice as many minutes as in the original version for young people) and thus to evaluate the CF in terms of speed of response (Figure 1).

For the test as a whole, the averages have a



Figure 2

Examples of the three blocks of the Finger Task: (a) Congruent trials, (b) Incongruent trials, and (c) Mixed trials.

Source: In Introzzi, I., Richard's, M., & Canet-Juric, L. (2019). TAC: Tareas de Autorregulación Cognitiva (www.tac.com.ar). *Revista Argentina de Neuropsicología*.

difficulty index of .57, a correlation index of .18, and a homogeneity index of .42. These results define the medium difficulty and good internal consistency of the test in a sample of Spanish young adults (Seisdedos, 2008). In our country, in a recent study (Vido et al., 2018), the test was applied to a sample of older people, obtaining reference values for this age group.

Finger Task from the Computerized Battery TAC (Introzzi & Canet-Juric, 2019). This is a modified version of the original task by Davidson et al. (2006) and adapted by Richard's et al. (2019) for older adults. The task consists of three rating blocks: Congruent, Incongruent and Mixed. Each rating block is preceded by a practice block of eight trials. A hand with an index finger (stimulus) appears on the screen indicating the direction in which the participant should press the key (right or left). The interval between stimuli is 1500 milliseconds, and each stimulus remains on the screen for 3000 milliseconds, during which the participant must respond. The (1) Congruent block contains 20 trials (ten stimuli presented on the left and ten on the right, randomly on the screen). The participant must press the key on the

same side as the stimulus (ipsilateral response). The (2) Incongruent block consists of 20 trials (ten stimuli presented on the right side of the screen and ten on the left side). The participant must press the key on the opposite side of the stimulus (contralateral response). The (3) Mixed block consists of all stimuli from the previous blocks which appear randomly (20 congruent and 20 incongruent trials). This block specifically measures cognitive flexibility as the participant must quickly and efficiently switch between two incongruent rules (ipsilateral or contralateral response). The performance indices obtained in the three blocks are the average percentage of correct responses (accuracy) and the average response times. Some authors suggest using a measure that combines speed and accuracy (Klein et al., 2004). For this purpose, the Inverse Efficiency (IE) index was used, which is the ratio of RT to accuracy, i.e. the value of RT divided by one - PE (proportion of errors). Since RT is expressed in milliseconds and divided by proportion (% correct), the index is also expressed in milliseconds. The time to complete the task is 15 minutes (Figure 2).

Procedure and data analysis

The techniques were administered to the participants under standardized conditions in a single session lasting 50/55 minutes, with a ten-minute break. First, they were informed about the purpose and importance of the study, and the voluntary nature and confidentiality of the data were guaranteed by current regulations (guidelines recommended by the American Psychological Association, National Law on the Protection of Personal Rights No. 25,326, Inter-American Convention on the Protection of Human Rights of the Elderly of the Organization of American States 2015, and guidelines for Social Sciences and Humanities of the Ethics Committee - CONICET 2857-06).

Descriptive and inferential statistical tests were used to analyze performance in CF (Cognitive Flexibility) in both assessment instruments (CAMBIOS and Finger Task). To obtain evidence of validity, two studies were performed: a) analysis of the evidence of external validity (concurrent validity) of the CAMBIOS test, based on correlations with the Cognitive Flexibility Finger Task of the computerized battery, and b) analysis of the evidence of age-related validity (lower performance with increasing age), considering the scores obtained in both instruments in younger and middle-aged older adults. To compare performance differences based on age, two groups were considered according to the criteria established in previous studies on cognitive development in old age (Salthouse, 2018): Group 1, younger older adults aged 60 to 69, and Group 2, middle-aged older adults aged 70 to 85. The data analysis plan was carried out in several steps. First, descriptive analyses of the study variables were conducted by age group. Second, a means test was administered to determine performance differences based on gender, and an analysis of

variance (ANOVA) was conducted to investigate whether performance in CF varied according to educational level. Third, correlation coefficients were calculated between the CAMBIOS test scores and the finger task scores to analyze the criteria of external validity (concurrent validity). Fourth, to determine compliance with age-related validity criteria, a mean difference analysis or T-test was applied based on the groups of younger and middle-aged older adults.

Results

Study 1: Analysis of CF performance during old age based on educational level and gender, and development of normative values in elderly from the Rio de la Plata region.

According to the scores obtained in both instruments (CAMBIOS and Finger Task), the performance of the middle-aged older adults group [$M = 11.12 (5.74)$] is lower than that of the younger older adults group [$M = 15.52 (5.50)$]. Considering the data from a previous local study conducted in 2019 in the city of Mar del Plata, the mean score achieved in the CAMBIOS test by the middle-aged older adults group was 9.13 (4.70), while the mean score for the younger older adults was 14.86 (6.30). This indicates that the higher performance of the younger age group is consistent across both studies (2019 and 2022) (Table 3).

An ANOVA was conducted to determine if performance in CF varied by educational level (primary, secondary, and university), but no significant differences were found ($p > .01$). Furthermore, a means test was conducted to determine if there were differences in performance based on gender, and the results indicated that there were no significant differences ($p > .01$).

Table 3

Statistical values of central tendency and variability of CF performance on the CAMBIOS test, discriminated by age groups, in two studies. (Years 2019 and 2022).

	Estudio actual (N= 200)		Estudio previo (N = 51)	
	Año 2022		Año 2019	
	Group 1	Group 2	Group 1	Group 2
N	126 (64)	74 (36 males)	28 (14 males)	23 (15 males)
Age	64,06 (3,01)	74,26 (3,65)	64,24 (3,39)	75,57 (3,92)
CAMBIOS Test	15,52 (5,50)	11,12 (5,74)	14,86 (6,30)	9,13 (4,70)

Note. Group 1: Younger older adults. Group 2: Middle-aged older adults.

Table 4

Percentile values for each level of performance in CF for the total sample (n = 200).

N	200
Mean (SD)	13,9 (5,97)
Minimum/Maximum:	2 / 27
Percentiles	
	10
	20
	25
	30
	40
	50
	60
	70
	75
	80

First, to obtain the normative values of the CAMBIOS test in older Rioplatense adults, the performance of the total sample (N = 200) is described, differentiated by quartiles (25, 50, 75 and 100). In this way, the CF performance of individuals is known across levels from low (quartile 1, 0 to 25) to high (quartile 4, 75 to 100). In addition, percentile values of performance level were obtained for the entire sample (Table 4). Second, percentile values of CF performance level differentiated by age groups were obtained (Table 5).

Table 5

Percentile values for each level of performance in CF, according to age groups.

PP	Group 1 (N= 126)	Group 2 (N= 74)
10	8-10	4-5
20	11	6
25	12	7
30	13	8
40	14	9
50	15-16	10-11
60	17	12-13
70	18-19	14
75	20	15
80	21-23	16-19
90		

Note. Group 1: Younger older adults. Group 2: Middle-aged older adults.

Study 2: Validity Evidence of the CAMBIOS Test for Cognitive Flexibility.

This study aims to provide evidence of the concurrent and age-related validity of the CAMBIOS test in a sub-sample of Rioplatense older adults.

2.1 External (concurrent) validity evidence of the CAMBIOS Test based on correlations with the Finger Task of the TAC computerized battery.

The Pearson correlation coefficient was applied between the following indices: number of

Table 6

Values of Centrality and Variability of cognitive flexibility performance in the Finger Task of the TAC in the sample of Study 2, discriminated by age group.

	Study 2 (N= 50)	
	Group 1	Group 2
N	32 (17 males)	18 (8 males)
Age - M (SD)	63,72 (2,59)	74,78 (3,42)
Finger Task		
RT (ms) - M (SD)	853,69 (112,24)	961,61 (241,18)
Accuracy (%) - M (SD)	96,31 (9,81)	87,94 (16,85)
IE- M (SD)	9,09 (2,53)	11,58 (4,42)

Note. Group 1: Younger older adults. Group 2: Middle-aged older adults. IE: Inverse Efficiency Index (IE): RT/Accuracy. The higher IE index value, the worse performance of CF.

correct responses in the CAMBIOS test, number of correct responses, average response times, and the Inverse Efficiency Index of the Finger Task in the TAC, within a subsample of 50 individuals. In all cases, statistically significant correlations were observed, albeit moderate in strength. Specifically, a) a negative correlation ($r = -.37$; $p < .01$) was found between correct responses on the CAMBIOS test and response times on the Finger Task, while b) a positive correlation ($r = .41$; $p < .01$) was found between correct responses on the CAMBIOS test and correct responses on the Finger Task. Finally, c) a negative correlation ($r = -.45$; $p < .01$) was found between correct responses on the CAMBIOS test and the Inverse Efficiency Index of the Finger Task. Thus, individuals who achieved a higher number of correct responses on the CAMBIOS test tended to have a lower Inverse Efficiency Index on the Finger Task, indicating better cognitive flexibility performance. Consistent with theoretical expectations, statistically significant correlations were found between the CAMBIOS test scores and the main performance indices of cognitive flexibility in the Finger Task. This suggests that both tests measure cognitive flexibility in the same direction, in a way that individuals who demonstrate high levels of cognitive flexibility on the CAMBIOS also

demonstrate high levels of cognitive flexibility on the TAC Finger Task. These findings provide empirical support for concurrent validity.

Regarding the intergroup analysis of cognitive flexibility performance on the finger task, the younger older adults group showed higher performance (shorter response times and higher number of correct responses) compared to the middle-aged older adults group. The average Inverse Efficiency Index for the group of younger older adults was 9.09 (2.53), while for the group of middle-aged older adults was 11.58 (4.42). It can be observed that the younger older adults were more efficient in their responses with shorter response times compared to the middle-aged older adults. In summary, the level of cognitive flexibility performance is higher in younger older adults than in middle-aged older adults, with an increase in response times and number of errors in the older age group (Table 6).

2.2. Evidence of age-related validity of the CAMBIOS Test in older Rioplatense subjects

To determine differences in CF performance on the CAMBIOS Test between the two age groups, an independent sample T-test was performed, which revealed statistically significant differences ($t = 5.37$; $df = 198$; $p < .01$). Similarly, in the performance on the Finger Task, statistical-

ly significant differences were found in all three indices: accuracy ($t = -2.22$; $df = 48$; $p < .01$), response times ($t = -2.16$; $df = 48$; $p < .01$), and inverse efficiency index ($t = -2.53$; $df = 48$; $p < .01$), with the younger elderly group achieving higher performance. Therefore, it is evident that the performance in CF is higher in younger elderly compared to middle-aged elderly.

Discussion

The present study has shown that the CAMBIOS test and the finger task of the computerized TAC battery are useful for assessing the level of cognitive flexibility in the elderly Rioplatense population, specifically and relatively independent of other executive functions. Older individuals who performed well on the TAC computerized task also performed better on the CAMBIOS test. Conversely, those older individuals who showed low levels of cognitive flexibility on the CAMBIOS also showed poorer performance on the computerized task. In conclusion, these results provide evidence of the external validity of the adaptation of the CAMBIOS Test in the Rioplatense population of elderly individuals and contribute to regional normative values for this instrument. Thus, the CAMBIOS test appears to be a valid tool for assessing cognitive flexibility in elderly individuals without neurocognitive pathology.

When comparing the results obtained from both age groups, statistically significant differences were found in all three indices assessed in the finger task (accuracy, response time, and inverse efficiency index). This is consistent with findings from recent years (Introzzi et al., 2020; Richard's et al., 2023; Richard's et al., 2019), where older adults are not considered a homogeneous group. Within this life stage, distinct subgroups with dif-

ferent characteristics emerge. In the case of the CAMBIOS test, statistically significant age-related differences were also observed. Younger older adults outperformed middle-aged adults, as expected given the specific age-related changes in brain activity during executive cognitive control (Spreng et al., 2017).

However, current research cautions that there are no uniform changes in all cognitive domains and functions with advancing age. Environmental factors, experience, and lifestyle can lead to significant differences in cognitive performance in old age (Barulli & Stern, 2013; Bartrés-Faz et al., 2018; Krzemien, 2016; Krzemien et al., 2018b; Oosterhuis et al., 2023). Given the interest in studying interindividual differences in cognitive aging trajectories that modulate the efficacy of executive functions, neurocognitive assessment techniques that allow the discrimination of specific performance indices related to each cognitive domain or function and that are sensitive to dynamic changes during aging are needed. In this regard, the CAMBIOS test makes it possible to obtain performance indices of cognitive flexibility that discriminate between age groups (young older adults and middle older adults) within the older age range. It would be of interest in future studies to compare performance in samples from other stages of the life span, such as young adulthood, middle adulthood, and late adulthood.

Both the Finger Task and the CAMBIOS test specifically assess cognitive flexibility, with increasing levels of complexity as the task progresses from simple to more complex stimulus presentation. The Finger Task provides specific indices of processing speed, while the CAMBIOS test considers responses given within a maximum time limit for older adults.

On the other hand, this study did not find statistically significant differences in CF performance related to gender and educational level

among older adults. Similarly, in the normative study of the Trail Making Test in the Buenos Aires population, no differences in educational level were found in the group of individuals aged 65 years and older (Margulis et al., 2018). Therefore, formal education does not seem to affect neuropsychological functioning in terms of CF in this sample. This could be because this process is more closely related to neurobiological and neurocognitive aspects and less influenced by educational learning and gender connotations.

The normative data provided in this study can be considered useful for the appropriate interpretation of the accuracy within a specific performance time for older adults in the Rioplatense region. The use of this CAMBIOS test can provide diagnostic guidance in gerontological professional practice. These results will contribute to the design of interventions and cognitive training programs aimed at optimizing cognitive flexibility, promoting the improvement of executive function and preventing the risk of neurodegenerative pathology in old age. It is recommended that future studies expand the sample to include diverse cultural contexts and further develop the adaptation of the CAMBIOS instrument as a gerontological cognitive assessment tool, which will contribute to increase the knowledge about this executive function in old age.

References

- Aguilar-Mateu, K., Arrabal-Guzman, M. A., & Herrera-Jimenez, F. (2014). Función ejecutiva en adultos mayores con patologías asociadas a la evolución del deterioro cognitivo. *Neuropsicología Latinoamericana*, 6(2), 7-14. https://neuropsicolatina.org/index.php/Neuropsicologia_Latinoamericana/index
- Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (2007). Life-span theory in developmental psychology. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical Models of Human Development* (6th ed.), (pp. 569-664). Wiley. <http://doi.org/10.1002/9780470147658.chpsy0111>
- Bartrés-Faz, D., Cattaneo, G., Solana, J., Tormos, J. M., & Pascual-Leone, A. (2018). Meaning in life: Resilience beyond reserve. *Alzheimer's Research & Therapy*, 10(1). <http://doi.org/10.1186/s13195-018-0381-z>
- Barulli, D., & Stern, Y. (2013). Efficiency, capacity, compensation, maintenance, plasticity: Emerging concepts in cognitive reserve. *Trends in Cognitive Sciences*, 17(10), 502-509. <http://doi.org/10.1016/j.tics.2013.08.012>
- Bell-McGinty, S., Podell, K., Franzen, M., Baird, A. D., & Williams, M. J. (2002). Standard measures of executive function in predicting instrumental activities of daily living in older adults. *Geriatric Psychiatry*, 17(9), 828-834. <http://doi.org/10.1002/gps.646>
- Betancourt-Zambrano, S. V., Tubay-Moreira, M. F., Cedeño-Yépez, M. E., & Caicedo-Chambers, K. M. (2020). Envejecimiento activo y las funciones ejecutivas en adultos mayores de un centro de salud. *Journal of Business and Entrepreneurial Studies*, 4(1). <https://www.journalbusinesses.com/index.php/revista/issue/archive>
- Bruno, D., Slachevsky, A., Fiorentino, N., Ruedaa, D. S., Bruno, G., Taglee, A. R., Olavarria, L., Flores, P., Lillo, P., Roca, M., & Torralva, T. (2020). Validación argentino-chilena de la versión en española del test Addenbrooke's Cognitive Examination III para el diagnóstico de demencia. *Neurología*, 35(2), 82-88. <http://doi.org/10.1016/j.nrl.2017.06.004>
- Burke, S. N., Mormino, E. C., Rogalski, E. J., Kawas, C. H., Willis, R. J., & Park, D. C. (2019). What are the later life contributions to reserve, resilience, and compensation? *Neurobiology of Aging*, 83, 140-144. <https://doi.org/10.1016/j.neurobiolaging.2019.03.023>
- Clare, L., Wu, Y., Teale, J. C., MacLeod, C., Matthews, F., Brayne, C., & Woods, B. (2017). Potentially modifiable lifestyle factors, cognitive reserve, and cognitive function in later life: A cross-sectional study. *PLOS*

- Medicine*, 14(3). <http://doi.org/10.1371/journal.pmed.1002259>
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44(11), 2037-2078. <http://doi.org/10.1016/j.neuropsychologia.2006.02.006>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64(1), 135-68. <http://doi.org/10.1146/annurev-psych-113011-143750>
- Dorbath, L., Hasselhorn, M., & Titz, C. (2013). Effects of education on executive functioning and its trainability. *Educational Gerontology*, 39(5), 314-325. <http://doi.org/10.1080/03601277.2012.700820>
- Ferguson, H. J., Brunsdon, V. E. A., & Bradford, E. E. F. (2021). The developmental trajectories of executive function from adolescence to old age. *Scientific Report*, 11(1). <https://doi.org/10.1038/s41598-020-80866-1>
- García-Coni, A., & Vivas, J. (2014). Estrategias ejecutivas de búsqueda, recuperación y cambio en la fluidez verbal. *Evaluar*, 14(1). <https://doi.org/10.35670/1667-4545.v14.n1.11520>
- George, D., & Mallery, M. (2003). *Using SPSS for Windows step by step: A simple guide and reference (4th ed.)*. Allyn & Bacon/Pearson.
- Grandi, F., & Tirapu-Ustárroz, J. (2017). Neurociencia cognitiva del envejecimiento: Modelos explicativos. *Revista Española de Geriatria y Gerontología*, 52(6), 326-331. <http://doi.org/10.1016/j.regg.2017.02.005>
- Hakun, J. G., Zhu, Z., Johnson, N. F., Gold, B. T. (2015). Evidence for reduced efficiency and successful compensation in older adults during task switching. *Cortex*, 64, 352-362. <https://doi.org/10.1016/j.cortex.2014.12.006>
- Herrera, D. M. & Eizaguirre, M. B. (2020, 25-27 de noviembre). *Flexibilidad cognitiva en los adultos mayores* [Presentación de Póster]. XII Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXVII Jornadas de Investigación. XVI Encuentro de Investigadores en Psicología del MERCOSUR. II Encuentro de Investigación de Terapia Ocupacional II Encuentro de Musicoterapia. Facultad de Psicología - Universidad de Buenos Aires. Buenos Aires, Argentina. <http://jimemorias.psi.uba.ar/index.aspx?anio=2020>
- Introzzi, I. & Canet Juric, L. (2019). TAC: Tareas de Autorregulación Cognitiva [Software y manual de usuario]. <https://tac.com.ar>
- Introzzi, I., Richard's, M., & Canet- Juric, L. (2019). *Diseño de una herramienta informatizada para la evaluación de las funciones ejecutivas durante el curso vital*. *Revista Argentina de Neuropsicología*, 35 (pp. 37-127). Resúmenes Trabajos libres-Formato Póster XIII Congreso Argentino de Neuropsicología 2018. Resúmenes Trabajos Libres XII Congreso Argentino de Neuropsicología, 4 al 6 de Octubre de 2018. Mendoza, Argentina.
- Introzzi, I. M., Richard's, M. M., García-Coni, A., Aydmune, Y., Stelzer, F., Canet-Juric, L., Zamora, E. V., Andrés, M. L., López Ramón, M. F., & Navarro-Pardo, E. (2020). Global versus controlled functioning throughout the stages of development. *Symmetry*, 12(12). <https://doi.org/10.3390/sym12121952>
- Ionescu, T. (2017). The variability-stability-flexibility pattern: A possible key to understanding the flexibility of the human mind. *Review of General Psychology*, 21(2), 123-131. <http://doi.org/10.1037/gpr0000110>
- Klein, R. M., Ivanoff, J., & Christie, J. (2004, 18-21 november). *Graphical and other methods for representing the speed and accuracy of performance* [Poster presentation]. The 45th annual meeting of the Psychonomic Society, Minneapolis, USA. <https://www.psychonomic.org/page/pastmeetings>
- Krzemien, D. (2013). El campo multidisciplinario de la Gerontología: Debate actual sobre demografía, desarrollo social e investigación del envejecimiento. *Académica Española*.
- Krzemien, D. (2016, 26-27 de agosto). *Estudio comparativo de la flexibilidad cognitiva, la reserva cognitiva y el pensamiento postformal en la adultez y vejez*.

- XV Congreso Argentino de Geriátría y Gerontología. Buenos Aires, Argentina.
- Krzemien, D., Richard's, M. M., & Biscarra, M. A. (2018a). Conocimiento experto y autorregulación en adultos mayores jubilados profesionales y no profesionales. *Avances en Psicología Latinoamericana*, 36(2), 331-344. <https://doi.org/10.12804/revistas.urosario.edu.co/apl/a.4793>
- Krzemien, D., Richard's, M. & Vido, V. (2018b, 30-31 de octubre). *Flexibilidad cognitiva en adultos y adultos mayores* [Presentación de póster]. I Jornadas Investigar. Universidad Nacional de Mar del Plata. Mar del Plata, Buenos Aires, Argentina. <https://www.mdp.edu.ar/index.php/investigacion/238-jornadas-investigar-unmdp>
- Lepe-Martínez, N., Cancino-Durán, F., Tapia-Valdés, F., Zambrano-Flores, P., Muñoz-Veloso, P., Gonzalez-San Martínez, I., & Ramos-Galarza, C. (2020). Desempeño en funciones ejecutivas de adultos mayores: Relación con su autonomía y calidad de vida. *Revista Ecuatoriana de Neurología*, 29(1), 92-103. <https://revecuatneurol.com>
- Margulis, L. E., Squillace, M. R., & Ferreres, A. R. (2018). Baremo del Trail Making Test para Capital Federal y Gran Buenos Aires. *Revista Argentina de Ciencias del Comportamiento*, 10(3), 54-63. <https://revistas.unc.edu.ar/index.php/racc/index>
- Martins, R., Joannette, Y., & Monchi, O. (2015). The implications of age-related neurofunctional compensatory mechanisms in executive function and language processing including the new temporal hypothesis for compensation. *Frontiers in Human Neuroscience*, 9. <http://doi.org/10.3389/fnhum.2015.00221>
- Mather, M. (2020). How do cognitively stimulating activities affect cognition and the brain throughout life? *Psychological Science in the Public Interest*, 21(1), 1-5. <http://doi.org/10.1177/1529100620941808>
- Meunier, D., Stamatakis, E. A., & Tyler, L. K. (2014). Age-related functional reorganization, structural changes, and preserved cognition. *Neurobiology Ageing*, 35(1), 42-54. <http://doi.org/10.1016/j.neurobiolaging.2013.07.003>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49-100. <http://doi.org/10.1006/cogp.1999.0734>
- Montero, L., & Leon, O. G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847- 862. <https://www.elsevier.es/es-revista-international-journal-clinical-health-psychology-355>
- Morales-Millán, K. L., Arroyo-Pérez, Y., González-Viruet, M., & Sánchez-Cardona, I. (2021). Relación entre medidas neuropsicológicas de ejecución y autoinforme de las funciones ejecutivas. *Evaluar*, 21(1), 53-72. <https://doi.org/10.35670/1667-4545.v21.n1.32832>
- Ojeda, V., Carvajal, C., Painevilu, S., & Zerpa, C. (2019). Desempeño de las funciones ejecutivas según estado cognitivo en adultos mayores. *Revista chilena de neuropsiquiatría*, 57(3), 207-214. <https://doi.org/10.4067/S0717-92272019000300207>
- Oosterhuis, E. J., Slade, K., May, P. J. C., & Nuttall, H. E. (2023). Toward an understanding of healthy cognitive aging: The importance of lifestyle in cognitive reserve and the scaffolding theory of aging and cognition. *The journals of Gerontology: Series B*, 78(5), 777-788. <https://doi.org/10.1093/geronb/gbac197>
- Padua, J. (Ed.) (1982). Muestreo. En *Técnicas de Investigación aplicadas a las Ciencias Sociales* (pp. 65-86). Fondo de cultura económica.
- Reuter-Lorenz, P. A., & Park, D. C. (2014). How does it STAC Up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychology review*, 24(3), 355-370. <http://doi.org/10.1007/s11065-014-9270-9>
- Richard's, M. M., Krzemien, D., Vido, V., Vernucci, S., Zamora, E. V., Comesaña, A., García-Coni, A., & Introzzi, I. (2019). Cognitive flexibility in adulthood and advanced age: Evidence of internal and external validity. *Applied neuropsychology: Adult*, 28(4), 464-478. <http://doi.org/10.1080/23279095.2019.1652176>

- Richard's, M., & Marino, J. (2021). El cambio y sus costos como indicadores de la Flexibilidad Cognitiva: Un proceso esencial para la autorregulación en nuestra vida. En I. M. Introzzi & L. Canet Juric (Eds.), *Funciones Ejecutivas. Definición conceptual, áreas de implicancia, evaluación y entrenamiento*. Neuroaprendizaje Infantil, 1ra ed. - ISBN 978-987-8910-09-3 (pp. 99-120) .
- Richard's, M. M., Zamora, E. V., Aydmune, Y., Comesaña, A., Krzemien, D., Introzzi, I., Lopez-Ramón, M. F., & Navarro-Pardo, E. (2023). Age-related switching costs in adulthood: "All or None Hypothesis" corollaries. *Current Psychology*, 2. <https://doi.org/10.1007/s12144-023-04340-7>
- Romo-Galindo, D. A., Ortiz-Jimenez, X. A., Garcia-Garcia, M. A., Ramirez-Tule, C. (2015). Análisis de la inhibición y flexibilidad cognoscitiva en el adulto mayor. *Ciencia UANL*, 18(76) 56-61. <https://cienciauanl.uanl.mx/?p=5183>
- Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition*, 52(3), 326-333. [http://doi.org/10.1016/S0278-2626\(03\)00170-2](http://doi.org/10.1016/S0278-2626(03)00170-2)
- Rosselli, M., Jurado, M. B., & Matute, E. (2008). Las funciones ejecutivas a través de la vida. *Revista Neuropsicología, Neuropsiquiatría y Neurociencia*, 8(1), 23-46. <http://revistaneuociencias.com/index.php/RNNN/index>
- Salthouse, T. A. (2018). Why is cognitive change more negative with increased age? *Neuropsychology*, 32(1), 110-120. <http://doi.org/10.1037/neu0000397>
- Seisdedos, N. (2008). *Test de Flexibilidad Cognitiva (CAMBIOS)*. TEA.
- Siagian, M. L., Indarwati, R., & Lestari, P. (2020). Non-pharmacological therapy for the elderly to prevent dementia through cognitive stimulation therapy: A systematic review. *Jurnal Ners*, 15(1Sp), 221-229. <http://doi.org/10.20473/jn.v15i1Sp.19018>
- Spreng, R. N., Shoemaker, L., & Turner, G. R. (2017). Executive functions and neurocognitive aging. In E. Goldberg (Ed.), *Executive functions in health and disease* (pp.169-196). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-803676-1.00008-8>
- Stieger, M., & Lachman, M. E. (2021). Increases in cognitive activity reduce aging-related declines in executive functioning. *Frontiers in Psychiatry*, 12. <http://doi.org/10.3389/fpsy.2021.708974>
- Tornimbeni, S., Pérez, E., & Baldo, M. (2004). *Introducción a los Tests Psicológicos*. Brujas.
- Uddin, L. Q. (2021). Cognitive and behavioural flexibility: Neural mechanisms and clinical considerations. *Nature Reviews Neuroscience*, 22(3), 167-179. <https://doi.org/10.1038/s41583-021-00428-w>
- Van den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Hulshoff Pol, H. E. (2009). Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience* 29(23), 7619-7624. <https://doi.org/10.1523/jneurosci.1443-09.2009>
- Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology and Aging*, 25(2), 343-355. <https://doi.org/10.1037/a0017729>
- Vido, V., Richard's M. & Krzemien, D. (2018, 04-06 de octubre). *Análisis de la Flexibilidad Cognitiva en adultos y adultos mayores. Evidencias de validez interna y externa* [Poster]. XIII Congreso Argentino de Neuropsicología. Sociedad de Neuropsicología Argentina (SONEPSA), Mendoza, Argentina.
- Wang, L., Li, Y., Metzack, P., He, Y., & Woodward, T. S. (2010). Age-related changes in topological patterns of large-scale brain functional networks during memory encoding and recognition. *NeuroImage*, 50(3), 862-872. <https://doi.org/10.1016/j.neuroimage.2010.01.044>
- Wecker, N. S., Kramer, J. H., Hallam, B. J., & Delis, D. C. (2005). Mental flexibility: Age effects on switching. *Neuropsychology*, 19(3), 345-352. <https://doi.org/10.1037/0894-4105.19.3.345>

Mobile-Related Experiences Questionnaire: Validation for Adults from Buenos Aires

Cuestionario de Experiencias Relacionadas con el Móvil: Validación para adultos de Buenos Aires

Franco Martín Ruscio * ¹, Juliana B. Stover ²

1 - Faculty of Psychology, University of Buenos Aires, Buenos Aires, Argentina.

2 - CONICET, Buenos Aires, Argentina; Faculty of Psychology, University of Buenos Aires, Buenos Aires, Argentina.

Introduction
Methods
Results
Discussion
References

Recibido: 11/04/2023 Revisado: 31/05/2023 Aceptado: 15/06/2023

Abstract

The Mobile-Related Experiences Questionnaire is a test designed to assess the experiences of the Spanish population with mobile phones. It consists of 10 items that evaluate the dimensions of Conflict and Communicational and Emotional Use. The aim of this study is to validate the scale in a sample of the local adult population. The sample was composed of 541 adults from Buenos Aires (19.0% male, 80.2% female, 0.7% non-binary), between the ages of 18 and 76 years ($M = 28.1$, $SD = 9.76$). Data were collected using a Sociodemographic Data Survey, the Mobile-Related Experiences Questionnaire, the Generalized Problematic Internet Use Scale 2, and the Symptom Checklist 27. The bifactorial structure was verified by means of a confirmatory factor analysis. The internal consistency was adequate. Evidence of concurrent validity is reported, and reference values are established. In conclusion, this study provides evidence of the psychometric quality of the questionnaire in a local sample.

Keywords: *ICT, Internet, smartphone, validity, reliability*

Resumen

El Cuestionario de Experiencias Relacionadas con el Móvil es un instrumento diseñado para su uso con población española. Consta de 10 ítems que evalúan las dimensiones de Conflicto y Uso Comunicacional y Emocional. El objetivo de este trabajo es adaptar dicha escala para evaluar población adulta local. Se trabajó con 541 adultos de Buenos Aires (19.0% Hombres, 80.2% Mujeres, 0.7% No Binario), de entre 18 y 76 años ($M = 28.1$, $DE = 9.76$). Se recolectaron datos con una Encuesta de datos socio-demográficos, el Cuestionario de Experiencias Relacionadas con el Móvil, la Generalized Problematic Internet Use Scale 2 y el Symptom Checklist 27. Mediante un análisis factorial confirmatorio se verificó la estructura bifactorial. La consistencia interna fue adecuada. Se reportan evidencias de validez concurrente. Se confeccionaron baremos. Se concluye que se aporta un instrumento de evaluación con evidencias de calidad psicométrica en una muestra local.

Palabras clave: *ICT, Internet, smartphone, validez, fiabilidad*

*Correspondence to: julianastover@psi.uba.ar, Gral. Juan Lavalle 2353 (C1052AAA), Buenos Aires, Argentina. Teléfono.: +54.11.4952.5481/5490.

How to cite: Ruscio, M. F., & Stover, B. J. (2023): Mobile-Related Experiences Questionnaire: Validation for Adults from Buenos Aires. *Revista Evaluar*, 23(3), 1-16. Retrieved from <https://revistas.unc.edu.ar/index.php/revaluar>

Participaron en la edición de este artículo: Agustina Mangieri, Stefano Macri, Eugenia Barrionuevo, Pablo Carpintero, Florencia Ruiz, Ricardo Hernandez, Jorge Bruera.

Introduction

The mobile phone has become an indispensable object in everyday life, not only due to its communication capabilities but also thanks to its easy access to the internet, its ability to take photographs, record videos, search and share information, perform transactions, and even work and study through the device (Herrera-Batista, 2009; van Velthoven et al., 2018). Its presence in people's daily lives has transformed the ways in which individuals interact with one another and their relationship with technology. The use of mobile phones has brought about changes and improvements in communication, social inclusion, economic activities, and productivity in various sectors (Pepper et al., 2012).

According to a survey conducted by the National Institute of Statistics and Census (INDEC) on the use and access to information and communication technologies (ICT) in Argentina, 88.1% of individuals use a mobile phone, while 87.2% have internet access. In the specific case of Buenos Aires, these numbers increase: Over 92% utilize the internet, and more than 94% of citizens use mobile phones, making them the most widely used technological device in the region, surpassing computers. These data indicate a high adoption of mobile devices among urban dwellers (INDEC, 2021).

In recent years, the focus has shifted towards the potential problems that excessive cellphone use can cause in people's daily lives, affecting academic, occupational, social, and family performance, and consequently, their relationships with others (Busch & McCarthy, 2021; Rodríguez-Ceberio et al., 2019).

In the field of Psychology, it is well-known that addiction or dependency is not solely limited to substance consumption, such as drugs or alcohol. Certain behavioral patterns can also lead to

dependency issues, and one of the most recent and highly specific forms of behavioral dependency is related to the dysfunctional use of ICT (Billieux et al., 2015; Chóliz et al., 2017). Several authors suggest refraining from using the term *addiction* when referring to *pathological* or *abusive* use of new technologies like the cellphone, as excessive and dependent technology use does not solely fit the addictive behavior model, as is the case with psychoactive substances (Carbonell et al., 2012b). Nevertheless, the excessive use of these technologies has caused numerous problems in individuals' family, academic, occupational, and personal domains (van Velthoven et al., 2018). As early as 1995, Griffiths addressed this issue proposing the concept of technological addictions, which are highly prevalent today through various mediums and forms. In the case of ICT, the problem lies not so much in the frequency of use, but in the dependency relationship it creates, leading to a loss of control over one's behavior and significant interference in daily life. Consequently, individuals with abusive cellphone use experience physical and psychological issues, such as anxiety, palpitations, and sweating, particularly when unable to use their device, for example, when forgetting it at home or experiencing low balance, coverage or battery (Echeburúa & de Corral, 2010).

Regarding the symptomatic manifestation of pathological mobile phone use, Chóliz et al. (2017) argue that, similar to pathological gambling and other dependencies, problematic use of these electronic devices is characterized by: a strong need to use the technology to achieve the same level of well-being as initially experienced (tolerance); various negative emotional reactions that arise when unable to use the device or during extended periods without using it (withdrawal); difficulties in daily life and across different domains due to the abusive use; and the inability to stop using the device, even though the individual

recognizes that its use may be inappropriate or potentially dangerous.

Currently, this problem is not considered a mental disorder according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Although a new category called “Addictive and Substance-Related Disorders” is included, only pathological gambling disorder has been recognized as a behavioral addiction (American Psychiatric Association [APA], 2014). The same applies to other diagnostic manuals, such as the International Classification of Diseases and Related Health Problems (ICD) by the World Health Organization (WHO, 2019). In any case, the pathological use of cell phones could be seen with the same characteristics of dependency disorders, but without intoxication by any substance (Redondo-Pacheco et al., 2016). However, despite not being classified as a disorder *per se*, due to the consequences of excessive technology use, researchers have shown interest in determining the incidence, prevalence, and causes of problematic ICT use, particularly in relation to internet and mobile phone use. Emphasis has been placed on the importance of having tools to identify where excessive use occurs and when it may become pathological (Simó-Sanz et al., 2017).

Various studies have been conducted on mobile phone use. Taking gender into account, the results indicate that women engage in more maladaptive use (Beranuy et al., 2009b; Carbonell et al., 2012a; Chóliz et al., 2009; Labrador-Encinas & Villadangos-González, 2010; López-Fernández et al., 2012; Sánchez-Martínez & Otero, 2009). In the case of Argentina, a study conducted in 2020 found that women scored significantly higher than men in aspects of withdrawal and loss of control in mobile phone use. While both men and women viewed the use of technological devices positive for staying connected with their families, women reported a higher frequency of

use than men and displayed greater dependence on the device (Duraó et al., 2021).

The abusive use of the device has been associated with different problems (e.g., Sohn et al., 2019), such as lower self-esteem (Bianchi & Phillips, 2005; Casale et al., 2022; Graben et al., 2020), feelings of loneliness (Menglong & Liya, 2017), anxiety (Foerster et al., 2015; Richardson et al., 2018), depression (Elhai et al., 2017; Güzeller & Coşguner, 2012; Rozgonjuk et al., 2018), sleep disorders (Xie et al., 2018) and lower psychological well-being (Horwood & Anglim, 2019). Other authors have investigated the harmful effects associated with different forms of mobile harassment (Turan et al., 2011), as well as an increased risk of occupational and traffic accidents related to inattention and sensory interference during its use (Backer-Grøndahl & Sagberg, 2011).

Assessment Instruments for Mobile Phone Use

Currently, there are various assessment instruments available for measuring variables related to mobile phone use (Harris et al., 2020). Toda et al. (2006) developed the Mobile Phone Dependence Questionnaire (MPDQ), consisting of 20 items that measure the degree of dependence on the cellphone. Bianchi and Phillips (2005) created the Mobile Phone Problem Use Scale (MP-PUS), comprising 27 items related to social problems arising from cell phone use. There is also an adaptation of the scale for use with Spanish adolescents (López-Fernández et al., 2012) and Ecuadorian university students (García-Umaña & Córdoba-Pillajo, 2020). Another instrument is the Problematic Mobile Phone Use Questionnaire (PMPUQ; Billieux et al., 2008), which assesses different types of mobile phone use through 30 items. Chóliz (2012) designed the Mobile Phone Dependency Test (TDM), originally consisting of

22 items that examine the main characteristics of cell phone dependence. In 2016, the same author selected the most representative items from the original TDM version and developed a reduced version (TDMB) consisting of 12 items. This version was adapted for the Argentine context (Duro et al., 2021) for use with students and has been validated for use with Peruvian (Gamero et al., 2016) and Brazilian populations (Flores-Robaina et al., 2013).

Another scale is the Problematic Use Mobile Phone (PUMP) developed by Merlo et al. (2013). It consists of 20 items that measure symptoms of problematic cellphone use, based on the substance addiction criteria of the DSM-5.

Lastly, the Mobile Phone-Related Experiences Questionnaire (CERM; Beranuy-Farques et al., 2009a) is an instrument that assesses problematic mobile phone use using 10 Likert scale items ranging from 1 (*Not at all*) to 4 (*Very much*). The items are related to negative effects on behavior, social relationships, reduction of activities due to mobile phone use, loss of control, and intense desire to be connected. As for validation, the instrument was administered to a sample of 1879 students from Catalan educational centers, including university and secondary education students aged 12 to 25. An exploratory factor analysis was conducted using the Generalized Least Squares (GLS) method with Oblimin rotation. The one-factor model showed an unsatisfactory fit, while the two-factor solution demonstrated acceptable data fit. Thus, the final version of the instrument is organized into two factors. The first factor (Conflicts) consists of 5 items explaining 34.6% of the variance, and the second factor (Communicational and Emotional Use) includes 5 items explaining 9.22% of the variance. The internal consistency of the dimensions estimated with Cronbach's alpha was .81 and .75, respectively. Concurrent validity studies were also

conducted with six dichotomous items used by de Gracia-Blanco et al. (2002) to assess problematic cellphone use, revealing positive and strong correlations for both the Conflict factor ($r = .44$) and the Communicational and Emotional Use factor ($r = .60$).

Psychometric studies of the scale were carried out in the Colombian population. A study was conducted with 639 students aged 12 to 25 (Redondo-Pacheco et al., 2016). The internal consistency of the total score yielded an α of .79. The same analyses were performed in the Chilean context (Inostroza et al., 2019). Working with 180 university students aged 18 to 30, Cronbach's alphas of .60 for Conflicts and .73 for Communicational and Emotional Use were found.

Due to its brevity, the CERM presents itself as a useful instrument for examining cellphone use through easy administration and evaluation. Since there are no studies available in the local population yet and considering the importance of having validated instruments in the country (Fernández et al., 2010), the objective of this study is to carry out the conceptual, linguistic, and metric adaptation of the CERM for its use with the adult population in Buenos Aires.

Methods

A non-experimental and cross-sectional design was conducted. The type of study carried out was correlational. A non-probabilistic sampling was employed.

Participants

The sample was composed of 541 adults residing in the City of Buenos Aires and its surroundings (19.0% males, 80.2% females, 0.7%

Table 1
Sociodemographic data. Frequencies.

	n	%
Gender		
Male	103	19.9%
Female	434	80.2%
Non-binary	4	0.7%
Marital status		
Single	270	49.9%
Dating (without living together)	201	37.2%
Dating (living together)	3	0.6%
Married	53	9.8%
Separated/Divorced	13	2.4%
Other	1	0.2%
Maximum educational level achieved		
Incomplete High-School Education	10	1.8%
Complete High-School Education	90	16.6%
Incomplete Tertiary Education	24	4.4%
Complete Tertiary Education	29	5.4%
Incomplete University Education	252	46.6%
Complete University Education	81	15.0%
Incomplete Postgraduate Education	20	3.7%
Complete Postgraduate Education	35	6.5%
Currently working		
Yes	371	68.6%
No	170	31.4%
Currently studying		
Yes	350	64.7%
No	191	35.3%
Socioeconomic status		
Low	19	3.5%
Lower-Middle	135	25.0%
Middle	308	56.9%
Upper-Middle	776	14.0%
High	3	0.6%
Residential location		
CABA	181	33.5%
Buenos Aires	360	66.5%

non-binary), ranging in age from 18 to 76 years ($M = 28.1$, $SD = 9.76$). Most participants were single at the time of the evaluation (49.9%). A high level of education was observed, with 81.6% of participants having at least some university or tertiary education. Additionally, 64.7% of the sample reported currently being enrolled in studies. Regarding their employment status, more than half of the participants were employed (68.6%). Many participants identified with a middle socio-economic level (56.9%). Table 1 provides a detailed overview of all socio-demographic data.

Additionally, a test-retest study was conducted with a subsample of 21 participants (85.7% females, 14.3% males; $M_{age} = 26.9$, $SD_{age} = 6.98$).

Material

Sociodemographic Data Survey. Designed specifically for this study, this survey collects information on gender, age, marital status, employment status, educational status, level of education, cohabitation group, and perceived socioeconomic level.

Mobile Phone-Related Experiences Questionnaire (CERM; Beranuy-Fargues et al., 2009a).

Consisting of 10 items with a Likert response format ranging from *Not at all* to *Very much*, this questionnaire assesses mobile phone-related problematic use. It includes two dimensions: Conflict and Communicational and Emotional Use. Adaptation of this instrument is the focus of this study.

Generalized Problematic Internet Use Scale 2 (GPIUS2; Caplan, 2010). This scale evaluates generalized problematic internet use. It consists of 15 items with a Likert response format ranging from *Strongly disagree* to *Strongly agree*. The scale is composed of four subscales: Preference for Online Social Interaction (POSI), Mood Regulation

(MR), Negative Outcomes (NO), and Deficient Regulation (DR), which includes the subscales of Compulsive Use (CU) and Cognitive Preoccupation (CP). An adaptation for use with the population of Buenos Aires was conducted (Stover et al., 2023). Previous studies have reported Cronbach's alphas ranging from .83 to .87 for its subscales, indicating good internal consistency (Stover & Tisocco, 2020).

Symptom Checklist 27 (SCL-27; de la Iglesia & Castro-Solano, 2019; Hardt & Gerbershagen, 2001). This scale assesses non-specific psychopathological symptoms experienced in the past seven days. It consists of 27 items grouped into a single dimension, which was confirmed through confirmatory factor analysis. The response format is a 5-point Likert scale ranging from *Not at all* to *Very much*. It has adequate internal consistency ($\alpha = .92$).

Procedure

Initially, a pilot administration was conducted with 10 adults from Buenos Aires to assess facies validity. As a result, the term "mobile" was modified to "cell phone" in the items. Subsequently, an expert judgment procedure was conducted with five experts in psychometrics to evaluate the proper operationalization of the construct through the items. The judges reached 80% to 100% agreement on all items, and no modifications were suggested. After these steps, the standardization sample was collected using a Google form. The form was disseminated through social networks and email lists of participants from previous research studies. Informed consent was obtained, guaranteeing anonymity and confidentiality of the responses, and the research objectives were explained.

Table 2
CERM items. Descriptive statistics.

Item	M	SD	Skewness	Kurtosis
1. Have you been at risk of losing an important relationship, a job or an academic opportunity due to the use of your cell phone?	1.27	.55	2.30	5.72
2. Do you think that your academic or work performance has been negatively affected by cell phone use?	2.12	.92	0.33	-0.85
3. Do you suffer from sleep disturbances due to aspects related to your cell phone?	2.25	.99	0.24	-1.02
4. Do you feel the need to invest more and more time on your cell phone to feel satisfied?	1.73	.90	0.96	-0.14
5. Do you stop going out with your friends because you spend more time using your cell phone?	1.15	.47	3.66	14.75
6. To what extent do you feel restless when you do not receive messages or calls?	1.87	.85	0.72	-0.19
7. When you get bored, do you use your cell phone as a form of distraction?	3.45	.72	-1.31	1.60
8. How often do you say things on your cell phone that you would not say in person?	2.16	.89	0.30	-0.73
9. Do you think that life without the cell phone is boring, empty and sad?	1.87	.87	0.76	-0.18
10. Do you get angry or irritated when someone bothers you while you use your cell phone?	1.70	.810	1	0.37

Table 3
Adjustment indices. CERM models.

	S-B χ^2	gl	S-B χ^2 /gl	CFI	IFI	RMSEA [90% CI]
2 Factors	132.228	34	3.889	.948	.949	.073 [.060, .086]
1 Factor	127.871	35	3.653	.872	.874	.070 [.057, .083]

As an initial step, univariate and multivariate normality were analyzed using descriptive statistics (mean, standard deviation, skewness, and kurtosis) and the Mardia coefficient to examine multivariate normality. Due to the violation of the multivariate normality assumption (Tabachnick & Fidell, 2013), a confirmatory factor analysis (CFA) was conducted using the Maximum Likelihood (ML) method with robust estimators, considering polychoric matrices due to the ordinal nature of the variables (Freiberg-Hoffmann et al., 2013; Muthén & Kaplan, 1985). CFA was chosen because previous information about the distribution of the data from the Spanish version was available and because CFAs are less dependent on sample characteristics than exploratory factor analyses (Richaud de Minzi, 2008). Unifactorial and bifactorial models were tested based on previous analyses of the original version (Beranuy-Fargues et al., 2009a). The analyses were performed using EQS 6.2 software (Bentler & Wu, 2012). Model fit was evaluated using various indices. Satorra-Bentler Chi-Square Correction (S-B χ^2 ; 2001) was estimated by dividing the degrees of freedom (S-B χ^2 /df), with values ranging from 1 to 3 considered appropriate (Carmines & McIver, 1981). Comparative Fit Index (CFI) and Bollen's Fit Index (IFI) cutoff was set at .90. Finally, Root Mean Square Error of Approximation (RMSEA) values between .05 and .08 were considered appropriate (Browne & Cudeck, 1993; Byrne, 2006; Hu & Bentler, 1999).

Ordinal alpha coefficients (Elosua-Oliden & Zumbo, 2008) and Cronbach's alphas were then estimated. The attenuation index between these

coefficients was calculated (Dominguez-Lara, 2018). Pearson correlations (r) were calculated between the factors of the CERM, GPIUS-2, and SCL-27. To examine the test-retest stability of the scores, the CERM was administered again to 21 participants after a 20-day interval. Pearson correlations were calculated between the scores of the dimensions. Effect size interpretations followed Cohen's guidelines: small = .10, medium = .30, large = .50 (Cohen, 1988). Independent samples t-tests were conducted to examine gender differences for the purpose of evaluating separate norms based on gender. Non-binary gender was not considered in this comparison due to the small number of participants (four individuals), representing 0.7% of the sample. The calculations were performed using Jamovi software (The Jamovi Project, 2023).

Results

Evidence of Construct Validity

Initially, skewness and kurtosis values of the items were calculated to evaluate univariate normality (Table 2). Most of the items had values within ± 1.96 . When assessing multivariate normality, the Mardia coefficient was 38.96. Therefore, robust estimators were chosen for the CFA (Tabachnick & Fidell, 2013).

The unifactorial and bifactorial models were compared by means of two CFAs carried out with the Maximum Likelihood method based on the polychoric matrices (Table 3). The S-B χ^2 /gl value was slightly lower in the unidimensional model.

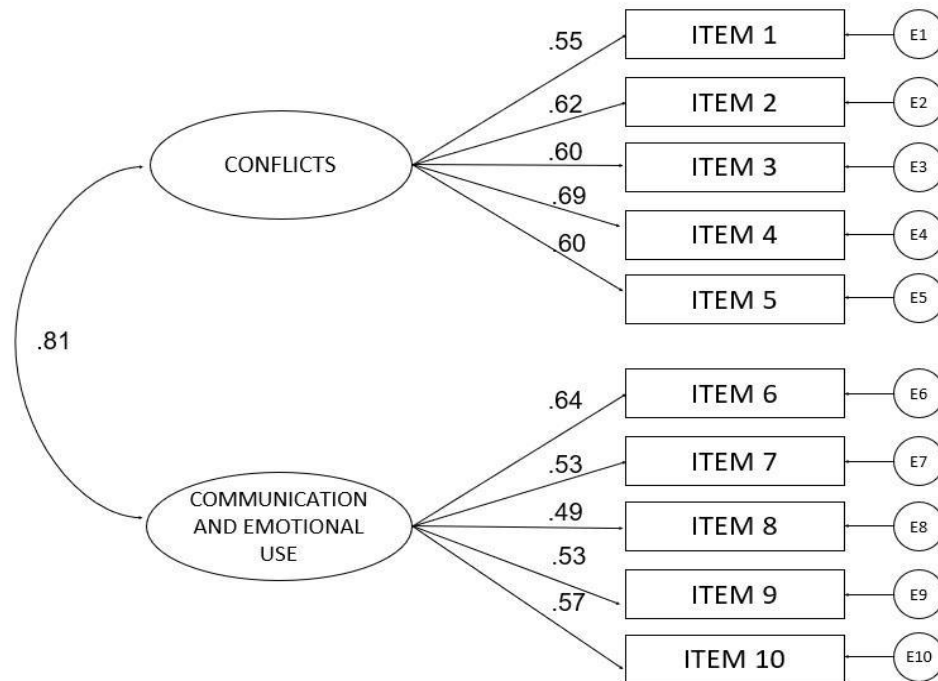


Figure 1
Parameters. CERM.

The CFI and IFI values exceeded the cutoff of .90 in the bifactorial model, but not in the unifactorial model. The RMSEA values were similar and within the expected range. Considering the CFI and IFI values, the decision was made to choose the two-factor model.

The parameters were statistically significant. The factor loadings ranged from .49 to .64. The correlation between the dimensions was .81 (Figure 1).

Internal Consistency

The internal consistency, estimated using Cronbach's alpha, was adequate, and when calculating ordinal alphas, it increased to .74 for Conflicts and .68 for Communication and Emotional Use. The attenuation index between both alphas was less than 30% (Table 4).

Table 4
CERM. Internal consistency.

	Ordinal Alpha	Cronbach's Alpha	Attenuation Index
Conflicts	.74	.69	8%
Communication and Emotional Use	.68	.66	4%

Table 5
CERM, GPIUS-2, SCL-27. Correlations.

	Conflicts	Communication and Emotional Use
Preference for Social Online Interaction	.30**	.38**
Negative Outcomes	.62**	.46**
Mood Regulation	.41**	.51**
Compulsive Use	.50**	.52**
Cognitive Preoccupation	.48**	.54**
SCL-27	.51**	.50**

Note. ** = $<.01$.

Evidence of Concurrent Validity with GPIU-2 and SCL-27

When calculating Pearson correlation between the dimensions of CERM and GPIUS-2, significant positive correlations were found in all cases. The highest effect sizes were observed between the Conflict factor and Negative Outcomes ($r = .62$) and Compulsive Use ($r = .50$). There were also significant positive correlations between the Communication and Emotional Use factor and Cognitive Preoccupation ($r = .54$), Compulsive Use ($r = .52$), and Mood Regulation ($r = .51$). When examining the relationship with

psychopathological symptoms, significant positive correlations with a strong effect size were found in both dimensions (Table 5).

Temporal Stability

Correlations were calculated between the total scores of each dimension at the two time points. A correlation coefficient of $r = .72$ was obtained for Conflict and $r = .82$ for Communication and Emotional Use.

Norms

Prior to establishing norms, an independent sample t-test was conducted to examine gender differences in cellphone use. No significant differences were found in either the Conflict dimension ($t = -1.30$, 535 *df*, $p = .195$, $d = -0.14$; $M_{\text{male}} = 8.20$ vs. $M_{\text{female}} = 8.59$) or the Communication and Emotional Use dimension ($t = -1.74$, 535 *df*, $p = .083$, $d = -0.19$; $M_{\text{male}} = 10.61$ vs. $M_{\text{female}} = 11.13$). Therefore, norms were developed for the entire sample (Table 6).

Table 6
CERM. percentiles.

Percentile	Conflicts	Communication and Emotional Use
10	5	8
20	6	9
25	6	9
30	7	10
40	7	10
50	8	11
60	9	12
70	10	12
75	10	13
80	11	13
90	12	15

Discussion

In the present study, the Mobile-Related Experiences Questionnaire (CERM; [Beranuy et al., 2009a](#)) was validated for use with the local adult population. When analyzing the dimensionality of the CERM and comparing unifactorial and bifactorial models through two confirmatory factor analyses (CFA), the latter was chosen. This structure replicates the original version ([Beranuy-Fargues et al., 2009a](#)), where better fit was found for the two-factor option compared to the unidimensional one. Regarding the internal consistency of the CERM, estimated through Cronbach's alpha, it was found to be adequate, and the ordinal alphas further increased. These findings are consistent with previous studies ([Beranuy-Fargues et al., 2009a](#); [Inostroza et al., 2019](#); [Redondo-Pacheco et al., 2016](#)).

To establish associations between mobile-related experiences and generalized problematic internet use, Pearson correlation coefficients were calculated between the CERM dimensions and the GPIUS-2, revealing positive correlations in all cases, providing evidence of concurrent criterion validity. These results align with previous studies that have also reported associations between cellphone use and internet use ([Beranuy et al., 2009a](#); [Inostroza et al., 2019](#)). Additionally, it is worth mentioning that one of the main functions that makes cellphones so attractive is its easy access to the internet ([Herrera-Batista, 2009](#)). Furthermore, when correlating CERM responses with those of the SCL-27 to establish concurrent validity evidence with non-specific psychopathological symptoms, significant positive correlations were also found. Various studies have previously linked cellphone use to psychopathological symptoms such as depression, social anxiety, and sleep disorders (e.g., [Elhai et al., 2017](#); [Foerster et al., 2015](#); [Güzeller & Coşguner,](#)

[2012](#); [Richardson et al., 2018](#); [Rozgonjuk et al., 2018](#); [Sohn et al., 2019](#); [Xie et al., 2018](#)). Finally, the temporal stability of the scores generated by the instrument was also adequate.

Due to previous research reporting gender differences ([Beranuy et al., 2009b](#); [Chóliz et al., 2009](#); [Durao et al., 2021](#); [Labrador-Encinas & Villadangos-González, 2010](#); [López-Fernández et al., 2012](#); [Sánchez-Martínez & Otero, 2009](#)), the means of men and women were compared for the two dimensions of the CERM. As no statistically significant differences were found, as the final step of this study, norms were established to enable the use of the adapted instrument in various applications.

In summary, this study contributes to technology transfer by providing a brief, easy-to-administer and evaluate instrument aimed at analyzing cellphone use in adults in Buenos Aires, a construct for which there is limited research in the country (e.g., [Cabañas & Korzeniowski, 2015](#); [Durao et al., 2021](#)). Given that cellphones are the most widely used technological device in the region ([INDEC, 2021](#)), it is of interest to understand their impact on daily life and people's health ([Busch & McCarthy, 2021](#); [Sohn et al., 2019](#); [van Velthoven et al., 2018](#)).

However, certain limitations should be taken into account when considering the results. Firstly, the sample is predominantly composed of women with a high level of education. Further studies are needed to confirm the psychometric quality of the instrument, such as examining factorial invariance across different samples.

Additionally, the fact that the questionnaires are self-administered could be controversial, as problematic behaviors tend to be concealed ([Redondo-Pacheco et al., 2016](#)). It would be interesting to expand the research with other forms of assessment, such as observation or the involvement of third parties to provide information on

the evaluated person's cellphone use, in order to analyze the behavior more comprehensively (e.g., Ryding & Kuss, 2020; Rozgonjuk et al., 2018). In future lines of research, it would be relevant to expand the study to other regions of the country to determine if the cellphone use situation in the Buenos Aires Metropolitan Area is replicated elsewhere in the country or to identify any differences that may exist.

References

- American Psychological Association [APA]. (2014). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. APA.
- Backer-Grøndahl, A., & Sagberg, F. (2011). Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. *Safety Science*, 49(2), 324-330. <https://doi.org/10.1016/j.ssci.2010.09.009>
- Bentler, P. M., & Wu, E. J. C. (2012). EQS 6.2 for Windows. [software de cómputo]. Multivariate Software.
- Beranuy-Fargues, M., Chamarro-Lusar, A., Graner-Jordania, C., & Carbonell-Sánchez, X. (2009a). Validación de dos escalas breves para evaluar la adicción a Internet y el abuso de móvil [Validation of two brief scales to assess Internet addiction and mobile phone abuse]. *Psicothema*, 21(3), 480-485. <https://www.psicothema.com/es>
- Beranuy, M., Oberst, U., Carbonell, X., & Chamarro, A. (2009b). Problematic Internet and mobile phone use and clinical symptoms in college students: The role of emotional intelligence. *Computers in Human Behaviour*, 25(5), 1182-1187. <https://doi.org/10.1016/j.chb.2009.03.001>
- Bianchi, A., & Phillips, J. G. (2005). Psychological predictors of problem mobile phone use. *CyberPsychology & Behavior*, 8(1), 39-51. <https://doi.org/10.1089/cpb.2005.8.39>
- Billieux, J., Maurage, P., Lopez-Fernandez, O., Kuss, D. J., & Griffiths, M. D. (2015). Can disordered mobile phone use be considered a behavioral addiction? An update on current evidence and a comprehensive model for future research. *Current Addiction Reports*, 2(2), 156-162. <https://doi.org/10.1007/s40429-015-0054-y>
- Billieux, J., Van der Linden, M., & Rochat, L. (2008). The role of impulsivity in actual and problematic use of the mobile phone. *Applied Cognitive Psychology*, 22(9), 1195-1210. <https://doi.org/10.1002/acp.1429>
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Sage.
- Busch, P. A., & McCarthy, S. (2021). Antecedents and consequences of problematic smartphone use: A systematic literature review of an emerging research area. *Computers in Human Behavior*, 114, 106414. <https://doi.org/10.1016/j.chb.2020.106414>
- Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming (2nd ed.)*. Lawrence Erlbaum.
- Cabañas, M., & Korzeniowski, C. (2015). Uso de celular e Internet: Su relación con planificación y control de la interferencia [Use of cell phones and the Internet: Their relationship with planning and control of interference]. *Revista Argentina de Ciencias Del Comportamiento [Argentine Journal of Behavioral Sciences]*, 7(1), 5-16. <https://revistas.unc.edu.ar/index.php/racc/index>
- Caplan, S. E. (2010). Theory and measurement of generalized problematic Internet use: A two-step approach. *Computers in Human Behavior*, 26(5), 1089-1097. <https://doi.org/10.1016/j.chb.2010.03.012>
- Carbonell, X., Chamarro, A., Griffiths, M., Oberst, U., Cladellas, R., & Talam, A. (2012a). Problematic Internet and cell phone use in Spanish teenagers and young students. *Anales de Psicología [Annals of Psychology]*, 28(3), 789-796. <https://doi.org/10.6018/analesps.28.3.156061>
- Carbonell, X., Fúster, H., Chamarro, A., & Oberst, U.

- (2012b). Adicción a Internet y móvil: Una revisión de estudios empíricos españoles [Internet and mobile addiction: A review of Spanish empirical studies]. *Papeles del Psicólogo [Psychologist Papers]*, 33(2), 82-89. <https://www.papelesdelpsicologo.es>
- Carmines, E. G., & McIver, J. P. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. W. Bohrnstedt & E. F. Borgatta (Eds.), *Social Measurement: Current Issues* (pp. 65-115). Sage.
- Casale, S., Fioravanti, G., Bocci-Benucci, S., Falone, A., Ricca, V., & Rotella, F. (2022). A meta-analysis on the association between self-esteem and problematic smartphone use. *Computers in Human Behavior*, 134. <https://doi.org/10.1016/j.chb.2022.107302>
- Chóliz, M. (2012). Mobile-phone addiction in adolescence: The Test of Mobile Phone Dependence (TMD). *Progress in Health Sciences*, 2(1), 33-44. https://progressinhealthsciences.publisherspanel.com/resources/html/cms/EDITORIAL_OFFICE
- Chóliz, M., Echeburúa, E., & Ferre, F. (2017). Screening tools for technological addictions: A proposal for the strategy of mental health. *International Journal of Mental Health and Addiction*, 15(2), 423-433. <https://doi.org/10.1007/s11469-017-9743-1>
- Chóliz, M., Villanueva, V., & Chóliz, M. C. (2009). Ellas, ellos y su móvil: Uso, abuso (¿y dependencia?) del teléfono móvil en la adolescencia [They and their mobile: Use, abuse (and dependence?) of the mobile phone in adolescence]. *Revista Española de Drogodependencias [Spanish Magazine on Drug Addiction]*, 34(1), 74-88. <https://www.aesed.com/es>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). LEA.
- de Gracia-Blanco, M., Vigo-Anglada, M., Fernández-Pérez, M. J., & Marcó-Arbonès, M. (2002). Problemas conductuales relacionados con el uso de Internet: Un estudio exploratorio [Behavioral problems related to the use of the Internet: an exploratory study]. *Anales de Psicología [Annals of Psychology]*, 18(2), 273-292. <https://revistas.um.es/analesps>
- de la Iglesia, G., & Castro-Solano, A. (2019). Academic achievement of college students: The role of the positive personality model. *Problems of Education in the 21st Century*, 77(5), 572-583. <https://doi.org/10.33225/pec/19.77.572>
- Dominguez-Lara, S. (2018). Fiabilidad y alfa ordinal [Reliability and ordinal alpha]. *Actas Urológicas Españolas [Spanish Urological Acts]*, 42(2), 140-141. <https://doi.org/10.1016/j.acuro.2017.07.002>
- Durao, M., Etchezahar, E., Ungaretti, J., & Calligaro, C. (2021). Propiedades psicométricas del Test de Dependencia al Teléfono Móvil (TDMB) en Argentina y sus relaciones con la impulsividad [Psychometric properties of the Mobile Phone Dependency Test (TDMB) in Argentina and its relationship with impulsivity]. *Actualidades en Psicología [News in Psychology]*, 35(130), 1-18. <https://doi.org/10.15517/ap.v35i130.41963>
- Echeburúa, E., & de Corral, P. (2010). Adicción a las nuevas tecnologías y a las redes sociales en jóvenes: Un nuevo reto [Addiction to new technologies and social networks in young people: a new challenge]. *Adicciones [Addictions]*, 22(2), 91-96. <https://doi.org/10.20882/adicciones.196>
- Elhai, J. D., Dvorak, R. D., Levine, J. C., & Hall, B. J. (2017). Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. *Journal of Affective Disorders*, 207, 251-259. <https://doi.org/10.1016/j.jad.2016.08.030>
- Elosua-Oliden, P., & Zumbo, B. D. (2008). Coeficientes de confiabilidad para escalas de respuesta categórica ordenada [Reliability coefficients for ordered categorical response scales]. *Psicothema*, 20(4), 896-901. <https://www.psicothema.com>
- Fernández, A., Pérez, E., Alderete, A. M., Richaud, M. C., & Fernández-Liporace, M. (2010). ¿Construir o Adaptar Tests Psicológicos? Diferentes respuestas a una cuestión controvertida [Build or Adapt Psychological Tests? Different Answers to a Controversial Question]. *Revista Evaluar [Evaluate Magazine]*,

- 10(1). <https://doi.org/10.35670/1667-4545.v10.n1.459>
- Freiberg-Hoffmann, A., Stover, J. B., de la Iglesia, G., & Fernández-Liporace, M. (2013). Correlaciones polichóricas y tetrachóricas en estudios factoriales exploratorios y confirmatorios [Polychoric and tetrachoric correlations in exploratory and confirmatory factorial studies in coexistence]. *Ciencias Psicológicas [Psychological Science]*, 7(2), 151-164. <https://doi.org/10.22235/cp.v7i1.1057>
- Flores-Robaina, N., Jenaro-Río, C., González-Gil, F., Martín-Pastor, E., & Poy-Castro, R. (2013). Adicción al móvil en alumnos de secundaria: Efectos en la convivencia [Mobile phone addiction in secondary school students: Effects on coexistence]. *European Journal of Investigation in Health, Psychology and Education*, 3(3), 215-225. <https://doi.org/10.3390/ejihpe3030019>
- Foerster, M., Roser, K., Schoeni, A., & Rössli, M. (2015). Problematic mobile phone use in adolescents: Derivation of a short scale MPPUS-10. *International Journal of Public Health*, 60(2), 277-286. <https://doi.org/10.1007/s00038-015-0660-4>
- Gamero, K., Flores, C., Arias, W. L., Ceballos, K. D., Román, A., & Marquina, E. (2016). Estandarización del Test de Dependencia al Celular para estudiantes universitarios de Arequipa [Standardization of the Cellular Dependency Test for Arequipa university students]. *Persona [Person]*, 19, 179-200. <https://doi.org/10.26439/persona2016.n019.979>
- García-Umaña, A., & Córdoba-Pillajo, É. (2020). Validación de escala MPPUS-A sobre el uso problemático del smartphone [Validation of scale MPPUS-A on the problematic use of the smartphone]. *Pixel-Bit. Revista de Medios y Educación [Pixel-Bit. Media & Education Magazine]*, 57, 173-189. <https://doi.org/10.12795/pixelbit.2020.i57.07>
- Graben, K., Doering, B. K., Jeromin, F., & Barke, A. (2020). Problematic mobile phone use: Validity and reliability of the Problematic Use of Mobile Phone (PUMP) Scale in a German sample. *Addictive Behaviors Reports*, 12, 100297. <https://doi.org/10.1016/j.abrep.2020.100297>
- Griffiths, M. D. (1995). Technological addictions. *Clinical Psychology Forum*, 1(76), 14-19. <https://doi.org/10.53841/bpscpf.1995.1.76.14>
- Güzeller, C. O., & Coşguner, T. (2012). Development of a Problematic Mobile Phone Use Scale for Turkish adolescents. *Cyberpsychology, Behavior and Social Networking*, 15(4), 205-211. <https://doi.org/10.1089/cyber.2011.0210>
- Hardt, J., & Gerbershagen, H. U. (2001). Cross-validation of the SCL-27: A short psychometric screening instrument for chronic pain patients. *European Journal of Pain*, 5(2), 187-197. <https://doi.org/10.1053/eujp.2001.0231>
- Harris, B., Regan, T., Schueler, J., & Fields, S. A. (2020). Problematic Mobile Phone and Smartphone Use Scales: A systematic review. *Frontiers in Psychology*, 11, 672. <https://doi.org/10.3389/fpsyg.2020.00672>
- Herrera-Batista, M. Á. (2009). Disponibilidad, uso y apropiación de las tecnologías por estudiantes universitarios en México: Perspectivas para una incorporación innovadora [Availability, use and appropriation of technologies by university students in Mexico: Perspectives for an innovative incorporation]. *Revista Iberoamericana de Educación [Ibero-American Magazine of Education]*, 48(6), 1-9. <https://doi.org/10.35362/rie4862130>
- Horwood, S., & Anglim, J. (2019). Problematic smartphone usage and subjective and psychological well-being. *Computers in Human Behavior*, 97, 44-50. <https://doi.org/10.1016/j.chb.2019.02.028>
- Hu, L., & Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Inostroza, T., Madrid, F., Salinas, M., Reyes, E., & Guerra, C. (2019). Estructura factorial del Cuestionario de Experiencias Relacionadas con Internet en universitarios chilenos [Factorial structure of the Inter-

- net-Related Experiences Questionnaire in Chilean university students]. *Interdisciplinaria [Interdisciplinary]*, 36(2), 55-68. <https://doi.org/10.16888/interd.2019.36.2.4>
- Instituto Nacional de Estadística y Censos [National Institute of Statistics and Censuses, INDEC]. (2021). *Acceso y uso de tecnologías de la información y la comunicación. Cuarto trimestre del 2021*. [Access and use of information and communication technologies. Fourth quarter of 2021] <https://www.indec.gob.ar>
- Labrador-Encinas, F. J., & Villadangos-González, S. M. (2010). Menores y nuevas tecnologías: Conductas indicadoras de posible problema de adicción [Minors and new technologies: Behaviors that indicate a possible addiction problem]. *Psicothema*, 22(2), 180-188. <https://www.psycothema.com>
- López-Fernández, O., Honrubia-Serrano, M. L., & Freixa-Blanxart, M. (2012). Adaptación española del “Mobile Phone Problem Use Scale” para población adolescente. *Adicciones*, 24(2), 123-130. <https://doi.org/10.20882/adicciones.104>
- Menglong, L., & Liya, L. (2017). La influencia de la adicción al teléfono móvil en la calidad de sueño de estudiantes secundarios dejados atrás: El rol mediador de la soledad [The influence of mobile phone addiction on left-behind middle school student’s sleep quality: The mediator role of loneliness]. *Revista Argentina de Clínica Psicológica [Argentine Journal of Clinical Psychology]*, 26(1), 71-81. <https://doi.org/10.24205/03276716.2017.1006>
- Merlo, L. J., Stone, A. M., & Bibbey, A. (2013). Measuring problematic mobile phone use: Development and preliminary psychometric properties of the PUMP Scale. *Journal of Addiction*, 912807. <https://doi.org/10.1155/2013/912807>
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189. <https://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Pepper, R., Solomon, G., & Williams, C. (2012). What is the impact of mobile telephony on economic growth? *GSM Association*. <https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2014/10/gsma-deloitte-impact-mobile-telephony-economic-growth.pdf>
- Redondo-Pacheco, J., Rangel-Noriega, K. J., Luzardo-Briceno, M., & Inglés-Saura, C. J. (2016). Experiencias relacionadas con el uso de Internet y celular en una muestra de estudiantes universitarios colombianos [Experiences related to the use of the Internet and cell phones in a sample of Colombian university students]. *Revista Virtual Universidad Católica Del Norte [Virtual Magazine Catholic University of the North]*, 49, 7-22. <https://revistavirtual.ucn.edu.co/index.php/RevistaUCN/index>
- Richardson, M., Hussain, Z., & Griffiths, M. D. (2018). Problematic smartphone use, nature connectedness, and anxiety. *Journal of Behavioral Addictions*, 7(1), 109-116. <https://doi.org/10.1556/2006.7.2018.10>
- Richaud de Minzi, M. C. (2008). Nuevas tendencias en psicometría [New trends in psychometrics]. *Revista Evaluar [Evaluate Magazine]*, 8(1), 1-19. <https://doi.org/10.35670/1667-4545.v8.n1.501>
- Rodríguez-Ceberio, M., Díaz-Videla, M., Agostinelli, J., & Daverio, R. (2019). Adicción y uso del teléfono celular [Cell phone addiction and use]. *Ajayu Órgano de Difusión Científica del Departamento de Psicología UCBSB [Ajayu Organ of Scientific Dissemination of the Department of Psychology UCBSB]*, 17(2), 211-235. <https://app.lpz.ucb.edu.bo/Publicaciones/Ajayu/revista.html>
- Rozgonjuk, D., Levine, J. C., Hall, B. J., & Elhai, J. D. (2018). The association between problematic smartphone use, depression and anxiety symptom severity, and objectively measured smartphone use over one week. *Computers in Human Behavior*, 87, 10-17. <https://doi.org/10.1016/j.chb.2018.05.019>
- Ryding, F. C., & Kuss, D. J. (2020). Passive objective measures in the assessment of problematic smartphone use: A systematic review. *Addictive Behaviors Reports*, 11, 100257. <https://doi.org/10.1016/j.abr.2020.100257>

abrep.2020.100257

- Sánchez-Martínez, M., & Otero, A. (2009). Factors associated with cell phone use in adolescents in the community of Madrid (Spain). *CyberPsychology and Behavior*, 12(2), 131-137. <https://doi.org/10.1089/cpb.2008.0164>
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514. <https://doi.org/10.1007/BF02296192>
- Simó-Sanz, C., Martínez-Sabater, A., Ballestar-Tarín, M. L., & Domínguez-Romero, A. (2017). Instrumentos de evaluación del uso problemático del teléfono móvil/Smartphone [Instruments for the evaluation of problematic use of mobile phones/Smartphones]. *Salud y Drogas [Health and Addictions]*, 17(1), 5-14. <https://ojs.haaj.org>
- Sohn, S. Y., Rees, P., Wildridge, B., Kalk, N. J., & Carter, B. (2019). Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: A systematic review, meta-analysis and GRADE of the evidence. *BMC Psychiatry*, 19(1), 356. <https://doi.org/10.1186/s12888-019-2350-x>
- Stover, J. B., & Tisocco, F. (2020). Generalized Problematic Internet Use Scale 2: Estudios psicométricos en adultos de Buenos Aires [Generalized Problematic Internet Use Scale 2: Psychometric studies in adults from Buenos Aires]. In *XII Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXVII Jornadas de Investigación. XVI Encuentro de Investigadores en Psicología del MERCOSUR. II Encuentro de Investigación de Terapia Ocupacional. II Encuentro de Musicoterapia*. Facultad de Psicología-Universidad de Buenos Aires. [XII International Congress of Research and Professional Practice in Psychology. XXVII Research Conference. XVI Meeting of Researchers in Psychology of MERCOSUR. II Occupational Therapy Research Meeting. II Music Therapy Meeting. Faculty of Psychology-University of Buenos Aires]. <http://jimemo-rias.psi.uba.ar/index.aspx?anio=2020>
- Stover, J. B., Fernández Liporace, M., & Castro Solano, A. (2023). Generalized Problematic Internet Use Scale 2: Adaptación para adultos de Buenos Aires. *Revista de Psicología*, 41(2), 1127-1151. <https://doi.org/10.18800/psico.202302.017>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Pearson.
- The Jamovi Project. (2023). *Jamovi*. (Version 2.3). [software de cómputo]. <https://www.jamovi.org>
- Toda, M., Monden, K., Kubo, K., & Morimoto, K. (2006). Mobile phone dependence and health-related lifestyle of university students. *Social Behavior and Personality: An International Journal*, 34(10), 1277-1284. <https://doi.org/10.2224/sbp.2006.34.10.1277>
- Turan, N., Polat, O., Karapirli, M., Uysal, C., & Turan, S. G. (2011). The new violence type of the era: Cyber bullying among university students: Violence among university students. *Neurology, Psychiatry and Brain Research*, 17(1), 21-26. <https://doi.org/10.1016/j.npbr.2011.02.005>
- van Velthoven, M. H., Powell, J., & Powell, G. (2018). Problematic smartphone use: Digital approaches to an emerging public health problem. *Digital Health*, 4, 1-9. <https://doi.org/10.1177/2055207618759167>
- World Health Organization. (2019). *International Statistical Classification of Diseases and Related Health Problems* (11th ed.). WHO.
- Xie, X., Dong, Y., & Wang, J. (2018). Sleep quality as a mediator of problematic smartphone use and clinical health symptoms. *Journal of Behavioral Addictions*, 7(2), 466-472. <https://doi.org/10.1556/2006.7.2018.40>

Funcionamiento de Distractores Semánticos en un Test Neuropsicológico de Relaciones Taxonómicas

Functioning of Semantic Distractors in a Neuropsychological Test of Taxonomic Relationships

Axel Fernández-Zaionz¹, Leticia Vivas^{1, 2} *

1 - Instituto de Psicología Básica, Aplicada y Tecnología (IPSIBAT), Mar del Plata, Argentina.

2 - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

Introducción
Metodología
Resultados
Discusión
Referencias

Recibido: 12/05/2023 Revisado: 06/06/2023 Aceptado: 13/06/2023

Resumen

Conocer el funcionamiento de las relaciones conceptuales es sumamente importante para la evaluación neuropsicológica, ya que se ven afectadas por muchos síndromes neuropsicológicos. La selección de los distractores adecuados resulta crucial para poder graduar la dificultad de las pruebas. En este trabajo, se estudió el efecto de la fuerza de correspondencia (FC) entre opciones de respuesta en una prueba de asociación semántica, denominada TAXON, que evalúa la capacidad de identificar relaciones taxonómicas. La muestra incluyó 43 adultos mayores de 65 años sin patología. Se analizó la correlación entre la diferencia en la FC de los estímulos y los tiempos (TR) y tasa de respuesta (TA). Además, se analizó si el dominio semántico de los ítems (vivo/no vivo) y el formato de presentación (verbal/pictórico) tenían influencia. Los resultados mostraron TR menores y TA mayores cuando la diferencia de correspondencia entre distractor y target era mayor. Además, se observó un efecto mayor en el formato pictórico.

Palabras clave: *evaluación, lenguaje, relaciones conceptuales, adultos mayores, asociación semántica*

Abstract

Knowing how conceptual relationships work is extremely important for neuropsychological evaluation, as many neuropsychological syndromes affect them. Selecting appropriate distractors is crucial for grading the difficulty of tests. This study examined the effect of correspondence strength (CS) between response options in a semantic association test called TAXON, which evaluates the ability to identify taxonomic relationships. The sample included 43 adults over 65 years old without pathology. The correlation between the difference in CS of stimuli and response times (RT) and response rate (RR) was analyzed, as well as if the semantic domain of items (living/non-living) and the presentation format (verbal/pictorial) had an influence. The results showed lower RT and higher RR when the correspondence difference between distractor and target was greater. Additionally, a greater effect was observed in the pictorial format.

Keywords: *evaluation, language, conceptual relationships, older adults, semantic association*

*Correspondencia a: axlram91@gmail.com, Dean Funes 3280, 7600. Teléfono: +54 0223 156-245-470. E-mail: lvivas@mdp.edu.ar, Rodríguez Peña 4272, 7600. Teléfono: +54 0223 155-224-532.

Cómo citar este artículo: Fernández-Zaionz, A., & Vivas, L. (2023): Funcionamiento de Distractores Semánticos en un test Neuropsicológico de Relaciones Taxonómicas. *Revista Evaluar*, 23(3) 31-44. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Participaron en la edición de este artículo: Mariana López, Gloria Nieve, Eugenia Barrionuevo, Abigail Pérez, Florencia Ruiz, Rodrigo Maderna, Jorge Bruera.

Introducción

Los déficits semánticos observados en la clínica neuropsicológica asumen perfiles diversos que dan cuenta de la complejidad de este sistema y recalcan la necesidad de incluir pruebas que abarquen las distintas aristas. Las relaciones conceptuales taxonómicas pueden verse afectadas en distintos síndromes neuropsicológicos. Entre ellos, en la afasia progresiva primaria variante semántica (tradicionalmente conocida como demencia semántica), donde se observa un déficit multimodal, es decir, que se evidencian dificultades con independencia de la vía de acceso (pictórica o verbal) y de salida (respuestas verbales, clasificaciones o emparejamiento de figuras) (Patterson et al., 2007). En estos casos se suelen conservar las relaciones superordinadas, pero perderse las coordinadas y subordinadas (Crutch & Warrington, 2008). Otro de los síndromes que suele presentar un déficit semántico es la demencia tipo Alzheimer (Comesaña & Vivas, 2015; Peraita-Adrados et al., 1999). Además, se estudiaron los déficits semánticos de categoría específica y se reportó una mayor cantidad de errores en el dominio de seres vivos (Peraita & Moreno, 2006). También existe una línea de investigación sobre la afasia semántica que evidencia que estos pacientes suelen presentar errores temáticos en tareas de denominación (Crutch & Warrington, 2008). En los últimos años se han desarrollado y validado localmente algunos instrumentos para evaluar la memoria semántica, aunque, respecto a la evaluación de las relaciones taxonómicas, presentan algunas limitaciones. Existe numerosa evidencia de que las relaciones taxonómicas y temáticas pueden afectarse de manera independiente (Cuetos-Vega & Castejón, 2005; Vivas et al., 2017), ya que tienen sustratos neuroanatómicos independientes (Kalénine et al., 2009; Sachs et al., 2008) y ejercen un efecto diferencial sobre distintos procesos cognitivos (Rose

& Abdel-Rahman, 2016), por lo que es importante contar con instrumentos de evaluación que contemplen ambas.

Considerando la importancia de contar con instrumentos de evaluación de la memoria semántica validados localmente que permitan clarificar el perfil de déficit semántico de los pacientes, se desarrolló hace unos años la prueba neuropsicológica (TAXON) para evaluar una de sus dimensiones: las relaciones taxonómicas. Esta prueba intenta contribuir al diagnóstico clínico de pacientes con afectación de la memoria semántica y brindar información específica acerca del reconocimiento de relaciones de tipo taxonómico. De esta manera, se complementa la evaluación de las relaciones conceptuales junto con pruebas que examinan el desempeño en relaciones temáticas. Esta prueba consta de cuatro opciones de respuesta: taxonómica fuerte, taxonómica débil, temática y no relacionada. Además, tiene dos formatos de presentación, pictórico y verbal, engloba 21 ítems para cada formato e incluye conceptos de los dominios vivos y no vivos.

Este trabajo se propone analizar si la diferencia entre las fuerzas de asociación entre el estímulo blanco y las opciones de respuesta taxonómica fuerte y débil inciden en los tiempos de respuesta y en la tasa de acierto de los participantes, además de analizar si hay un efecto diferencial por formato de presentación (verbal o pictórico) y por el dominio al que pertenecen los estímulos (vivos o no vivos).

Memoria semántica

La memoria semántica nos permite identificar y nominar los conceptos y los conocimientos con los cuales organizamos el mundo que percibimos. Además de almacenar dicha información, debemos recurrir a diferentes aspectos del conoci-

miento para guiar el comportamiento en diferentes circunstancias (Hoffman et al., 2018). A su vez, utilizamos nuestro conocimiento semántico cuando iniciamos la producción del lenguaje (Lambon-Ralph, 2014).

Según Lambon-Ralph (2014), los conceptos no se almacenan como representaciones unitarias en una región específica del cerebro, sino que reflejan la acción masiva de múltiples fuentes de información específicas de la modalidad, cada una de las cuales está codificada en diferentes regiones corticales. Este autor señala que, más allá de las cortezas de asociación secundarias específicas de modalidad, existen áreas corticales "terciarias" intermedias que no están vinculadas a ninguna modalidad particular, para la que Luria (1976) adoptó el término corteza "transmodal". Mediante el uso de un modelo computacional, Rogers et al. (2004) demostró que un centro interconectado que reúne información específica de la modalidad se podría comportar como un sistema de representación transmodal. El modelo Hub-and-Spoke (Centro y Radios) de Rogers es capaz de codificar con éxito el conocimiento de palabras abstractas y concretas, relaciones asociativas y taxonómicas, y los múltiples significados de los homónimos, dentro de un único espacio de representación.

Relaciones conceptuales

Lo que caracteriza a los conceptos es que estos se encuentran relacionados de tal manera que la activación de unos genera la activación de otros. Estos vínculos particulares se denominan relaciones conceptuales, existen dos tipos ampliamente estudiados: las relaciones temáticas y las relaciones taxonómicas (Lin & Murphy, 2001). Las primeras se entienden como cualquier relación funcional, temporal, espacial o causal entre objetos o personas. Los conceptos están relacionados

temáticamente si realizan funciones complementarias en el mismo escenario o los mismos eventos (Golonka & Estes, 2009), por ello se afirma que las relaciones temáticas son contextuales y externas a las propiedades de la cosa o el objeto en sí mismo. Por ejemplo, (perro-correa) están asociados a la actividad habitual de pasear a la mascota. Por otro lado, las relaciones taxonómicas se conforman con base en las características propias e inherentes al concepto, es decir, se basan en las propiedades internas de los conceptos y cuando estos tienen propiedades compartidas forman parte de la misma categoría taxonómica (Hampton, 2006). Estos dos tipos de relaciones parecen implicar diferentes procesos, tal como muestran diferentes estudios realizados con sujetos con alteraciones neurológicas y sujetos sanos (Barsalou, 2005, 2009; Lin & Murphy, 2001).

Relaciones taxonómicas

Las relaciones taxonómicas son aquellas que vinculan conceptos de la misma categoría semántica y pueden ser de tres tipos: supraordinadas (perro-animal), subordinadas (perro-caniche) y coordinadas (perro-gato) (Lin & Murphy, 2001). Los objetos de la misma categoría taxonómica usualmente comparten el nombre genérico (ej., animales) y tienen propiedades perceptivas y no perceptivas similares. Dado que los componentes de este tipo de relaciones tienen rasgos comunes, las vinculaciones se establecen comparando las propiedades de ambos conceptos y, en función de ello, se decide cuál es su grado de semejanza (Estes et al., 2011). Se asume que cuantos más atributos en común compartan los conceptos entre sí, mayor asociación semántica tendrán. Para representar estas relaciones, el sistema semántico debe ser sensible a patrones de co-ocurrencia espacio-temporal entre palabras y objetos (Hoffman et al., 2018).

Efecto de Interferencia Semántica

En el contexto de la investigación de la producción del lenguaje, el Efecto de Interferencia Semántica (SIE) ha sido interpretado durante mucho tiempo como evidencia de la naturaleza competitiva de la selección léxica (La Heij et al., 2006; Roelofs, 1992; Rose et al., 2019). Muchos modelos de acceso léxico asumen que la facilidad con la que se selecciona un nodo léxico depende no solo de su nivel de activación, sino también del de otros nodos (Costa & Caramazza, 2002; Roelofs, 1992; Starreveld & La Heij, 1995). Si en el momento de la selección otros nodos léxicos están altamente activados, la selección del nodo léxico objetivo se retrasará. Se supone que el SIE revela la mayor competencia léxica producida por distractores relacionados que no relacionados. Los distractores relacionados interfieren más porque, hipotéticamente, están más activados que los distractores no relacionados. Por ejemplo, en el paradigma de Interferencia Imagen-Palabra (PWI), nombrar una imagen (ej., perro) se retrasa en presencia de una palabra distractora relacionada semánticamente (ej., gato) en relación con una palabra no relacionada (ej., pluma) (Damian & Bowers, 2003; Hantsch et al., 2005). Este nivel diferencial de activación surge porque la representación semántica de la imagen (perro) activa el nodo léxico del distractor relacionado (gato) pero no el del distractor no relacionado (pluma) (Roelofs, 2018; Rose et al., 2019; Vieth et al., 2014). Un supuesto fundamental de los modelos de competencia léxica es que la activación competitiva depende del grado de similitud semántica (o distancia) entre representaciones. Este factor determina la cantidad de activación diseminada dentro de las etapas conceptual y conceptual léxica, donde las representaciones coactivadas fuertemente relacionadas deberían competir más que las representaciones débilmente relacionadas debido a la alta superposición de caracterís-

ticas (Roelofs, 1992; Vigliocco et al., 2004). De acuerdo con los modelos competitivos tradicionales de selección léxica (Vigliocco et al., 2004), los efectos graduales deben observarse en tiempos de denominación más rápidos para distractores no relacionados.

Este efecto se ha estudiado habitualmente en tareas de denominación, pero ha sido escasamente estudiado en tareas de asociación semántica como en la que nos centraremos en este estudio.

Efectos de modalidad de presentación: verbal y pictórico

El efecto de modalidad de presentación de los estímulos en pruebas de evaluación del lenguaje ha sido estudiado por varios autores, quienes han presentado diversos modelos sobre el procesamiento de palabras y dibujos. Bright et al. (2004) aluden a dos tipos de modelos que han sido tratados ampliamente en la literatura científica. Por una parte, están aquellos modelos que plantean la existencia de un solo sistema semántico capaz de procesar diferentes entradas de información, sean estas palabras o dibujos (Caramazza et al., 1990). Mientras que por otra parte, están aquellos que defienden la existencia de dos subsistemas independientes encargados de procesar palabras y dibujos (Paivio, 1991). Respecto a los que sostienen que el sistema semántico es amodal, estudios de neuroimagen han brindado evidencia de un papel crítico de la extensión anterior de la circunvolución fusiforme en la representación del conocimiento conceptual que no parece estar modulada por la modalidad de entrada (imágenes o palabras), lo que sugiere que contiene representaciones semánticas amodales formadas a través de entradas convergentes de áreas posteriores (Bright et al., 2004; Tyler et al., 2003). Por otro lado, respecto a la propuesta de un sistema semántico multimodal, varios estudios

han presentado evidencia acerca de un costo en el procesamiento perceptual, que se traduce en tiempos de reacción más lentos al hacer un cambio de modalidad (ej., auditiva a pictórica) en pruebas de categorización semántica (Marques, 2006; Pecher et al., 2003). En la misma línea, otras investigaciones han evidenciado que los tiempos de respuesta son más rápidos cuando los estímulos que deben ser codificados son dibujos y no palabras. Este hecho se conoce como el efecto de superioridad de los dibujos (Paivio, 1991). Paivio (1991) elaboró el modelo de codificación dual, el cual involucra la actividad de dos subsistemas distintos, uno verbal que codifica información lingüística y conceptos abstractos y un sistema no-verbal encargado del procesamiento de información visual y conceptos concretos. Los sistemas están compuestos por unidades de representación internas que se activan cuando uno reconoce, manipula, o simplemente piensa en palabras o cosas. El funcionamiento de ambos mecanismos explicaría el efecto de superioridad de los dibujos, ya que estos obtienen representación en ambos sistemas. Las palabras, en cambio, solo se representan en el sistema visual cuando son muy concretas o bien la demanda de la tarea hace que se deban transformar en imágenes (Bajo & Cañas, 1989).

Efectos de dominio semántico: vivos y no vivos

Una de las consecuencias más intrigantes de ciertos tipos de daño cerebral es que el conocimiento conceptual de los pacientes puede verse afectado selectivamente para categorías específicas de objetos, mientras que otros permanecen relativamente preservados. La disociación informada con más frecuencia es un déficit de seres vivos como frutas, verduras y animales en relación con los seres no vivos, tales como herramientas (Sheridan & Humphreys, 1993; Warrington & Shallice, 1984).

También se ha observado el patrón inverso (Moss & Tyler, 2000; Warrington & McCarthy, 1987). Se han propuesto varias teorías para explicar esta doble disociación. Martin (2007) propone que el principio organizador del sistema conceptual es el tipo de rasgo. La versión más simple de este relato propone dos subsistemas semánticos, uno para características sensoriales y otro para características funcionales. La idea central es que el conocimiento del objeto está organizado por características sensoriales (ej., forma, movimiento, color) y propiedades motoras asociadas con el uso del objeto. Desde este punto de vista, los trastornos del conocimiento específicos de una categoría ocurren cuando una lesión altera la información sobre una propiedad particular o un conjunto de propiedades críticas para definir esa categoría de objeto y para distinguir entre sus miembros. Adicionalmente, algunos autores plantean que la categorización de estímulos biológicos muestra tiempos de reacción más cortos y menor número de errores debido a la gran cantidad de correlaciones entre propiedades compartidas, que se traduce en una activación más distribuida y rápida de los conceptos (McRae et al., 1999).

Limitaciones en la evaluación de relaciones conceptuales taxonómicas

En los últimos años se han desarrollado y validado localmente algunos instrumentos para evaluar la memoria semántica. Entre los más reconocidos se encuentra la Batería 64 (Adlam et al., 2010), que cuenta con una adaptación argentina (Martínez-Cuitiño et al., 2009). Algunas subpruebas de la Escala de Wechsler permiten valorar el conocimiento semántico. (Wechsler, 2002). En español también se ha desarrollado la Batería de Evaluación de la Memoria Semántica para la Demencia tipo Alzheimer (Peraita-Adrados et al.,

2000; Grasso & Peraita, 2011). El método Distsem (Lizarralde et al., 2011; Vivas, 2008) constituye un instrumento para la evaluación de la memoria semántica basado en la distancia entre conceptos. Otro instrumento ampliamente empleado en la evaluación de la memoria semántica es el Test de Pirámides y Palmeras (Howard & Patterson, 1992), que también cuenta con una adaptación argentina (Martínez-Cuitiño & Barreyro, 2010).

Sin embargo, las pruebas desarrolladas hasta el momento para evaluar relaciones taxonómicas cuentan con algunas limitaciones. En general, el formato de presentación de estímulos difiere radicalmente del usado en pruebas de relaciones temáticas, como Pirámides y Faraones. Esto dificulta la comparación entre ambos tipos de relaciones (taxonómicas y temáticas). A la vez, ninguna de las pruebas dedicadas a las relaciones taxonómicas posee dos vías de presentación (pictórica y verbal). La vía pictórica resulta fundamental para su empleo con pacientes afásicos, en los que numerosos trabajos evidencian que suelen presentarse desórdenes conceptuales no verbales (Gainotti, 2014), mientras que la verbal podría facilitar la evaluación de pacientes con dificultades en la vía visual. A su vez, las opciones de respuesta de las pruebas existentes no permiten medir claramente el tipo de organización conceptual que predomina en las elecciones del paciente.

Metodología

Muestra

La muestra estuvo compuesta por 43 adultos mayores de 65 años ($M = 74.4$, $DE = 9.8$) sin patología neurológica ni psiquiátrica. Del total de la muestra el 79.1% eran mujeres, mientras que el 20.9%, varones. Respecto al nivel educativo, el 20.9% tenía estudios primarios, el 39.5% estudios secundarios y el 39.5% restante terciarios/univer-

sitarios. Se contó con acceso a participantes de los talleres socio-preventivos del PAMI.

Materiales e Instrumentos

La tarea administrada fue la prueba TAXON (Vivas & Fernández-Liporace, 2017). Esta prueba se desarrolló para evaluar la capacidad de identificar relaciones conceptuales taxonómicas. Consta de cuatro opciones de respuesta: taxonómica fuerte, taxonómica débil, temática y no relacionada. Tiene dos formatos de presentación, pictórico y verbal. Hay 21 ítems para cada formato, incluidos los conceptos de los dominios vivos y no vivos. Se presentan cinco estímulos: uno arriba, en el centro, y cuatro debajo. La persona debe indicar cuál de los cuatro estímulos de abajo pertenece a la misma clase que el objeto ubicado arriba. La respuesta requerida es de elección forzada (hay opciones restringidas de respuesta) (Hogan, 2015) y emparejamiento con la muestra. Se diseñaron ítems para ambas escalas mediante la selección de estímulos de seis categorías semánticas, y se priorizaron las tradicionalmente utilizadas en evaluación del sistema conceptual (Fumagalli et al., 2015; Martínez-Cuitiño, Barreyro y Jaichen-co, 2009; Peraita-Adrados et al., 2000): animales, frutas/verduras, objetos manipulables, prendas de vestir, muebles, vehículos. Los estímulos se eligieron a partir del set de imágenes de Cykowicz et al. (1997), y se incluyeron estímulos con grados medio y alto de familiaridad de acuerdo con normas psicolingüísticas argentinas (Manoiloff et al., 2010). Para extraer los valores de fuerza asociativa entre los estímulos, se recurrió a las Normas de Producción de Atributos Semánticos en español para adultos mayores (Vivas et al., 2022). Las mismas consisten en la recolección de los atributos semánticos de mayor importancia para la definición de un concepto. Para ello se solici-

ta a un grupo de personas que digan los atributos que, a su entender, mejor definen un conjunto de conceptos. A partir de esos datos, se calculan numerosas variables dentro de las que se encuentra la correspondencia entre conceptos, calculada mediante una comparación de vectores entre las listas de atributos de pares de conceptos. Para realizar este cálculo se lleva a cabo un análisis mediante la técnica geométrica de comparación de vectores en un espacio euclidiano y dimensional a partir del ángulo formado entre ellos, tal que el paralelismo representa la mayor similitud y la ortogonalidad, la mayor diferencia (Kintsch, 2001). El producto de esto es una matriz concepto-concepto donde se ve el grado de correspondencia (que aquí denominaremos fuerza de correspondencia) entre conceptos en función de los atributos que tienen en común. Para el presente proyecto se seleccionaron los valores de Fuerza de Correspondencia (FC) entre el estímulo de referencia y dos de las opciones de respuesta: taxonómico fuerte y taxonómico débil. Luego se calculó la Diferencia entre las Fuerzas de Correspondencia entre ambos (DFC).

Procedimiento

Luego de haber obtenido el correspondiente consentimiento informado, la prueba se administró de manera individual mediante una computadora con una pantalla de 15" y el software de presentación de estímulos E-Prime 2.0. Cada formato de la prueba, verbal y pictórica, se administró en una sesión independiente. Se registraron tanto las respuestas como los tiempos.

Análisis de datos

En primer lugar, se analizaron el sentido y la magnitud de la asociación entre la DFC y los

tiempos y tasa de respuesta mediante una prueba de correlación de Pearson considerando un valor p crítico de .05. En segundo lugar, se calcularon dos modelos lineales generales mediante el programa SPSS 23 en el que se consideraron como variables dependientes a los TR y los aciertos respectivamente, para determinar si la DFC tiene un efecto significativo en dichas medidas. La DFC se incluyó como covariable por ser una variable continua y se incluyeron como factores el formato de presentación (pictórico vs. verbal) y el dominio (vivo no vivo), para ver si aportaban a la variabilidad de los TR y TA. Se consideró un valor p crítico de .05. También se calculó el tamaño del efecto mediante un η^2 parcial al cuadrado para medir la magnitud de la influencia de aquellas variables que resultaran significativas.

Resultados

Se presenta a continuación la Tabla 1 correspondiente a los tiempos de respuesta (TR) y tasa de aciertos (TA) promedio para el formato de presentación pictórico y verbal.

En relación al dominio al que pertenecen los estímulos, se presenta la Tabla 2 correspondiente a los TR y TA promedio para los vivos y no vivos.

Tras el análisis de la correlación entre la diferencia en la fuerza de correspondencia (DFC) y los tiempos de respuesta, se encontró una asociación negativa y estadísticamente significativa ($r = -.24, p < .01$). Se observó que a mayor diferencia de correspondencia entre las opciones de respuesta taxonómico fuerte y taxonómico débil, menores fueron los tiempos de respuesta.

Por otro lado, se analizó la correlación entre DFC y la tasa de aciertos y se encontró una asociación positiva y estadísticamente significativa ($r = .20, p < .05$). A mayor diferencia de correspondencia entre las opciones de respuesta taxonómico fuerte y taxonómico débil, mayor fue la tasa de aciertos.

Tabla 1

Valores de TR y TA según el formato de presentación.

Formato de presentación	Tiempos de respuesta* (media y desvío)	Tasa de aciertos (me- dia y desvío)	N
Pictórico	4962,07 (878,19)	34,33 (5,36)	60
Verbal	5313,87 (769,78)	34,48 (5,01)	60

Nota. * en milisegundos.**Tabla 2**

Valores de TR y TA según el dominio de los estímulos.

Formato de presentación	Tiempos de respuesta* (media y desvío)	Tasa de aciertos (me- dia y desvío)	N
Vivos	4987,63 (589,35)	34,75 (4,92)	40
No vivos	5213,13 (936,43)	34,24 (5,30)	80

Nota. * en milisegundos.

En segundo lugar, se realizaron análisis univariados de varianza considerando al tiempo y a la tasa de respuestas como variables dependientes, a la DFC como covariable y como factores al formato de presentación (pictórico, verbal) y el dominio (vivo, no vivo).

En relación al análisis univariado sobre los TR, se puede observar que la variable DFC es estadísticamente significativa ($p < .05$), con una magnitud del efecto moderada ($np^2 = 0,052$). Se puede observar que la variable dominio no es estadísticamente significativa ($p > .05$). En relación al formato de presentación, observamos que muestra un efecto significativo ($p < .05$) y la magnitud del efecto es relativamente baja ($np^2 = 0,036$). Los resultados pueden verse en la Tabla 3.

En relación con análisis univariado sobre la TA, se puede observar que la variable DFC es estadísticamente significativa ($p < .05$), y la magnitud del efecto es relativamente baja ($np^2 = 0,039$). Se observa que tanto la variable dominio como la variable formato de presentación no son estadísticamente significativas ($p > .05$), por lo que otras variables podrían explicar mejor la varianza. Los resultados pueden verse en la Tabla 4.

Discusión

Este trabajo se propuso analizar de qué manera la diferencia entre las fuerzas de correspondencia (DFC) entre el estímulo blanco y las opciones de respuesta taxonómica fuerte y débil inciden en los tiempos de respuesta y en la tasa de aciertos de los participantes en una prueba de relaciones taxonómicas. Respecto a los TR, el análisis de los datos indica una asociación negativa estadísticamente significativa y se observó que, a mayor diferencia de correspondencia entre las opciones de respuesta taxonómico fuerte y taxonómico débil, menores fueron los tiempos de respuesta. Además, el análisis de varianza arroja que la variable DFC es estadísticamente significativa, con una magnitud del efecto moderada para los TR. Estos datos son consistentes con los modelos de competencia léxica anteriormente presentados, los que plantean que las representaciones fuertemente relacionadas compiten más que las débilmente relacionadas por la alta superposición de características entre las mismas, por lo tanto el resultado es un retraso en la selección del estímulo blanco.

Adicionalmente, se observó que el formato de presentación tiene un efecto significativo para

Tabla 3

Valores ANOVA de los tiempos promedio.

Origen	Tipo III de suma de cuadrados	gl	Media cuadrática	F	Sig.	Eta parcial al cuadrado	Potencia observada ^b
Modelo corregido	9361170,23 ^a	4	2340292,55	3,59	0,008	0,11	0,86
Intersección	1585292430,12	1	1585292430,12	2436,74	0,000	0,95	1,00
DFC	4092681,08	1	4092681,08	6,29	0,014	0,05	0,70
Dominio*	564001,61	1	564001,61	,86	0,354	0,00	0,15
Formato de presentación	2781501,61	1	2781501,61	4,27	0,041	0,03	0,53
Dominio * presentación	199560,69	1	199560,69	0,30	0,581	0,00	0,08
Error	74816506,36	115	650578,31				
Total	3252027401,74	120					
Total corregido	84177676,60	119					

Nota. a. R al cuadrado = 0,111 (R al cuadrado ajustada = 0,080). b. Se ha calculado utilizando $\alpha = ,05$.

los TR, aunque la magnitud del efecto es relativamente baja, lo cual nos obliga a ser cautos en nuestras conclusiones acerca de la magnitud en la que esta variable está incidiendo en los TR. Lo que se observó es que el efecto de proximidad semántica fue mayor cuando los estímulos pertenecían al formato de presentación pictórico, el cual presentó menores tiempos de respuesta en relación con el verbal. Diversos autores arribaron a la conclusión de que los tiempos de respuesta son más rápidos cuando los estímulos que deben ser codificados son imágenes en vez de palabras. Este hecho es conocido como el efecto de superioridad de las imágenes (Paivio, 1991). Este se debe a que las imágenes obtienen representación en dos subsistemas (verbal y pictórico); en cambio, las palabras solo se representan en el sistema visual cuando son muy concretas o bien la demanda hace que se deban transformar en imágenes (Bajo & Cañas, 1989).

Respecto a la TA, el análisis de datos indica una asociación positiva con DFC y se observó que a mayor diferencia de correspondencia entre las opciones de respuesta taxonómico fuerte y taxonómico débil, mayor fue la tasa de aciertos. En este caso, la variable formato de presentación

no fue estadísticamente significativa y los datos arrojados por el análisis de varianza indican una magnitud del efecto relativamente baja para DFC, por lo que es probable que otras variables no contempladas en este estudio estén incidiendo en la varianza de la TA. Sin embargo, la asociación positiva entre DFC y TA podría aportar evidencia en favor de los modelos de competencia léxica. Una forma de interpretar estos datos refiere/apunta a una mayor facilidad en la selección de los nodos léxicos correspondientes al estímulo blanco cuando las opciones de respuesta tienen mayor distancia semántica en el momento de la selección.

Por último, el análisis de datos nos indica que la variable dominio no es estadísticamente significativa ni para los TR ni para la TA. Por lo cual, no se observa un efecto diferencial según el dominio al que pertenecen los estímulos. Una posible interpretación de los datos podría hacerse a la luz de modelos que explican la estructura del conocimiento semántico como un procesamiento de dominio general con una especificidad de dominio graduada en áreas específicas. La disociación documentada, entre los dominios vivos y no vivos, refieren a modelos de dominio específico, que proponen que el conocimiento semántico está

Tabla 4

Valores ANOVA de las tasas de acierto.

Origen	Tipo III de suma de cuadrados	gl	Media cuadrática	F	Sig.	Eta parcial al cuadrado	Potencia observada ^b
Modelo corregido	138,30 ^a	4	34,57	1,30	0,271	0,04	0,86
Intersección	63683,59	1	63683,59	2408,53	0,000	0,95	1,00
DFC	122,18	1	122,18	4,62	0,034	0,03	0,56
Dominio	0,20	1	0,20	0,00	0,929	0,00	0,05
Formato de presentación	0,03	1	0,03	0,00	0,970	0,00	0,05
Dominio * presentación	8,43	1	8,43	0,31	0,573	0,00	0,08
Error	3040,68	115	26,44				
Total	145251,00	120					
Total corregido	3178,99	119					

Nota. a. R al cuadrado = 0,044 (R al cuadrado ajustada = 0,010). b. Se ha calculado utilizando $\alpha = 0,05$.

sustentado en redes neuronales especializadas, ya sea por dominio (Santos & Caramazza, 2002) o por características sensorio-motoras (Martin, 2007; McRae et al., 2005). Según estos estudios, cabría esperar TR más cortos y mayor TA para los estímulos biológicos debido a la gran cantidad de correlaciones entre propiedades sensoriales compartidas (McRae et al., 1999). Por otro lado, existen modelos de dominio general que plantean que el conocimiento semántico se encuentra distribuido y que todos sus componentes contribuyen a todos los dominios de conocimiento. En este sentido, Chen et al. (2017) desarrollaron una tercera propuesta que integra los modelos de dominio específico y los de dominio general. Esta propuesta expone un modelo hub-and-spokes (Centro y Radios) en la que el Lóbulo Temporal Anterior (ATL) constituye un centro semántico de dominio general en la representación semántica, mientras que otras regiones cerebrales (radios) presentan sensibilidad específica de modalidad. Así, el modelo ilustra cómo pueden surgir patrones específicos de dominio dentro de los "radios" de una red de este tipo, pero el procesamiento en el ATL es de dominio general. Ahora bien, TAXÓN es una prueba de asociación semántica y el objeti-

vo de estas tareas es que el paciente asocie una imagen o palabra en función de alguna relación semántica. Este tipo de tarea demanda procesos de control ejecutivo que regulan la activación del conocimiento para satisfacer las demandas situacionales (Hoffman et al., 2018). Si bien el centro en el ATL recibe información de otros componentes de la red, al codificar los estímulos en un procesamiento de modalidad general, es esperable que no se observen diferencias por dominio.

Finalmente, cabe mencionar una limitación de este trabajo, ya que se realizó en función de datos obtenidos a partir del estudio de una prueba para evaluar relaciones taxonómicas. Esto quiere decir que no surgieron de un experimento especialmente diseñado para demostrar hipótesis respecto del efecto de las distancias semánticas. Ahora que ya se encuentran publicadas las normas de producción de atributos semánticos en español para personas mayores (Vivas et al., 2022), es posible diseñar un experimento que tenga mayor control de varias medidas de distancias semánticas obtenidas empíricamente de personas de nuestra comunidad.

Referencias

- Adlam, A.-L., Patterson, K., Bozeat, S., & Hodges, J. R. (2010). The Cambridge Semantic Memory Test Battery: Detection of semantic deficits in semantic dementia and Alzheimer's disease. *Neurocase*, 16(3), 193-207. <https://doi.org/10.1080/13554790903405693>
- Bajo, M. T., & Cañas, J. J. (1989). Phonetic and semantic activation during picture and word naming. *Acta Psychologica*, 72(2), 105-115. [https://doi.org/10.1016/0001-6918\(89\)90038-3](https://doi.org/10.1016/0001-6918(89)90038-3)
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophy Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1284-1289. <https://doi.org/10.1098/rstb.2008.0319>
- Barsalou, L. W. (2005). Situated conceptualization. En H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 619-650). Elsevier. <https://doi.org/10.1016/b978-008044612-7/50083-4>
- Bright, P., Moss, H., & Tyler, L. K. (2004). Unitary vs. multiple semantics: PET studies of word and picture processing. *Brain and Language*, 89(3), 417-432. <https://doi.org/10.1016/J.BANDL.2004.01.010>
- Caramazza, A., Hillis, A. E., Rapp, B. C., & Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology*, 7(3), 161-189. <https://doi.org/10.1080/02643299008253441>
- Chen, L., Lambon-Ralph, M. A., & Rogers, T. T. (2017). A unified model of human semantic knowledge and its disorders. *Nature Human Behaviour*, 1(3), 0039. <http://dx.doi.org/10.1038/s41562-016-0039>
- Comesaña, A., & Vivas, J. (2015). Evolución de la categorización semántica en adultos mayores con diagnóstico de DCL-A y DTA y sin patología neurológica. *Interdisciplinaria: Revista de Psicología y Ciencias Afines*, 32(1), 7-29. <https://doi.org/10.16888/interd.2015.32.1.1>
- Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding in speech production. *Journal of Memory & Language*, 46(1), 178-198. <https://doi.org/10.1006/jmla.2001.2804>
- Crutch, S. J., & Warrington, E. K. (2008). Contrasting patterns of comprehension for superordinate, basic-level, and subordinate names in semantic dementia and aphasic stroke patients. *Cognitive Neuropsychology*, 25(4), 582-600. <https://doi.org/10.1080/02643290701862290>
- Cuetos-Vega, F., & Castejón, L. (2005). Disociación de la información conceptual y lingüística a partir de un estudio de caso. *Revista de Neurología*, 41(8), 469-474. <https://doi.org/10.33588/rn.4108.2004617>
- Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, 65(2), 171-237. <https://doi.org/10.1006/jecp.1996.2356>
- Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review*, 10(1), 111-117. <https://doi.org/10.3758/bf03196474>
- Estes, Z., Golonka, S., & Jones, L. L. (2011). Thematic thinking. The apprehension and consequences of thematic relations. In B. H. Ross (Ed.), *Psychology of Learning and Motivation: Advances in research and theory* (Vol. 54, pp. 249-294). Elsevier Academic Press. <https://doi.org/10.1016/b978-0-12-385527-5.00008-5>
- Fumagalli, J., Shalóm, D., Soriano, F., Carden, J., Cabañas-Fale, P., Tomio, A., Borovinsky, G., & Martínez-Cutiño, M. (2015). Normas categoriales para una muestra de hablantes adultos del español de Argentina. *Revista Evaluar*, 15(1), 1-40. <https://doi.org/10.35670/1667-4545.v15.n1.14907>
- Gainotti, G. (2014). Old and recent approaches to the problem of non-verbal conceptual disorders in aphasic patients. *Cortex*, 53, 78-89. <https://doi.org/10.1016/j.cortex.2014.01.009>
- Golonka, S., & Estes, Z. (2009). Thematic relations affect similarity via commonalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,

- 35(6), 1454-1464. <https://doi.org/10.1037/a0017397>
- Grasso, L., & Peraita, H. (2011). Adaptación de la Batería de Evaluación de la Memoria Semántica en la Demencia de tipo Alzheimer (EMSDA) a la población de la Ciudad de Buenos Aires. *Interdisciplinaria: Revista de Psicología y Ciencias Afines*, 28(1), 37-56. <https://dialnet.unirioja.es/servlet/revista?codigo=2511>
- Hampton, J. A. (2006). Concepts as prototypes. In B. H. Ross (Ed.), *Psychology of Learning and Motivation: Advances in Research and Theory*. Elsevier Academic Press. [https://doi.org/10.1016/S0079-7421\(06\)46003-5](https://doi.org/10.1016/S0079-7421(06)46003-5)
- Hantsch, A., Jescheniak, J. D., & Schriefers, H. (2005). Semantic competition between hierarchically related words during speech planning. *Memory & Cognition*, 33(6), 984-1000. <https://dx.doi.org/10.3758/BF03193207>
- Hoffman, P., McClelland, J. L., & Lambon-Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review*, 125(3), 293-328. <https://doi.org/10.1037/rev0000094>
- Hogan, T. (2015) Pruebas psicológicas. Una introducción práctica (2^{da} ed.). Ciudad de México, México. El Manual Moderno. ISBN: 978-607-448-498-4
- Howard, D., & Patterson, K. (1992). *Pyramids and palm trees: A Test of Semantic Access from Pictures and Words*. Thames Valley Test Company: Bury St. Edmunds, UK. ISBN: 187-426-115-6
- Kalénine, S., Peyrin, C., Pichat, C., Segebarth, C., Bonthoux, F., & Baciú, M. (2009). The sensory-motor specificity of taxonomic and thematic conceptual relations: A behavioral and fMRI study. *NeuroImage*, 44(3), 1152-1162. <https://doi.org/10.1016/j.neuroimage.2008.09.043>
- Kintsch, W. (2001). Predication. *Cognitive Science*, 25(2), 173-202. https://doi.org/10.1207/s15516709cog2502_1
- La Heij, W., Kuipers, J.-R., & Starreveld, P. A. (2006). In defense of the lexical-competition account of picture-word interference: A comment on Finkbeiner and Caramazza. *Cortex*, 42(7), 1028-1031. [http://dx.doi.org/10.1016/S0010-9452\(08\)70209-0](http://dx.doi.org/10.1016/S0010-9452(08)70209-0)
- Lambon-Ralph, M. A. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120392. <https://doi.org/10.1098/rstb.2012.0392>
- Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, 130(1), 3-28. <https://doi.org/10.1037/0096-3445.130.1.3>
- Lizarralde, F., Huapaya, C., & Vivas, J. (2011). Memoria semántica. Evaluación del conocimiento en carreras de ingeniería. *Revista Evaluar*, 11(1), 40-57. <https://doi.org/10.35670/1667-4545.v11.n1.2843>
- Luria, A. R. (1976). *The working brain: An introduction to neuropsychology*. Penguin.
- Manoiloff, L., Artstein, M., Canavoso, M., Fernández, L., & Seguí, J. (2010). Expanded norms for 400 experimental pictures in an Argentinean Spanish-speaking population. *Behavior Research Methods*, 42(2), 452-460. <https://doi.org/10.3758/brm.42.2.452>
- Marques, J. F. (2006). Specialization and semantic organization: Evidence for multiple semantics linked to sensory modalities. *Memory & Cognition*, 34(1), 60-67. <https://doi.org/10.3758/BF03193386>
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, 58, 25-45. <https://doi.org/10.1146/annurev.psych.57.102904.190143>
- Martínez-Cuitiño, M., & Barreyro, J. P. (2010). ¿Pirámides y palmeras o pirámides y faraones? Adaptación y validación de un test de asociación semántica al español rioplatense. *Interdisciplinaria*, 27(2), 247-260. <https://dialnet.unirioja.es/servlet/revista?codigo=2511>
- Martínez-Cuitiño, M., Barreyro, J. P., & Jaichenco, V. (2009). Adaptación y validación en español de una herramienta de evaluación semántica: La Batería 64. *Neuropsicología Latinoamericana*, 1(1), 24-31. <https://www.neuropsicolatina.org>

- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4), 547-559. <https://doi.org/10.3758/BF03192726>
- McRae, K., Cree, G. S., Westmacott, R., & de Sa, V. R. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology*, 53(4), 360-373. <https://doi.org/10.1037/H0087323>
- Moss, H. E., & Tyler, L. K. (2000). A progressive category-specific semantic deficit for non-living things. *Neuropsychologia*, 38(1), 60-82. [https://doi.org/10.1016/S0028-3932\(99\)00044-5](https://doi.org/10.1016/S0028-3932(99)00044-5)
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 45(3), 255-287. <https://doi.org/10.1037/H0084295>
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8, 976-987. <http://dx.doi.org/10.1038/nrn2277>
- Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, 14(2), 119-124. <https://doi.org/10.1111/1467-9280.t01-1-01429>
- Peraita-Adrados, H., Galeote-Moreno, M. Á., & González-Labra, M. J. (1999). Deterioro de la memoria semántica en pacientes de Alzheimer. *Psicothema*, 11(4), 917-937. <http://www.psicothema.com>
- Peraita-Adrados, H., González-Labra, M. J., Sánchez-Bernardos, M. L., & Galeote-Moreno, M. Á. (2000). Batería de evaluación del deterioro de la memoria semántica en Alzheimer. *Psicothema*, 12(2), 192-200. <http://www.psicothema.com>
- Peraita, H., & Moreno, F. J. (2006). Análisis de la estructura conceptual de categorías semánticas naturales y artificiales en una muestra de pacientes de Alzheimer. *Psicothema*, 18(3), 492-500. <http://www.psicothema.com>
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42(1-3), 107-142. [https://doi.org/10.1016/0010-0277\(92\)90041-F](https://doi.org/10.1016/0010-0277(92)90041-F)
- Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition*, 172, 59-72. <http://dx.doi.org/10.1016/j.cognition.2017.12.007>
- Rogers, T. T., Lambon-Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, 111(1), 205-235. <http://dx.doi.org/10.1037/0033-295X.111.1.205>
- Rose, S. B., & Abdel-Rahman, R. (2016). Cumulative semantic interference for associative relations in language production. *Cognition*, 152, 20-31. <http://dx.doi.org/10.1016/j.cognition.2016.03.013>
- Rose, S. B., Aristei, S., Melinger, A., & Abdel-Rahman, R. (2019). The closer they are, the more they interfere: Semantic similarity of word distractors increases competition in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(4), 753-763. <https://doi.org/10.1037/xlm0000592>
- Sachs, O., Weis, S., Krings, T., Huber, W., & Kircher, T. (2008). Categorical and thematic knowledge representation in the brain: Neural correlates of taxonomic and thematic conceptual relations. *Neuropsychologia*, 46(2), 409-418. <https://doi.org/10.1016/j.neuropsychologia.2007.08.015>
- Santos, L., & Caramazza, A. (2002). The domain-specific hypothesis: A developmental and comparative perspective on category-specific deficits. In E. M. Forde & G. W. Humphreys (Eds), *Category specificity in brain and mind* (pp. 1-24). Psychology Press.
- Sheridan, J., & Humphreys, G. W. (1993). A verbal-semantic category-specific recognition impairment. *Cognitive Neuropsychology*, 10(2), 143-184. <https://doi.org/10.1080/02643299308253459>

- Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21(3), 686-698. <https://doi.org/10.1037/0278-7393.21.3.686>
- Tyler, L. K., Bright, P., Dick, E., Tavares, P., Pilgrim, L., Fletcher, P., Greer, M., & Moss, H. (2003). Do semantic categories activate distinct cortical regions? Evidence for a distributed neural semantic system. *Cognitive Neuropsychology*, 20(3-6), 541-559. <https://doi.org/10.1080/02643290244000211>
- Vieth, H. E., McMahon, K. L., & de Zubicaray, G. I. (2014). Feature overlap slows lexical selection: Evidence from the picture-word interference paradigm. *Quarterly Journal of Experimental Psychology*, 67(12), 2325-2339. <http://dx.doi.org/10.1080/17470218.2014.923922>
- Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48(4), 422-488. <http://dx.doi.org/10.1016/j.cogpsych.2003.09.001>
- Vivas, J. R. (2008). Distsem. Un método de captura y graficación de redes semánticas. Aplicaciones a educación y a neuropsicología. *Revista Mexicana de Psicología*, (Esp. Oct.), 27-31. <https://www.researchgate.net>
- Vivas, J., Vivas, L., Comesaña, A., García-Coni, A., & Vrano, A. (2017). Spanish semantic feature production norms for 400 concrete concepts. *Behavior Research Methods*, 49(3), 1095-1106. <https://doi.org/10.3758/s13428-016-0777-2>
- Vivas, L., & Fernández-Liporace, M. (2017). TAXON: Un nuevo instrumento para evaluar déficits semánticos. *Neurología Argentina*, 9(1), 10-18. <https://doi.org/10.1016/j.neuarg.2016.08.004>
- Vivas, L., García-García, R., Perea-Bartolomé, M. V., D'Almeida, A. L., & Ladera-Fernández, V. (2016). Recognition of thematic and taxonomic conceptual relations in patients with aphasia. *Aphasiology*, 30(6), 657-677. <https://doi.org/10.1080/02687038.2015.1111996>
- Vivas, L., Yerro, M., Romanelli, S., García-Coni, A., Comesaña, A., Lizarralde, F., Passoni, I., & Vivas, J. (2022). New Spanish semantic feature production norms for older adults. *Behavior Research Methods*, 54(2), 970-986. <https://doi.org/10.3758/s13428-021-01660-z>
- Warrington, E. K., & McCarthy, R. A. (1987). Categories of knowledge: Further fractionations and an attempted integration. *Brain*, 110(5), 1273-1296. <https://doi.org/10.1093/BRAIN/110.5.1273>
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107(3), 829-854. <http://dx.doi.org/10.1093/brain/107.3.829>
- Wechsler, D. (2002). *Escala de inteligencia para adultos (WAIS-III)* (3^{ra} ed.). Paidós Ibérica.

Elementos metodológicos de thinking aloud para la obtención de evidencias de validez de contenido

Thinking aloud methodological elements to obtain evidence of content validity

Graciela Ordóñez-Gutiérrez * 1

1 - Universidad de Costa Rica.

Introducción
Método
Discusión
Referencias

Recibido: 04/07/2023 Revisado: 20/09/2023 Aceptado: 27/09/2023

Resumen

Los test *thinking aloud* se han implementado en la investigación educativa, principalmente, para obtener las estrategias de solución de los sujetos en diferentes tareas. En particular, se han empleado con la finalidad de determinar las estrategias que los examinados implementan para dar solución a ítems en pruebas educativas. Sin embargo, los elementos metodológicos para elaborar un *thinking aloud* paso a paso y obtener evidencias de validez no son explícitos en la literatura. Sin embargo, es importante obtener esos resultados, por lo que el objetivo de este artículo es proporcionar los pasos metodológicos, desde una perspectiva práctica de implementación en la obtención de habilidades de razonamiento cuantitativo. Para llevar a cabo los test *thinking aloud*, se realizó una revisión exhaustiva de la literatura y se propusieron las siguientes etapas: 1) definición del propósito, 2) elaboración de los procesos de solución de los ítems desde una perspectiva teórica y práctica, 3) selección de la muestra, 4) proceso de simulación, 5) recolección de los datos, 6) transcripción y análisis. Al seguir estas etapas, se pueden efectuar los test *thinking aloud* que resulten exitosos y que permitan obtener evidencias válidas y confiables.

Palabras clave: *thinking aloud*, evidencias de validez, selección de la muestra, índice de consistencia de expertos, habilidades de razonamiento cuantitativas

Abstract

Think-aloud has been implemented in educational research mainly to define the solution strategies of the subjects on different tasks. In particular, it has been used to determine the strategies the subjects implement to find solutions to items in educational tests. However, since the steps of the methodological think-aloud elements followed to obtain valid evidence are not explicit and are important to define, this article aims to provide the methodological steps, from a theoretical perspective so they can be implemented. To develop the steps, a comprehensive literature review was carried out, and the following steps were proposed: 1) define the purpose, 2) elaborate the solution processes of the items from a theoretical perspective, 3) select the thinking-aloud sample, 4) simulation process, 5) data recollection, 6) think-aloud transcript and analysis. These steps can lead to a successful think-aloud process that produces reliable, valid evidence.

Keywords: *thinking aloud*, valid evidence, select sample, expert consistency index, quantitative reasoning skill

*Correspondencia a: graciela.ordonez@ucr.ac.cr

Cómo citar este artículo: Ordóñez-Gutiérrez, G. (2023): Elementos metodológicos de *thinking aloud* para la obtención de evidencias de validez de contenido. *Revista Evaluar*, 23(3), 45-60. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Participaron en la edición de este artículo: Vanesa Mariela Toledo, Stefano Macri, Juan Cruz Balverdi Nieto, Pablo Carpintero, Florencia Ruiz, Rodrigo Maderna, Jorge Bruera.

Introducción

En los *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) para la elaboración de pruebas educativas y psicológicas, se establece que, para determinar la calidad de medida de los test, es indispensable recolectar evidencias de validez en las que se justifiquen las inferencias que se realizan sobre las puntuaciones que obtiene un grupo de personas. Una de las evidencias requeridas para realizar inferencias válidas sobre las puntuaciones de dichos test hace referencia a la recolección de los procesos de respuestas, o bien estrategias que emplean los sujetos para dar solución a los ítems que componen la prueba (Green, 1998; Embretson, 2017; Keehner et al., 2017; Padilla & Benítez, 2014; Leighton, 2013) y que forman parte de las evidencias relativas a la validez del constructo.

Con la recolección y análisis de las estrategias, se obtendrían evidencias de validez de contenido (Padilla & Leighton, 2017; Fonteyn et al., 1993) y al mismo tiempo permitiría verificar si la prueba realmente mide el constructo que se pretende medir. Además, con los test *thinking aloud*, también llamados en la literatura como “protocolos de pensamiento en voz alta”, “protocolos verbales” e incluso “entrevistas cognitivas”, se pueden detectar posibles fuentes de varianza irrelevante al constructo (Brizuela et al., 2016) y el Funcionamiento Diferencial de los Ítems (DIF, por sus siglas en inglés) (Ercikan et al., 2010). Por lo tanto, los test *thinking aloud* se han empleado para determinar habilidades de razonamiento de las personas y juegan un papel importante en la obtención de evidencias de validez de instrumentos de medición, así como en la elaboración de metodologías de evaluación (Green, 1998). En es-

te sentido, y según Keehner et al. (2017), los test *thinking aloud* aplicados a las tareas de evaluación brindan una perspectiva única sobre el procesamiento individual de los sujetos incluyendo la información referente a los conceptos erróneos, las debilidades en las habilidades y el uso de estrategias para la resolución de problemas.

En otro sentido, Padilla y Leighton (2017) manifiestan que la falta de experiencia de los investigadores, la consolidación de las mejores prácticas para efectuar los test *thinking aloud* y las recomendaciones de cómo obtener evidencia de los procesos de respuesta pueden generar que se pierda la oportunidad de recolectar datos valiosos y sustantivos y resultar en evidencias que lleven a interpretaciones incorrectas sobre las puntuaciones de los examinados. Por otro lado, a pesar de la importancia y de la existencia de una amplia literatura sobre las maneras de efectuar reportes verbales, estos no brindan elementos explícitos y que son indispensables a la hora de realizar los test *thinking aloud* desde la perspectiva de las pruebas educativas para obtener evidencias de validez. Por ello, el objetivo de este artículo es proporcionar las etapas metodológicas, desde una perspectiva teórica y práctica, para efectuar los test *thinking aloud* y obtener evidencias de validez de contenido.

Método

Para elaborar la perspectiva teórica se realizó una búsqueda en sitios web. Primeramente, se escribió en el sitio *scholar.google.com* la expresión “*thinking aloud in education testing*”, la cual arrojó 41 mil resultados en julio del 2019. Luego, se realizó la búsqueda empleando la frase “*thinking aloud in education items testing*” y la cantidad de resultados se redujo a 12. Por otra parte, también se buscó la frase “*verbal protocols*

in education testing” y esto llevó a 13 resultados. Asimismo, se efectuó una búsqueda en las bases de datos *ProQuest* y *Erick*.

Es importante mencionar que, la mayoría de estos escritos tienen como contexto el deportivo, la neurociencia, la psicología, la enfermería, la neuropsicología, entre otras, que están más asociadas a la Ciencias Médicas que a la educación (Ercikan et al., 2010). Por esta razón, la cantidad de escritos que consideran los test *thinking aloud* para la recolección de evidencias desde la perspectiva educativa en general, más precisamente en pruebas educativas, es escasa.

Ahora bien, los documentos que hacen referencia a alguna prueba educativa están enfocados en los procesos de solución de ítems en el área de idiomas (*language*) y área verbal (en el caso del español). Cabe resaltar que, los artículos encontrados se analizaron con detalle y de estos, se extrajo y se compararon las etapas que los investigadores siguieron para realizar los test *thinking aloud*, lo que dio como resultado una cantidad de 6 etapas para su elaboración.

Para elaborar la perspectiva práctica, se realizaron los test *thinking aloud* a 13 personas (7 mujeres y 6 hombres), con el objetivo de determinar habilidades de razonamiento cuantitativo mediante la aplicación de un test. Para esta aplicación, se utilizó el formulario de la Prueba de Habilidades Cuantitativas, la cual está compuesta por 40 ítems de selección única con cuatro opciones de respuesta. Cada aplicación se realizó bajo estricta confidencialidad y de manera individual. Además, se aplicó de manera concurrente y después retrospectiva.

Etapas teóricas de los test thinking aloud

El test *thinking aloud* es un método empleado en estudios asociados a la psicología cognitiva

para recolectar información acerca de los procesos de información humana en la resolución de un problema que requiere de la manipulación de información para encontrar la solución de un estado de cosas complejo (Padilla & Leighton, 2017; Fonteyn et al., 1993). Actualmente, a este método lo emplean algunos investigadores para explorar la validez de los ítems en pruebas de rendimiento (Leighton, 2004; Padilla & Leighton, 2017; Embretson, 2017), pues proporciona información sobre los procesos cognitivos generados por los examinados, y por ende, para determinar evidencias de validez que permitan inferencias sobre las puntuaciones del test. Además, permite fortalecer el constructo que se pretende medir con la prueba. Por ello, la manera de llevar a cabo los test *thinking aloud* permitirá recolectar la información suficiente y necesaria de forma rigurosa. Por este motivo, a continuación, se explican las etapas que permitirán la elaboración de reportes verbales exitosos en el contexto de pruebas educativas.

Es importante mencionar que, existe una coincidencia entre la literatura sobre la preocupación de la confiabilidad de los datos obtenidos con los test *thinking aloud*, ya que las verbalizaciones de los sujetos pueden reflejar lo que un grupo de examinados piensa, en lo que cree que están haciendo y no sus procesos de pensamiento reales (Ercikan et al., 2010). Ante esto, en Leighton (2004) se exponen algunos problemas relacionados con la confiabilidad de las verbalizaciones y, también, se pueden encontrar recomendaciones al respecto. Específicamente, Leighton (2004) hace referencia a tres factores que afectan el uso exitoso de los test *thinking aloud*: 1) el momento en que se recolectan los informes, 2) la manera en que se recolectan, y 3) las tareas, o bien los ítems, seleccionados para la recolección de los procesos. Por eso, se recomienda seguir una detallada y rigurosa metodología en la que se resguarde la confiabilidad de los reportes. A continuación, se

detallan los pasos sugeridos en la literatura para realizar los test *thinking aloud*.

Definir el propósito de efectuar los test thinking aloud

Esta primera etapa es indispensable y no se encuentra explícita en la literatura revisada, por lo que su planteamiento se establece a partir de la experiencia de la autora de este trabajo. Pues, definir *qué y para qué* realizar los test *thinking aloud* permite al investigador tener un panorama claro de lo que realmente se requiere y se pretende con ello. Además, dicha definición le genera los indicadores que necesitará en el momento de implementar los test *thinking aloud* con las personas examinadas. Luego, se deberá elaborar una guía con los elementos que se incluyeron en la definición del *qué y para qué*.

Por otra parte, se deberán elegir los ítems de la prueba que presenten las mejores medidas métricas según alguna teoría de medición (esta puede ser a partir de la Teoría Clásica de los Test, Teoría de Respuesta a los Ítems, u otra), según la dificultad y la discriminación. Está claro que, esto depende de la finalidad de efectuar los test *thinking aloud* en la obtención de evidencias de validez del constructo, ya sea que con el constructo se quieran medir habilidades o conocimientos. Si la finalidad es recolectar información con respecto a las dificultades o errores que comenten los examinados, entonces, se deberá elaborar un listado de los posibles errores para, luego, determinar qué elementos son considerados por el grupo de personas a la hora de cometerlos. Si la finalidad es determinar las habilidades, entonces, se deberá contar con una serie de indicadores que guíen la identificación de esas habilidades.

Elaboración de los procesos de cada ítem

Luego de seleccionar los ítems, la persona investigadora deberá considerar jueces expertos en el tema de interés, para que resuelva cada uno de los reactivos. De ser posible, cada ítem deberá estar enmarcado en alguna dimensión, según el constructo a medir, o bien, según la finalidad de medición mediante la prueba. En esta etapa, cada juez debe valorar si el reactivo pertenece o no a la dimensión del constructo seleccionado y resolverlo proporcionando la solución, según sus requerimientos. Es importante aclarar que, los jueces deben ser expertos en el área de interés.

Después de elaborar los procesos dados por los jueces, la persona investigadora deberá realizar el análisis de congruencia sobre los procesos generados por los expertos y verificar que todos siguen el mismo proceso de solución, o bien, llegar a un consenso entre ellos para establecer el mejor camino de solución y que responda a las dimensiones del constructo. Dicho análisis se podrá realizar empleando algún índice de congruencia entre jueces, como por ejemplo, Kappa de Cohen (Landis & Koch, 1977); Kappa de Fleiss (Fleiss, 1971; Cohen, 1960), el índice de Rovinelli y Hambleton (1977), *Many-Facet-Rasch Measurement* (Prieto-Adánez, 2011), el índice de consistencia de *Hierarchy* (Leighton et al., 2009), la metodología *Rule-Space* (Artavia-Medrano, 2015) o la Teoría de la Generalizabilidad (Martínez-Arias et al., 2006).

Elección del tamaño de la muestra

La elección del tamaño de la muestra, o de la cantidad de personas que se elegirán para efectuar los test *thinking aloud* es indispensable, ya que con ello se determinará la cantidad de datos que se producirán (Padilla & Leighton, 2017; Leighton,

2004). Además, dependiendo de la personalidad de cada participante, se pueden obtener datos fructuosos en información, o por lo contrario, se pueden obtener datos bajos en información. Desde esta perspectiva, Fonteyn et al. (1993) argumenta que, lo mejor es recolectar datos ricos y profundos de una pequeña muestra de sujetos.

Con respecto a la muestra de sujetos, varios estudios (Fonteyn et al., 1993; Virzi, 1992) sugieren que una cantidad de 5 o 6 personas puede producir resultados estables. No obstante, otros investigadores indican que una muestra pequeña y heterogénea de sujetos afecta la relación entre el tamaño de los datos y su estabilidad (Van Den Haak et al., 2003). Cabe aclarar que el tamaño de la muestra dependerá de la finalidad de realizar los test *thinking aloud*. En este sentido, Padilla y Benítez (2014) manifiestan que, para determinar el número total de los participantes, se deben tener dos criterios: 1) la saturación teórica y 2) la relevancia de la información. Ambos criterios afectan el muestreo porque el número y las características de los sujetos dependen del análisis de las entrevistas. De esta manera, Padilla y Benítez (2014) indican que la saturación teórica hace referencia a que los investigadores deberán seguir aplicando los test *thinking aloud* hasta que no surjan nuevos hallazgos, mientras que la relevancia teórica hace referencia a la selección de los entrevistados, de acuerdo con la teoría del constructo de la investigación.

Con respecto al primer criterio, lo que se pretende es construir teoría, mientras que con el segundo se pretende comprobar una teoría. Ahora bien, si se quiere construir una teoría, entonces, la cantidad de sujetos a considerar en el muestreo es entre 20 y 50, pero si se requiere comprobar la teoría, basta con considerar 10 sujetos, ya que los patrones de respuestas en función de la teoría se vuelven repetitivas a partir del séptimo sujeto (Padilla & Benítez, 2014). Es importante mencio-

nar que, estos criterios deben ser considerados en términos del objetivo del estudio y de la complejidad de la evaluación propuesta.

Entrenamiento de las personas participantes

Antes de realizar los test *thinking aloud* cuyos datos formarán parte del estudio, es importante entrenar a las personas participantes del estudio; esto permitirá al grupo de sujetos tener una visualización de lo que se quiere que se realice en el estudio. Además, se puede anticipar qué dificultades tendrá el participante en la verbalización de los procesos de respuesta, y así, poder controlarla en la recolección de los datos reales (Padilla & Leighton, 2017). En este sentido, Fonteyn et al. (1993) manifiestan que, el investigador puede hacer una valoración sobre la relevancia de las verbalizaciones de los sujetos, de la misma tarea a proporcionar, de la preselección de la tarea, o bien, de la escogencia del ítem, de la estandarización en cuanto a la forma de llevar a cabo el trato con el sujeto y, además, una valoración de aspectos emergentes a la hora de realizar los test *thinking aloud* necesarios para su investigación. Además, anticipará formas adicionales de abordar un mismo tema y de controlar el tiempo real con la finalidad de estandarizarlos para todos los examinados (Fonteyn et al., 1993). Ahora bien, es crucial tomar en cuenta que, aunque los entrenamientos permiten establecer una presentación de un conjunto estándar de circunstancias, estas no son iguales para todo el conjunto de participantes, pues son diferentes a las que se pueden presentar en las condiciones reales. Sin embargo, sí proporcionan una idea inicial de las diferentes circunstancias que se puedan presentar.

Recolección de los datos

Padilla y Leighton (2017) y Fonteyn et al. (1993) argumentan que, cada persona participante tiene que ser convocado en un horario en el cual la persona investigadora y participante puedan. Las sesiones deben darse de manera individual y el lugar debe ser un entorno tranquilo que facilite el test *thinking aloud* de los sujetos y en las que no se generen distracciones.

Además, todas las sesiones de los test *thinking aloud* deben ser grabadas en audio, seguida de la transcripción de los datos verbales producidos por los entrevistados. Por otra parte, se les debe decir a las personas participantes que piensen constantemente en voz alta, sin dejar que se generen pausas prolongadas (entre dos o más minutos), ya que con esto se perdería información sobre los procesos de pensamiento. Si las personas participantes realizan una pausa un poco prolongada, entonces el investigador debe recordarles que deben resolver los ítems o la tarea expresando en voz alta la manera en que van resolviéndola. No obstante, se debe recordar que la interacción entre el investigador y el sujeto debe ser mínima mientras se estén efectuando los test *thinking aloud* con mínimas interrupciones.

En Green (1998), se indican algunas recomendaciones para la recolección de los datos antes de iniciar las sesiones de los test *thinking aloud*. Algunas de estas son:

1 - Preparar instrucciones sin ambigüedades en las que se explicita lo que se debe hacer con el objeto, la tarea o el ítem; es decir, una instrucción clara.

2 - Informar a las personas sobre lo que realmente se requiere de ellos y explicar claramente cuál es el procedimiento y las condiciones de este.

3 - Practicar con la persona al menos un par de veces para que adquiera confianza y no se intimide ante la tarea que debe ejecutar.

4 - Aclarar a la persona participante que debe mantenerse hablando; si se mantiene en silencio por periodos prolongados, realice la siguiente pregunta: ¿puedes indicarme qué estás pensando?

5 - Durante la sesión, estar atento a los silencios prolongados de las personas participantes.

6 - Permitir cierto tiempo adicional si no se ha terminado la tarea, puesto que la verbalización requiere de más tiempo que el simple hecho de escribir y elegir una respuesta correcta.

7 - Utilizar un equipo adicional como videocámaras, o grabaciones en audio, pues no se puede recolectar todo con solo escuchar y escribir en un reporte.

8 - Finalmente, después de las sesiones, efectúe preguntas que le permitan corroborar lo expresado por la persona participante a la hora de realizar la tarea que se indicó. En este caso, se le puede pedir que explique el proceso realizado para resolver la tarea.

Desde otra perspectiva, Fonteyn et al. (1993), recomiendan que las verbalizaciones de los participantes se deben dividir en dos partes: una concurrente y otra retrospectiva. Si la información se verbaliza en el momento en que el sujeto resuelve la tarea o el ítem, esta será una verbalización concurrente. Pero, si se le indica a la persona participante que explique el proceso llevado a cabo para los ítems, o la tarea, luego de haberlos resuelto, esta será una verbalización retrospectiva (Ericsson & Simon, 1987; Russo, Johnson & Stephens, 1989; Sapsirin, 2016). Ambos tipos de verbalización son importantes para determinar evidencias de validez en pruebas educativas y también de instrumentos de medición. En este sentido, la información que se obtenga del proceso retrospectivo complementa los datos concurrentes.

Transcripción, codificación y análisis de los thinking aloud

La transcripción de los datos es la parte más larga y lenta de efectuar. Sobre todo, si no se cuenta con los softwares modernos que permitan la transcripción de los datos de audio a datos escritos. Sin embargo, es una de las etapas que se deben de realizar para efectuar la codificación y, posteriormente, el análisis de los datos. La transcripción y codificación de los test *thinking aloud* deben ser un proceso riguroso y estandarizado, que incluya múltiples evaluadores y el cálculo de un índice de confiabilidad, como los mencionados en el apartado de elaboración de los procesos de solución de los ítems (Padilla & Leighton, 2017; Ericsson & Simon, 1987; Green, 1998).

La persona investigadora debe realizar las transcripciones de manera individualizada; esto es, por cada sujeto participante, de forma tal que se respete el lenguaje y las terminologías empleadas por cada uno. Luego de la transcripción, las verbalizaciones de los participantes se pueden dividir en secciones o en líneas de unidades o segmentos etiquetados, de acuerdo con la teoría establecida previamente, o según la finalidad de la prueba. Estas divisiones facilitan la identificación de partes del texto que son importantes para determinar puntos o palabras claves de la investigación (Joseph & Patel, 1990)

Ahora bien, durante la etapa de transcripción, es importante tener en cuenta: 1) la creación de una codificación de los test *thinking aloud* basada en un modelo teórico del constructo a medir sobre la tarea o proceso desarrollado por las personas participantes, 2) la inclusión del conjunto de conocimientos y habilidades esperadas, y ejemplos de los tipos de respuestas que se esperaría como evidencia de los conocimientos y habilidades; 3) si fuera el caso, la capacitación a los evaluadores que efectuarán la revisión y va-

loración de los procesos de solución de cada uno de los ítems realizados por los examinados en los test *thinking aloud* y los esperados teóricamente; 4) en el entrenamiento de los evaluadores, la indicación de que no valoren la dificultad del ítem, la discriminación, ni tampoco el funcionamiento diferencial de los ítems, sino que se centren en los procesos de respuesta brindado por cada uno de los participantes; 5) el cálculo del acuerdo inicial entre los evaluadores y determinación de la fuerza de concordancia entre las puntuaciones otorgadas en los procesos realizados versus los procesos esperados; 6) si la fuerza de concordancia entre los jueces es baja, entonces se debe volver a realizar una valoración con otros jueces, con la finalidad de obtener una puntuación de concordancia aceptable (mayor a 0.60 en Kappa por ejemplo).

En otro sentido, Ericsson y Simon (1984) manifiestan que, el análisis de los protocolos conlleva tres pasos: 1) un análisis de las frases de referencia; 2) un análisis afirmativo; 3) un análisis de guiones. En el primer paso, el investigador debe identificar todos los nombres y frases nominales en los datos verbales de cada sujeto y codificar las frases con el nombre del concepto de referencia. Estos códigos definen e identifican el vocabulario que los sujetos han verbalizado durante la ejecución de la tarea. Luego, cada concepto es identificado y codificado, de acuerdo con la identificación del investigador y según la naturaleza de la tarea. Ahora bien, el análisis de las frases de referencia continúa, según Fonteyn et al. (1993), hasta que todos los conceptos de referencia en las transcripciones hayan sido codificados y todas las frases de referencia hayan sido examinadas para garantizar que no queden conceptos indefinidos.

Generación de informes

Cuando se ha concluido cualquier proyecto investigativo, es indispensable otorgar a la población de estudio un informe con respecto a los resultados que surgieron de la investigación; esto permitiría a las personas participantes conocer sus fortalezas y debilidades en el contenido o en el constructo, e implementar programas de mejora. Igualmente, se demostraría la transparencia con respecto a la investigación realizada y, por ende, se ayudaría a motivar a las personas a participar en estudios futuros. Cabe destacar que, lo importante en la generación de los informes es el mantenimiento de la ética: ser honestos y respetuosos con respecto a los datos presentados, analizados y otorgados por los demás.

Evidencias de habilidades de razonamiento cuantitativo mediante la aplicación de los test *thinking aloud*

Definir el propósito de efectuar los test thinking aloud

Este estudio forma parte de una investigación principal que inició en el año 2020. El principal objetivo del estudio fue analizar las evidencias de validez, que permiten realizar inferencias sobre las habilidades de razonamiento cuantitativo demostradas por un grupo de examinados en la Prueba de Habilidades Cuantitativas y las requeridas en los cursos de Química General I e Introducción a la Química en la Universidad de Costa Rica.

Para cumplir con el objetivo principal, el estudio se dividió en siete etapas y en una de ellas se realizaron los test *thinking aloud* con la finalidad de obtener evidencias de validez de contenido y del constructo de la Prueba de Habilidades Cuantitativas (PHC), a partir de los procesos de respuesta realizado por un grupo de examinados y,

de esta manera, identificar las habilidades de razonamiento cuantitativo demostrado por las personas a la hora de resolver los ítems.

Cabe mencionar que, la PHC es un requisito de admisión en carreras que lo solicitan y está formada por 40 ítems de selección única con cuatro opciones de respuesta. Cada año, desde el 2015, las personas se inscriben para luego postularse a las carreras que la tienen como requisito de admisión. Además, la dificultad TRI promedio es igual a 0, y se busca maximizar la precisión del nivel de habilidad ($\theta=0$) (Rojas-Torres & Ordóñez-Gutiérrez, 2019).

Elaboración de los procesos de cada ítem

Para llevar a cabo los test *thinking aloud*, primeramente, se elaboró un marco teórico de referencia en el que se define el constructo razonamiento cuantitativo que se mide mediante la PHC. Además, dicho constructo se explica por las siguientes dimensiones: cuantificar, relacionar, clasificar, ejemplificar, validar y generalizar.

Luego, se solicitó a un grupo de expertos que resolvieran los ítems de la PHC que se aplicó en el año 2020, para los ingresos a las carreras en el año 2021. Cada experto los resolvió y los clasificó según las dimensiones propuestas del constructo. Además, se solicitó analizar si los ítems podían medir distintas dimensiones al mismo tiempo, según el proceso de solución realizado, o bien, el camino escogido para resolver el reactivo. Las soluciones fueron codificadas de acuerdo con las dimensiones del constructo.

Elaboración de los procesos de cada ítem

Para llevar a cabo los test *thinking aloud*, primeramente, se elaboró un marco teórico de

referencia en el que se define el constructo razonamiento cuantitativo que se mide mediante la PHC. Además, dicho constructo se explica por las siguientes dimensiones: cuantificar, relacionar, clasificar, ejemplificar, validar y generalizar.

Luego, se solicitó a un grupo de expertos que resolvieran los ítems de la PHC que se aplicó en el año 2020, para los ingresos a las carreras en el año 2021. Cada experto los resolvió y los clasificó según las dimensiones propuestas del constructo. Además, se solicitó analizar si los ítems podían medir distintas dimensiones al mismo tiempo, según el proceso de solución realizado, o bien, el camino escogido para resolver el reactivo. Las soluciones fueron codificadas de acuerdo con las dimensiones del constructo.

Elección del tamaño de la muestra

Para elegir a los participantes se consideró lo que teóricamente está establecido. Esto es, se escogieron las personas que realizaron la PHC en el año 2020, que ingresaron a alguna de las carreras en el año 2021 y que tuvieran en su malla curricular el curso Introducción a la Química o Química General I. Las personas seleccionadas estaban empadronadas para el año 2021 en alguna de las siguientes carreras: Farmacia, Química, Ingeniería Mecánica, Ingeniería Eléctrica, Física, Meteorología y Geología. Además, las personas seleccionadas obtuvieron una calificación mayor e igual a 90 en la PHC que se aplicó en el año 2020 para los ingresos a las carreras porque se quería sustentar la teoría y no establecer niveles de razonamiento.

La cantidad de participantes fue de 13 personas, 7 mujeres y 6 varones, la cual se consideró suficiente, ya que la finalidad fue comprobar la teoría y no su elaboración. Además, a partir del quinto participante, se logró determinar los patrones de respuesta en función de la teoría, tal y como mencionan Padilla y Benítez (2014).

Entrenamiento

Antes de que la persona participante realizara el proceso de los test *thinking aloud*, se leyó el consentimiento informado y después se brindó una explicación para llevar a cabo el entrenamiento. Para esto, la persona investigadora efectuó el protocolo de resolución para que la persona participante observara cómo debía proceder. Cada participante tuvo que resolver dos ítems que estaban en una ficha llamada *Ficha de entrenamiento para realizar la técnica thinking aloud*, la cual contenía dos ítems muy semejantes a los que se proporcionan en la PHC. Luego de resolver los ítems de entrenamiento, la investigadora determinaba si la persona participante lograba captar qué se requería en el ejercicio o no; si la persona lograba tener la fluidez en el ejercicio, entonces, iniciaba el test *thinking aloud* con los reactivos de la prueba meta, y si no, debía volver a realizar dicho ejercicio.

Es importante comentar que la mayoría de las personas participantes lograron captar, de manera inmediata, qué era lo que debían realizar en el primer intento; solamente una de las participantes tuvo que realizar el ejercicio dos veces, en gran medida debido a la ansiedad y nerviosismo que presentó. Ahora bien, cuando se evidenciaba exceso de nerviosismo, la investigadora procedió a entablar una conversación, sin que la persona participante se sintiera amenazada o estresada; después, la misma persona optaba por iniciar con los test *thinking aloud*.

Recolección de los datos

Con respecto a la recolección de los datos, se siguió lo establecido por Padilla y Leighton (2017), Fonteyn et al. (1993) y Green (1998), quienes indican que los test *thinking aloud* se deben realizar en un lugar tranquilo, que deben

efectuarse de manera individual y que se deben grabar en audio o en video. Para dicho proceso de recolección de los datos, se prepararon instrucciones haciendo explícito lo que se debía realizar. Además, los test *thinking aloud* se realizaron en las habitaciones propias de cada uno de los participantes porque a nivel país, se encontraban en confinamiento debido a la pandemia por COVID-19. También, se les explicó a las personas qué se esperaba de cada uno, el procedimiento y la duración de los test. Se realizó un entrenamiento antes de iniciar con los ítems del estudio y, cuando se generaban períodos de silencio, la investigadora les decía: “Recuerda que debes verbalizar mientras estás resolviendo los ítems”.

Las verbalizaciones se realizaron, primeramente, de manera concurrente, y luego, de manera retrospectiva. En el caso del modo concurrente, las personas participantes debían verbalizar lo que estaban pensando en el momento de resolver el ítem, mientras que, para el modo retrospectivo, la persona debía explicar a la investigadora qué fue lo que empleó para resolver el ítem (Padilla & Leighton, 2017).

Transcripción, codificación y análisis de los thinking aloud

Luego de las verbalizaciones de los *test thinking aloud*, se procedió a la transcripción de las verbalizaciones de las 13 personas participantes y de los 40 ítems que contenía el formulario de la PHC. Las transcripciones se realizaron en *Word* y luego, se limpiaron quitando las verbalizaciones del enunciado de los ítems y solamente se dejó el proceso de respuesta del grupo de participantes.

Los análisis se realizaron mediante el software *Atlas ti.9* y se consideraron las postulaciones teóricas con respecto al constructo de razonamiento cuantitativo y sus dimensiones. Luego,

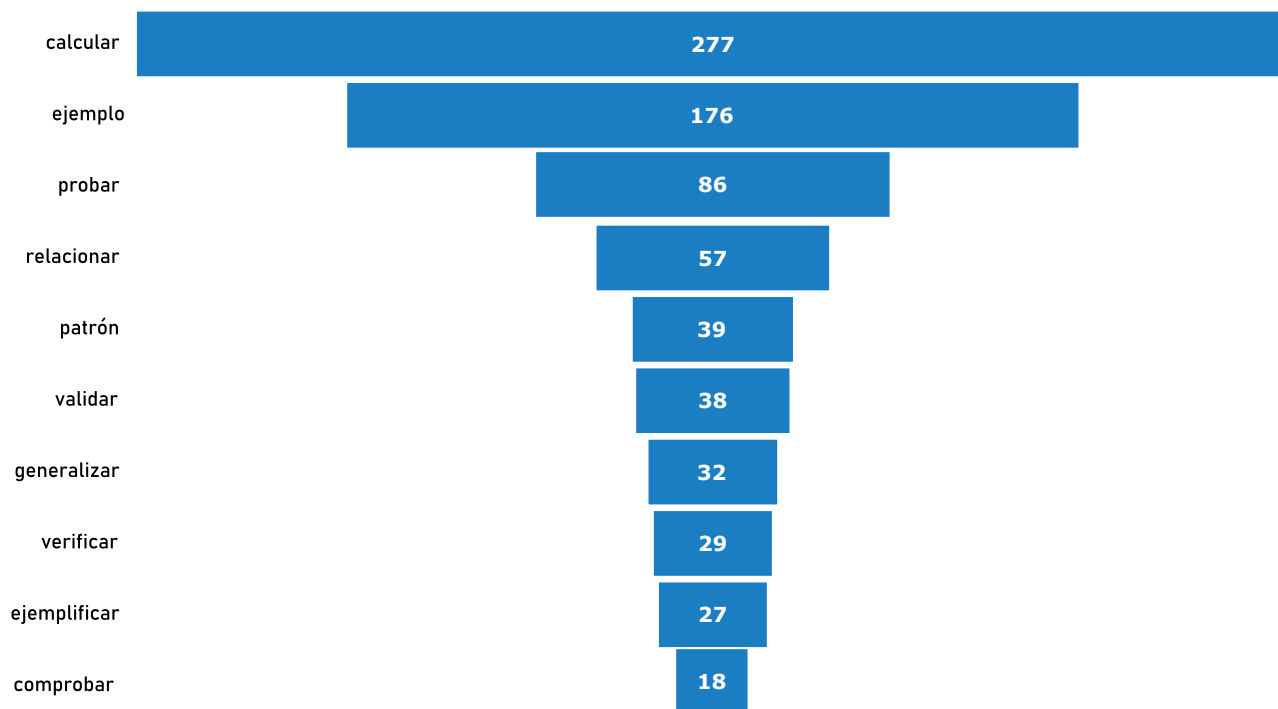
se realizó un análisis de frecuencia de palabras. Posteriormente, se tomaron frases y un grupo de personas expertas en el constructo realizó un juzgamiento sobre los procesos de solución producidos por las personas participantes codificando y clasificando los procesos en cada una de las dimensiones del constructo.

Como cada persona experta debía juzgar el proceso de solución y clasificarlo, dicho juzgamiento se analizó mediante el índice de *Kappa de Fleiss*. De esta manera, se evaluó la estabilidad de acuerdos entre los jueces al asignar calificaciones categóricas a un ítem determinado. Cabe mencionar que, al grupo de jueces se le entregó un instrumento donde debía indicar: *(-1)* si el proceso efectuado por el conjunto de examinados no correspondía a lo solicitado en el ítem, según categoría; *(0)* si el proceso efectuado por el conjunto de examinados correspondía medianamente a lo solicitado en el ítem, según categoría; y *(1)* si el proceso efectuado por el conjunto de examinados correspondía fuertemente a lo solicitado en el ítem, según categoría.

Generación de informes

La generación de informes implica proporcionar los resultados del estudio. De esta manera, como primer resultado de la aplicación de los *test thinking aloud*, tenemos el análisis de la frecuencia de palabras que relacionan las verbalizaciones con lo establecido teóricamente. Esta frecuencia se establece en la Figura 1.

Se puede notar que la palabra más frecuente entre el grupo de personas participantes fue “calcular”. Sin embargo, se debe aclarar que las personas, cuando están ante la resolución de una prueba que tiene contenido matemático, a todo le establecen el sello de “calcular”. También, se pueden observar las palabras “relacionar”, “patrón”,

**Figura 1**

Análisis exploratorio de la frecuencia de palabras empleadas por el grupo de personas participantes en la resolución de los ítems de la PHC, 2020.

“ejemplo”, “validar”, “generalizar”, “comprobar” y “verificar”. Al revisar la literatura correspondiente a las dimensiones del razonamiento cuantitativo, se puede notar que las personas participantes emplean estas palabras en la resolución de los ítems y que, además, forman parte de las habilidades de razonamiento cuantitativo que se requieren en los cursos de Química indicados.

Como segundo resultado, se obtuvo la congruencia entre jueces sobre el análisis de los procesos de respuesta y su clasificación en las dimensiones del constructo. El índice de congruencia calculado fue *Kappa de Fleiss*, según la categoría del constructo y el formulario analizado, y se presenta en la Tabla 1.

Se puede notar que los valores más bajos corresponden a la categoría “relacionar”. Sin embargo, Ruiz y Luciano (2012) explican que en la Teoría del Marco Relacional existen diferentes tipos de relaciones de acuerdo con los procesos que

se deban emplear en la resolución de un problema, y que estas son: coordinación (es, es igual a, es semejante, es similar), oposición (es opuesto de, es lo contrario a, es diferente de), espaciales (arriba, abajo, lejos, cerca), deícticos (yo, tu, aquí, allí, antes, después), y que la visualización dependerá del contexto de formación de cada individuo. Ahora bien, el índice de congruencia sobre los procesos de solución para el formulario en general fue de 0.77, 0.75, 0.75 y 0.78 respectivamente y se consideró un valor muy bueno.

Por otro lado, se logró establecer cuáles habilidades de razonamiento cuantitativo lograban demostrar el grupo de examinados a la hora de resolver los ítems de la Prueba de Habilidades Cuantitativas y se tomó en cuenta la teorización del constructo y sus dimensiones. Estas se establecen a continuación en la Tabla 2.

Finalmente, se identificó en los patrones de respuestas aquellas habilidades que el grupo de

Tabla 1

Valor *Kappa de Fleiss* para la congruencia entre jueces según categorización de los ítems de la PHC 2020 por cada cuadernillo (o fórmula).

Categorías	<i>Kappa de Fleiss</i>			
	F1	F2	F3	F4
Calcular	0.73	0.72	0.72	0.73
Relacionar	0.68	0.63	0.63	0.68
Clasificar	1.00	1.00	1.00	1.00
Ejemplificar	0.85	0.85	0.85	0.85
Validar	0.79	0.76	0.76	0.78
Generalizar	1.00	1.00	1.00	1.00

Nota. * p -value = 0, z = 14.8.

personas participantes logró demostrar en resolución de los ítems de manera concurrente y retrospectiva. Este informe se les devolvió y se brindaron sugerencias de mejora.

Discusión

Este artículo muestra las etapas para realizar los test *thinking aloud* mediante una aplicación teórico-práctica en la determinación de habilidades de razonamiento cuantitativo que conlleva revisar rigurosa y sistemáticamente elementos indispensables del constructo a considerar y así, no generar vacíos a la hora de analizar los datos que se obtienen.

Ahora bien, es indispensable efectuar los test *thinking aloud* de manera planificada y con un objetivo claro. Además, es necesario tomar en cuenta todos los aspectos importantes para el estudio; por ejemplo, el tiempo destinado para la ejecución de los reportes verbales en los que tanto el investigador como los participantes tengan el tiempo suficiente para abarcar todos los puntos que se presentan en los ítems y de manera tal que no se vuelva una carrera por terminar la tarea otorgada.

Es importante indicar que, realizar los test *thinking aloud* al estudiantado permitirá que el

cuerpo docente innove, estructure y realice mejoras en su labor docente. En consecuencia, conocer cómo piensan los estudiantes su materia permitirá saber la manera en que los conceptos son aprendidos. Igualmente, se genera conciencia pedagógica, y se desarrollan habilidades, en particular, habilidades de razonamiento cuantitativo que resultan imprescindibles para el funcionamiento adaptativo en la sociedad actual, tal y como lo plantea [Stelzer et al. \(2020\)](#).

Es importante mencionar que, en este escrito, se pone a disposición una guía general de las etapas que se deberían seguir, como mínimo, para efectuar los test *thinking aloud* de manera tal que permitan a los investigadores obtener evidencias de validez de contenido con respecto al constructo de interés. Por otro lado, se aporta una bibliografía a la que se puede acudir para efectuar una revisión en profundidad. La aplicación de los test *thinking aloud* realmente no es reciente; sin embargo, se ha desarrollado poco para obtener evidencias de validez de contenido. Además, las investigaciones en las que se mencionan no establecen una estructura evidente de los pasos por seguir para una obtención exitosa de los datos y de los análisis que se deberían realizar al respecto.

Por otra parte, las etapas plasmadas en este artículo no son estáticas, ya que pueden variar se-

Tabla 2

Habilidades de razonamiento cuantitativo demostrado por un grupo de personas en la resolución de los ítems de la PHC.

Dimensión	Subcategoría	Verbalización
Cuantificar	Obtención de medidas	“Lo que hice fue <i>sacar las medidas</i> y con esas medidas grandes fui partiendo para encontrar las medidas más chiquititas y así, saber las medidas de los lados” (Exa1). “Empecé a sacar medidas y al final, hice la resta” (Exa4).
	Verificación de las respuestas	“Primero, hice un mini cálculo para ver, como, para saber si es negativo o positivo; luego, verifico que eso se dé...” (Exa5).
	Simplificación de expresiones	“Voy a simplificar un poco esta expresión” (Exa12).
	Efectuar operaciones	“Está multiplicando esto, entonces, como veo que este producto de la suma “ (Exa1).
Relacionar	Coordinación	“Agrupo términos semejantes...” (Exa4). “Me daban <i>relaciones</i> entre diferentes variables...como no había una <i>relación</i> directa entre...” (Exa2).
	Oposición	“Es lo <i>opuesto</i> a lo que me piden” (Exa7). “ <i>R</i> sería <i>mayor a p</i> ...” (Exa1).
	Espaciales	“Entonces, comprobamos la de arriba” (Exa11). “Multiplicando arriba y abajo”(Exa12).
	Equivalencias	“Bueno, por la equivalencia que me están dando” (Exa3). “Yo saqué la <i>equivalencia</i> de lo que es el” (Exa8).
	Conexiones o transformación de expresiones	“Estoy <i>comparando</i> la opción D con esto, a ver si puedo ver alguna <i>conexión</i> de este tipo...” (Exa2). “Es la misma regla de tres, pero...” (Exa3).
	Clasificar	Identificación de características numéricas
Clasificación de figuras		“Lo que se me ocurre es dibujar esta línea y verlo (sic) como 4 triángulos (Exa4).
Ubicación de elementos dentro de un conjunto		“W pertenece con total certeza al intervalo” (Exa3). “Los términos que tienen Y se agrupan en” (Exa1).
Ordenar y caracterizar objetos		“No me dicen nada de cómo están ordenados” (Exa13). “Cuando ordenamos los datos como el...” (Exa12).
Ejemplificar	Identificación de patrones	“Usé ejemplos y busqué excepciones” (Exa10). “Lo que hice fue coger un ejemplo, bueno, varios ejemplos para ver si lograba ver algún patrón al (sic) algo parecido...” (Exa9).
	Verificación de respuestas	“Lo confirmó con varios números para estar segura” (Exa2). “Fui probando cada uno de ellos [las opciones de respuesta] hasta llegar a...” (Exa7).
	Dar solución al problema planteado	“Se me ocurre que la manera más fácil es que se haga un ejemplo” (Exa2). “Utilicé ejemplos hipotéticos y analicé que todo dependía” (Exa10).
	Encontrar de manera directa la solución	“Lo que hice fue hacer casos para ver cuál era la solución” (Exa6). “Entonces si S es 100 por ejemplo en este caso” (Exa3).

Dimensión	Subcategoría	Verbalización
Validar	Determinar el valor de verdad de un enunciado	“Voy a <i>poner un ejemplo</i> 1, 2, 3, 4, 5 porque cumplen las condiciones...” (Exa10). “Si asumiera que es un cero... pero esto no me indica que alguna de las respuestas sea cierta” (Exa9).
	Plantear relaciones	“Utilicé ejemplos específicos para ver la relación” (Exa6). “Por ejemplo, digamos que la x sea mayor que la y” (Exa1).
	Recordar conceptos	“Voy a darme valores para recordar...” (Exa4). “Estoy pensando en un ejemplo primero para ver...” (Exa5).
	Obtención del valor de verdad mediante ejemplos específicos	“Eso no es cierto porque ya sabemos que es” (Exa5, Exa7).
	Conjeturar mediante juicios de valor	“ <i>Bajo ese escenario</i> , las otras tres opciones se descartan y la C no” (Exa13). “Entonces, voy a volver a comprobar la opción A y la opción D” (Exa4).
	Confirmación de proposiciones	“Voy a confirmar con las otras opciones” (Exa1). “Se me ocurre cómo confirmar que eso puede ser” (Exa9).
Generalizar	Identificación de patrones	“Debería ver el patrón que sigue” (Exa2). “Aquí, lo que estoy haciendo es tratando (sic) de identificar un patrón”(Exa3).
	Encontrar una solución mediante un patrón establecido	“Cómo hago para generalizar el 4n” (Exa8). “Estoy intentando de (sic) encontrar la manera lógica de generalizar el n” (Exa13).
	Identificación de propiedades	“Veo un tipo de patrón en los resultados” (Exa13). “Me doy cuenta (sic) que todo producto de consecutivos debería ser la generalización de que todo...” (Exa2).

gún la finalidad de la investigación y el objetivo que se persigue con el instrumento de medición. Por ello, se debe tener en cuenta *qué y para qué* efectuar los test *thinking aloud*, además de saber cuál es el constructo para considerar. Igualmente, la manera de efectuar los análisis puede variar, ya que dependen de los avances en las tecnologías y de las nuevas propuestas que surjan para recolectar los datos de manera rigurosa, de manera tal que se obtengan mejores resultados al respecto.

Referencias

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- Artavia-Medrano, Á. (2015). Interpretación y análisis de pruebas psicológicas y educativas con el método Rule-Space. *Revista Actualidades en Psicología: Medición y Psicometría*, 29(119), 63-77. <https://doi.org/10.15517/ap.v29i119.18724>
- Brizuela, A., Jiménez-Alfaro, K., Pérez-Rojas, N., Rojas-Rojas, G. (2016). Autorreportes verbales en voz alta para la identificación de procesos de razonamiento en pruebas estandarizadas. *Revista Costarricense de Psicología*, 35(1), 17-30. <http://dx.doi.org/10.22544/rcps.v35i01.02>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Embretson, S. (2017). An Integrative Framework for Construct Validity. En A. A. Rupp & J. P. Leighton

- (Eds.), *The Handbook of Cognition and Assessment: Frameworks, Methodologies and Applications* (pp. 102-123). Wiley Blackwell.
- Ercikan, K., Arim, R., Law, D., Domene, J., Gagnon, F., & Lacroix, S. (2010). Application of think aloud protocols for examining and confirming source of Differential Item functioning identified by expert reviews. *Educational Measurement: Issues and Practice*, 29(2), 24-35. <https://doi.org/10.1111/j.1745-3992.2010.00173.x>
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. The MIT Press.
- Ericsson, K. A., & Simon, H. A. (1987). Verbal reports on thinking. En C. Faerch & G. Kasper (Eds.), *Multilingual matter: Introspection in second language research* (pp. 24-53). Multilingual Matters.
- Fleiss, J. L. (1971). Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin*, 76(5), 378-382. <https://doi.org/10.1037/h0031619>
- Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A Description of Think Aloud Method and Protocol Analysis. *Qualitative Health Research*, 3(4) 430-441. <http://dx.doi.org/10.1177/104973239300300403>
- Green, A. (1998). *Verbal Protocol Analysis in Language Testing Research: A Handbook*. Cambridge University Press.
- Joseph, G., & Patel, V. (1990). Domain knowledge and hypothesis generation in diagnostic reasoning. *Medical Decision Making*, 10(1), 31-46. <https://doi.org/10.1177/0272989X9001000107>
- Keehner, M., Gorin, J. S., Feng, G., & Katz, I. R. (2017). Developing and validating cognitive models in assessment. En A. A. Rupp & J. P. Leighton (Eds.). *The Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications*. (pp. 75-101). Wiley Blackwell. <https://doi.org/10.1002/9781118956588.ch4>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. <https://doi.org/10.2307/2529310>
- Leighton, J. P. (2004). Avoiding misconception, misuse and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice*, 23(4), 6-15. <https://doi.org/10.1111/j.1745-3992.2004.tb00164.x>
- Leighton, J. P. (2013). Item difficulty and interviewer knowledge effects on the accuracy and consistency of examinee response processes in verbal reports. *Applied Measurement in Education*, 26(2), 136-157. <https://doi.org/10.1080/08957347.2013.765435>
- Leighton, J. P., Cui, Y., Ken-Cor, M. (2009). Testing expert-based and student-based cognitive models: An application of the attribute Hierarchy method and Hierarchy Consistency Index. *Applied Measurement in Education*, 22(3), 229-254. <https://psycnet.apa.org/doi/10.1080/08957340902984018>
- Martínez-Arias, M. R., Hernández-Lloreda, M. J., & Hernández-Lloreda, M. V. (2006). *Psicometría*. Alianza Editorial.
- Padilla, J. L., & Leighton, J. P. (2017). Cognitive interviewing and think aloud methods. En B. D. Zumbo, & A. M. Hubley (Eds.), *Understanding and investigating response processes in validation Research* (pp. 211-228). Springer. https://doi.org/10.1007/978-3-319-56129-5_12
- Padilla, J. L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136-144. <https://doi.org/10.7334/psicothema2013.259>
- Prieto-Adánez, G. (2011). Evaluación de la ejecución mediante el modelo Many-Facet Rasch Measurement. *Psicothema*, 23(2), 233-238. <http://www.psicothema.com/pdf/3876.pdf>
- Rojas-Torres, L., & Ordóñez-Gutiérrez, G. (2019). Proceso de construcción de pruebas educativas: El caso de la Prueba de Habilidades Cuantitativas. *Revista Evaluar*, 19(2), 15-29. <https://doi.org/10.35670/1667-4545.v19.n2>
- Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Tijdschrift voor Onderwijsresearch*, 2(2), 49-60. <https://files.eric.ed.gov/fulltext/ED121845.pdf>
- Ruiz, F.J. & Luciano, C. (2012). Relacionar relacio-

nes como modelo analítico-funcional de la analogía y la metáfora. *Revista Latina de Análisis de Comportamiento*, 20. <https://www.redalyc.org/articulo.oa?id=274525194014>

Russo, J. E., Johnson, E. J. and Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, 17, 759 – 769. <https://link.springer.com/article/10.3758/BF03202637#Bib1>

Sapsirin, S. (2016). The application of verbal protocol analysis in second/foreign language testing research. *Revista de revisión del idioma - Universidad de Chulalongkorn*, 31. <https://www.culi.chula.ac.th/en/pasaa-paritat/view/9>

Stelzer, F., Vernucci, S., Aydmune, Y. S., del Valle, M. V., & Andrés, M. L. (2020). Diseño y validación de una escala de actitudes hacia las matemáticas. *Revista Evaluar*, 20(2), 51-68. <https://doi.org/10.35670/1667-4545.v20.n2.30109>

Van Den-Haak, M., De Jong, M., & Jan-Schellens, P. (2003). Retrospective vs. Concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*, 22(5), 339-351. <http://dx.doi.org/10.1080/0044929031000>

Virzi, R. A. (1992). Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Human Factors*, 34(4), 457-468. <https://doi.org/10.1177/001872089203400407>

Validación española de la Escala de Agencia Personal y Empoderamiento (ESAGE-Sp-19)

Spanish Validation of the Personal Agency and Empowerment Scale (ESAGE-Sp-19)

José M. Rodríguez-Pellejero ¹, Juan L. Núñez ¹, Angeles Martel ²,
Desirée González ², Manuel Hernández-Rodríguez ², Debora Hernández ² *

1 - Department of Psychology, Sociology and Social Work, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain.

2 - Las Palmas de Gran Canaria Town Hall Social Services. Spain.

Introducción
Método
Resultados
Discusión
Referencias

Recibido: 04/07/2023 **Revisado:** 21/09/2023 **Aceptado:** 28/09/2023

Resumen

La agencia personal y el empoderamiento son claves para el cambio y la promoción social. El objetivo de este estudio es desarrollar una versión corta y adaptada al español de la Escala de Agencia Personal y Empoderamiento (ESAGE) y analizar sus propiedades psicométricas. Los participantes fueron 599 estudiantes universitarios (edad, $M = 24.89$, $DT = 6.69$; 63.43% mujeres). La estructura del modelo se probó a través de enfoques factoriales exploratorios y confirmatorios. El Análisis Factorial Confirmatorio (AFC) mostró un ajuste óptimo para el modelo con un Factor de segundo orden (Agencia Personal) y tres Factores de primer orden (Autoconfianza, Control Interno y Empoderamiento). Los valores alfa de Cronbach fueron aceptables. La regresión múltiple por pasos proporcionó evidencias de validez predictiva y el análisis factorial exploratorio utilizando las escalas de la ESAGE-Sp-19 y de ISP-20 reveló una solución unifactorial clara e interpretable. La ESAGE-Sp-19 se mostró fiable y válida en la población adulta española.

Palabras clave: *agencia personal, empoderamiento, disfunción ejecutiva, exclusión social, propiedades psicométricas*

Abstract

Personal agency and empowerment are keys to change and social promotion, measuring them is essential for intervening with people at risk of social exclusion. The aim of this study is to develop a Spanish-short and adapted version of the Personal Agency and Empowerment Questionnaire (ESAGE) and analyze its psychometric properties. Participants were 599 university students (age, $M = 24.89$, $SD = 6.69$; 63.43% women). The model's structure was tested through exploratory and confirmatory factorial approaches. The Confirmatory Factor Analysis showed an optimal fit for the model with a second-order factor (Personal Agency) and three first-order factors (Self-confidence, Internal Control, and Empowerment). Cronbach's alpha values were acceptable. Stepwise multiple regression provided evidence of predictive validity, and exploratory factor analysis using the ESAGE-Sp-19 and ISP-20 scales revealed a clear and interpretable unifactorial solution. The ESAGE-Sp-19 proved to be reliable and valid in the Spanish adult population.

Keywords: *personal agency, empowerment, executive dysfunction, social exclusion, psychometric properties*

* **Correspondencia a:** José M. Rodríguez-Pellejero. Facultad de Ciencias de la Educación c/ Santa Juana de Arco, 1 -Las Palmas de Gran Canaria 35004, España. Email: jose.pellejero@ulpgc.es. Teléfono: + 34 928 451 775.

Cómo citar este artículo: Rodríguez-Pellejero, M., J., Núñez L., J., Martel, A., González D., Hernández-Rodríguez, M., & Hernández, D. (2023). Validación española de la Escala de Agencia Personal y Empoderamiento (ESAGE-Sp-19). *Revista Evaluar*, 23(3) 61-76. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Participaron en la edición de este artículo: Débora Camponetti, Andrea Suárez, Eugenia Barrionuevo, Florencia Ruiz, Rodrigo Maderna, Jorge Bruera.

Introducción

La pobreza y la exclusión social siguen siendo grandes desafíos para la Unión Europea (UE), ya que los planes contra la pobreza y las políticas aplicadas no han logrado resolver el problema. En este contexto, la gestión que las personas en situación de riesgo y exclusión social hacen de las ayudas económicas y de las oportunidades de promoción social suelen ser puestas en cuestión.

Medir la competencia de los individuos para cambiar su propia situación personal y promocionar socialmente es una necesidad de los profesionales que intervienen con personas en situación de pobreza y exclusión social, esto permitiría diseñar itinerarios personalizados de inclusión social.

En la UE el riesgo de pobreza y de exclusión social se miden utilizando la tasa *At Risk Of Poverty or Exclusion* (AROEPE), un índice de medida multidimensional que considera no sólo indicadores monetarios, sino también otros indicadores sociales. Este índice ha sido elaborado por EUROSTAT. Según EUROSTAT, en el año 2020, el 21.9% de la población de la UE se encontraba en riesgo de pobreza o exclusión social, en términos absolutos unos 96.5 millones de personas. En ese mismo informe se advierte que el riesgo de pobreza o exclusión social fue mayor para las mujeres que para los hombres (22.9% frente a 20.9%). Existen, también, importantes diferencias territoriales en cuanto a pobreza y exclusión dentro de la propia UE. España, con un índice AROPE del 26.4% se sitúa por encima de la media europea, solo Rumanía, Bulgaria y Grecia presentan índices superiores a España en pobreza y exclusión (Eurostat, 2023). A pesar de la gravedad del problema, especialmente en el territorio español, no existen instrumentos de medida validados en población española que permitan apreciar las diferencias individuales en cuanto a capacidad de promoción social.

Existe un consenso generalizado sobre la multidimensionalidad de la exclusión social y sobre el origen social, y no psicológico, de la pobreza. En este mismo sentido, Daminger et al. (2015, p.14), concluyen que la cuestión central no es qué hacen mal las personas que están en situación de pobreza, sino qué le está haciendo la pobreza a las personas, o qué características contextuales están distorsionando sus elecciones y acciones. En esta línea, se ha encontrado una bibliografía extensa sobre el papel mediador que tiene la pobreza en el desarrollo cognitivo (Fracchia et al., 2020; Johnson et al., 2016; Pepper & Nettle, 2017; Segretin et al., 2016; Yoshikawa et al., 2012) y concretamente sobre la capacidad de control de los individuos. La pobreza y la exclusión social estarían provocando un cambio en la estrategia de control, debilitando el control proactivo en favor del control reactivo (Rodríguez-Pellejero & Nuñez, 2018; Xu et al., 2020). Autores como Sheehy-Skeffington y Haushofer (2014) sostienen que la pobreza afecta a la capacidad de control de los individuos sobre su entorno, produciendo a su vez: disminución de la auto-eficacia percibida, estrés crónico, agotamiento de recursos cognitivos, mayor probabilidad de ceder ante pensamientos sesgados y reducción de la capacidad cognitiva, esto lleva a los individuos a tomar decisiones que, a largo plazo, perpetúan la situación de pobreza de estas personas.

En los estudios sobre los efectos que tiene la pobreza en el control cognitivo del individuo es posible encontrar dos grupos de teorías. Unas centradas en el aprendizaje social, que suelen referirse a las creencias de las personas y a una cierta “cultura de la pobreza” y otras centradas en el desarrollo neurocognitivo. El primer grupo de teorías enfatiza el papel de: la autonomía (Kagitcibasi, 2005), el locus de control (Rotter, 1966), la autoeficacia (Bandura et al., 1999; Bandura, 2000), o la agencia personal y el empoderamiento (Alkire, 2002,

2008; Pick et al., 2007; Pick & Sirkin, 2010). El segundo grupo de teorías enfatiza el papel de la carga cognitiva (Schilbach et al., 2016), la memoria (Farah et al., 2006), la respuesta regulatoria del estrés (Christian, 2015; Doom & Gunnar, 2013) y las funciones ejecutivas (Dickerson & Popli, 2016; Mani et al., 2013; Rodríguez-Pellejero & Nuñez, 2018; Shah et al., 2012).

En cualquier caso, muchos investigadores han relacionado determinadas variables cognitivas con la pobreza y algunos de ellos afirman que cuando los pobres actúan como agentes, reducen su propia pobreza de forma eficaz (Alkire, 2008; Bandura, 2001; Pick et al., 2007; Sen, 1985). A pesar de estas evidencias, no se tiene la constancia de la existencia de escalas o instrumentos validados en población española para medir agencia personal, y las escalas que miden agencia personal son útiles para medir las capacidades personales que tienen los individuos para salir de la pobreza, para fomentar la propia inclusión social o conseguir la propia sustentabilidad. Una escala diseñada con este propósito es la ESAGE (Pick et al., 2007), cuya versión original se validó en estudiantes mexicanos.

Pick et al. (2007) llevaron a cabo una investigación para validar la ESAGE y reportaron la existencia de dos constructos bien diferenciados pero relacionados: La Agencia personal y el Empoderamiento. Estos autores proponen el concepto de Agencia Personal (AP) para describir el proceso que lleva a una persona, informada y de manera auto-controlada, a tomar decisiones motivadas intrínsecamente y dirigirse hacia las metas propias, previamente definidas de forma autónoma (Pick & Sirkin, 2010). La AP se refiere a la competencia individual para cambiar la propia situación personal en un determinado contexto social (Pick et al., 2007). Las intenciones producidas por el sujeto, los significados y las motivaciones son centrales en la formación y el estado

de la AP. Se trata de un concepto adaptado del economista Amartya Sen (1985) y basado en la Teoría de la Agencia Humana de Bandura (2006). Este último define la agencia como el deseo de influir en el funcionamiento y las circunstancias de la vida. Sen, por su parte, se refiere a la agencia personal como “lo que una persona es libre de hacer y conseguir en la búsqueda de cualquier meta o valores que él o ella tome como importantes” (1985, p. 203).

El término empoderamiento ya se usaba antes de los años 70, pero fue a partir de entonces que empezó a usarse de forma más extendida, especialmente en entornos del movimiento feminista, de la educación popular y en movimientos de minorías que demandaban representación política (Úcar-Martínez et al., 2017). A partir de ahí, el término se ha usado con mucha frecuencia, especialmente en el campo de las ciencias sociales, incluso por el Banco Mundial (Chambers, 2001). Desde una perspectiva psicológica el empoderamiento viene impulsado por los mismos factores internos que se agrupan bajo el concepto de AP. “La agencia es un proceso interno que se define como empoderamiento cuando empieza a impactar al contexto” (Pick et al., 2007, p.299). Alkire (2005) sostiene que el concepto de agencia se incluye en el concepto de empoderamiento, de tal forma que, si queremos incrementar el empoderamiento, tendremos que incrementar la agencia, aunque esta no siempre implique empoderamiento. En esta misma línea, Pick et al. (2007) hacen una diferencia entre dos formas de empoderamiento, el impulsado internamente y el provocado por factores externos.

Concretamente, la ESAGE mide el Empoderamiento Intrínseco (EI), un constructo que hace referencia a la capacidad interna del individuo para influir en su contexto de manera sostenible, el EI es la capacidad de un individuo para cambiar el contexto, para promover el desarrollo a

nivel contextual, más allá del cambio en la propia situación personal. El EI puede definirse también como el impacto que tiene el individuo en su entorno cercano, en su familia, amistades, en sus organizaciones y, por extensión, el poder que tiene para generar un cambio social (Alkire, 2005; Pick et al., 2007). De esta forma, Pick et al. (2007), concluyeron que la ESAGE unía dos disciplinas a distintos niveles: psicología y economía. Los conceptos EI y AP unen, de abajo hacia arriba, la psicología (nivel micro) con la economía (nivel macro), ambos constructos enfatizan la centralidad del individuo como factor determinante en el éxito de los programas para la inclusión social y el desarrollo social, ambos son ingredientes importantes para el cambio y la promoción social (Pick & Sirkin, 2010).

Para el desarrollo de la ESAGE los autores realizaron una revisión de la literatura sobre las definiciones de Agencia en psicología, constatando la dificultad para su medición debido fundamentalmente a que la AP puede tomar distintas formas a la hora de manifestarse y a que se nombra usando diferentes conceptos en la literatura específica. La agencia puede tomar la forma de: toma de decisiones, negociación, asertividad, reflexión, análisis, etc. (Kabeer, 1999; Pick et al., 2007).

Por otro lado, en la literatura psicológica, la AP se nombra usando distintos conceptos: autoeficacia, autonomía, control, autodeterminación, etc. Dadas las similitudes teóricas de estos conceptos, el grupo de investigación de Pick propone el término AP como término paraguas que aglutina diferentes constructos psicológicos, aquellos que definen un mismo funcionamiento competente e individual (Pick et al., 2007).

La ESAGE se construye a partir de un conjunto inicial de ítems utilizados para medir: autoeficacia (Bandura et al., 1999; Bandura, 1998, 2001; Sherer & Adams, 1983); autonomía (Kagitcibasi, 2005); autodeterminación (Deci &

Ryan, 1985, 2000); locus de control (Rotter, 1966) y autorregulación (Metcalf & Mischel, 1999). A partir de estos conceptos Pick y sus colaboradores extraen una primera versión de la escala formada por 42 reactivos agrupados en las siguientes variables: 1) Auto-eficacia, 2) Autodeterminación, 3) Control sobre mis conductas, 4) Pensamiento independiente, 5) Identificación de necesidad de cambio, 6) Miedo al éxito, 7) Reconocimiento de mi aprendizaje, 8) Percepción de mi contexto y 9) Control sobre mi entorno (Pick et al., 2007). Posteriormente sometieron estos ítems a un análisis factorial exploratorio, de ejes principales (*principal axis*), con rotación oblicua (*oblimin*), en donde se asume que las dimensiones teóricas estaban relacionadas entre sí. En este primer análisis exploratorio los ítems se agruparon en dos factores prácticamente independientes entre sí, un Factor de 35 ítems que medía las competencias de cambio en la persona, denominado Agencia Personal y un segundo Factor conformado por 7 reactivos que medían la percepción que tiene la persona sobre su incidencia en su contexto y que denominaron Empoderamiento. Posteriormente confirmaron que ambos factores se encontraban leve, pero significativamente, correlacionados ($r = -.131, p < .01$). Para analizar la confiabilidad de la escala, los autores originales utilizaron el alfa de Cronbach y encontraron coeficientes de confiabilidad aceptables para cada una de las subescalas (Agencia Personal, $\alpha = .718$; Empoderamiento, $\alpha = .749$). Las conclusiones de este primer estudio apuntaban a que la ESAGE permitía evaluar, con adecuada validez y confiabilidad, el impacto conjunto producido por la agencia personal y el empoderamiento de las personas.

Hasta donde sabemos, los profesionales que trabajan con población en exclusión social en España no cuentan con instrumentos fiables para medir la AP ni el EI. Sin esta herramienta, no es posible conocer en qué medida los destinatarios

de los programas podrán gestionar con éxito sus propios proyectos de vida. Atendiendo a esta necesidad de la práctica profesional, nuestro objetivo fue desarrollar una versión adaptada al español de la ESAGE (Pick et al., 2007) y estudiar sus propiedades psicométricas. Además, la validación original de la ESAGE propuso la utilización de la puntuación global de la escala a partir de dos únicos factores, sin embargo, aún no se ha probado la interpretabilidad de la puntuación total de la ESAGE, ni se ha realizado un análisis confirmatorio de su estructura. Por lo tanto, teniendo en cuenta el origen multidimensional del constructo AP, nuestro objetivo fue probar las estructuras del modelo con dos, tres y cuatro dimensiones.

Para lograr nuestros propósitos, realizamos un análisis de ítems y una estimación de confiabilidad de las puntuaciones de la escala en nuestra muestra. Como evidencia de validez y para probar diferentes modelos factoriales de la ESAGE, realizamos un análisis factorial confirmatorio. Usamos el análisis factorial exploratorio para proporcionar evidencia de validez convergente entre agencia personal y disfunción ejecutiva.

Método

Participantes

La muestra de estudio estuvo compuesta por 599 estudiantes de la Universidad de Las Palmas de Gran Canaria (España) seleccionados por un muestreo de conveniencia en tres titulaciones; Grado de Maestro en Educación Primaria (22.4%), Grado de Relaciones Laborales (37.7%) y Máster Universitario de Formación del Profesorado (49.8%). La edad media de los participantes fue de 24.89 años (DE = 6.69) y el 63.43% eran mujeres. La edad media de las mujeres fue de 24.22 años (DE = 6.19) y la de los hombres de 26.06 (DE = 7.34).

Instrumentos

Una versión reducida y adaptada al castellano de la Escala de Agencia Personal y Empoderamiento (ESAGE; Pick et al., 2007) fue el cuestionario objeto de estudio. La versión de 19 ítems se obtuvo seleccionando los ítems más discriminativos y con mayores cargas factoriales de la versión completa adaptada. Se utilizó un formato de respuesta del tipo Likert de cinco opciones, codificadas entre 0 y 4 (*nunca* y *siempre*).

La versión española y abreviada del Inventario de Síntomas Prefrontales (ISP-20; Pedrero-Pérez & Ruiz-Sánchez de León, 2019) se incluyó en el estudio para el análisis de la validez convergente. El ISP-20 es un inventario de cribado que consta de 20 ítems que se responden en una escala del tipo Likert de cinco opciones (*nunca* y *siempre*) y miden tres factores: el primero (cuatro ítems) mide problemas para el control de la conducta social; el segundo (cuatro ítems) mide problemas para el control emocional y el tercero (12 ítems) mide problemas para el control ejecutivo. Según sus autores, los tres factores encontrados formaban parte de un suprafactor al que denominaron: sintomatología prefrontal. La consistencia interna de esta prueba en su conjunto ofrece, según estos mismos autores, valores adecuados (Ω de McDonald = .86; α de Cronbach = .86).

Procedimiento

Se adaptó la ESAGE (Pick et al., 2007) de acuerdo con la *International Test Commission* (2017). Inicialmente, exploramos la equivalencia entre las culturas mexicana y española con respecto a los constructos subyacentes a la ESAGE y consideramos las diferencias lingüísticas, psicológicas y culturales. Como resultado, concluimos que, tanto la ESAGE como sus dimensiones po-

drían extrapolarse a la cultura española.

Posteriormente, el proceso de adaptación cultural comenzó con dos miembros del equipo de investigación y cuatro profesionales de los Servicios Sociales Municipales con experiencia directa en el trabajo con personas excluidas socialmente. Investigadores y profesionales de atención directa adaptaron de forma independiente los elementos de la ESAGE. En esta versión preliminar, los seis miembros estuvieron de acuerdo en mantener la redacción de 23 ítems de los 42 ítems originales. A su vez, todos estuvieron de acuerdo en cambiar los 4 ítems que hacían referencia a “pena” (expresión mexicana que en España se identifica con vergüenza) y también en cambiar el término original “colonia”, que aparecía en los 7 ítem del factor empoderamiento, por el término “barrio”, de mayor uso en español. Hubo una discrepancia entre los miembros en la interpretación de otros 6 ítems, que fueron cambiados aceptando las propuestas alternativas que alcanzaron mayor consenso. La propuesta final contó con la aprobación del total de los seis miembros. Debido a diferencias culturales, en la versión final de la ESAGE-Sp-19, la redacción de los ítems 2, 5, 9 y 11 se modificó ligeramente (ver Tabla 1). La versión final se sometió a estudio empírico de carácter psicométrico.

Las personas que aceptaron participar proporcionaron su consentimiento informado por escrito y completaron el cuestionario autoadministrado en grupos de 20 a 60 personas. Se garantizó la confidencialidad y el anonimato, los estudiantes completaron el cuestionario en sus aulas, en presencia de un entrevistador.

Análisis de datos

Se realizaron análisis descriptivos (medias, desviaciones típicas, asimetría y curtosis) para

los 42 ítems de la ESAGE original adaptado. Los coeficientes de correlación ítem-subescala de los ítems seleccionados para la versión reducida se proporcionan como evidencia de la adecuación del ítem. Para probar la distribución normal de los datos, se utilizó la prueba de Kolmogorov-Smirnov. Además, llevamos a cabo un análisis paralelo optimizado, basado en un análisis de rangos mínimos para determinar el número de factores que se debía retener. Sobre la mejor solución obtenida se efectuó una rotación *Oblimin* y se estimaron criterios de simplicidad y estimación de residuos. La consistencia interna de la escala se estimó utilizando el alpha de Cronbach, puesto que son varios los autores que argumentan que alfa es el indicador menos sesgado de la consistencia de los ítems del tipo Likert (p. ej., [Zumbo et al., 2007](#)).

Para estudiar la estructura interna de la versión reducida y adaptada de la ESAGE se realizó un análisis factorial confirmatorio (AFC). Como método de estimación de parámetros se eligió, por ser un método robusto a la falta de normalidad, el DWLS (*Diagonally weighted least squares*). Se trata de uno de los algoritmos más recomendados cuando se usan escalas del tipo likert y cuando no se cumple la normalidad multivariada ([Jöreskog et al., 2001](#); [Morata-Ramírez et al., 2015](#)). Se probaron cuatro estructuras factoriales y para determinar el ajuste del modelo se ofrecen, además del Chi cuadrado, los siguientes índices de ajuste, escogidos por ser los que se ven menos afectados por el tamaño de la muestra: índice de bondad de ajuste (índice de ajuste GFI), comparativo (índice de ajuste CFI), no normalizado (NNFI), error cuadrático medio de aproximación (RMSEA), el índice de la raíz del cuadrado medio de los residuos (SRMR) y el índice de validación cruzada esperada (ECVI). Cuando GFI, CFI y NNFI toman valores cercanos a .90 o .95 indican un buen ajuste; RMSEA y SRMR < .05 indica un buen

ajuste y los valores entre .05 y .08 indican un ajuste aceptable; el índice de validación cruzada esperada (ECVI) es una aproximación a la bondad del ajuste que conseguiría el modelo estimado en otra muestra del mismo tamaño. Cuando se están comparando modelos, el menor valor de ECVI indica el modelo con mejor ajuste (Schermelleh-Engel et al., 2003).

Con el fin de proporcionar evidencia de validez basada en la relación con otras variables se realizó una regresión lineal múltiple. Mediante el procedimiento paso a paso de selección de variables se probó el poder explicativo de la nueva versión ESAGE-Sp-19 con respecto a tres resultados relacionados con el funcionamiento ejecutivo: problemas de control de la conducta social, problemas de control emocional y problemas con el control ejecutivo. El supuesto de linealidad se evaluó mediante los factores de inflación de la varianza (VIF) y la inspección visual de los residuos frente a las parcelas ajustadas resultantes de los modelos de regresión probados. Se realizó un análisis factorial exploratorio (utilizando factorización del eje principal con rotación oblicua) para proporcionar evidencia de la validez convergente, incluyendo la versión adaptada y reducida de la ESAGE y el ISP-20.

Resultados

Análisis de los ítems

Los resultados de las estadísticas descriptivas se pueden ver en la Tabla 1. No se observó estructura bimodal en los ítems. En general, los valores de asimetría y curtosis estuvieron por debajo del valor absoluto de 1 para todos los ítems, excepto en el ítem 37 del factor Empoderamiento, que fue eliminado de la versión corta ESAGE-Sp-19 por presentar los valores más altos de asimetría y curtosis. La prueba de Kolmogorov-

Smirnov indicó que los ítems tenían una distribución no normal. El valor más alto de la correlación ítem-test fue .51.

Dimensionalidad y consistencia interna

Los índices de ajuste obtenidos en el AFC de la escala original de dos factores y 42 ítems fueron los siguientes: $\chi^2 = 3083,42$ ($gl = 818$, $N = 599$, $p < 001$), $RMSEA = 0,068$, $NNFI = 0,82$, $CFI = .83$, $GFI = .90$, $SRMR = .079$, $ECVI = 5.440$. Los valores de las cargas factoriales oscilaron entre .10 y .66. Siguiendo a Marsh et al. (2010), se seleccionaron los siguientes ítems de la escala original: 2, 6, 9, 12, 13, 15, 17, 18, 19, 22, 23, 32 y 33 (factor AP) y 36, 38, 39, 40, 41 y 42 (factor E). Como se puede observar en la Tabla 1, los ítems seleccionados mostraron valores elevados de pesos factoriales ($\geq .41$) y correlación ítem-test ($\geq .25$). A pesar de tener altos pesos factoriales y correlaciones ítem-test, los ítems 4, 20 y 24 no fueron seleccionados por motivos teóricos. Una vez obtenidos los ítems de la versión corta de la escala (ESAGE-Sp-19), se probaron diferentes estructuras factoriales. La Tabla 2 muestra los resultados de las pruebas de estructuras de cuatro factores. El AFC no apoyó estructuras de un solo factor.

El modelo original de dos factores y 42 elementos tenía una consistencia interna aceptable, pero las cargas de los elementos del factor AP oscilaban entre .10 y .61, de estos elementos, 11 tenían cargas factoriales $< .30$. Además, los índices de ajuste cuestionaron el ajuste del modelo. Se probó un segundo modelo factorial con una solución de dos factores. El AFC mostró un buen ajuste, la correlación de Pearson entre los dos factores fue significativa ($r = .11$). Todas las cargas factoriales estuvieron por encima de .49 (AP) y .71 (E). El análisis de consistencia interna reveló valores óptimos para cada una de las subescalas

Tabla 1

Estadísticas descriptivas, cargas factoriales y correlación ítem-test corregida de la ESAGE.

Escala AP	M	SD	Skewness	Kurtosis	Factor loading	Ítem-test <i>r</i>
Me es difícil expresar mi opinión públicamente.	3.4	1.14	-.29	-.67	.51	.47
Me siento inseguro cuando tomo decisiones*.	3.27	1.02	-.16	-.52	.59	.48
Dejo las cosas a medias.	3.91	.98	-.66	-.21	.36	.24
Tengo iniciativa para hacer las cosas.	3.84	.87	-.41	-.12	.43	.47
Me cuesta trabajo terminar lo que estoy haciendo.	3.61	.93	-.23	-.17	.37	.24
Me es difícil saber qué esperar de la vida*.	3.28	1.09	-.22	-.62	.60	.49
Exijo mis derechos, aunque otros no estén de acuerdo.	3.58	.98	-.36	-.31	.19	.01
Busco la solución a un problema, aunque otros me digan que no hay.	4.04	.88	-.61	-.26	.25	.30
Me da vergüenza equivocarme*.	3.16	1.21	-.19	-.88	.54	.44
Cumplir con mis planes está fuera de mi control.	4.05	.88	-.71	.15	.38	.27
Cuando tengo un problema, sé lo que necesito para solucionarlo.	3.41	.80	-.07	.15	.36	.39
Me da vergüenza hablar en público*.	3.34	1.23	-.30	-.80	.49	.43
Tapo mis errores para que nadie se dé cuenta*.	3.56	1.05	-.44	-.36	.42	.31
Encuentro soluciones novedosas a problemas difíciles.	3.17	.88	-.03	-.23	.30	.38
Me desespero ante situaciones difíciles*.	3.26	1.02	-.24	-.34	.61	.51
Dejo que las cosas sucedan en vez de planearlas.	3.36	.92	-.05	-.11	.39	.28
Siento que tengo poco control sobre lo que me pasa*.	3.68	.94	-.55	.07	.53	.43
Hago menos cosas de las que soy capaz de hacer*.	3.06	1.17	0	-.84	.47	.36
Me siento incapaz de cumplir lo que me propongo*.	4.05	.87	-.66	-.01	.54	.43
Me es difícil saber con quién puedo contar cuando me ocurre algo negativo.	3.85	1.2	-.88	-.18	.43	.33
Lo que ocurre en mi vida está bajo mi control.	3.46	.88	-.08	-.07	.25	.24
Sólo empiezo con ganas algo cuando me parece fácil*.	3.62	1.01	-.49	-.25	.47	.40
Le doy demasiada importancia a las opiniones de los demás*.	3.38	1.09	-.27	-.55	.55	.46
Me da vergüenza que me elogien.	3.09	1.18	-.07	-.81	.42	.32
Me gusta tener responsabilidades.	3.72	.95	-.49	.1	.27	.34
Me quejo a las autoridades cuando hay un abuso.	3.15	1.2	-.12	-.9	.10	.25
Es mejor tomar decisiones que esperar a ver lo que pasa.	3.92	.88	-.34	-.54	.29	.28
Sé por qué me pasan las cosas.	3.58	.93	-.28	-.12	.21	.25
Me gusta ser el primero en hacer cosas nuevas.	3.26	1.02	-.22	-.26	.19	.25
Me es fácil tomar decisiones.	3.19	1.02	-.11	-.35	.31	.29
Hago lo que creo que es mejor para mí sin importar lo que otros crean.	3.26	1.07	-.15	-.59	.25	.22
Me da vergüenza cobrar lo que me deben*.	3.2	1.33	-.14	-1.13	.41	.29
Tengo que aguantarme con la vida que me tocó*.	3.82	1.14	-.76	-.13	.45	.36
Conozco las leyes de mi país.	3.27	1.01	-.18	-.54	.39	.27
Pienso que este mundo lo dirigen aquellos que tienen poder.	2.31	1.31	.78	-.54	.29	.16

Escala E	M	SD	Skewness	Kurtosis	Factor loading	Ítem-test <i>r</i>
En mi barrio/comunidad ayudo a resolver los conflictos que se presentan*.	2.02	1.08	.81	-.24	.66	.34
En mi barrio/comunidad participo en las asambleas o juntas vecinales.	1.50	.92	1.86	2.75	.57	.32
Opino sobre lo que debe hacerse para mejorar mi barrio/comunidad*.	2.88	1.28	.03	-1.11	.65	.34
En mi barrio/comunidad conozco a las autoridades que me representan*.	2.95	1.41	.05	-1.28	.50	.31
Se cuáles son los problemas de mi barrio/comunidad*.	2.48	1.18	.31	-.88	.62	.34
Estoy enterado de los planes que el gobierno tiene para mi barrio /comunidad*.	2.32	1.13	.46	-.71	.65	.37
Quiero lograr cambios en mi barrio/comunidad*.	3.00	1.22	-.05	-.98	.58	.25

Nota. Ítems seleccionados para la versión reducida ESAGE-Sp-19.

por separado, AP ($\alpha = .82$; $\Omega = .82$), E ($\alpha = .80$; $\Omega = .80$) y para el total de la ESAGE-Sp-19 ($\alpha = .80$; $\Omega = .77$). El modelo 3 probó una solución de un factor de segundo orden (AP) y tres factores de primer orden. Todas las cargas factoriales estuvieron por encima de .62 en el factor denominado Autoconfianza (AC), por encima de .55 en el factor denominado Control Interno (CI) y por encima de .66 en el factor E. Las correlaciones de Pearson de la variable latente (AP) con las variables de primer orden fueron: AP-AC = .89; AP-CI = .88, AP-E = .11. Las correlaciones entre las variables de primer orden fueron: AC-E = .10; CI-E = .11; AC-CI = .57. Este modelo presenta un buen

ajuste según los índices de ajuste evaluados, con el valor más bajo en el índice ECVI. El análisis de consistencia interna reveló valores aceptables para los factores de primer orden; E ($\alpha = .82$; $\Omega = .82$), AC ($\alpha = .74$; $\Omega = .75$) y CI ($\alpha = .71$; $\Omega = .71$). Finalmente, se probó un cuarto modelo factorial. Estos 4 factores de primer orden mostraron bajos coeficientes de confiabilidad en las puntuaciones de los factores de primer orden: AC ($\alpha = .74$; $\Omega = .75$), CI ($\alpha = .62$; $\Omega = .62$), Autodeterminación ($\alpha = .62$; $\Omega = .62$) y E ($\alpha = .82$; $\Omega = .82$). Además, mostró índices de ajuste adecuados pero un índice ECVI ligeramente superior al del modelo 3.

Tabla 2

Índices de bondad de ajuste para los modelos probados.

Modelo	NNFI	CFI	GFI	RMSEA	SRMR	ECVI	$\chi^2(df)$
Modelo 1	.82	.83	.90	.068	.079	5.440	3083.42 (818) ***
Modelo 2	.98	.98	.99	.030	.045	.584	233.19 (151) ***
Modelo 3	.99	.99	.99	.022	.042	.457	191.19 (149) **
Modelo 4	.99	.99	.99	.022	.040	.520	185.13 (146) **

Nota.

Modelo 1 = Modelo original de dos factores y 42 ítems.

Modelo 2 = Modelo de dos factores y 19 ítems.

Modelo 3 = Tres factores de primer orden, un factor de segundo orden y 19 ítems.

Modelo 4 = Cuatro factores de primer orden y 19 ítems.

*** $p < .001$.

** $p < .05$.

Validez predictiva y convergente

Para evaluar la validez predictiva se realizó un modelo de regresión múltiple por pasos en donde se controló la variable edad. La variable dependiente fue la disfunción ejecutiva, evaluada por los tres factores de la escala ISEP-20. Se introdujeron como variables independientes los tres factores de primer orden de la ESAGE-Sp-19 obtenidos en el modelo 3: AC, CI y E. Los resultados mostraron que un modelo formado por AC, CI y E explicaba el 30% de las variaciones observadas en el puntaje total del ISP-20 (ver Tabla 3); el CI emerge como la variable que tiene mayor capacidad predictiva ($\beta = -.47$). El CI mostró poder predictivo sobre la disfunción ejecutiva y todas sus dimensiones. La AC presentó poder predictivo sobre problemas de control emocional y problemas de control conductual, pero no sobre problemas de comportamiento social. La E mostró poder predictivo solo sobre los problemas de control conductual y sobre la puntuación total de disfunción ejecutiva. No se observaron patrones de multicolinealidad en ninguno de estos modelos de regresión. Los resultados mostraron valores

bajos del factor de inflación de varianza (VIF).

En cuanto a las evidencias de validez convergente, el análisis factorial exploratorio con los factores de primer orden de ambas escalas reveló una solución de un único factor claro e interpretable (Tabla 4), con un polo positivo formado por los factores AC y CI de la ESAGE-Sp-19 y un polo negativo formado por los tres factores de la ISP-20, con cargas factoriales superiores a .30. La escala E no formó parte de la solución factorial.

Discusión

El presente estudio exploró las propiedades psicométricas de la versión española y reducida de la ESGAE (Pick et al., 2007). Hasta donde sabemos, este es el primer estudio que presenta una versión adaptada a la población española de un instrumento para medir agencia personal y empoderamiento; también, es el primer estudio que realiza un análisis psicométrico de las propiedades de la ESAGE utilizando el análisis factorial confirmatorio.

Nuestros resultados muestran el adecuado funcionamiento psicométrico de la escala con una

Tabla 3

Modelo de regresión múltiple por pasos de la variable dependiente disfunción ejecutiva y sus dimensiones.

DV	Model	R ²	β	T	F	P
Disfunción Ejecutiva (Total ISP-20)	(Intercept)			28.741		< .001
	Autoconfianza		-.10	-2.499		.013
	Autocontrol	.30	-.47	-11.139	84.464	< .001
	Empoderamiento		-.09	-2.561		.011
Problemas de control Emocional	(Intercept)			16.995		< .001
	Autoconfianza	.09	-.19	-4.143	31.744	< .001
	Control Interno		-.15	-3.317		< .001
Problema de control ejecutivo	(Intercept)			27.303		< .001
	Autoconfianza		-.09	-2.211		.027
	Control Interno	.30	-.48	-11.459	87.912	< .001
	Empoderamiento		-.11	-3.157		.002
Problema de control conducta social	(Intercept)			12.436		< .001
	Control Interno	.07	-.26	-6.713	45.070	< .001

Tabla 4
Solución Factorial para subescalas ISP-20 y subescalas ESAGE-Sp-19.

	Factor 1
ISP-20 Problemas de Conducta Social	-.329
ISP-20 Problemas de Control Emocional	-.375
ISP-20 Problemas de control Comportamental	-.643
ESAGE-Sp-19 Autoconfianza (AC)	.640
ESAGE-Sp-19 Control Interno (CI)	.840

adecuada consistencia interna. Aportamos evidencias de la validez basada en la estructura interna de la escala y su relación con otras variables de su red nomotética. Los valores estimados de consistencia interna de las escalas de Agencia personal y Empoderamiento son adecuados y ligeramente superiores a los reportados anteriormente (Pick et al., 2007). Para las dimensiones de Agencia Personal y Empoderamiento, los valores son adecuados y superiores a los de la versión de 42 ítems (Pick et al., 2007).

El análisis factorial exploratorio realizado para la versión de 42 ítems no respaldó la estructura factorial de esta escala y 23 de los 42 ítems mostraron cargas factoriales muy bajas. En su versión original, la ESAGE (Pick et al., 2007) se construyó a partir de las medidas de autoeficacia, autonomía, control, autorregulación y autodeterminación, por lo que algunos teóricos posteriores señalan que esta escala brinda una estructura poco explicativa de la agencia personal (Maytorena-Noriega, 2020).

Nuestro estudio probó varios modelos a partir de los 19 ítems seleccionados por sus cargas factoriales, en donde se encontró que la estructura original de dos factores, agencia personal y empoderamiento, tenía valores de ajuste aceptables y explicaban el 32% de la varianza. Pero fue un modelo con una estructura factorial de un factor de segundo orden (Agencia Personal), formado por dos factores de primer orden, y un tercer factor independiente de primer orden (Empoderamiento), el cual mostró los mejores valores de ajuste y una

varianza explicada del 35%. Dado que el modelo unidimensional no encajó, recomendamos que la puntuación total de la ESAGE no se calcule a partir de los ítems originales y proponemos la utilización de la versión reducida y adaptada de la ESAGE-Sp-19 para la población española. Se recomienda calcular la Agencia Personal a partir de los 13 ítems de la ESAGE-Sp19 que componen las dimensiones AC-CI y, por otro lado, calcular el Empoderamiento a partir de los 6 ítems seleccionados.

Aunque la suma de la puntuación total de la ESAGE-Sp-19 puede ser de utilidad en la práctica y para determinar los apoyos que las personas pueden necesitar en la autogestión de proyectos de vida, debemos señalar que se trata de dimensiones distintas, aunque relacionadas.

El análisis convergente reveló que las dimensiones de la ESAGE-Sp-19 no difieren de las dimensiones de disfunción ejecutiva medidas con el ISP-20. En sentido contrario, la agencia personal converge con la disfunción ejecutiva y emerge como el polo positivo de un continuo entre agencia personal y disfunción ejecutiva (en el polo negativo). Los dos constructos están relacionados y muestran coherencia teórica, diferenciándose ambos del Empoderamiento, un constructo relacionado pero independiente.

Está bien documentado que, tanto la agencia personal como el funcionamiento ejecutivo se asocian con la toma de decisiones, también con el control de la conducta, la planificación y la consecución de objetivos (Bandura et al., 1999;

Bandura, 2001; Pick & Sirkin, 2010; Wolf & Jonker, 2020). De los análisis multivariados de nuestro estudio emerge un modelo predictivo de la disfunción ejecutiva formado por los tres factores de primer orden de la ESAGE-Ssp-19. Algunos autores han señalado la relación existente entre las funciones ejecutivas, la autorregulación y la agencia personal (Sanders et al., 2019) y otros han relacionado la agencia personal y las funciones ejecutivas con el empoderamiento (Wolf & Jonker, 2020) Nuestros hallazgos apoyan dichos estudios. La agencia personal correlacionó de forma significativa con las llamadas funciones ejecutivas intelectuales y, también, con las llamadas funciones ejecutivas emocionales, sugiriendo que la Agencia Personal es un constructo que incluye habilidades cognitivas y emocionales, relacionado con la inteligencia, pero no equiparable a esta (Ardila, 2018).

Existe un debate teórico en la literatura científica sobre los factores que están detrás del deterioro que produce la exclusión social sobre el control cognitivo. Algunos investigadores sostienen la hipótesis del aprendizaje social, situando la auto-eficacia en la base del proceso de control o de la agencia personal (Alkire, 2002, 2008; Bandura et al., 1999; Pick et al., 2007; Pick & Sirkin, 2010). Otros investigadores sostienen la hipótesis de la incapacidad, enfatizando el déficit en el funcionamiento ejecutivo (Lurquin et al., 2014; Shilling & Brown, 2016).

En nuestra investigación, la regresión múltiple reveló que el CI tenía los coeficientes de regresión más altos en los modelos que tomaron como variable explicada los problemas en el control ejecutivo y de conducta social, también fue el predictor con mayor peso para la puntuación total en Disfunción Ejecutiva y el único estadísticamente significativo para explicar todas y cada una de las dimensiones del ISP-20. Dado que nuestros resultados evidencian un mayor peso del Control

Interno, una variable relacionada con el funcionamiento ejecutivo sobre el resto de las variables de carácter más social, nuestros resultados suponen una aportación en favor de la hipótesis de la capacidad cognitiva deteriorada, en concordancia con las investigaciones que sostienen que la exclusión social provoca un cambio en la estrategia de control, debilitando el control proactivo en favor del control reactivo (Xu et al., 2020; Rodríguez-Pellejero & Nuñez, 2018).

Si bien este estudio proporciona apoyo empírico para el uso de la ESAGE-Sp-19, deben tenerse en cuenta ciertas limitaciones. En comparación con los estudiantes universitarios, las personas en situación de exclusión social cronificada pueden tener más problemas de agencia personal (Njozela et al., 2018), de disfunción ejecutiva (Rodríguez-Pellejero & Nuñez, 2018) y de empoderamiento (Wolf & Jonker, 2020). Nuestro trabajo se realizó con una muestra de estudiantes universitarios, lo que limita el uso de la escala en otras subpoblaciones. Por lo tanto, la investigación futura debería probar las propiedades psicométricas de la ESAGE-Sp-19 en otras subpoblaciones, particularmente en muestras no universitarias de adultos jóvenes y en personas en situación de exclusión social leve, moderada y grave. Además, la mayoría de los participantes en este estudio eran mujeres. Los autores del ISP-20 reportaron diferencias de género en las puntuaciones obtenidas en esa escala; las mujeres puntuaron significativamente más alto en problemas de control emocional y los hombres lo hicieron en problemas de control de la conducta social (Pedrero-Pérez & Ruiz-Sánchez de León, 2019). No está claro si el género puede afectar la invarianza de la estructura de la ESAGE-Sp-19. Los estudios futuros (con muestras grandes) deberían explorar las posibles diferencias de género en agencia personal y empoderamiento y realizar análisis de equivalencia psicométrica, ya que algunos autores los consideran un requisito

fundamental previo a la comparación de resultados entre grupos (Peña-Contreras et al., 2020).

Finalmente, obtuvimos evidencias de validez de forma transversal a partir de la relación con otras variables. Dado que la utilidad de la ESAGE-Sp-19 radica en su capacidad para predecir la capacidad del sujeto de gestionar su propio proyecto de vida y promocionar socialmente, los estudios futuros deberían probar la validez predictiva de la ESAGE-Sp-19 en relación con la capacidad de las personas de cambiar su circunstancias personales y sociales. Explorar las posibles correlaciones entre ESAGE-Sp-19 y otros cuestionarios como el Inventario de los Problemas Interpersonales (IIP-32) desarrollado por Gómez-Penedo et al. (2022) o la Escala de Solución de Problemas en Directivos (ESOP-D) desarrollada por Vega-Valero et al. (2019), resultarían de especial interés para comprobar la validez convergente de nuestra Escala.

Teniendo en cuenta la evidencia de validez y confiabilidad proporcionada en este estudio, nuestros hallazgos apoyan el uso de la estructura original de dos factores de la ESAGE propuesta por Pick et al. (2007), considerando que la Agencia Personal no se puede considerar como factor de primer orden, sino el resultado de otros factores de primer orden como son la AC y el CI. Además, a la luz de nuestros resultados, el uso de la puntuación total de la escala podría usarse en la práctica para determinar el nivel de apoyo que necesita una persona en proceso de inclusión social. A mayor puntuación total en la ESAGE-Sp-19, menor será la necesidad de acompañamiento profesional y mayor la probabilidad de éxito en condiciones de igualdad de oportunidades. Del mismo modo, el ISP-20 se presenta como un cuestionario alternativo y de aplicación para medir dificultades con la agencia personal en personas en exclusión social.

Referencias

- Alkire, S. (2002). *Valuing Freedoms: Sen's capability approach and poverty reduction*. Oxford University Press. <https://doi.org/10.1093/0199245797.001.0001>
- Alkire, S. (2005). Subjective quantitative studies of human agency. *Social Indicators Research*, 74, 217-260. <https://doi.org/10.1007/s11205-005-6525-0>
- Alkire, S. (2008). Concepts and measures of agency. In K. Basu & R. Kanbur (Eds.), *Arguments for a better world: Essays in honor of Amartya Sen. Vol. 1, ethics, welfare and measurement* (pp. 455-474). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199239115.003.0025>
- Ardila, A. (2018). Is intelligence equivalent to executive functions? *Psicothema*, 30(2), 159-164. <https://www.psicothema.com>
- Bandura, A. (1998). Health promotion from the perspective of social cognitive theory. *Psychology & Health*, 13(4), 623-649. <https://doi.org/10.1080/08870449808407422>
- Bandura, A. (2000). Self-efficacy: The foundation of agency. In W. J. Perrig & A. Grob (Eds.), *Control of human behavior, mental processes, and consciousness: Essays in honor of the 60th birthday of August Flammer* (pp. 17-33). Lawrence Erlbaum.
- Bandura, A. (2001). Social Cognitive Theory: An agentic perspective. *Annual Review of Psychology*, 52, 1-26. <https://doi.org/10.1146/annurev.psych.52.1.1>
- Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science*, 1(2), 164-180. <https://doi.org/10.1111/j.1745-6916.2006.00011.x>
- Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-efficacy: The exercise of control. *Journal of Cognitive Psychotherapy*, 13(2), 158-166. <https://doi.org/10.1891/0889-8391.13.2.158>
- Chambers, R. (2001). The world development report: Concepts, content and a chapter 12. *Journal of International Development*, 13(3), 299-306. <https://doi.org/10.1002/jid.784>
- Christian, L. M. (2015). Stress and immune func-

- tion during pregnancy: An emerging focus in mind-body medicine. *Current Directions in Psychological Science*, 24(1), 3-9. <https://doi.org/10.1177/0963721414550704>
- Daminger, A., Hayes, J., Barrows, A., & Wright, J. (2015). *Poverty interrupted: Applying behavioral science to the context of chronic scarcity* (Informe de investigación). Ideas42. <https://www.ideas42.org/project/poverty-interrupted>
- Deci, E. L., & Ryan, R. M. (1985). The General Causality Orientations Scale: Self-determination in personality. *Journal of Research in Personality*, 19(2), 109-134. [https://doi.org/10.1016/0092-6566\(85\)90023-6](https://doi.org/10.1016/0092-6566(85)90023-6)
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268. http://doi.org/10.1207/S15327965PLI1104_01
- Dickerson, A., & Popli, G. K. (2016). Persistent poverty and children’s cognitive development: Evidence from the UK Millennium Cohort Study. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 179(2), 535-558. <http://doi.org/10.1111/rssa.12128>
- Doom, J. R., & Gunnar, M. R. (2013). Stress physiology and developmental psychopathology: Past, present, and future. *Development and Psychopathology*, 25(4pt2), 1359-1373. <https://doi.org/10.1017/S0954579413000667>
- Eurostat. (Junio 2023). *Living conditions in Europe - Poverty and social exclusion*. [Artículo estadístico en línea] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Living_conditions_in_Europe_poverty_and_social_exclusion&oldid=549030#Key_findings
- Farah, M. J., Shera, D. M., Savage, J. H., Betancourt, L., Giannetta, J. M., Brodsky, N. L., Malmud, E. K., & Hurt, H. (2006). Childhood poverty: Specific associations with neurocognitive development. *Brain Research*, 1110(1), 166-174. <https://doi.org/10.1016/j.brainres.2006.06.072>
- Fracchia, C. S., Segretin, M. S., Hermida, M. J., Prats, L., & Lipina, S. J. (2020). Mediating role of poverty in the association between environmental factors and cognitive performance in preschoolers. *Argentinean Journal of Behavioral Sciences*, 12(2), 24-38. <https://doi.org/10.32348/1852.4206.v12.n2.26482>
- Gómez-Penedo, J. M., Fernández-Álvarez, J., Maristany, M., & Freiberg-Hoffmann, A. (2022). Inventario de Problemas Interpersonales-32 (IIP-32): Propiedades psicométricas y datos normativos en una muestra clínica de Argentina. *Revista Evaluar*, 22(2), 64-75. <https://doi.org/10.35670/1667-4545.v22.n2.38688>
- International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests* (2^{da} ed.). <https://www.InTestCom.org>
- Johnson, S. B., Riis, J. L., & Noble, K. G. (2016). State of the art review: Poverty and the developing brain. *Pediatrics*, 137(4), e20153075. <https://doi.org/10.1542/peds.2015-3075>
- Jöreskog, K. G., Sörbom, D., du Toit, S., & du Toit, M. (2001). *LISREL 8: New statistical features*. Scientific Software International.
- Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women’s empowerment. *Development and Change*, 30(3), 435-464. <https://doi.org/10.1111/1467-7660.00125>
- Kagitcibasi, C. (2005). Autonomy and relatedness in cultural context: Implications for self and family. *Journal of Cross-Cultural Psychology*, 36(4), 403-422. <https://doi.org/10.1177/0022022105275959>
- Lurquin, J. H., McFadden, S. L., & Harbke, C. R. (2014). An electrophysiological investigation of the effects of social rejection on self-control. *The Journal of Social Psychology*, 154(3), 186-197. <https://doi.org/10.1080/00224545.2014.881768>
- Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, 341(6149), 976-980. <https://doi.org/10.1126/science.1238041>
- Marsh, H. W., Martin, A. J., & Jackson, S. (2010). Introducing a short version of the physical self description questionnaire: new strategies, short-form evaluative criteria, and applications of factor analy-

- ses. *Journal of Sport and Exercise Psychology*, 32(4), 438-482.
- Maytorena-Noriega, M. de los A. (2020). Escala de Agencia Personal en Educación Superior: Diseño y valoración: Escala de Agencia Personal. *Psicumex*, 10(1), 39-60. <https://doi.org/10.36793/psicumex.v10i1.338>
- Metcalf, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review*, 106(1), 3-19. <https://doi.org/10.1037/0033-295X.106.1.3>
- Morata-Ramírez, M. A., Holgado-Tello, F. P., Barbero-García, I., & Mendez, G. (2015). Análisis factorial confirmatorio: Recomendaciones sobre mínimos cuadrados no ponderados en función del error Tipo I de Ji-Cuadrado y RMSEA. *Acción Psicológica*, 12(1), 79-90. <https://doi.org/10.5944/ap.12.1.14362>
- Njozela, L., Burns, J., & Langer, A. (2018). The effects of social exclusion and group heterogeneity on the provision of public goods. *Games*, 9(3), 55. <https://doi.org/10.3390/g9030055>
- Pedrero-Pérez, E. J., & Ruiz-Sánchez de León, J. M. (2019). Síntomas prefrontales en la vida diaria: Normalización y estandarización del Inventario de Síntomas Prefrontales Abreviado en población española. *Revista de Neurología*, 68(11), 459-467. <https://doi.org/10.33588/rn.6811.2018394>
- Pedrero-Pérez, E. J., Ruiz-Sánchez de León, J. M., Morales-Alonso, S., Pedrero-Aguilar, J., & Fernández-Méndez, L. M. (2015). Sintomatología prefrontal en la vida diaria: Evaluación de cribado mediante el Inventario de Síntomas Prefrontales Abreviado (ISP-20). *Revista de Neurología*, 60(9), 385-393. <https://doi.org/10.33588/rn.6009.2014545>
- Peña-Contreras, E. K., Lima-Castro, S. E., Arias-Medina, W. P., Bueno-Pacheco, G. A., Aguilar-Sizer, M. E., & Cabrera-Vélez, M. M. (2020). Propiedades psicométricas de la Escala Breve de Resiliencia (BRS) en el contexto ecuatoriano. *Revista Evaluar*, 20(3), 83-98. <https://doi.org/10.35670/1667-4545.v20.n3.31715>
- Pepper, G. V., & Nettle, D. (2017). The behavioural constellation of deprivation: Causes and consequences. *Behavioral and Brain Sciences*, 40, e314. <https://doi.org/10.1017/S0140525X1600234X>
- Pick, S., & Sirkin, J. (2010). *Breaking the poverty cycle: The human basis for sustainable development*. Oxford University Press.
- Pick, S., Sirkin, J., Ortega, I., Osorio, P., Martínez, R., Xocolotzin, U., & Givaudan, M. (2007). Escala Para Medir Agencia Personal y Empoderamiento (ESAGE). *Revista Interamericana de Psicología/ Interamerican Journal of Psychology*, 41(3), 295-304.
- Rodríguez-Pellejero, J. M., & Núñez, J. L. (2018). Relationship between attachment and executive dysfunction in the homeless. *Social Work in Health Care*, 57(2), 67-78. <https://doi.org/10.1080/00981389.2017.1344754>
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80(1), 1-28. <https://doi.org/10.1037/h0092976>
- Sanders, M. R., Turner, K. M. T., & Metzler, C. W. (2019). Applying self-regulation principles in the delivery of parenting interventions. *Clinical Child and Family Psychology Review*, 22, 24-42. <https://doi.org/10.1007/s10567-019-00287-z>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 8(2), 23-74.
- Schilbach, F., Schofield, H., & Mullanathan, S. (2016). The psychological lives of the poor. *American Economic Review*, 106(5), 435-440. <https://doi.org/10.1257/aer.p20161101>
- Segretin, M. S., Hermida, M. J., Prats, L. M., Fracchia, C. S., Ruetti, E., & Lipina, S. J. (2016). Childhood poverty and cognitive development in Latin America in the 21st Century. *New Directions for Child and Adolescent Development*, 152, 9-29. <http://doi.org/10.1002/cad.20162>
- Sen, A. (1985). Well-being, agency and freedom: The

- Dewey Lectures 1984. *The Journal of Philosophy*, 82(4), 169-221. <https://doi.org/10.2307/2026184>
- Shah, A. K., Mullainathan, S., & Shafir, E. (2012). Some consequences of having too little. *Science*, 338(6107), 682-685. <https://doi.org/10.1126/science.1222426>
- Sheehy-Skeffington, J., & Haushofer, J. (2014). The Behavioural Economics of Poverty. In United Nations Development Programme Istanbul International Center for Private Sector in Development (Ed.), *Barriers to and Opportunities for Poverty Reduction* (pp. 96-112).
- Sherer, M., & Adams, C. H. (1983). Construct validation of the Self-Efficacy Scale. *Psychological Reports*, 53(3), 899-902. <https://doi.org/10.2466/pr0.1983.53.3.899>
- Shilling, A. A., & Brown, C. M. (2016). Goal-driven resource redistribution: An adaptive response to social exclusion. *Evolutionary Behavioral Sciences*, 10(3), 149-167. <https://doi.org/10.1037/ebs0000062>
- Úcar-Martínez, X., Jiménez-Morales, M., Soler-Masó, P., & Trilla-Bernet, J. (2017). Exploring the conceptualization and research of empowerment in the field of youth. *International Journal of Adolescence and Youth*, 22(4), 405-418. <https://doi.org/10.1080/02673843.2016.1209120>
- Vega-Valero, C. Z., Hernández-Toledano, R. A., García-Arreola, O., Nava-Quiroz, C., & Ruíz-Méndez, D. (2019). Escala de Habilidades de Solución de Problemas en Directivos: Desarrollo y validación. *Revista Evaluar*, 19(3), 68-85. <https://doi.org/10.35670/1667-4545.v19.n3.26814>
- Wolf, J. R. L. M., & Jonker, I. E. (2020). Pathways to empowerment: The social quality approach as a foundation for person-centered interventions. *International Journal of Social Quality*, 10(1), 29-56. <https://doi.org/10.3167/IJSQ.2020.100103>
- Xu, M., Li, Z., Qi, S., Fan, L., Zhou, X., & Yang, D. (2020). Social exclusion modulates dual mechanisms of cognitive control: Evidence from ERPs. *Human Brain Mapping*, 41(10), 2669-2685. <https://doi.org/10.1002/hbm.24970>
- Yoshikawa, H., Aber, J. L., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioral health of children and youth: Implications for prevention. *American Psychologist*, 67(4), 272-284. <https://doi.org/10.1037/a0028015>
- Zumbo, B.D., Gadermann, A.M., & Zeisser, C. (2007). Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales. *Journal of Modern Applied Statistical Methods*, 6, 4. <https://doi.org/10.22237/jmasm/1177992180>

Conceptual and Metric Adaptation of the Professional Quality of Life Scale in a Population of Argentinian Psychologists

Adaptación Conceptual y Métrica de la Escala Professional Quality of Life Scale en Población de Psicólogos Argentinos

Leandro Eidman^{1,2,3}, Daiana Zaragoza-Seratti^{2,3} *

1 - Universidad Nacional del Chaco Austral, Saenz Peña, Chaco, Argentina.

2 - Universidad de Ciencias Empresariales y Sociales, Resistencia, Chaco, Argentina.

3 - Instituto de Neurociencias Argentina (INAR), Resistencia, Chaco, Argentina.

Introduction
Methodology
Results
Discussion
References

Recibido: 16/06/2023 Revisado: 24/08/2023 Aceptado: 04/09/2023

Abstract

The quality of professional life, especially in clinical psychologists, is a relevant issue due to the high incidence of burnout and emotional exhaustion. A non-experimental, cross-sectional, instrumental study was conducted to adapt and validate the Pro-QOL R-IV scale in Argentinian psychologists. Data were obtained through a voluntary and non-probabilistic sampling. A sample of 560 psychologists of mean age 40.7 (SD = 28.37; Min = 23, Max = 67; 67% women, 33% men) was used. In all cases, adequate percentage values and Aiken's V coefficients were obtained. The exploratory factor analysis grouped 30 items into three latent variables, showing adequate values in the validity tests. Confirmatory factor analysis supported a good fit for the three-factor model. It is concluded that the Pro-QOL-Psi Inventory is a valid and reliable instrument to assess quality of life in Argentine psychologists.

Keywords: *quality of life, Argentinian psychologists, conceptual adaptation, metric adaptation*

Resumen

La calidad de vida profesional, especialmente en psicólogos clínicos, es un tema relevante debido a la alta incidencia de agotamiento y desgaste emocional. Se realizó un estudio no experimental transversal de tipo instrumental para adaptar y validar la escala Pro-QOL R-IV en psicólogos argentinos. Los datos se obtuvieron a través de un muestreo voluntario y no probabilístico. Se utilizó una muestra de 560 psicólogos de edad promedio 40.7 (DE = 28.37; Mín = 23, Máx = 67; 67% mujeres, 33% hombres), en la cual, se obtuvieron adecuados valores de porcentajes y coeficientes V de Aiken en todos los casos. El análisis factorial exploratorio agrupó 30 ítems en tres variables latentes, mostrando valores apropiados en las pruebas de validez. El análisis factorial confirmatorio respaldó un buen ajuste para el modelo de tres factores. Se concluye que el Inventario Pro-QOL-Psi es un instrumento válido y confiable para evaluar la calidad de vida en psicólogos argentinos.

Palabras clave: *calidad de vida, psicólogos argentinos, adaptación conceptual, adaptación métrica*

*Correspondence to: leaneidman@gmail.com. Teléfono: 3624653979.

How to cite: Eidman, L., & Zaragoza-Seratti, D. (2023) Conceptual and Metric Adaptation of the Professional Quality of Life Scale in a Population of Argentinian Psychologists. *Revista Evaluar*, 23(3), 77-93. Retrieved from <https://revistas.unc.edu.ar/index.php/revaluar>

Participaron en la edición de este artículo: María Eugenia Maiorana, Juan Cruz Balverdi Nieto, Abigail Pérez, Florencia Ruiz, Rodrigo Maderna, Jorge Bruera.

Introduction

In recent years, quality of life has been a key topic of study in the case of health professionals. Numerous research studies have found increased levels of emotional, physical and social depletion in this population (Ramírez et al., 2020; Salgado-Roa & Leria-Dulčić, 2020; Vidotti et al., 2019; Yslado-Méndez et al., 2019).

Professional quality of life was defined as the feeling of well-being derived from the balance a person experiences between the demands or the load of a challenging, intense and complex job, and the psychological, organizational and linking resources the person has to face them (García-Sánchez, 1993). Therefore, it will depend on multiple factors: some in the personal area, such as age, gender, personality; others in the family area including social status and support network; any others in the work context, for example the retribution in a professional career; and lastly, elements related to professional recognition, work circumstances and management style (Abouzeid et al., 2020; Cazana-Vásquez, 2017; Loli et al., 2018; Salgado-Roa & Leria-Dulčić, 2020).

In the professional field of mental health, particularly in the case of psychologists, there is no single characteristic that synthesizes the work performed with pain, since the influences on its efficacy and wellbeing are multiple (Larsen & Stamm, 2008). So much so, that the concept of professional quality of life is a result of the dynamic interaction of the positive and gratifying factors, as well as the negative and harmful ones, both essential in order to understand the impact that working with trauma has on the professionals of mental health. Among them, it is possible to find the significant risks that are inherent to the work performed by care providers, such as compassion fatigue and/or burnout, as well as its benefits, which can be grouped under the name

“compassion satisfaction” (Leonardo Granados & Chocó Cedillos, 2018).

Given that quality of life encompasses the assessment of subjective factors, a reliable method is needed in order to obtain information from the individual (Velarde Jurado & Avila Figueroa, 2002). That is where the interest of adapting and validating the *Professional Quality of Life Scale, Compassion Satisfaction and Fatigue Subscales-Revision IV* (ProQOL R-IV; Stamm, 2005), created as a tool to detect positive and negative aspects of the work among professionals dedicated to helping (Stamm, 2010), arises in benefit of the population of Argentinian psychologists. This is the instrument par excellence to assess professional quality of life, and, as its name indicates, it consists of three discreet subscales which allow the studying of compassion satisfaction, burnout and compassion fatigue (Suárez, 2012).

Compassion is manifested in contexts of someone else’s particular pain or distress, and the intention of providing a relief response originated by empathy (Klos & Lemos, 2018). Related to this, compassion satisfaction was defined as the feeling of pleasure therapists derive from doing their job well (Stamm, 2002), and also as the positive stimulus of collaborating (Stamm, 2010), associated with the pleasure of taking the role of helping others and the process of mutual healing between the professional and the patient, the internal self-reflection, the connection with peers, the increased feeling of spirituality and a strong level of empathy, as was mentioned before (Benito et al., 2010).

Regarding burnout syndrome, it was delimited as a psychological disorder arising when the mechanisms of adaptation to prolonged stressful situations fail (Otero et al., 2013). It consists of three factors: a) emotional burnout, that is, the feeling of lack of energy and chronic toll; b) depersonalization, which means that the per-

son experiences an attitude of dehumanization or lack of interest towards the person to whom he or she is providing the service, and c) the reduction of personal fulfillment, which is the tendency to value oneself negatively and to feel displeased and dissatisfied with the professional results (Edwards, 2006; Illera, 2006; Maslach et al., 2001; Rupert & Morgan, 2005). As a precedent, it is worth mentioning that the origins of this concept are associated with the intense speed with which socio-economic and cultural changes were produced, which have transformed the industrial society into a society of service economy (González-Riviera et al., 2022).

Lastly, compassion fatigue was described as the natural behaviors and affections, consequences of knowing a traumatic event experienced by a person who is valuable to the individual. It is the stress that derives from assisting or trying to help someone who is traumatized or hurt. It may also be described as a state of biological, psychological and social exhaustion and dysfunction (Figley, 2013). This is a current concept which, after evolving over these twenty years, became a concept on its own, used as a synonym of secondary traumatic stress (Cuartero et al., 2020).

The ProQOL (Stamm, 2005), originally developed in English, has been translated into approximately 27 languages, including Spanish (Morante-Benadero et al., 2005). At international level, studies have been conducted on psychometric properties in China, Korea, Japan, Turkey, Spain and Brazil, with adequate fits and reliability of estimates (Dinç & Ekinçi, 2019; Fukumori et al., 2016; Galiana et al., 2017; Kim & Choi, 2019; Shen et al., 2015). In the same way, the creators of this inventory encourage its diffusion and translation into other languages (Stamm, 2005).

Argentina has the largest number of psy-

chologists per capita at world level, with a significant predominance in the exercise of the clinical area over the others (Alonso et al., 2011). Although there are studies focused on the assessment of the quality of life of healthcare professionals (Chavez, 2017; Hauser & García, 2017), and also on the adaptation of instruments dedicated to obtaining feedback from patients receiving psychotherapy (Gómez Penedo et al., 2021), there is not an instrument specifically focused on assessing the quality of life of populations exclusively made up of psychologists yet, which also takes into consideration the native language that Argentinians use and which allows them to receive valuable information, both for their personal development and their optimum performance in the clinical exercise.

This is striking, since the exercise of psychology exposes professionals to psychosocial risks by stating a very close bond with the people they help. This type of relationship frequently demands their involvement in emotional conflicts which often impact the personal life of the psychologist (Benavides Pereira et al., 2002). Considering that the practice of psychotherapy becomes a demanding task, it is essential that psychologists carry out self-care strategies (Fernández Álvarez, 2008). In order to contribute to this, it is necessary for them to know about their mental health, implied in the way in which they conduct their activity.

For these reasons, the general aim has been to conduct the conceptual, linguistic and metric adaptation of the *Professional Quality of Life. Compassion Satisfaction and Fatigue Subscales-Revisión IV* (Pro-QOL R-IV; Stamm, 2005) in the population of Argentinian clinical psychologists. The specific objectives were: (1) to examine evidence of content validity; (2) to contribute with evidence of face validity; (3) to contribute with evidence of construct validity, and (4) to

analyze the internal consistency of scores.

Methodology

Sample

The sample consisted of 560 Argentinian clinical psychologists. The mean age was 40.07 (SD = 28.37, Min = 23, Max = 67). 67% ($n = 375$) were women and 33% ($n = 185$) were men. Regarding the place of residence, 33.8% ($n = 189$) reported living in the province of Buenos Aires, 12.9% ($n = 72$) in the City of Buenos Aires, 11.1% ($n = 62$) in Córdoba, 10% ($n = 56$) in Chaco, 7.9% ($n = 44$) in Santa Fe, and the remaining 24.3% ($n = 137$) were distributed between the other provinces of the country. With regards to work context, 62.3% ($n = 349$) reported working in the private area, 7.0% ($n = 39$) in the public area, and 30.7% ($n = 172$) in both. Concerning work experience, the mean was 10.64 years, and, regarding daily workload, the average was of 7.56 hours. In relation to the theoretical orientation, 35.9% ($n = 199$) reported being psychoanalysts; 28.2% ($n = 156$), behavioral-cognitive; 19% ($n = 104$), integrative; and the remaining 17% ($n = 101$) were found among the systemic, humanistic, eclectic, EMDR, neuropsychological and Gestalt orientations. Regarding the attention modality, 14.3% ($n = 80$) reported working in person; 12.3% ($n = 69$), online; and 73.4% ($n = 411$) both. With respect to the type of population assisted, 51.1% ($n = 286$) reported assisting the population in general; 37.5% ($n = 210$), adult patients; 10.7% ($n = 60$), children and adolescents; .7% ($n = 4$) older adults. Finally, in connection with supervised practice, 83.8% ($n = 469$) reported its inclusion, while 16.3% ($n = 91$) said they did not contemplate it in their professional practice.

Instruments

Record of socio-demographic and professional variables. Through this instrument, data were collected about place of residence, gender, age, work experience, daily workload, work context, attention modality, theoretical orientation and type of population assisted, type of problems approached, and supervision among professionals.

Professional Quality of Life Scale. Compassion Satisfaction and Fatigue Subscales-Revisión IV (ProQOL R-IV; Stamm, 2005). This scale consists of 30 items reflecting different sensations towards daily work, and each of them indicates how often these situations present themselves in the professional life of the person. Each answer is qualified as *never*, *rarely*, *sometimes*, *frequently*, *almost always* and *always*. In addition, it is made up of three subscales in which it is possible to study Compassion Satisfaction, Burnout and Compassion Fatigue. The reliability of the subscales is $\alpha = .87$ for Compassion Satisfaction; $\alpha = .72$ for Burnout and $\alpha = .80$ for Compassion Fatigue.

Procedure

A non-experimental, cross-sectional, instrumental study was designed (Ato et al., 2013). The author of the *Professional Quality of Life Scale. Compassion Satisfaction and Fatigue Subscales-Revisión IV* (Pro-QOL R-IV; Stamm, 2005) was asked for authorization. Data were obtained through a voluntary and non-probabilistic sampling. The techniques used to measure the variables were distributed across social media under the modality of *google forms*©. Participants did not receive any compensation for their collaboration. As a mandatory field, the form contained

the acceptance of an informed consent in the front page, in which details of the objectives of the research were specified, along with the guarantee of confidentiality enforced by law 25.326, of personal data protection, which deals with the ethical implications of research in the health field in which human beings participate, in order to protect their fundamental rights, weighing the need of promoting research.

Data analysis

The epistemological foundation was classical psychometrics. In order to validate the content of the instrument, two researchers participated in the process of translation, following the recommendations of the process of expert judgment suggested by [Escobar-Pérez and Cuervo-Martínez \(2008\)](#). Both researchers hold a PhD in Letters and showed good command of the native tongue. Therefore, they translated the Pro-QOL R-IV ([Stamm, 2005](#)) independently and with no debate and, according to the judges, the versions did not differ. After that, there were some adjustments made to ensure comprehensibility and psychological equivalence. The criteria to select the judges were: (a) they have previous experience to act as expert judges, (b) they have expertise in psychometry and psychological assessment, and (c) they show evidence of knowledge about professional quality of life. After the selection of judges, a spreadsheet with instructions was created, where the aims of the study were made explicit, as well as the indications stating what people were expected to do. In order to assess semantic and syntactic clarity, they used a four-point Likert scale, where 1 indicated *different*; 2, *quite different*; 3, *quite similar*; and 4, *similar*. This is how the understanding of an item in our cultural context was assessed. In order to assess the coherence

of paraphrase, people were asked to use a four-point Likert scale, where 1 indicated that it did not match the criteria; 2, low level of criteria; 3, moderate level of criteria; and 4, high level, in order to see if the statement showed logic regarding the dimension or indicator it measured. Finally, to assess the relevance of the paraphrased statements, they used a four-point Likert scale, where 1 indicated *irrelevant*; 2, *low level of relevance*; 3, *moderate level of relevance*; and 4, *relevant*. In this way, the statement was valued as essential, as important or as to be excluded.

Additionally, according to [Tornimbeni et al. \(2008\)](#), a section for observations was displayed, indicating the possibility of contributing with comments about the congruence of the statement with the dimension and the syntactic aspects that should be highlighted. After obtaining the results, a spreadsheet with the corresponding assessments was created, through which the percentage of agreement of the judgment that had been conducted ([Tinsley & Weiss, 1975](#)) and Aiken's V coefficient were estimated ([Aiken, 1985](#)). These indicators are represented by values ranging from 0 to 1; the closest the statement to 1, the higher the content validity.

After that, a pilot study was developed before administering the instrument to the whole sample. Once the definitive sample was obtained, the univariate normality of data was acquired through the indicators of skewness and kurtosis, for which scores between +/- 2 are desirable ([Tabachnick & Fidell, 2013](#)). The discriminatory power of the items was studied through the corrected item-total correlation, for which r scores equal or higher than .30 are expected ([Nunnally & Bernstein, 1995](#)). Following that, studies tending to verify construct validity through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were conducted. For the exam of factorial loads, standardized loads higher than

the $> .30$ limit (Hair et al., 2006; Nunnally & Bernstein, 1994) were considered as *acceptable*, and, regarding the correlations between factors, scores $> .19$ were considered as *very low*; between $> .20$ and $< .39$, as *low*; between $> .40$ and $< .59$, as *moderate*; between $> .60$ and $< .79$, as *high*; and $< .80$, as *very high* (Brown, 2006; Evans, 1996).

While some authors criticize the joint use of these analyses (Pérez-Gil et al., 2000), numerous works opt for conducting both procedures (Martorell et al., 2011). The EFA was calculated through a robust method of maximum likelihood (MLR) using Varimax rotation in a polychoric matrix. The notion that a larger sample size generally provides more accurate and reliable estimates is a widely accepted concept in the field of statistics and research methodology. However, the suggestion that smaller samples could be enough in specific cases, where factorial relations are evident, is an idea that has been discussed by various experts in the field of psychometry and statistics (Galbraith et al., 2002; George & Mallery, 2019; Hair et al., 2014). Criteria to determine the number of factors was developed through the *Scree Test*. In addition, factors, such as theoretical interpretation, coherence with previous research and practical relevance were considered when deciding the number of factors to be kept in the analysis. The CFA was conducted using a robust estimator of weighted least squares mean and variance adjusted (WLSMV-R), the estimation method used was MLR, and, given that the variables were ordinal, the polychoric matrix was applied, since it is the most appropriate one for this type of data (Freiberg Hoffmann et al., 2013; Múthen & Kaplan, 1985).

The following indices of goodness of fit were considered: chi-square (χ^2), comparative fit index (*CFI*), Bollen's incremental fit index (*IFI*) and root mean squared error of approximation

(*RMSEA*) (Hu et al., 1999). Regarding the criteria of acceptable fit values, a score of $.90$ in *CFI* (Kline 2011; Schumacker & Lomax, 2016) is considered, as well as scores lower than or equal to $.08$ in *RMSEA* (Browne & Cudeck, 1993).

Finally, the Omega statistical was used, (McDonald, 1999) as a favorable alternative to the limitations of Cronbach's alpha coefficient (Cho, 2016) to know the reliability of a scale from the point of view of its internal consistency by offering an approximation to the reliability based in the factor structure of the instrument. The acceptable fit value had to be higher than $.70$ (Kline, 2011; Raykov & Shrout, 2002). In the same way, Cronbach's alpha coefficients were calculated to compare them with other studies.

Results were processed using R (Version 3.6.0) and the interface R Studio (Version 1.1.463) through the ggplot2 packages for the visualization of data (Wickham, 2009), psych (Revelle, 2018), psychometric (Fletcher & Fletcher, 2013) and psycho (Revelle, 2018), in order to estimate some psychometric properties. While lavaan (Rosseel, 2012), semPlot (Epskamp et al., 2019) and semTools (Jorgensen et al., 2018) were used to calculate and trace the Structural Equation Modeling.

Results

Evidence of Content Validity and Face Validity

Based on the experts' judgment, the paraphrased scales that were considered as the most adequate, along with the Likert scale, where, from the indices of percentage and Aiken's V statistic, judges suggested that a scale from *Nothing* to *Always* was conceptually adequate to measure the professional quality of life in Argentinian psychologists. In Table 1, adequate values of percentages in the items according to a range of $.80$ and 1 can be seen. The same happened with Aiken's V

Table 1
Expert Judgment: Percentage of agreement and Aiken's V.

	Clarity		Coherence		Relevance	
	% Acceptance	Aiken's V	% Acceptance	Aiken's V	% Acceptance	Aiken's V
CVPA1 ^a	.92	.92	1.00	1.00	1.00	1.00
CVPB1 ^b	1.00	1.00	1.00	1.00	1.00	1.00
CVPC1 ^c	.90	.96	.90	.96	1.00	1.00
CVPD1 ^d	1.00	1.00	.96	.92	1.00	1.00
CVPE1 ^e	.90	.96	.97	.97	.90	.95
CVPF1 ^f	1.00	1.00	1.00	1.00	1.00	1.00
CVP 1 ^g	1.00	1.00	1.00	1.00	1.00	1.00
CVP 2 ^g	.84	.93	.93	.92	.80	.87
CVP 3 ^g	1.00	1.00	1.00	1.00	1.00	1.00
CVP 4 ^g	.80	.87	1.00	1.00	.80	.80
CVP 5 ^g	1.00	1.00	1.00	1.00	1.00	1.00
CVP 6 ^g	.84	.93	.83	.92	.80	.87
CVP 7 ^g	.84	.90	1.00	1.00	.83	.86
CVP 8 ^g	.84	.93	1.00	1.00	.80	.87
CVP 9 ^g	.83	.96	.83	.96	.83	.96
CVP 10 ^g	1.00	1.00	1.00	1.00	1.00	1.00
CVP 11 ^g	.94	.97	1.00	1.00	.88	.96
CVP 12 ^g	.86	.90	.94	.90	.95	.98
CVP 13 ^g	.90	.93	.96	.98	1.00	1.00
CVP 14 ^g	.92	.96	.81	.93	.83	.96
CVP 15 ^g	.90	.92	.87	.94	.80	.87
CVP 16 ^g	1.00	1.00	1.00	1.00	1.00	1.00
CVP 17 ^g	.84	.93	.83	.92	.80	.87
CVP 18 ^g	1.00	1.00	1.00	1.00	1.00	1.00
CVP 19 ^g	1.00	1.00	1.00	1.00	1.00	1.00
CVP 20 ^g	.83	.96	.83	.96	.83	.96
CVP 20 ^g	1.00	1.00	1.00	1.00	1.00	1.00
CVP 21 ^g	1.00	1.00	1.00	1.00	1.00	1.00
CVP 22 ^g	.80	.87	1.00	1.00	.80	.80
CVP 23 ^g	.85	.93	1.00	1.00	.88	.92
CVP 24 ^g	.90	.97	.87	.90	.93	.95
CVP 25 ^g	1.00	1.00	.80	.93	.80	.93
CVP 26 ^g	.85	.90	.88	.93	.82	.95
CVP 27 ^g	.80	.87	1.00	1.00	.80	.80
CVP 28 ^g	.85	.93	1.00	1.00	.88	.92
CVP 29 ^g	1.00	1.00	1.00	1.00	1.00	1.00
CVP 30 ^g	.84	.93	.93	.92	.80	.87

Note. CVP = calidad de vida profesional (professional quality of life); ^a = primer valor de la escala Likert (first value of the Likert scale); ^b = segundo valor de la escala likert (second value of the Likert scale); ^c = tercer valor de la escala Likert (third value of the Likert scale); ^d = cuarto valor de la escala Likert (fourth value of the Likert scale); ^e = quinto valor de la escala Likert (fifth value of the Likert scale); ^f = sexto valor de la escala Likert (sixth value of the Likert scale); ^g = ítems del test (test items).

coefficients, oscillating between expected scores of .80 and 1 in all cases (Hyrkäs et al., 2003).

From the suggestions and observations pointed out by the judges, modifications were made to the items which did not contemplate the female suffix in their statement, so “/a” was added. In addition, the original statement of item 4 *Me siento vinculado a otras personas con ocasión de mi trabajo (I feel linked to other people in the occasion of my work)*, was changed to *Me siento conectado/a con personas que no forman parte de mi trabajo (I feel connected to people who are not part of my job)*; item 7 *Encuentro difícil separar mi vida personal de mi vida profesional (I find it difficult to separate my personal life from my professional life)* was changed to *Me resulta difícil separar mi vida personal de mi vida profesional (It is difficult for me to separate my personal life from my professional life)*; item 8 *Pierdo el sueño por las experiencias traumáticas de las personas a las que he ayudado (I lose sleep over the traumatic experiences of people I have helped)* was changed to *Las experiencias traumáticas de las personas a las que he ayudado me han provocado alteraciones en el sueño (The traumatic experiences of people I have helped have provoked alterations in my sleep)*; and finally item 23 *Evito ciertas actividades o situaciones porque me recuerdan a las experiencias espantosas de la gente a la que he ayudado (I avoid certain activities or situations because they remind me of the awful experiences of the people I have helped)* was changed to *Evito ciertas actividades o situaciones porque me recuerdan a las experiencias negativas de la gente a la que he ayudado (I avoid certain activities or situations because they remind me of the negative experiences of the people I have helped)*.

Pilot Study

Prior to administering the Pro-QOL R-IV to the definitive sample for the adaptation of the instrument, a pilot study was conducted on a reduced scale ($n = 60$) with the purpose of examining the correct function of the scale. The examined criteria were: a) understanding of the items, b) understanding of the answer options, c) understanding of the codes for the answers, d) length of the instrument, e) language and vocabulary, f) complexity of the instrument, g) motivation to answer, and h) appearance of possible ego-defensive answers. The series of analyses reported that no difficulties were found in any of the assessed criteria. Additionally, in the pilot test, they consulted about the modifications suggested by expert judgment. In all cases, individuals mentioned they have understood the modified versions better, after the observations made by the judges.

Evidence of construct validity

First, an EFA of the data tending to assess the existence of absent scores, the detection of atypical cases and the underlying statistical assumptions (normality, linearity of relationships, multicollinearity) was carried out. Variables with more than 5% of missing data were not observed, nor the existence of univariate atypical cases (scores outside the range $z \pm 3$). In Table 2, the corrected item-total correlation and a distribution close to normality are presented, with scores of skewness and kurtosis lower than ± 2 . In order to run the internal structure studies, the sample was divided, having, then, an estimated sample ($n = 150$) to conduct the exploratory factor analysis (EFA) and a validation sample ($n = 410$) to conduct the CFA.

In Table 3, EFA can be observed through the robust method of maximum likelihood, which,

Table 2
Descriptive Statistics: Pro-QOL-Psi in Argentinian Psychologists.

	M	SD	Skewness	Kurtosis	Corrected total r
Item 1	2.69	.89	.47	-.43	.37
Item 2	3.53	.95	.34	-.16	.41
Item 3	5.20	.89	-1.10	.68	.30
Item 4	2.64	1.29	.44	-.55	.55
Item 5	2.65	1.28	.84	.18	.41
Item 6	4.64	1.07	-.55	-.34	.42
Item 7	2.54	1.15	.94	.63	.42
Item 8	1.77	.87	1.11	.97	.49
Item 9	1.81	.89	.89	.43	.48
Item 10	2.64	1.32	.46	-.63	.60
Item 11	2.41	1.25	.61	-.37	.63
Item 12	5.42	.71	-1.26	1.78	.33
Item 13	1.84	.95	1.11	.98	.46
Item 14	1.40	.70	1.87	1.27	.32
Item 15	4.13	1.81	-.51	-1.13	.63
Item 16	4.46	1.30	-.54	-.72	.41
Item 17	2.70	1.11	.58	-.14	.40
Item 18	4.70	.99	-.95	1.17	.54
Item 19	3.68	1.18	.105	-.54	.61
Item 20	4.51	.98	-.43	-.15	.42
Item 21	3.21	1.27	.328	-.48	.58
Item 22	4.26	1.11	-.31	-.43	.42
Item 23	1.47	.80	1.87	1.26	.36
Item 24	4.97	1.13	-1.12	1.04	.49
Item 25	1.87	.97	1.15	1.27	.47
Item 26	3.16	1.42	.31	-.68	.45
Item 27	4.49	.91	-.30	-.37	.33
Item 28	1.61	.83	1.89	1.45	.36
Item 29	3.60	1.21	-.41	-.25	.42
Item 30	5.16	1.01	-1.32	1.85	.35

from the Varimax rotation, determined the grouping of 30 items in 3 latent variables. The factor solution showed scores similar to .87 for the Káiser Meyer Olkin (KMO) index and for Barlett's Test of Sphericity ($\chi^2= 584.62$; $SD = .07$; $p < .000$), which were considered adequate.

Following, a CFA was conducted to determine the factor structure of the Pro-QOL R-IV. Due to the categorical nature of the items (Likert

scale), the chosen estimation method was the Robust Maximum Likelihood (MLR), using a polychoric matrix. In order to assess the fit of the model, different fit indices obtained, by the robust method, were examined. As way to evaluate the goodness of fit of the model, different indices were analyzed: square-chi (χ^2), normed fit index (*NFI*), comparative fit index (*CFI*), Bollen's incremental fit index (*IFI*) and root mean square er-

Table 3
Exploratory Factor Analysis of the Pro-QOL-Psi Inventory in Argentinian Psychologists.

Items	Factor 1	Factor 2	Factor 3
cvp22	.78		
cvp20	.74		
cvp30	.70		
cvp27	.64		
cvp18	.62		
cvp06	.56		
cvp03	.40		
cvp12	.49		
cvp16	.47		
cvp24	.42		
cvp21		.86	
cvp19		.85	
cvp26		.80	
cvp10		.79	
cvp29		.77	
cvp01		.68	
cvp04		.56	
cvp08		.45	
cvp15		.40	
cvp17		.36	
cvp25			.41
cvp05			.41
cvp23			.36
cvp14			.34
cvp11			.34
cvp09			.33
cvp07			.32
cvp02			.32
cvp13			.31
cvp28			.31
% of variance explained for each factor	15.2	29.3	36.3
M	47.8	30.27	21.12
DE	6.52	6.11	5.69

ror of approximation (*RMSEA*). In Table 4, the trial of different models can be observed. Different models were tested: (1) a model of a single factor, with a single measure of professional quality of life, (2) a model of two factors, where one latent dimension represented aspects of compassion satisfaction, and the other dimension, the aspects of compassion fatigue, (3) a model of three factors, as proposed by Stamm (2005). The three-factor structure is the model that best fits and replicates the structure of the Professional Quality of Life Scale. In Figure 1, the regression weights of the statement for each dimension can be observed.

Internal consistency

Regarding the internal consistency of the scale, McDonald's omega and Cronbach's alpha coefficients were calculated. In Table 5, it can be observed that the obtained scores in the Compassion Satisfaction dimension have high internal consistency, while the scores for the Burnout and Compassion Fatigue dimensions were lower, although it is acceptable in all analyses.

Table 4
Indices of Fit of the Models of the Pro-QOL-Psi Inventory in Argentinian Psychologists.

Models	χ^2	<i>NFI</i>	<i>CFI</i>	<i>IFI</i>	<i>RMSEA (IC 90%)</i>
Single Factor	2573.299	.71	.87	.87	.12 (.12 - .11)
Two Factors	2281.836	.71	.72	.72	.22 (.21 - .22)
Three Factors	1873.826	.90	.92	.92	.05 (.05 - .06)

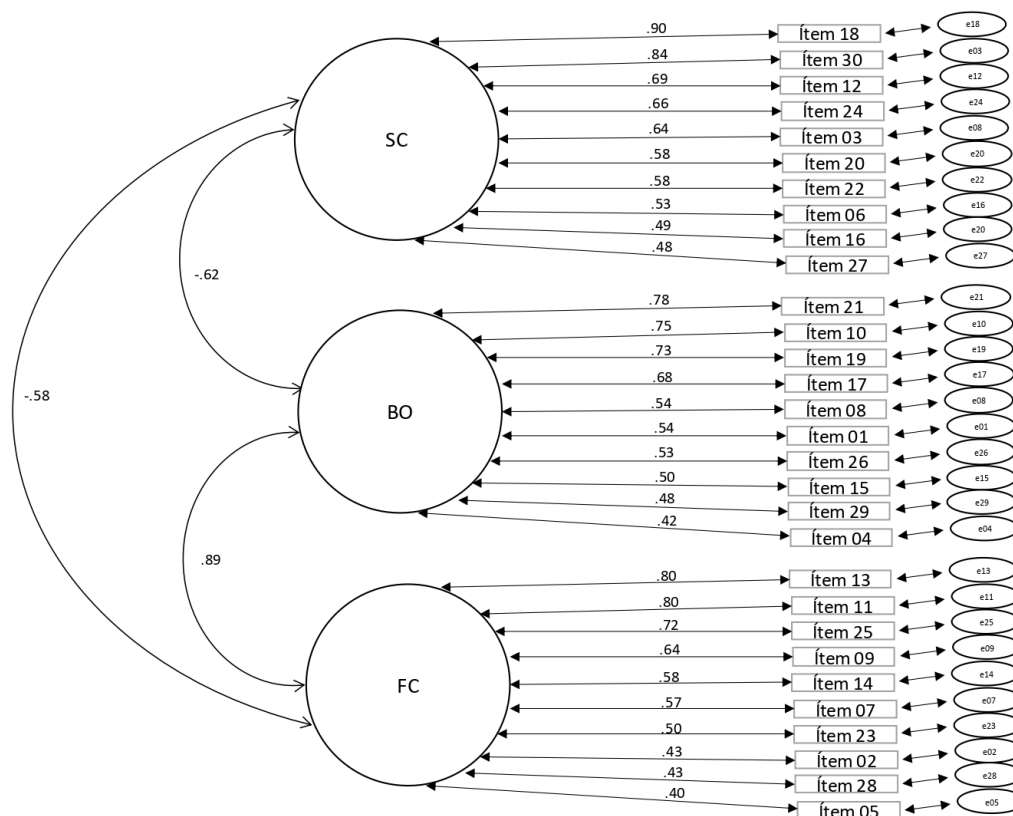


Figure 1
Confirmatory Factor Analysis of the Pro-QOL-Psi Inventory in Argentinian Psychologists.

Table 5
Internal Consistency of the Pro-QOL-Psi Inventory in Argentinian Psychologists.

Dimensions	Omega	Alfa
Compassion Satisfaction	.93	.90
Burnout	.78	.74
Compassion Fatigue	.82	.79

Discussion

With the aim of finding an adaptation of the *Professional Quality of Life. Compassion Satisfaction and Fatigue Subscales-Revision IV* (Pro-QOL R-IV; Stamm, 2005) with linguistic, conceptual and metric equivalence in the population of Argentinian clinical psychologists, a translation by expert judges was completed. That is how evidence of content validity was obtained from the statements that make up the instrument. According to the judges, linguistic modifications were made to items 4, 7, 8 and 23, with the purpose of reaching a better perception by the subjects of the sample. Later, adjustments were made to ensure the understanding and psychological equivalence. Based on the results obtained, those paraphrased scales, selected as the most adequate ones, along with the Likert scale with options going from *Nothing* to *Always* proposed in the original scale (Stamm, 2005), were selected. Then, through a pilot study it was possible to test the adequate understanding of the items, the answer choices, the code of the answers, the length of the instrument, its complexity and the motivation to answer.

Afterwards, the analyses of construct validity and internal consistency were conducted, which were found to be in agreement with the antecedents that contributed with empirical evidence about the instrument (Dinç & Ekinici, 2019; Fukumori et al., 2016; Galiana et al., 2017; Kim & Choi, 2019; Morante Baradero et al., 2005;

Shen et al., 2015; Stamm, 2005). The preliminary analysis of the items allowed the obtaining of evidence of univariate normality of the statements in the scale. In addition, the indices of skewness and kurtosis were in agreement with the scores recommended by Tabachnick and Fidell (2013).

The EFA determined the grouping of 30 items in the three dimensions reported in the original scale (Stamm, 2005), which can be explained through the theory sustaining each construct (Benito et al., 2010; Edwards, 2006; Illera, 2006; Maslach et al., 2001; Otero et al., 2013; Suárez, 2012; Rupert & Morgan, 2005). That is to say that certain statements exist that are explained by Compassion Satisfaction, Burnout and Compassion Fatigue. This is how it can be understood that the dynamic interaction of dimensions is essential to take in the impact that working with trauma has on the professionals of mental health. Moreover, the Compassion Satisfaction dimension explained the 15.2 % of the variance; while Burnout, the 29.3 %; and the Compassion Fatigue, the 36.3 %, data resembling the original version and the adaptations made in other countries (Dinç & Ekinici, 2019; Fukumori et al., 2016; Galiana et al., 2017; Kim & Choi, 2019; Morante Baradero et al., 2005; Shen et al., 2015; Stamm, 2005). Additionally, all the factor weights were higher than .30, defined as moderate and high in agreement with what was reported in previous research (Brown, 2006; Evans, 1996; Hair, et al., 2006; Nunnally & Bernstein, 1994), which determined that the 30 items distributed in the three proposed dimensions contribute to measure those dimensions in a consistent way and in a combined way for the Professional Quality of Life of Argentinian Psychologists.

Three models of CFA were tested. The three-dimension model, just like the original scale (Stamm, 2005), portrays a better combination of fit coefficients than other models. This explains

that the professional quality of life of a psychologist in Argentina can be assessed through the feeling of pleasure obtained by therapists when doing their job well, from emotional depletion and behaviors and natural affections deriving from knowing about the suffering of the patient.

Finally, the obtained indices of internal consistency were similar to those found in the original version. These results indicate that, with this instrument, a reliable overall measure of quality of life of the professional psychologist in Argentina can be reached. This is underpinned by the fact that all the dimensions have excellent internal consistency, in agreement with the conducted analyses of reliability. It is necessary to highlight that in order to assess internal consistency, the McDonald's omega and Cronbach's alpha indices of reliability (McDonald, 1999) were calculated, with the purpose of comparing the psychometric properties of the instrument with studies conducted in other countries.

Now, it is also necessary to consider a number of limitations in this study: first, the use of a non-probabilistic, intentional sampling; and second, the fact that there were no studies conducted on positive and negative convergent validity (Coulacoglou & Saklofske, 2017). In future research, it would be advisable to use probabilistic sampling to analyze a factor structure in a similar population, and to investigate positive and negative convergent validity in order to obtain more evidence regarding the psychometric properties of the instrument. On the other hand, the reliability of the measure was only examined from the perspective of internal consistency, whereas it would be desirable in future studies to report data regarding its stability in time (test-retest).

To conclude, it could be reported that the Pro-QOL-Psi Inventory has adequate evidence of content, face and construct validity, and excellent scores of internal consistency, which determines

that it is a valid and reliable instrument to assess the quality of life in Argentinian psychologists. Hopefully, this instrument will be useful and it will function as a starting point in the exploration of a variable that, in spite of its relevance, is studied very little in our environment. Finally, it is accurate to clarify that the version offered in this study should not be used in other Hispanic countries without prior and careful analysis of its cultural relevance, since new adaptations might be necessary before it is used in other regions of Latin America, especially in other continents.

References

- Abouzeid, G. K. A., Montesinos, M. J. L., & Alarcón, L. M. (2020). Consecuencias de los factores psicosociales en la calidad de vida laboral. *RECIEN: Revista Electrónica Científica de Enfermería*, (19), 24-39. <https://doi.org/10.14198/recien.2020.19.03>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and psychological measurement*, 45(1), 131-142. <https://doi.org/10.1177/0013164485451012>
- Alonso, M., Klinar, D., & Gago, P. (2011). Psicólogas/os en Argentina. Actualización cuantitativa 2010. *III Congreso Internacional de Investigación y Práctica Profesional en Psicología; XVIII Jornadas de Investigación; VII Encuentro de Investigadores de Psicología del Mercosur*. Buenos Aires.
- Ato, M., López-García, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología/Annals of Psychology*, 29(3), 1038-1059.
- Benavides Pereira, A. M., Moreno-Jiménez, B., Garrosa Hernández, E., & González Gutiérrez, J. L. (2002). La evaluación específica del síndrome de burnout en psicólogos: El inventario de burnout de psicólogos. *Clínica y salud*, 13(3), 257-283.
- Benito, E., Arranz, P., & Cancio, H. (2010). Herramientas

- para el autocuidado de profesionales que atienden a personas que sufren. FMC (Formación Médica Continuada). *Atención Primaria*.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford.
- Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. En K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Sage.
- Cazana Vásquez, Z. E. (2017). *Factores asociados a la calidad de vida profesional en enfermeros de un hospital público de Lima Norte*. [Tesis de Grado, Universidad Católica Sedes Sapientiae. Facultad de Ciencias de la Salud]. Repositorio Institucional Digital UCSS. <http://repositorio.ucss.edu.pe>
- Chavez, L. P. (2017). Percepción de Salud y Calidad de Vida en Profesionales de la Salud. *Anuario de Investigaciones de la Facultad de Psicología*, 3(2), 448-463.
- Cho, E. (2016). Making reliability reliable. *Organizational Research Methods*, 19(4), 651-682. <https://doi.org/10.1177/1094428116656239>
- Coulacoglou, C., & Saklofske, D. H. (2017). *Psychometrics and Psychological Assessment Principles and Applications*. Elsevier.
- Cuartero, M. E., Riera, J. A., & Casado, T. (2020). Comunicaciones Orales-La fatiga por compasión: La consecuencia de trabajar con personas que sufren. II *Jornadas Doctorales de la Universidad de Murcia*.
- Dinç, S., & Ekinçi, M. (2019). Turkish adaptation, validity and reliability of compassion fatigue short scale. *Psikiyatriye Guncel Yaklasimlar*, 11, 192-202. <https://doi.org/10.18863/pgy.590616>
- Edwards, D., Burnard, P., Hannigan, B., Cooper, L., Adams, J., Juggessur, T., & Coyle, D. (2006). Clinical supervision and burnout: the influence of clinical supervision for community mental health nurses. *Journal of clinical nursing*, 15(8), 1007-1015. <https://doi.org/10.1111/j.1365-2702.2006.01370.x>
- Epskamp, S., Stuber, S., Nak, J., Veenman, M., & Jorgensen, T. D. (2019). *semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output (Version 1.1.2)* [Computer software].
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en medición*, 6(1), 27-36.
- Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Brooks/Cole Publishing.
- Fernández-Álvarez, H. (2008). *Integración y salud mental actualización bibliográfica*. Bilbao.
- Figley, C. R. (2013). *Compassion fatigue: Coping with secondary traumatic stress disorder in those who treat the traumatized*. Routledge..
- Fletcher, T. D. & Fletcher, M. T. D. (2013). Package psychometric. <http://cran.rproject.org>
- Freiberg Hoffmann, A., Stover, J. B., de la Iglesia, G. & Fernández Liporace, M. (2013). Correlaciones Policóricas y Tetracóricas en Estudios Exploratorios y Confirmatorios. *Ciencias Psicológicas*, 7(2), 151-164. <https://cutt.ly/agwkKzd>
- Fukumori, T., Goto, T., Sato, H., & Sakamoto, H. (2016). Development, reliability, and validation of a Japanese nurse version of the Professional Quality of Life Scale (ProQOL-JN). In Links] *Poster session presented at the meeting of the 31st International Congress of Psychology, Yokohama, Japan*.
- Galbraith, J. I., Moustaki, I., Bartholomew, D. J., & Steele, F. (2002). *The analysis and interpretation of multivariate data for social scientists*. Crc Press.
- Galiana, L., Arena, F., Oliver, A., Sansó, N., & Benito, E. (2017). Compassion satisfaction, compassion fatigue, and burnout in Spain and Brazil: ProQOL validation and cross-cultural diagnosis. *Journal of pain and symptom management*, 53(3), 598-604. <https://doi.org/10.1016/j.jpainsymman.2016.09.014>
- García Sánchez, S. (1993). La qualitat de vida professional com a avantatge competitiu. *Revista de Qualitat*, 11, 4-9.
- George, D., & Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge.
- Gómez-Penedo, J. M., Areas, M. A., Manubens, R., Babl, A. M., Challú, L., Juan, S., Muñños, R., Fernández-

- Álvarez, J., Alalú, N., Roussos, A., Lutz, W., & Martin grosse-Holtforth, M. grosse-H. (2021). Propiedades psicométricas del Hopkins Symptom Checklist (HSCL-11) en Argentina: Un instrumento para monitoreo y feedback en psicoterapia. *Revista Evaluar*, 21(2), 33–47. <https://doi.org/10.35670/1667-4545.v21.n2.34393>
- González-Riviera, J. A., Álvarez-Alatorre, Y., Rosario-Hernández, E., Sepúlveda-López, V., Torres-Rivera, N., Ortiz-Santiago, T., Tirado de Alba, M., & González-Malavé, C. M. (2022). Escala de Cansancio Emocional: Análisis psicométrico en estudiantes de posgrado en Puerto Rico. *Revista Evaluar*, 22(2), 47-63. <https://doi.org/10.35670/1667-4545.v22.n2.38687>
- Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., & Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly*, 21(4), 431-454. <https://doi.org/10.1016/j.ecresq.2006.09.005>
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2014). Pearson new international edition. *Multivariate data analysis, Seventh Edition*. Pearson Education Limited Harlow, Essex.
- Hauser, M. P., & García, H. D. (2017). Satisfacción laboral y personalidad en trabajadores de salud pública de Pueyrredón. *Revista De Psicología Universidad De Antioquia*, 9(2), 9-25. <https://doi.org/10.17533/udea.rp.v9n2a02>
- Hu, L. T., & Bentler, P. M. (1999). Criterios de corte para los índices de ajuste en el análisis de la estructura de covarianza: Criterios convencionales versus nuevas alternativas. *Modelado de ecuaciones estructurales*, 6(1), 1-55.
- Hyrkäs, K., Appelqvist-Schmidlechner, K., y Oksa, L. (2003). Validating an instrument for clinical supervision using an expert panel. *International Journal of nursing studies*, 40(6), 619-625. [https://doi.org/10.1016/S0020-7489\(03\)00036-1](https://doi.org/10.1016/S0020-7489(03)00036-1)
- Illera, D. (2006). Síndrome de Burnout, aproximaciones teóricas. Resultado de algunos estudios en Popayán. *Revista Facultad Ciencias de la Salud: Universidad del Cauca*, 8(3), 21-29.
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., Rosseel, Y., Miller, P., Quick, C., & Garnier-Villareal, M. (2018). semTools: Useful tools for structural equation modeling. *R package version 0.5*, 1.
- Kim, J. I., & Choi, H. S. (2019). Validation of Korean version of Professional quality of life (K-ProQoL) for nurses through confirmatory factor analysis. *Journal of the Korea Academia-Industrial Cooperation Society*, 20(3), 243-251. <https://doi.org/10.5762/KAIS.2019.20.3.243>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford Press.
- Klos, M. C., & Lemos, V. N. (2018). Adaptación y validación de un instrumento para evaluar el constructo compasión. *Revista Evaluar*, 18(2), 31-44. <https://doi.org/10.35670/1667-4545.v18.n2.20801>
- Larsen, D., & Stamm, B. H. (2008). Professional quality of life and trauma therapists. En S. Joseph & P. A. Linley (Eds.), *Trauma, recovery, and growth: Positive psychological perspectives on posttraumatic stress* (pp. 275-293). John Wiley & Sons Inc.
- Leonardo-Granados, M., & Chocó-Cedillos, A. (2018). Calidad de vida profesional y Síndrome de Burnout en personal de enfermería del Departamento de Medicina Interna del Hospital Roosevelt. *Revista científica*, 27(2), 10-18.
- Loli, A. E., Danielli, J., Navarro, V., & Cerón, F. (2018). La calidad de vida laboral y el soporte institucional para el trabajo en profesionales de las entidades públicas y privadas de Lima. *Revista de investigación en psicología*, 21(1), 27-50. <https://doi.org/10.15381/rinvp.v21i1.15111>
- Martorell, M. C., González, R., Odoñez, A., & Gómez, O. (2011). Estudio confirmatorio del cuestionario de conducta antisocial (CCA) y su relación con variables de personalidad y conducta antisocial. *Revista Iberoamericana de Diagnóstico y Evaluación – e*

Avaliação Psicológica, 31(2), 35-52.

- Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual review of psychology*, 52(1), 397-422. <https://doi.org/10.1146/annurev.psych.52.1.397>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum Associates.
- Morante-Benadero, M. E., Moreno-Jiménez, B., & Rodríguez-Muñoz, A. (2006). Professional satisfaction and fatigue subscales—Version IV (ProQOL). *Universidad Autónoma de Madrid (Spain): Universidad Autónoma de Madrid. Available online: http://www.proqol.org/uploads/ProQol_vIV_Spanish_Oct05.pdf (accessed on 10 December 2019)*.
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189. <https://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory*. McGraw Hill.
- Nunnally, J. C., & Bernstein, I. J. (1995). *Psychometric theory*. McGraw-Hill.
- Otero, B. G. I., Grajeda, G. G., Canto, V. B., Venegas, J. M., & Luque, C. M. (2013). Burnout en psicólogos de la salud: características laborales relacionadas. *Psicología y salud*, 23(2), 217-226. <https://doi.org/10.25009/pys.v23i2.503>
- Pérez-Gil, J. A., Moscoso, S. C., & Rodríguez, R. M. (2000). Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema*, 12(Suplemento), 442-446.
- Ramírez, M. K. G. F., Escalante-Hernández, C., Ortiz-López, G., & Chico-Barba, L. G. (2020). Calidad de vida en profesionales de enfermería que laboran en áreas críticas en una institución de tercer nivel de atención. *Revista de Enfermería Neurológica*, 19(2), 53-65. <https://doi.org/10.51422/ren.v19i2.299>
- 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*, 9(2), 195-212. https://doi.org/10.1207/S15328007SEM0902_3
- Revelle, W. (2018). Psych: Procedures for psychological, psychometric, and personality research. *R package version, 1*(10).
- Roa, J. A., & Leria-Dulčić, F. (2018). Síndrome de burnout y calidad de vida profesional percibida según estilos de personalidad en profesores de educación primaria. *CES Psicología*, 11(1), 69-89. <https://doi.org/10.21615/cesp.11.1.6>
- Rosseel, Y. (2012). Iavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Rupert, P. A., & Morgan, D. J. (2005). Work setting and burnout among professional psychologists. *Professional Psychology: Research and Practice*, 36(5), 544-550. <https://doi.org/10.1037/0735-7028.36.5.544>
- Salgado-Roa, J. A., & Leria-Dulčić, F. J. (2020). Burnout, satisfacción y calidad de vida laboral en funcionarios de la salud pública chilenos. *Universidad y Salud*, 22(1), 06-16. <https://doi.org/10.22267/rus.202201.169>
- Schumacker, R. E., & Lomax, R. G. (2016). *A beginner's guide to structural equation modeling*. psychology press.
- Shen, J., Yu, H., Zhang, Y., & Jiang, A. (2015). Professional quality of life: A cross-sectional survey among Chinese clinical nurses. *Nursing & health sciences*, 17(4), 507-515. <https://doi.org/10.1111/nhs.12228>
- Stamm B. H. (2005). *The ProQOL Manual: The Professional Quality of Life Scale: Compassion Satisfaction, Burnout & Compassion Fatigue/Secondary Trauma Scales*. Carrollton, TX: Sidran Press.
- Stamm, B. H. (2002). Measuring compassion satisfaction as well as fatigue: Developmental history of the compassion satisfaction and fatigue test. *Treating compassion fatigue*, 1, 107-119.
- Stamm, B. H. (2010). *The concise ProQOL manual*. Eastwoods
- Suárez, S. (2012). *Estresores y satisfacción laboral del personal de enfermería de las Unidades de Cuidados*

Intensivos del Principado de Asturias [Tesis de maestría, Universidad de Oviedo] <https://digibuo.uniovi.es/dspace>

- Tabachnick, B., & Fidell, L. S. (2013). *Using Multivariate Statistics*. Pearson.
- Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22(4), 358. <https://doi.org/10.1037/h0076640>
- Tornimbeni, S., Pérez, E., Olaz, F., de Kohan, N. C., Fernández, A., & Cupani, M. (2008). *Introducción a la psicometría*. Paidós.
- Velarde-Jurado, E., & Avila-Figueroa, C. (2002). Evaluación de la calidad de vida. *Salud pública de México*, 44(4), 349-361.
- Vidotti, V., Martins, J. T., Galdino, M. J. Q., Ribeiro, R. P., & Robazzi, M. L. D. C. C. (2019). Síndrome de burnout, estrés laboral y calidad de vida en trabajadores de enfermería. *Enfermería Global*, 18(55), 344-376. <https://doi.org/10.6018/eglobal.18.3.325961>
- Wickham, H. (2009). *Elegant graphics for data analysis (ggplot2)*. Springer.
- Yslado-Méndez, R. M., Norabuena-Figueroa, R. P., Loli-Poma, T. P., Zarzosa-Marquez, E., Padilla-Castro, L., Pinto-Flores, I., & Rojas-Gamboa, A. (2019). Síndrome de burnout y la satisfacción laboral en profesionales de la salud. *Horizonte Médico* 19(4), 41-49. <https://doi.org/10.24265/horizmed.2019.v19n4.06>

Adaptación de la Escala de Liderazgo en el Deporte (LSS-1) en deportistas Argentinos

Adaptation of the Leadership in Sport Scale (LSS-1) in Argentinian Athletes

Marzorati, Ariela ^{1*}; Pagano, Alejandro, E. ²; Caicedo Cavagnis, Estefanía ³;
Fiotti, Julieta, A. ²; Fernández, Clara ⁴; Lorusso, Leandro, J. ⁵;
Lehman, Federico

1 - Facultad de Psicología y Ciencias Sociales, Universidad de Flores, Buenos Aires, Argentina.

2 - Facultad de Psicología, Universidad de Buenos Aires, Buenos Aires, Argentina.

3 - Instituto de Investigaciones Psicológicas (IIPsi-UNC-CONICET), Córdoba, Argentina.

4 - Facultad de Ciencias Humanas y de la Conducta, Universidad Favaloro, Buenos Aires, Argentina.

5 - Facultad de Psicología, Universidad Nacional de Mar del Plata, Buenos Aires, Argentina.

Recibido: 26/06/2023 Revisado: 29/08/2023 Aceptado: 10/09/2023

Introducción
Metodología
Resultados
Discusión
Agradecimientos
Referencias
Apéndice

Resumen

El objetivo de este estudio fue adaptar la versión el LSS-1 en deportistas argentinos. La investigación se realizó sobre 333 deportistas residentes en 12 provincias argentinas, con edades entre 16 y 63 años ($M = 26,37$ $DE = 8,51$) de 13 disciplinas deportivas distintas. Los resultados muestran que la traducción es clara y congruente. Los reactivos presentan un nivel de discriminación adecuado, a excepción de los ítems 6, 27, 34 y 40. El análisis factorial confirmatorio (AFC) no converge en la solución de 5 factores originales, por lo que se realizó un análisis factorial exploratorio (AFE) en el cual se descartaron los ítems 16 y 33 y se evidenció una estructura de 3 factores (comportamiento autocrático $\alpha = .54$, entrenamiento e instrucción $\alpha = .94$, comportamiento democrático $\alpha = .87$) que explican el 43% de la varianza. Se presentan valores normativos, los cuales dan cuenta de la obtención de una versión adaptada con propiedades psicométricas aceptables que permite avanzar en los estudios sobre la medición del liderazgo en deportistas argentinos.

Palabras clave: liderazgo, adaptación LSS-1, propiedades psicométricas, deportistas

Abstract

The aim of this study was to adapt the LSS-1 version in Argentinian athletes. The research was carried out on 333 athletes residing in 12 Argentine provinces, aged between 16 and 63 years old ($M = 26.37$ $SD = 8.51$) from 13 different sports disciplines. The results show that the translation is clear and congruent. The items present an adequate level of discrimination except for items 6, 27, 34, and 40. The AFC did not converge in the solution of 5 original factors, so an AFE was performed, discarding items 16 and 33 and evidencing a structure of 3 factors (autocratic behavior $\alpha = .54$, training and instruction $\alpha = .94$, democratic behavior $\alpha = .87$) that explained 43% of the variance. Normative values are presented, which show that an adapted version with acceptable psychometric properties has been obtained, which allows progress in studies on the measurement of leadership in Argentine athletes.

Keywords: leadership, LSS adaptation, psychometric properties, athletes

*Correspondencia a: Lic. Ariela Marzorati. Dirección: Pedemera 275, Universidad de Flores, Facultad de Psicología y Ciencias Sociales, Buenos Aires, Argentina.
E-mail: ariela.marzorati@ufouniversidad.edu.ar

Cómo citar este artículo: Marzorati, A., Pagano, A., E., Caicedo Cavagnis, E., Fiotti, J., A., Fernández, C., Lorusso, L., J., & Löhman, F. (2023): Adaptación de la Escala de Liderazgo en el Deporte (LSS-1) en deportistas argentinos. *Revista Evaluar*, 23(3), 94-111. Recuperado de <https://revistas.unc.edu.ar/index.php/revaluar>

Participaron en la edición de este artículo: Vanesa Mariela Toledo, Gloria Rebeca Nieve, Eugenia Barrionuevo, Florencia Ruiz, Francisco Madema, Jorge Bruera.

Introducción

Históricamente, uno de los temas de interés e intervención de la Psicología Deportiva ha sido el liderazgo. Algunos autores (Barrow, 1977; Northouse 2009, citados en Marcén-Muñio et al., 2016) definen al liderazgo como un proceso grupal donde un individuo ejerce cierta influencia sobre otros con el fin de alcanzar una meta en común. En este intercambio, existen factores que determinarán el impacto de la influencia en procesos grupales como la comunicación, la cohesión y la motivación (Torrado-Quintela, 2012). El liderazgo se configura, entonces, como un proceso interaccional en el cual participan un/a líder, quienes siguen a ese/a líder y los objetivos establecidos (Bohórquez-Gómez-Millán & Checa-Esquiva, 2020).

Con relación a ello, en el deporte, es el/la entrenador/a quien ejerce el rol de líder (Crespo et al., 1994), lo que da cuenta de su personalidad y de las habilidades que tiene al manejar un grupo (Urrea-Tobar, 2015). Por ello, un elemento fundamental para alcanzar el éxito deportivo es la relación que se establece entre el/la deportista y el/la entrenador/a, ya que el ejercicio del liderazgo por parte de este último es esencial para la consecución de los objetivos deportivos (Horn, 1992; Ralston, White & Wilson, 1975, citados en Zhang et al., 1997). Las investigaciones en el ámbito deportivo se han centrado en la figura del líder formal (Torrado-Quintela, 2012) y han tenido diversos enfoques a través del tiempo, por ejemplo: a) características de la personalidad del/de la entrenador/a; b) comportamientos asociados al liderazgo en diversos contextos y c) enfoque interaccionista entre el/la líder y la situación (Marcén-Muñio et al., 2016; Zhang et al., 1997).

Una de las teorías más estudiadas y de mayor consenso (Cox, 2009, citado en Marcén-Muñio et al., 2016) en el estudio del liderazgo ha

sido el Modelo Multidimensional de Liderazgo (Chelladurai & Carron, 1978) que dio lugar a un nuevo enfoque que contempla que la satisfacción de quienes integran un grupo y el rendimiento colectivo estarían relacionados con las características de la situación de los/as atletas y del entrenador/a. En este sentido, la conducta real del/de la líder, la preferida por los/as deportistas y las tareas requeridas por la situación determinarán el rendimiento y la satisfacción de ese grupo (Bohórquez-Gómez-Millán & Checa-Esquiva, 2020). Por esta razón, es ideal la congruencia entre las tres conductas: preferida, requerida y real (Chelladurai & Carron, 1978; Chelladurai, 2012; Jun & Kim, 2011, citados en Marcén-Muñio et al., 2016).

Un modo de evaluar la congruencia entre estas tres conductas ha sido desarrollada por Chelladurai y Saleh (1980) con la Escala de Liderazgo en el Deporte (LSS, por sus siglas en inglés), la cual surge con el objeto de evaluar el modelo propuesto por Chelladurai y Carron (1978) en el deporte. Esta escala posee tres versiones: la que evalúa la conducta preferida por los deportistas (LSS-1), la que evalúa la conducta real del líder percibida por los/as deportistas (LSS-2) y por último, la percepción del entrenador/a respecto de sus propios comportamientos (LSS-3).

Esta escala en sus tres versiones sondea cinco factores relacionados con el ejercicio del liderazgo: un factor relacionado con la tarea (entrenamiento e instrucción), dos factores relacionados con la toma de decisiones (comportamiento democrático y comportamiento autocrático) y dos factores motivacionales (apoyo social y feedback positivo), a partir de 40 ítems con una escala tipo Likert de 5 puntos.

La LSS es útil para “analizar el comportamiento del/de la entrenador/a y su eficacia” (Chelladurai & Saleh, 1980, p.43), ya que proporciona una herramienta cuya estructura factorial se diferencia de otras propuestas que brindan

categorías conceptuales sobre las conductas del entrenador/a, en las cuales cada dimensión es relativamente confiable. Los autores afirman que la administración a diferentes equipos dentro de un mismo deporte podría dar cuenta de las diferencias sobre las percepciones o preferencias relacionadas con el rendimiento y la satisfacción. Por otra parte, distinguir entre el estilo de liderazgo y la esencia de los comportamientos es una herramienta útil para la toma de decisiones en situaciones deportivas, lo que permite identificar aquellos comportamientos que favorecen el rendimiento o la satisfacción (Antunes et al., 1998 citado en Ruiz-Barquín & de la Vega-Marcos, 2015)

La escala de liderazgo en el deporte, ha sido traducida a diferentes idiomas tales como el japonés (Chelladurai et al., 1988), portugués (Serpa et al., 1988), inglés (Zhang et al., 1997) y el castellano (Crespo et al., 1994; Marcén-Muñio et al., 2016). Cabe destacar que, la mayoría de las adaptaciones mencionadas no han podido replicar la estructura factorial original de cinco factores, a excepción de la versión japonesa.

Con implicaciones especiales en el objetivo de este estudio, resulta interesante observar las versiones adaptadas al español (Crespo et al., 1994; Marcén-Muñio et al., 2016). Por una parte, la versión española de Crespo et al. (1994) toma como muestra entrenadores en el ámbito del tenis, lo que adapta la versión LSS-3. Como resultado, los investigadores obtienen un instrumento de 31 ítems compuestos por 4 factores: orientación hacia las relaciones (factor apoyo social de la escala original), orientación hacia la tarea (factores entrenamiento e instrucción y *feedback* positivo del LSS), conducta democrática y conducta autocrática (tal como en la escala original).

Por otra parte, Marcén-Muñio et al. (2016) toman como muestra 207 deportistas y 27 entrenadores de 15 modalidades deportivas pertenecientes al Centro Nacional de Alto Rendimiento

de México a los cuales administran las escalas LSS-2 y LSS-3. En dicha investigación, la estructura factorial de la LSS-2 dio como resultado una solución de dos factores: implicación en el desarrollo del deportista (compuesto por 33 ítems) y conducta autocrática (7 ítems).

Como puede observarse, las investigaciones previas muestran dos características: 1) se han orientado hacia la versión de percepción del/ de la entrenador/a y preferencia de los/as deportistas y 2) se ha trabajado mayormente con población de deportistas universitarios (Fenoy & Campoy, 2012; Álvarez, Castillo & Falcó, 2010, citados en Marcén-Muñio et al., 2016).

Cabe destacar que, como indican Franco et al. (2017), la cultura podría tener una influencia sobre el comportamiento de las variables psicológicas, por lo que es necesario adaptar los instrumentos y sus teorías subyacentes a las poblaciones donde serán aplicados y así evitar la generalización de conocimientos que puedan contener sesgos culturales que afecten su comportamiento.

Por ello, el objetivo del presente estudio es analizar las propiedades psicométricas de la versión de preferencia (LSS-1) de los/las deportistas argentinos/as, comprobar su estructura factorial y propiedades psicométricas en nuestra población, desarrollar en Argentina instrumentos de liderazgo que permitan realizar evaluaciones específicas de sus manifestaciones conductuales y analizar su eficacia en relación con el deporte y con herramientas válidas, fiables (Mikulic, 2007, citado en Pagano & Vizioli, 2020) con la adecuación lingüística apropiada a la cultura meta de la adaptación.

Metodología

Se realizó un estudio instrumental (Montero & León, 2007) del *Leadership in Sport Scale*

(Chelladurai & Saleh, 1980) en deportistas argentinos/as. Se siguieron las normas de la *International Test Commission* (2017) y la *American Educational Research Association* (2014).

Primero se realizó la traducción de la escala siguiendo un proceso de traducción inversa, luego un estudio de jueces para evaluar la validez de contenido de la escala y, finalmente, un estudio de entrevistas cognitivas para evaluar la validez basada en el proceso de respuesta. A continuación, se evaluó la validez factorial y consistencia interna del instrumento.

Este último estudio fue desarrollado de manera *online*, a través de un formulario de Google. Se contactó a clubes, asociaciones, deportistas, entrenadores/as o personas vinculadas a la actividad deportiva a través de redes sociales, correo electrónico y, en algunos casos presencialmente, para invitarles a participar del estudio.

A quienes aceptaron participar, se les envió el link del formulario donde se indicaban tanto las condiciones de participación como la duración aproximada del estudio (20 minutos). En estas condiciones, se expresó que la participación era anónima, confidencial, voluntaria y que podían abandonar el estudio en cualquiera de sus fases, sin que ello ocasionara perjuicio alguno, para lo cual las personas daban el consentimiento de su voluntad de participación. A lo largo del estudio, se siguieron las normas éticas para la investigación en humanos de la *American Psychological Association* (APA, 2017).

Participantes

Mediante un muestreo de tipo auto-elegido (Romero & Bologna, 2011), participaron 333 deportistas de 12 provincias de Argentina (Tabla 1), mientras que Buenos Aires y Córdoba aportaron la mayor cantidad de participantes (76% -253- y

17.1% -57- respectivamente). El rango etario de la muestra osciló entre los 16 y 63 años de edad ($M = 26,37$ $DS = 8,51$). El 53% (179) de quienes participaron se autopercebieron como mujer, 45.3% como varón, 0.3% (1) como No Binario y 0.6% (2) como Otros.

Del total de la muestra, el 61.9% (206) trabajan y de estos solo un 14.1% (47) trabaja en relación con el deporte o ejercicio físico; el 29.1% (97) no trabaja, un 1.2% (4) se quedó sin empleo a partir de la pandemia y un 7.8% (26) recibe remuneración económica por su actividad deportiva (beca, sponsors, etc.).

En relación con el deporte (Tabla 2), 236 deportistas (70.9%) pertenecen a alguna Federación, mientras que 97 (29.1%) no. Los días de entrenamiento semanal oscilan entre 1 y 7 días a la semana ($M = 4,08$ $DS = 1,40$). Cada entrenamiento tiene una duración promedio de 117,13 minutos ($DS = 54,02$). Compiten a nivel provincial/metropolitano 188 (56.5%) personas, a nivel nacional 78 (23.4%), internacional 23 (6.9%) y no compiten/no poseen competencias 44 (13.2%). Al respecto, 62 deportistas (18.6%) reportaron no haber retomado las competencias desde la interrupción

Tabla 1
Frecuencia de participantes por Provincia.

Provincia	Frecuencia	Porcentaje
Buenos Aires	253	76,0
La Pampa	3	,9
Mendoza	2	,6
Neuquén	1	,3
Santa Cruz	1	,3
Catamarca	1	,3
Santa Fe	9	2,7
Tucumán	2	,6
Córdoba	57	17,1
Corrientes	2	,6
Entre Ríos	1	,3
Jujuy	1	,3

Nota. $n = 333$.

Tabla 2
Frecuencia de participantes por deporte.

Deporte	Frecuencia	Porcentaje
Atletismo	24	7,2
Natación	13	3,9
Rugby	14	4,2
Básquet	17	5,1
Vóley	27	8,1
Otros	23	6,9
Deportes de Combate	27	8,1
Deportes Artísticos y Acrobáticos	15	4,5
Fútbol	62	18,6
Futsal	29	8,7
Handball	47	14,1
Hockey sobre césped	35	10,5

Nota. n = 333. Deportes de combate: *tae kwon do*, *muay thai*, *karate*, *jiu jitsu*, *mma*, *karate*, *boxeo*, *judo* y *sambo*. Deportes Acrobáticos y Artísticos: *gimnasia artística*, *patín artístico*, *acrobacia*, *pole sport*.

de las actividades deportivas a causa de la pandemia por COVID-19. Cabe destacar que, del total de la muestra, solo 80 deportistas (24%) trabajan para su preparación deportiva con psicólogo/a del deporte, mientras que 253 (76%) no.

Instrumentos

Planilla de datos socio-demográficos y deportivos. Se indagaron una serie de dimensiones socio-demográficas y deportivas acorde con los análisis por realizar y que pudieran estar vinculados a la variable en estudio. Así, se exploró: identidad de género, edad, si actualmente las personas poseían trabajo o no y provincia de origen. A nivel deportivo: días y tiempo de entrenamiento, deporte de práctica, nivel de competición actual y si actualmente utiliza servicios de psicología del deporte.

Leadership in Sport Scale. Se utilizó la versión de la escala de Chelladurai y Saleh (1980), la cual es una versión revisada de la escala original de

1978 desarrollada por estos mismos autores y que no comprobaron buenas evidencias de validez y fiabilidad (Chelladurai & Saleh, 1978). Esta escala posee 40 ítems subdivididos en 5 dimensiones: 1) Entrenamiento e instrucción (13 ítems), refleja una de las funciones importantes de un/a entrenador/a para mejorar el nivel de rendimiento del/de la deportista. El entrenador/a entrena e instruye al/a la atleta para ayudarles a alcanzar su máximo potencial físico. También se espera que él o ella les instruya sobre cómo adquirir las habilidades necesarias y enseñarles las técnicas y las tácticas de los deportes. Además, en el caso de los deportes de equipo, el/la entrenador/a coordina las actividades de los/as integrantes del equipo; 2) Comportamiento democrático (9 ítems), refleja hasta qué punto el/la entrenador/a permite la participación de los/as atletas en la toma de decisiones. Estas decisiones pueden relacionarse con el establecimiento de metas grupales o las formas en que estas metas se deben alcanzar; 3) Comportamiento autocrático (5 ítems), indica la medida en que un/a entrenador/a se mantiene alejado/a de los/as atletas y enfatiza su autoridad al tratar con ellos/as. En tales situaciones se esperaría que el entrenador/a exigiera el estricto cumplimiento de sus decisiones; 4) Apoyo social (8 ítems), grado en que el/la entrenador/a está involucrado/a en la satisfacción de las necesidades interpersonales de sus atletas. El comportamiento del/de la entrenador/a puede satisfacer directamente tales necesidades o puede crear un clima en el que los/as integrantes satisfagan mutuamente sus necesidades interpersonales. Cabe señalar que el apoyo social se proporciona independientemente del desempeño de los/as integrantes; y 5) *Feedback* positivo (5 ítems), grado en que el/la entrenador/a expresa su agradecimiento y felicitación a los/as atletas por su desempeño y contribución en una competencia.

Las personas deben responder en una escala tipo Likert de 5 puntos (1 *Nunca* a 5 *Siempre*)

las preferencias personales respecto de comportamientos específicos que puede tener un/a entrenador/a. A nivel psicométrico, los indicadores de consistencia interna de cada dimensión fueron de .45 (comportamiento autocrático) a .93 (entrenamiento e instrucción), .71 a .82 *test-retest*, estabilidad de las dimensiones a lo largo de las muestras (validez factorial), validez de contenido.

Análisis de datos

En relación con los análisis de datos, una vez realizada la prueba piloto se procedió con el análisis de ítems, para esto se utilizó el software SPSS 24.0. Se evaluaron valores perdidos y se identificaron los casos atípicos. Se calcularon puntuaciones *z* para cada variable respecto de los univariados (atípicos = $z > \pm 3.29$). Luego, se utilizó distancia de Mahalanobis (*D*) para los multivariados ($p < .01$). Consecutivamente, se evaluó la distribución de las puntuaciones mediante los índices de asimetría y curtosis. Los índices de asimetría y curtosis entre ± 1.00 se consideraron excelentes y los valores inferiores a ± 2.00 , adecuados (George & Mallery, 2013).

Una vez realizado este proceso, para analizar el poder de discriminación de los reactivos, se utilizaron los métodos de comparación entre extremos y el índice de homogeneidad (*r*) (Muñiz, 2005). Se dividió la muestra en cuartiles respecto de las puntuaciones en cada una de las cinco dimensiones propuestas en el modelo original del LSS (Chelladurai & Saleh, 1980). Luego, se comparó el cuartil 1 y 4 según el valor obtenido en cada uno de los ítems. Para esto, se utilizó el estadístico *t* de Student y *U* de Mann Whitney según el comportamiento de los ítems cumpliera o no con el supuesto de normalidad. Para calcular el índice de homogeneidad, (*r*) se calculó la correlación ítem/dimensión. Muñiz (2005) expone que, si un

ítem discrimina correctamente entre las personas que puntúan bajo y alto, como consecuencia, debería existir una correlación positiva entre la puntuación obtenida en el ítem y la del *test*, o de su dimensión correspondiente.

Para calcularla validez factorial del LSS, se realizó un Análisis Factorial Confirmatorio (AFC) y se evaluó el ajuste de los datos al modelo inicial de cinco factores propuesto por Chelladurai y Saleh (1980). El AFC se realizó utilizando el software Mplus 7.11 (Muthén & Muthén, 2012). Para esto, se utilizó el método mínimos cuadrados no ponderados (*Unweighted Least Squares*, ULS, Jöreskog, 1977). Se consideraron los siguientes índices de bondad de ajuste: χ^2 dividido por los grados de libertad (valores $\leq 5,0$ indican un buen ajuste); índice Tucker-Lewis (TLI); y *Comparative Fit Index* (CFI) con valores mayores o iguales a .90 y valores menores o iguales a .06 en *Root Mean Square Error of Approximation* (RMSEA, Kline, 2011; Schumacker & Lomax, 2016). Por último, se consideraron adecuadas las cargas factoriales mayores a .30 (Nunnally & Bernstein, 1994).

Debido a que el modelo no tuvo convergencia con las 5 dimensiones propuestas por los autores, se realizaron dos Análisis Factoriales Exploratorios (AFE) con el programa Factor versión 9.2 (Lorenzo-Seva & Ferrando, 2006). Para esto, se utilizó el método de extracción mínimos cuadrados no ponderados (ULS), con rotación *promin*, mientras que los factores fueron determinados por el método de análisis paralelo recomendado por Timmerman y Lorenzo-Seva (2011). Para evaluar la pertinencia a realizar el AFE, se utilizó el índice de adecuación muestral Kaiser-Meyer-Olkin (KMO), que consideró adecuado un valor superior a .6 (Tabachnick & Fidell, 1989), y la prueba de esfericidad de Bartlett ($p < .01$). Se consideraron adecuadas las cargas factoriales mayores a .30 (Nunnally & Bernstein, 1994).

Para evaluar la fiabilidad del instrumento, se utilizó el programa *R* versión 3.6.0, mediante el cual se obtuvieron los coeficientes alfa ordinal de los tres factores finales. Se evaluaron los coeficientes siguiendo el criterio de [George y Mallery \(2003\)](#): puntuaciones mayores a $\geq .9$ se consideraron *excelentes*, $\geq .8$ *buenas*, $\geq .7$ *aceptables*, $\geq .6$ *cuestionables*, $\geq .5$ *malas* y $\leq .5$ *inaceptables*. Por último, a través del programa SPSS 24.0 se obtuvieron percentiles de la muestra para poder brindar valores normativos.

Resultados

Se realizaron dos traducciones del instrumento del inglés al español y se integraron posteriormente ambas traducciones en una única versión. Sucesivamente, se tradujo la versión de español al inglés y se contactó al autor original de la técnica para que evalúe si el contenido conceptual de los ítems se mantenía en la versión retro-traducida.

La evaluación de jueces de la traducción en términos de claridad y congruencia encontró que los índices de Kappa oscilaron entre 0.73 y 0.82 (entrenamiento e instrucción = .73, comportamiento democrático = .82, comportamiento autocrático = .74 apoyo social = .82 *Feedback* positivo = .73), acuerdos entre moderados y fuertes para todos los factores ([Mc Hugh, 2012](#)). Sobre la base de las observaciones realizadas por los/as jueces y, a partir de los desacuerdos respecto de la congruencia de los ítems con el factor, se realizó el proceso de entrevistas cognitivas. Durante las entrevistas se preguntó por los ítems 4, 7, 8, 12, 22, 32 y 33.

En función de los datos obtenidos, se decidió modificar el término atleta por deportista en todos los ítems por ser más adecuado a la cultura meta. Por otro lado, en el ítem 4, se realizó una modificación: “*se asegure que su papel en el equipo sea entendido por todos los atletas*” fue modificado por

“*se asegure que su propio rol en el equipo sea entendido por todos los deportistas*”. Por su parte, el ítem 12 “*explique cómo la contribución de cada atleta encaja en el cuadro final*” se modificó por “*explique cómo la contribución de cada deportista encaja en el funcionamiento general del equipo*”.

En lo referente al análisis de ítems, la evaluación del comportamiento de los reactivos indicó la presencia de 44 casos atípicos que se decidió conservar, ya que no generaron distorsiones en los análisis estadísticos ([Hair et al., 1999](#)). Respecto de las pruebas de asimetría y curtosis, estas dan cuenta de que 6 de los 40 reactivos no se comportan conforme al modelo normal (5, 8, 11, 17, 34, 10).

En cuanto al método de comparación entre extremos, sobre los resultados de la comparación de los ítems según el cuartil 1 y el cuartil 4 obtenidos por la puntuación total de cada caso, se evidencia que las diferencias fueron significativas en todos los casos $p < .01$ ($\alpha = .01$).

La última prueba realizada para analizar los ítems correspondió a la correlación ítem-test, también denominado índice de homogeneidad (r). En esta línea, de los resultados obtenidos se observa que los valores de correlación ítem/dimensión son adecuados $> .30$ ([Muñiz, 2005](#)). Sin embargo, los ítems 6, 27, 34, 40 obtuvieron valores de correlación ítem/dimensión a $< .30$; estos 4 reactivos corresponden al factor comportamiento autocrático. Si bien la comparación entre extremos aporta evidencia del poder de discriminación del reactivo, esta falta de correlación podría dar cuenta de problemas de consistencia interna en la escala que deban ser abordados.

Validez estructural

Para [Chelladurai \(1990\)](#) la evaluación de las cinco dimensiones de la LSS debe ser entendida bajo un análisis factorial confirmatorio. Siguiendo

esta sugerencia, se probó la estructura factorial original de cinco factores propuesta por los autores. Los resultados del AFC dieron cuenta de que el modelo tuvo convergencia en una solución factorial esperada y excedió el número de iteraciones como sucedió en otras investigaciones realizadas (Serpa et al., 1988; Zhang et al., 1997; Crespo et al, 1994; Ruiz-Barquín & de la Vega-Marcos, 2015; Marcén-Muñio et al., 2016). Debido a esto, se procedió a evaluar la estructura factorial mediante un AFE.

En primer lugar, a partir del método de extracción de factores de análisis paralelo se evidenció la presencia de tres factores (factor 1= comportamiento autocrático; factor 2 = entrenamiento e instrucción; factor 3 = comportamiento democrático) que explican el 43% de la varianza. Posteriormente, para evaluar la adecuación de la muestra al análisis propuesto, se obtuvo en la prueba de esfericidad de Bartlett un Chi cuadrado de 3659.2 ($p < .001$) que daba cuenta de que la matriz de correlación se diferencia de forma significativa respecto de la matriz de unidad. Con respecto a la medida de adecuación, se obtuvo un valor KMO = .846 (adecuado). Luego, se procedió a realizar el AFE con rotación promin, en el mismo se observaron cargas factoriales $>.30$ en todos los ítems, excepto en el ítem 33 “*Deje a los deportistas trabajar a su propio paso*” (.230) y en el ítem 16 “*Vea que un deportista sea recompensado por buen desempeño*”, que obtuvo una doble carga factorial para los factores 1 y 3.

Se procedió a realizar un segundo AFE y se descartaron los ítems 16 y 33. En la prueba de esfericidad de Bartlett, se obtuvo un Chi cuadrado de 3666.9 ($p < .001$) y un valor KMO = .848 (adecuado). En la tabla 3, se muestran las cargas factoriales de la versión final del instrumento con una estructura de 3 factores que explican el 44% de la varianza (factor 1= comportamiento autocrático 30% de la varianza; factor 2 = entrenamiento

Tabla 3

Cargas Factoriales del Análisis Factorial Exploratorio Final.

Ítem	F1	F2	F3
1		.735	
5		.648	
8		.790	
10		.514	
11		.781	
14		.567	
17		.659	
20		.770	
23		.616	
25		.676	
26		.766	
28		.680	
29		.699	
32		.790	
35		.689	
37		.510	
38		.665	
2			.533
3			.573
4			.320
7			.518
9			.448
13			.550
15			.750
18			.627
19			.658
21			.493
22			.581
24			.379
30			.671
31			.491
36			.507
39			.404
6	.345		
12	.501		
27	.510		
34	.545		
40	.469		

Nota. F1= comportamiento autocrático; F2 = entrenamiento e instrucción; F3 = comportamiento democrático.

e instrucción 8% de la varianza; factor 3 = comportamiento democrático 6% de la varianza).

Por otra parte se observó una correlación entre el factor 1 y 3 de $-.075$, entre el factor 1 y 2 de $-.168$ y entre el factor 2 y 3 de $.622$.

Análisis de fiabilidad

Para evaluar la consistencia interna del LSS, se procedió a realizar análisis de alfa ordinal. Respecto del factor entrenamiento e instrucción, presentó un alfa ordinal $.94$ (*excelente*), el

comportamiento democrático evidenció un alfa ordinal de $.87$ (*bueno*) y el comportamiento autocrático obtuvo un alfa ordinal de $.57$ (*malo*). Por último, en la Tabla 4 se presentan los valores normativos de la muestra utilizada para evaluar las propiedades psicométricas.

Valores obtenidos por los/as deportistas en la población argentina

Presentan una elevada preferencia por conductas orientadas hacia el entrenamiento e instrucción ($M = 4,20$), seguido por el comportamiento

Tabla 4
Valores normativos LSS versión adaptada a población argentina.

Percentil	Comportamiento autocrático	Entrenamiento e instrucción	Comportamiento democrático
5	6	50	36
10	6	58	42
15	7	63	43
20	8	64	45
25	8	66	47
30	9	68	48
35	9	69	49
40	10	71	51
45	10	73	52
50	11	74	54
55	11	75	56
60	12	76	57
65	12	77	58
70	12	78	60
75	13	79	61
80	14	80	62
85	14	82	64
90	15	83	66
95	16	84	69
N	333	333	333
M	10,80	71,51	53,44
DE	3,31	10,71	10,18
PP	2,16	4,20	3,34
MPP	2,16	4,20	3,34

Nota. M = Media; DE = Desvío estándar; PP = Puntaje promedio de la dimensión; MPP = Puntaje promedio de la escala.

democrático ($M = 3,34$) y por último, las conductas autocráticas ($M = 2,16$).

Las conductas preferidas por los/as deportistas respecto de la figura del/ de la entrenador/a se expresan en la Tabla 5. Dentro de ellas, podemos destacar: “*Explique a cada deportista las técnicas y tácticas del deporte*” ($M = 4,52$; $SD = 0,897$), “*Planee por adelantado lo que se debe hacer*” ($M = 4,47$; $SD = 0,87$), “*Preste particular atención a corregir los errores de los deportistas*” ($M = 4,43$; $SD = 0,921$), “*Anime a los deportistas a confiar en sí mismos*” ($M = 4,41$; $SD = 0,844$) y “*Se asegure que su propio rol en el equipo sea entendido por todos los deportistas*” ($M = 4,38$; $SD = 0,925$) pertenecientes al Factor 2 del presente estudio. Por otra parte, aquellos comportamientos infravalorados por los/as deportistas han sido: “*Se distancie del resto*” ($M = 1,65$; $SD = 0,932$); “*Se niegue a comprometerse en un punto*” ($M = 1,98$; $SD = 1,092$); “*Invite a los deportistas a su casa*” ($M = 2,01$; $SD = 1,143$); “*No explique su accionar*” ($M = 2,01$; $SD = 1,156$); “*Hable de tal manera que no se lo pueda cuestionar*” ($M = 2,17$; $SD = 1,319$) y “*Haga favores personales a los deportistas*” ($M = 2,18$; $SD = 1,08$), cuatro de ellos pertenecientes al factor de comportamiento autocrático y dos al factor de comportamiento democrático.

Adicionalmente a los estudios psicométricos, se indagaron diferencias individuales según género y modalidad deportiva. Se encontraron diferencias significativas en el factor comportamiento autocrático en función al sexo de los/as deportistas ($M_{\text{mujeres}} = 10,08$ $M_{\text{varones}} = 11,66$; $t(328) = -4,40$; $p < .05$). En lo que respecta a la modalidad deportiva, los/as deportistas de modalidades individuales presentan puntuaciones significativamente mayores en los factores entrenamiento e instrucción ($M_{\text{individual}} = 73,92$ $M_{\text{equipo}} = 70,27$; $t(302) = 2,49$) y comportamiento democrático ($M_{\text{individual}} = 56,02$ $M_{\text{equipo}} = 52,30$; $t(302) = 2,75$) en comparación con los/as deportistas de modalidades de equipo (Tabla 6).

Discusión

En el presente estudio, se realizó la traducción de la Escala de Liderazgo en el Deporte (LSS-1) de [Chelladurai y Saleh \(1980\)](#) al castellano-argentino y se han analizado sus propiedades psicométricas para la población deportiva argentina.

La escala no convergió con la estructura factorial de 5 factores propuesta por los autores lo que coincide con estudios previos realizados en contextos culturales y poblaciones diferentes: la estructura factorial original de la escala no se mantiene ([Crespo et al., 1994](#); [Ruiz-Barquín, 2007](#); [Ruiz-Barquín & de La Vega-Marcos, 2015](#); [Marcén-Muñio et al., 2016](#)).

Por esta misma razón, se procedió a analizar los datos mediante un AFE. Los resultados dan cuenta de que la escala de liderazgo en el deporte en su versión adaptada a Argentina presenta una estructura factorial compuesta por tres factores: comportamiento autocrático, entrenamiento e instrucción y comportamiento democrático. A su vez, han sido descartados 2 ítems por baja carga factorial (16 y 33 correspondientes a la escala de *feedback* positivo y comportamiento democrático). La estructura resultante se asemeja más a la agrupación de dos factores obtenida por [Marcén-Muñio et al. \(2016\)](#) quienes los denominaron: implicaciones en el desarrollo del deportista, compuesto por 33 ítems, y comportamiento autocrático, compuesto por 7 ítems de la escala original.

La interpretación del contenido de las escalas, queda determinado de la siguiente manera: Factor I, comportamiento autocrático, formado por 5 ítems de la escala original, que es el único factor que no sufrió modificaciones; el Factor II, entrenamiento e instrucción, formado por 17 ítems que en la escala original se distribuyen de la siguiente manera: 13 pertenecen al factor “entrenamiento e instrucción”, 3 pertenecen al fac-

Tabla 5
Estadísticos descriptivos ítems versión LSS-1 versión original.

Ítem	Media	Desvío Estándar
1. EI. Se fije que cada deportista esté trabajando dentro de su capacidad.	4,21	1
2. CD. Pida la opinión de los deportistas sobre estrategias para competencias específicas.	3,59	1,18
3. SS. Ayude a los deportistas con sus problemas personales.	3,5	1,22
4. FP. Felicite a un deportista por su desempeño en frente a otros.	4,02	1,16
5. EI. Explique a cada deportista las técnicas y tácticas del deporte.	4,52	0,9
6. CA. Trabaje relativamente independiente de los deportistas.	2,99	1,33
7. SS. Ayude a los miembros del grupo a solucionar sus conflictos.	3,72	1,18
8. EI. Preste particular atención a corregir los errores de los deportistas.	4,43	0,92
9. CD. Consiga la aprobación del grupo en temas importantes antes de seguir adelante.	3,54	1,23
10. FP. Le diga a un deportista cuando hace un trabajo particularmente bueno.	4,35	0,94
11. EI. Se asegure que su propio rol en el equipo sea entendido por todos los deportistas.	4,38	0,93
12. CA. No explique su accionar.	2,01	1,16
13. SS. Se preocupe por el bienestar personal de los deportistas.	4,19	1,01
14. EI. Instruya a cada deportista individualmente en las habilidades del deporte.	4,23	0,95
15. CD. Incluya a los deportistas en la toma de decisiones.	3,61	1,19
16. FP. Vea que un deportista sea recompensado por buen desempeño.	3,71	1,16
17. EI. Planee por adelantado lo que se debe hacer.	4,47	0,87
18. CD. Incentive a los deportistas a hacer sugerencias en la manera de llevar a cabo las prácticas.	3,74	1,17
19. SS. Haga favores personales a los deportistas.	2,18	1,08
20. EI. Explique a cada deportista lo que debe y no debe hacer en su práctica deportiva.	4,16	1,06
21. CD. Deje que el grupo ponga sus propias metas.	3,27	1,23
22. SS. Expresé el afecto que tiene por sus deportistas.	3,41	1,14
23. EI. Espere que cada deportista realice su tarea hasta el último detalle.	3,89	1,09
24. CD. Deje a los deportistas intentar hacer las cosas a su manera, incluso cuando se equivocan.	2,99	1,21
25. SS. Anime a los deportistas a confiar en él.	4,41	0,84
26. EI. Señale las fortalezas y debilidades de cada deportista.	4,17	1,04
27. CA. Se niegue a comprometerse en un punto.	1,98	1,09
28. FP. Reconozca cuando un deportista se desempeña bien.	4,31	0,85
29. EI. Dé instrucciones específicas a cada deportista sobre qué hacer en cada situación.	4,17	1,03
30. CD. Pida la opinión de los deportistas sobre temas de entrenamiento importantes.	3,56	1,26
31. SS. Fomente relaciones cercanas e informales con deportistas.	3,02	1,19
32. EI. Vea que los esfuerzos sean coordinados.	4,09	0,93
33. CD. Deje a los deportistas trabajar a su propio paso.	3,17	1,14
34. CA. Se distancie del resto.	1,65	0,93
35. EI. Explique cómo la contribución de cada deportista encaja en el funcionamiento general del equipo.	4,01	1,09
36. SS. Invite a los deportistas a su casa.	2,01	1,14
37. FP. Dé crédito cuando sea merecido.	4	1,09
38. EI. Especifique en detalle qué se espera de cada deportista.	3,71	1,15
39. CD. Deje que los deportistas decidan las jugadas a usar en el partido.	3,08	1,15
40. CA. Hable de manera tal que no se lo pueda cuestionar.	2,17	1,32

Tabla 6
Comparación de medias según Género y Modalidad deportiva.

	Género				<i>p</i>	Modalidad deportiva				<i>p</i>
	Varón		Mujer			Individual		Equipo		
	N=179		N=151			N=73		N=231		
	M	SD	M	SD		M	SD	M	SD	
Comportamiento autocrático	11,66	3,02	10,08	3,37	0,000*	10,64	3,36	10,92	3,29	0,531
Entrenamiento e instrucción	71,58	10,48	71,49	10,93	0,939	73,92	9,20	70,27	11,36	0,013
Comportamiento democrático	54,43	10,17	52,65	10,18	0,115	56,03	9,73	52,30	10,23	0,006*

Nota. * nivel de significación $p < .05$

tor “*feedback positivo*” y 1 pertenece al factor de “apoyo social”. Por último al Factor III, comportamiento democrático, formado por 8 ítems de la escala original, se añaden 7 ítems de “apoyo social” y un ítem de “*feedback positivo*”.

En relación con los resultados obtenidos, podemos observar que existiría una congruencia a nivel teórico respecto de las dimensiones de entrenamiento e instrucción y *feedback positivo*. Se podría asumir que, brindar *feedback* al atleta sobre su desempeño en entrenamientos formaría parte de conductas de liderazgo relacionadas con brindar instrucción al deportista, mediante las cuales el entrenador/a ayuda a los/as deportistas a alcanzar su máximo potencial a nivel físico, técnico, táctico y desarrollo de habilidades. Por otra parte, la escala de *feedback positivo* hace referencia a la conducta que mantiene un/a entrenador/a al comunicar a los/as deportistas la correcta realización de los ejercicios y ejecuciones solicitadas. En ese sentido, ambas dimensiones forman parte de un estilo de liderazgo que se caracteriza por la implicación en la tarea. Brindar instrucciones precisas respecto de cómo adquirir habilidades físicas, técnicas y tácticas, comunicar el progreso que el/la deportista va consiguiendo, a partir del *feedback positivo*, son comportamientos que tienden a generar un clima de trabajo que se orien-

ta hacia la maestría (Duda, 1989; Maehr, 1984; Nicholls, 1984).

En ese sentido, hallamos deportistas que prefieren un estilo de entrenador/a que valore el esfuerzo y la mejora y que considere que el rol de cada atleta es importante para conseguir el éxito y que valore la importancia a la cooperación entre quienes integran un equipo (Newton et al., 2000).

A su vez, se ha evidenciado en el presente estudio que este tipo de conductas son preferidas por los/as deportistas argentinos y que esto coincide con estudios que destacan que los/as deportistas presentan mayor preferencia por los estilos de liderazgo orientados al entrenamiento e instrucción y al *feedback positivo* (Antunes et al., 1998, citado en Ruiz-Barquín & de la Vega-Marcos, 2015; Walach-Bista, 2013).

Por otro lado, a nivel teórico sucede algo similar con las dimensiones de comportamiento democrático y apoyo social. La dimensión de comportamiento democrático refleja la preferencia por actitudes del/de la entrenador/a que permitan la participación de los/as deportistas en la toma de decisiones, mientras que la dimensión de apoyo social hace referencia a la preferencia por el entrenador/a que se involucre de forma interpersonal con sus deportistas. La aparición de un único factor que agrupe estas dimensiones indica

la preferencia del/la deportista por el entrenador/a que se aproxime para establecer un vínculo y le permita su participación en la toma de decisiones.

No obstante, es relevante informar que, los ítems que hacen referencia a una preferencia a que el entrenador/a establezca un vínculo muy cercano con el/la deportista son los que menor puntuación han obtenido en la muestra de referencia (ítem 19 “*haga favores personales a los deportistas*” $M = 2,18$ $DE = 1,08$; ítem 36 “*invite a los deportistas a su casa*” $M = 2,01$ $DE = 1,14$).

El presente estudio evidencia que los/as deportistas argentinos han presentado mayor preferencia por el factor de entrenamiento e instrucción y comportamiento democrático que por comportamientos autocráticos. En este sentido, los resultados indican una asociación positiva entre el factor 2 (entrenamiento e instrucción) orientado hacia la tarea y el 3 (comportamiento democrático) que hace referencia a un estilo de decisión por parte del entrenador (Chelladurai, 1990). La satisfacción de los/as integrantes de un equipo estaría dada por un ambiente cálido y participativo donde el líder brinde la posibilidad de participación en la toma de decisiones, el reconocimiento de los esfuerzos y la capacitación e instrucción para fomentar una relación entre el aprendizaje, el esfuerzo y el desempeño para mejorar el rendimiento del equipo. Así, la asociación entre estas variables se puede explicar a partir de que ambas preferencias implican actitudes de apertura al diálogo por parte de los/as entrenadores/as con el equipo, lo que vuelve esperable que una conducta vinculada a brindar *feedback* e instrucción sobre las ejecuciones implique mayor participación de los/as deportistas en la comprensión y sentido de las tareas por realizar.

Por otra parte, es sabido que los estilos de liderazgo de entrenamiento e instrucción y comportamiento democrático, *feedback* positivo y apoyo social están asociados a un clima motivacional

orientado hacia la cohesión grupal y hacia la tarea (Leo-Marcos et al., 2013, citado en Marcén-Muñío et al., 2016). En ese sentido, Losada et al. 2012 (citado en Marcén-Muñío et al., 2016) hallaron que los deportistas individuales tienen una preferencia por este tipo de comportamientos en mayor medida que los deportistas de equipo.

Según la presente investigación, las diferencias en la preferencia se evidencian únicamente en la variable comportamiento democrático, a favor de deportistas de modalidad individual. Tal como lo refieren Marcén-Muñío et al. (2016), este estilo de liderazgo da lugar a una mayor confianza mutua y lazos afectivos entre entrenador/a-deportista. Se podría hipotetizar que en las modalidades individuales, el vínculo entrenador/a-deportista posee propiedades socioafectivas muy específicas en comparación con el que se pone en juego en deportes de equipo, ya que es un vínculo de mucha más intimidad, escucha y trabajo conjunto, que probablemente afecten esa preferencia a favor de un liderazgo democrático que le dé lugar a la opinión y percepción del/la deportista en la toma de decisiones, en comparación con lo que ocurre en los deportes de equipo.

Por otra parte, a diferencia de los estudios precedentes (Marcén-Muñío et al., 2016), en el estudio actual existen diferencias significativas en cuanto al género, ya que los varones tienen una mayor preferencia por conductas autocráticas en los entrenadores que las mujeres. Esto podría deberse a una construcción social del género (Bonilla, 1998; García, 1999, citados en Hernández-Mendo & Canto-Ortiz, 2003) donde las mujeres se encontrarían orientadas hacia un estilo más democrático y las relaciones sociales, mientras que los hombres estarían orientados hacia la tarea y un estilo de liderazgo más autoritario (Hernández-Mendo & Canto-Ortiz, 2003).

Por último, el factor de comportamiento autocrático, en términos de orientación a las relacio-

nes marca conductas que establecen una distancia incluso a nivel interpersonal (ej.: “se distancie del resto”; “se niegue a comprometerse en un punto”). Es este factor el de menor preferencia en la muestra de referencia. Chelladurai 2012 (citado en Marcén-Muñío et al., 2016) refiere que este tipo de conductas adoptadas por el líder están asociadas a una baja satisfacción de los deportistas.

En cuanto a la fiabilidad obtenida, el factor de comportamiento autocrático obtuvo un alfa ordinal de .57. Chelladurai (1990) señala que la consistencia interna de la escala de comportamiento autocrático en la versión de preferencia de los deportistas suele ser baja. Cabe destacar que, este factor es el único que presenta ítems enunciados con una connotación negativa, se podría pensar que el comportamiento autocrático estaría definido por la ausencia de conductas de apoyo y cercanía interpersonal hacia los deportistas. A su vez, esta escala está compuesta por una baja cantidad de reactivos, motivo que podría afectar su fiabilidad.

De acuerdo con estos resultados y los obtenidos en estudios anteriores sobre la fiabilidad de la dimensión, sería adecuado, en un futuro, reestructurarla agregando ítems que abarquen en mayor proporción el universo de comportamientos autocráticos que un/a entrenador/a puede presentar. Con relación a ello, Chelladurai (1990) destaca que los ítems de la LSS se han construido a través de escalas relacionadas con el comercio y la industria y sugiere la creación de ítems significativos para las subescalas existentes en la población deportiva.

Entre algunas de las aportaciones prácticas de este estudio, se encuentran, por un lado, el estudio del comportamiento de la LSS en un contexto deportivo y cultural específico, con lo cual la herramienta puede aplicarse a poblaciones similares a la aquí estudiada. Por otra parte, es necesario recordar que los estudios empíricos de

la LSS son escasos (Marcén-Muñío et al., 2016); por lo tanto, aquí radica la importancia de validar la escala en sus tres versiones para crear programas de intervención tomando como premisa la congruencia entre las preferencias de los/as deportistas, la conducta percibida y la conducta real del/ de la entrenador/a en. Futuras investigaciones deberían contemplar que los resultados en las tres escalas pueden variar en función del género, el nivel competitivo (Ruiz-Barquín & de la Vega-Marcos, 2015) o el tipo de deporte (Losada et al., 2012, citado en Marcén-Muñío et al., 2016). En ese sentido, entre las limitaciones presentes en esta investigación, se destaca que las interpretaciones de los resultados obtenidos respecto de las comparaciones según género deben ser asumidas con cautela, debido a que no se han realizado estudios de invarianza factorial según género.

Por otra parte, se considera que se debe profundizar el estudio de la estructura factorial del instrumento en distintas muestras para obtener resultados que permitan generalizar la validez del instrumento en la población de referencia. No obstante, se destaca el valor de la presente investigación para avanzar con estudios psicométricos en el contexto deportivo y argentino, ya que los procesos de validación de un instrumento son continuos (Richaud de Minzi, 2008).

Agradecimientos

A todos los/as deportistas y entrenadores/as que hicieron posible la realización de este estudio.

Referencias

- American Educational Research Association. (2014). *Validity*. En *Standards for educational and psychological testing*, pp.11-31. Washington

- American Psychological Association. (2017). Ethical principles of psychologist and code of conduct. Recuperado de: <https://www.apa.org/ethics/code>
- Bohórquez-Gómez-Millán, M. R., & Checa-Esquiva, I. (2020). Escala de Liderazgo en entrenadores de fútbol: Estudio psicométrico y descriptivo. *Informació Psicològica*, 119, 65-77. <https://doi.org/10.14635/IPSIC.2020.119.3>
- Chelladurai, P. & Carron, A. V. (1978). Leadership. Ottawa, Canada: Sociology of sport Monograph Series, Canadian Association for Health, Physical Education and Recreation.
- Chelladurai, P., & Saleh, S. D. (1978). Preferred leadership in sports. *Canadian Journal of Applied Sport Sciences*, 3, 85-92.
- Chelladurai, P., & Saleh, S. D. (1980). Dimensions of leader behavior in sport: Development of a leadership scale. *Journal of Sport and Exercise Psychology*, 2(1), 34-45. <https://doi.org/10.1123/jsep.2.1.34>
- Chelladurai, P., Imamura, H., Yamaguchi, Y., Oinuma, Y., & Miyauchi, T. (1988). Sport leadership in a cross-national setting: The case of Japanese and Canadian university athletes. *Journal of Sport and Exercise Psychology*, 10(4), 374-389. <https://doi.org/10.1123/jsep.10.4.374>
- Chelladurai, P. (1990). Leadership in sport: A review. *International Journal of Sport Psychology*, 21(4), 328-354.
- Crespo, M., Balaguer, I., & Atienza, F. (1994). Análisis psicométrico de la versión española de la Escala de Liderazgo en el Deporte de Chelladurai y Saleh en la versión entrenadores. *Revista de Psicología Social Aplicada*, 4(1), 5-28.
- Duda, J. L. (1989). Relationship between task and ego orientation and the perceived purpose of sport among high school athletes. *Journal of Sport and Exercise Psychology*, 11(3), 318-335. <https://doi.org/10.1123/jsep.11.3.318>
- Franco, E., Coterón, J., Gómez, V., & Brito, J. (2017). Propiedades psicométricas de la BPNES adaptada a la Educación Física en Argentina, España y Colombia. *Revista Evaluar*, 17(2), 97-112. <https://doi.org/10.35670/1667-4545.v17.n2.18724>
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4^{ta} ed.). Allyn & Bacon.
- George, D., & Mallery, P. (2013). *IBM SPSS Statistics 21 step by step: A simple guide and reference* (13^a ed.). Pearson Higher Education.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1999). *Análisis multivariante* (5^{ta} ed.). Pearson Prentice Hall.
- Hernández-Mendo, A., & Canto-Ortiz, J. (2003). El liderazgo en los grupos deportivos. En A. Hernández-Mendo (Coord.), *Psicología del Deporte (vol.1): Fundamentos 2* (pp. 6-28). Tulio Guterman.
- International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests* (2^{da} ed.). Recuperado de: www.InTestCom.org
- Jöreskog, K. G. (1977). Factor analysis by least-squares and maximum-likelihood methods. En K. Enslein, A. Ralston, & H. S. Wilf (Eds.), *Statistical methods for digital computers* (pp. 125-153). Wiley.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3^{ra} ed.). The Guilford Press.
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, 38(1), 88-91. <https://doi.org/10.3758/bf03192753>
- Maehr, M. L. (1984). Meaning and motivation: Toward a theory of personal investment. En C. Ames & R. Ames (Eds.), *Research on motivation in education, Vol. 1: Student motivation* (pp. 115-144). Academic Press.
- Marcén-Muñío, C., Gimeno-Marco, F., & Gómez-Bahillo, C. (2016). Adaptación de la Escala de Liderazgo para el Deporte (LSS) para deportistas y entrenadores de un centro de tecnificación. *Cuadernos de Psicología del Deporte*, 16(3), 21-32. <https://revistas.um.es/cpd>
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276-282.
- Montero, I., y León, O. G. (2007). Guía para nombrar los

- estudios de investigación en psicología. *International Journal of Clinical and Health Psychology*, 7(3), 847-862.
- Muñiz, J. (2005). *Análisis de los ítems*. La Muralla.
- Muthén, L. K., y Muthén, B. O. (2012). *Mplus User's Guide. Seventh Edition*. Los Ángeles, CA: Muthén & Muthén
- Newton, M., Duda, J. L., & Yin, Z. (2000). Examination of the psychometric properties of the Perceived Motivational Climate in Sport Questionnaire-2 in a sample of female athletes. *Journal of Sports Sciences*, 18(4), 275-290. <https://doi.org/10.1080/026404100365018>
- Nicholls, J. G. (1984). Conceptions of ability and achievement motivation. En R. Ames & C. Ames (Eds.), *Research on motivation in education, Vol. 1: Student motivation* (pp. 39-73). Academic Press.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3^{ra} ed.). McGraw-Hill.
- Pagano, A. E., & Vizioli, N. A. (2020). Adaptación del Cuestionario de Características Psicológicas Relacionadas con el Rendimiento Deportivo (CPRD). *Revista Evaluar*, 20(3), 51-67. <https://doi.org/10.35670/1667-4545.v20.n3.31713>
- R Core Team. (2019). R: A language and environment for statistical computing (Versión, 3.6.0) [software de cómputo]. R Foundation for Statistical Computing. <https://www.r-project.org>
- Richaud de Minzi, M. C. (2008). Nuevas tendencias en psicometría. *Revista Evaluar*, 8(1), 1-19. <https://doi.org/10.35670/1667-4545.v8.n1.501>
- Romero, W. y Bologna, E. (2011) Técnicas de muestreo en *Estadística para Psicología y Educación*, 251-376.
- Ruiz-Barquín, R. (2007). Características de liderazgo en el deporte del judo. *Revista de Psicología del Deporte*, 16(1), 9-24.
- Ruiz-Barquín, R., & de la Vega-Marcos, R. (2015). Adaptación de la escala de liderazgo LSS-3 al fútbol. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte*, 60, 677-700. <https://doi.org/10.15366/rimcafd2015.60.005>
- Schumacker, R., & Lomax, R. (2016). *A beginner's guide to structural equation modeling*. Routledge.
- Serpa, S., Lacoste, P., Antunes, I., Pataco, V., & Santos, F. (1988). *Methodology of translation and adaptation of a specific sport test: A Leadership Scale for Sports* [Presentación de artículo]. II National Symposium on Psychology Research, Lisbon, Portugal.
- Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics* (2da ed.). Harper & Row.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209-220. <https://doi.org/10.1037/a0023353>
- Torrado-Quintela, J. (2012). Liderazgo entre iguales en equipos deportivos: Una revisión camino a la integración. *Escritos de Psicología - Psychological Writings*, 5(2), 12-19. <https://doi.org/10.24310/espiescpsi.v5i2.13296>
- Urra-Tobar, B. A. (2015). Análisis del liderazgo preferido, percibido y observado por técnicos y deportistas en fútbol formativo: Un estudio de caso. *Cuadernos de Psicología del Deporte*, 15(1), 197-209 <https://doi.org/10.4321/S1578-84232015000100019>
- Wałach-Biśta, Z. (2013). A Polish Adaptation of Leadership Scale for Sports – A Questionnaire Examining Coaching Behavior. *Human Movement*. 14(3) <https://doi.org/10.2478/humo-2013-0032>
- Zhang, J., Jensen, B. E., & Mann, B. L. (1997). Modification and revision of the leadership scale for sport. *Journal of Sport Behavior*, 20(1), 105-122.

Apéndice 1

Escala de Liderazgo en el Deporte (LSS-1)

Se presenta la versión definitiva de la adaptación de la LSS (Chelladurai y Saleh, 1980) en deportistas argentinos.

Cada uno de los ítems del siguiente cuestionario describe comportamientos específicos que puede tener un entrenador. Por favor, indica **tu preferencia** utilizando las siguientes opciones:

1	2	3	4	5
Siempre	A menudo (75% del tiempo)	De vez en cuando (50% del tiempo)	Raramente (25% del tiempo)	Nunca

Por ejemplo: “Yo prefiero que mi entrenador...”

	1	2	3	4	X	
Felicite a un deportista por su desempeño en frente a otros.						<i>Esto significa que Yo prefiero que mi entrenador NUNCA felicite a un deportista por su desempeño frente a otros.</i>

Lo que te pedimos aquí son tus preferencias personales. Contesta a todos los ítems, tené en cuenta que no hay respuestas correctas o incorrectas. Tus respuestas sinceras y espontáneas son importantes.

Yo prefiero que mi entrenador/a...	1	2	3	4	5
1. Se fije que cada deportista esté trabajando dentro de su capacidad.					
2. Pida la opinión de los deportistas sobre estrategias para competencias específicas.					
3. Ayude a los deportistas con sus problemas personales.					
4. Felicite a un deportista por su desempeño en frente a otros.					
5. Explique a cada deportista las técnicas y tácticas del deporte.					
6. Trabaje relativamente independiente de los deportistas.					
7. Ayude a los miembros del grupo a solucionar sus conflictos.					
8. Preste particular atención a corregir los errores de los deportistas.					
9. Consiga la aprobación del grupo en temas importantes antes de seguir adelante.					
10. Le diga a un deportista cuando hace un trabajo particularmente bueno.					
11. Se asegure que su propio rol en el equipo sea entendido por todos los deportistas.					
12. No explique su accionar.					
13. Se preocupe por el bienestar personal de los deportistas.					
14. Instruya a cada deportista individualmente en las habilidades del deporte.					
15. Incluya a los deportistas en la toma de decisiones.					
16 (17). Planee por adelantado lo que se debe hacer.					
17 (18). Incentive a los deportistas a hacer sugerencias en la manera de llevar a cabo las prácticas.					
18 (19). Haga favores personales a los deportistas.					
19 (20). Explique a cada deportista lo que debe y no debe hacer en su práctica deportiva.					
20 (21). Deje que el grupo ponga sus propias metas.					
21 (22). Expresé el afecto que tiene por sus deportistas.					
22 (23). Espere que cada deportista realice su tarea hasta el último detalle.					
23 (24). Deje a los deportistas intentar hacer las cosas a su manera, incluso cuando se equivocan.					

Yo prefiero que mi entrenador/a...	1	2	3	4	5
24 (25). Anime a los deportistas a confiar en él.					
25 (26). Señale las fortalezas y debilidades de cada deportista.					
26 (27). Se niegue a comprometerse en un punto.					
27 (28). Reconozca cuando un deportista se desempeña bien.					
28 (29). Dé instrucciones específicas a cada deportista sobre qué hacer en cada situación.					
29 (30). Pida la opinión de los deportistas sobre temas de entrenamiento importantes.					
30 (31). Fomente relaciones cercanas e informales con deportistas.					
31 (32). Vea que los esfuerzos sean coordinados.					
32 (34). Se distancie del resto.					
33 (35). Explique cómo la contribución de cada deportista encaja en el funcionamiento general del equipo.					
34 (36). Invite a los deportistas a su casa.					
35 (37). Dé crédito cuando sea merecido.					
36 (38). Especifique en detalle qué se espera de cada deportista.					
37 (39) Deje que los deportistas decidan las jugadas a usar en el partido.					
38 (40). Hable de manera tal que no se lo pueda cuestionar.					

Claves de corrección:

La puntuación de cada uno de los ítems es la siguiente:

Siempre = 5

A menudo = 4

De vez en cuando = 3

Raramente = 2

Nunca = 1

Puntaje promedio de los factores: la suma de las puntuaciones de los elementos de una dimensión se divide por el número de elementos de esa dimensión para obtener el puntaje promedio de cada factor.

Obtención de percentil: se traslada la suma de las puntuaciones brutas a la Tabla 4.

Factor	Ítems	Puntuación bruta	Percentil
Comportamiento autocrático	6; 12; 26; 32; 38		
Entrenamiento e instrucción	1; 5; 8; 10; 11; 14; 16; 19; 22; 24; 25; 27; 28; 31; 33; 35; 36		
Comportamiento democrático	2; 3; 4; 7; 9; 13; 15; 17; 18; 20; 21; 23; 29; 30; 34; 37		